

CHAPTER ONE

Intention

Suppose that you have been wondering what colour to paint your front door. You have narrowed the options to two: dark red, or dark blue. Both would be nice; both are available. But time is pressing, and you need to decide. So you make your choice. Blue is it.

As a result of your choice you are in further mental state to those you were in before. You still think that both would be nice; you still think that both are available. In addition though, you are now in a state that does not look like either a belief or a desire. You have an *intention* to paint the door blue.

We ascribe intentions freely, both to ourselves and to others.¹ Yet for most of the Twentieth Century neither philosophy nor psychology made much reference to them. Admittedly theorists were happy to speak of actions being performed intentionally; but this was standardly understood just to mean that they flowed from the agent's beliefs and desires. Little attention was given to the idea that intentions might be self-standing mental states.

Things have changed. In philosophy in the last twenty years, a great deal of work has been done exploring the nature of intentions; and work in psychology has provided good evidence that the philosophical account is basically right. My aim here is to give a sketch of the core philosophical work, show how the empirical work bears on it, and then extend the core account to include a particular kind of intention that I call *resolutions*. I conclude this chapter by returning to the issue of whether intentions should be understood as self-standing states, or whether they can be reduced to beliefs and desires.

THE FEATURES OF INTENTION

Let us start with Michael Bratman's influential account. It has three parts.² Bratman identifies the characteristics that a state needs to have if we are to count it as an intention; he explains the useful role that states with these characteristics would have in our lives; and he gives reasons for thinking that beliefs and desires on their own could not fulfil this role. Taken together these provide the basis for the case that we should acknowledge the existence of intentions. For if we seem to have them, and if having them would play a useful role, and if indeed they do seem to play this useful role, then the burden of proof rests in their favour. We will need compelling arguments before giving them up.

Let us postpone for now the third element of Bratman's argument—the contention that beliefs and desires cannot play the role allotted to intentions—and focus instead on the other two elements: the account of the distinguishing marks of intentions, and of the useful role that intentions can play. I start with the distinguishing marks; in particular, with the distinguishing marks of *future directed* intentions, that is, intentions to do something at some future time.

¹ For an exploration of how this works see (Malle and Knobe, 2001).

² The main statement is in (Bratman, 1987)

A future directed intention to perform a certain action typically has the following characteristics. The agent forms the intention at one time either by making an explicit conscious decision to perform the action or by some less deliberate, more automatic, process. Then, unless it is revised, the intention will lead the agent to perform the action *directly*; it is, as Bratman says, *controlling*. Moreover, it is relatively immune to reconsideration and hence to revision. Once formed, intentions have a tendency to persist. They have what Bratman calls *stability*. Stability is not, of course, absolute. Sometimes we revise our intentions, and it is quite rational that we do so. (“When the facts change,” Keynes reputedly said when accused of inconsistency, “I change my mind. What do you do, sir?”) Stability can best be understood as a shift in the threshold of relevance of information: some information that would have been relevant in forming an intention will not be sufficient to provoke rational reconsideration once an intention has been formed.

What need do we have of states that are stable and controlling in this way? Part of the answer stems from the observation that we are epistemically limited creatures. Information is scarce, and costly to obtain. Reasoning on the information that we have takes time and effort. It is rational then to allocate our scarce resources by limiting the amount of time we spend looking for information and reasoning on it. We should do a certain amount of searching and reasoning, and then stop and make a decision on the basis of that. Now if we were to act immediately on making the decision, then this would not in itself give rise to a need for intentions. But suppose our decision is a decision to act tomorrow. Then we need some way of storing our decision so that we act on it tomorrow without reconsidering it (for to reconsider it would violate the requirement that we not reason about it further). We need an intention.

This in turn raises another question: why should we want to make a decision now about how we will act tomorrow? It could be that now is a more propitious moment to reason. Tomorrow there will be some factor that will stop us from reasoning well: we will be short of time, or distracted, or under the spell of some temptation. But a more common reason for doing our reasoning ahead of time is that all sorts of other actions will be dependent upon what we decide to do; and we will need to perform some of these actions in the meanwhile. Thus suppose I will not paint my front door till tomorrow. There might still be many reasons for deciding today what colour to paint it. It could be that tomorrow I will start at the crack of dawn when there will be too little light to see the colour charts properly; that will give me a reason of the first kind for doing my deliberation today. More likely, tomorrow the paint shop will be shut, so I should buy the paint today; and since I want only to buy the colour that I will use, I will need to have decided today what colour that will be. I need to form my intention ahead of time. Thus intentions are necessary in the intra-personal case for agents to co-ordinate their activities. They are also important in the inter-personal case. Suppose some friends are coming to visit me next week, and the easiest way for them to recognize my house is from the colour of the front door. I shall be speaking with them today, but not again before they visit. Then once again I have a reason for forming today an intention about which colour I shall paint my door tomorrow, so that I can tell them what to look for.

It will be important to my argument to realize that there can be good reason for forming intentions even in cases where rational deliberation does not provide sufficient grounds for my decision. One such case involves so-called Buridan examples, examples in which I am indifferent

between options.³ If I am really indifferent between red and blue for my door, it is nonetheless important that I decide on one of them, and form the corresponding intention. Otherwise I won't know what paint to buy, or what to tell my visitors. A more interesting case concerns options that the agent finds incommensurable.⁴ Suppose (taking an example at random) that I do not know how to compare the demands of leaving home to fight fascism with the demands of staying to look after my mother.⁵ Here the problem is not that each outcome fares equally well on the same scale; I do not know how to place them on the same scale. Nevertheless I need to decide on one course, and to form the corresponding intention; both intra- and inter-personal co-ordination require this. I need to know whether to start winding up my affairs and packing my bag; and I need to know what to tell my mother so that she too can make her plans accordingly.⁶

PSYCHOLOGICAL EVIDENCE

The account I have sketched shows us why it is that having intentions would be beneficial. But that does not by itself show us that we have them. After all, wings may well be beneficial, but we certainly lack them. Admittedly, we don't come with a prior commitment to the view that we have wings, whereas we are committed to intentions. Still, it would be good to know if there is solid psychological evidence that we have intentions along something like these lines.

There is. Indeed there are two distinct bodies of evidence supporting his account, both stemming from the work of Peter Gollwitzer. The first provides evidence of something like the stability brought on by forming intentions. The second provides evidence of their controlling nature. And while both basically support the account of intention that we antecedently hold, both require refining that account in an interesting way.

Stability

To get an understanding of the first body of evidence it will be helpful to step back to consider a famous article published by Alloy and Abramson in 1979.⁷ Their aim was to understand how much control subjects think that they have over the world. They set up a simple apparatus consisting of a button and a light, and asked subjects how to estimate the degree of control that the button gave them over the illumination of the light. The start of each trial would be marked by a secondary light coming on, at which point the subjects would either press the

³ Named after the example of Buridan's ass, who starves, midway between two piles of hay and unable to choose either. Although the example is standardly attributed to Jean Buridan, the attribution is not substantiated. See the discussion in (Ullman-Margalit & Morgenbesser, 1977).

⁴ I don't say that they need *be* incommensurable, only that the agent finds them so. I return to this in Chapter Three.

⁵ (Sartre, 1945)

⁶ Bratman points out that cases like these show the utility of intentions even in cases where information is not limited. For we have reason to think that Buridan cases at least will arise even for agents who know everything. He also argues, I think convincingly, that they provide grounds for thinking that we cannot reduce talk of intentions to talk of beliefs and desires. For, by hypothesis, I do not desire that I take one course rather than the other; and my intention to take one course cannot be seen as a belief that I will take it. See (Bratman, 1985) at pp. 22-3; (Bratman 1978) pp. 11, 22ff.

⁷ (Alloy and Abramson 1979)

button or not; this was followed by the main light either coming on or not. Subjects were asked to judge how responsive the main light was to the pushing of the button. Alloy and Abramson found that normal subjects were pretty good at realizing that they had no control when the light only came on rarely, i.e. when it came on in 25% of trials. But once the light started coming on frequently (in 75% of trials) they hugely overestimated how much control they had.

Much of the interest of Alloy and Abramson's article came from their finding that depressed subjects were not vulnerable to this illusion: unlike normal subjects, they did not overestimate their degree of control. And this gave rise to a large literature suggesting that over-confidence in one's ability to control the world, together with over-confidence in one's abilities more generally, could be beneficial: people who are over-confident are more likely to persist, and ultimately to succeed, when more realistic people would give up.⁸

This raises an interesting question. For whilst over-confidence can bring certain benefits in enabling agents to persist in the face of adversity, it can obviously also be a liability if it makes agents choose projects that they are unable to complete. It is here that Gollwitzer's work comes in. He found that in non-depressed agents, over-confidence comes and goes depending on their attitude. If the agent is in what he called a *deliberative mindset*, i.e. a mindset that is focused on deciding what to do, they are much less prone to illusions of control. In contrast, if they are in an *implemental mindset*, i.e. a mindset that is focused on implementing a prior decision, they are far more prone to illusions of control. Gollwitzer found that it was enough to get subjects to focus either on a unresolved personal problem, or on a fixed personal goal, to get radically different results in the Alloy and Abramson set up.⁹

The difference between deliberative and implemental mindsets is manifested in a host of further ways. Those in an implemental mindset are more likely to over-estimate their other abilities in addition to control,¹⁰ are more likely to focus on the advantages of achieving their goal than the disadvantages,¹¹ and are less receptive to new or peripheral information.¹²

These last features take us back to the idea of the stability of intention. For the idea there was exactly that agents are less ready to consider new information once an intention is formed than they would have been when they were in the process of forming the intention. However, we need to be a little careful before concluding that this provides empirical support for the stability of intentions. For stability was proposed as a feature of the intentions themselves, whereas Gollwitzer has shown that insensitivity to evidence is a feature of the mindsets of agents, something that can be then applied to other states that are in no way connected to an intention. When agents assess the control they have over the light they are not implementing an intention. So how do the two pictures relate?

I think that there are two possible interpretations of Gollwitzer's findings. The first is that stability is a feature of the intentions themselves, but that the effect of contemplating such an intention is to cause agents to impute the stability to other states: there is a form of stability contagion from intentions to other mental states. The second interpretation is that stability is

⁸ See especially (Taylor and Brown 1988), and for later review (Taylor and Brown 1994).

⁹ (Gollwitzer & Kinney, 1989); for background to this piece and summary of subsequent work see (Gollwitzer, 2003). For a summary of the differences between deliberative and implementational mindsets more generally see (Gollwitzer & Bayer, 1999).

¹⁰ (Taylor & Gollwitzer, 1995)

¹¹ (Taylor & Gollwitzer 1995)

¹² (Heckhausen & Gollwitzer, 1987); (Gollwitzer, Bayer & Wasel 1998)

not a feature of intentions themselves, but is solely a feature of the mindsets that agents bring to intentions; however, the stable (i.e. implemental) mindset is one that is engendered by focusing on intentions. (A compromise sees stability in both the mindset and the intentions.)

I don't know which of these interpretations is right. Gollwitzer writes as though the second is, but, so far as I can see his data are all consistent with the first. But interesting though the distinction is, I don't think that it need worry us. I will go on speaking of intentions as stable, meaning this to be understood either that they have a degree of intrinsic stability, or that they engender a mindset that treats them as stable. Either way we will get much the same results.

Control

Let us turn now to the second characteristic imputed to intentions on our account, their controlling nature. The idea was that intentions can move agents to action *directly*; the agents do not need to reconsider whether to perform the act.

If intentions were controlling in this way, one would expect to find that an agent who formed an intention to perform a certain action would be more likely to perform it than would a similarly motivated agent who had so far failed to form that intention. This is indeed the case, but to understand the findings here, and to see that they really do support the idea that (at least *some*) intentions are controlling in Bratman's sense, we need to understand a second distinction made by Gollwitzer, that between goal intentions and implementation intentions. Goal intentions are categorical in form: for instance, one intends to visit the Taj Mahal. In contrast implementation intentions are, explicitly or implicitly, conditional: one intends, if one sees a bus to the Taj Mahal, to get on it; or one intends to leave for the Taj Mahal at two o'clock.¹³ Typically implementation intentions work in the service of goal intentions: having decided to go to the Taj Mahal one forms the further intention to get the bus to it, or to go at a certain time.¹⁴

Gollwitzer found that it is implementation intentions, rather than mere goal intentions, that increase the likelihood that subjects will act as they plan. Forming an intention to write a report of what they did on Christmas Eve sometime during the following 48 hours was enough to get a third of subjects to do so; but forming an additional implementation intention to write the report at some specific time within that 48 hours was enough to get three quarters of subjects to do so. Similarly a simple goal intention to do breast self examination was acted upon by half of the women in the study; where this was augmented with an implementation intention about exactly when to do it, all the women did it. Numerous other studies have found similar effects.¹⁵

Clearly these results are very striking, but they should not be too surprising. If all one has is a goal intention, then this cannot be completely controlling: one will have to do more deliberation about how to implement it. And one may either forget to do that deliberation, or having started it, one may decide to give up on the intention altogether. Alternatively, one may decide to procrastinate, and then, having procrastinated, one may never come back. In contrast an implementation intention, providing it is well chosen, does not require the agent to reopen

¹³ I.e. one intends, if it is two o'clock, to leave for the Taj Mahal.

¹⁴ (Gollwitzer 1993); (Gollwitzer 1996); (Gollwitzer, 1999) 3; for summary see (Gollwitzer, Fujita, & Oettingen, 2004).

¹⁵ (Gollwitzer, Fujita, & Oettingen, 2004), pp. 214ff.

deliberation. The relevant cue can simply serve to trigger the action. In Gollwitzer's phrase, one is 'passing the control of one's behavior on to the environment'.¹⁶

An array of further research confirms that this is what is happening. Implementation intentions cause the agent to be sensitized to the relevant cues; indeed the response can happen even if the cues are presented subliminally.¹⁷ So they can certainly involve the short-circuiting of explicit deliberation. Nonetheless, the process is not blind. Implementation intentions have little effect if the goal intention that they serve is weak, and none at all if it has been abandoned.

So we have good evidence that intentions can be controlling. This is not in the mechanical sense that we are locked into an action come what may. Rather, provided we have formulated the implementation intentions in ways that tie them to perceptible cues, intentions enable us to act in ways that do not require us to deliberate further.

RESOLUTIONS

I have sketched the main lines of an argument for the utility of intentions, and have provided some empirical grounds for thinking that we actually have them. Once we have got this far it is easy to see that they might be useful in other roles too. Suppose that there is an act that I now believe I should perform at some time in the future; indeed currently I actively want to perform it. Suppose though that I know that when the time comes I shall not want to perform it. I shall be tempted to do something else. Then it would be useful to form an intention now, an intention that will lead me directly to act when the time comes, and that will provide some resistance to reconsideration in the light of the inclinations I shall have then. Similarly, suppose that I know that my future reasoning will go awry: after a few glasses of wine my confidence in my own abilities will be absurdly high. Then again it would be good to form intentions now that are somewhat resistant to reconsideration in the light of those beliefs. In short, it would be good to have a specific type of intention that is designed to stand firm in the face of future contrary inclinations or beliefs: what I shall call a *resolution*.

Philosophers have been keen on the idea that sometimes an intention will not be enough; that we will need to bind ourselves in some further way to fulfil our intentions, perhaps by placing temptation out of our reach, or by telling others of our plans so that fear of their disapproval should we fall short will provide an extra incentive.¹⁸ But we should not let our interest in these exotic methods blind us to the fact that very often intention *is* enough. People get up on cold dark mornings, leave enjoyable lunches to return to work, give up nicotine, or alcohol, or other drugs; and they frequently use no mechanism other than an intention to overcome their contrary desires.

¹⁶ (Gollwitzer 1993) p. 173.

¹⁷ (Gollwitzer, Fujita & Oettinger 2004), p. 213.

¹⁸(Elster, 1979). I suspect that very often the point of telling others of our resolutions is not to incur their scorn if we revise them. Rather it is simply to remind us of what our resolutions were, or of how seriously we made them. A midwife told me of the sole entry on one woman's birth plan: "If I ask for pain relief in labour, show me this birth plan". The woman did ask for pain relief, the midwife showed her the birth plan, and the woman decided against taking any. Presumably this was exactly a case in which the woman didn't want to publicize her resolution, but still wanted others to remind her of it.

The use of intention to overcome desire becomes especially important in cases that involve repeated actions. I reason that since smoking forty cigarettes a day for the rest of my life will make a considerable difference to my chance of getting lung cancer, I should give it up. But should I deny myself the cigarette I was about to have? Smoking one cigarette will make very little difference to my chances of getting lung cancer, and anyway, that is all in the future. In contrast, one cigarette can give me considerable pleasure now. So why deny myself? Unfortunately the same argument will work just as well forty times a day for the rest of my life. What we need here is not just an intention to perform—or, in this case, to refrain from performing—a specific action. We need a general intention concerning a certain type of action; what Bratman calls a *policy*.¹⁹

It is no easy matter to see quite how intentions can work to overcome future desires or beliefs. After all, although intentions are stable, they don't lock us in to a course of action. We can imagine a being that did work in this way: a being that, having formed an intention, would be inexorably moved to action by it. But we are not like that. We can revise an intention in the light of a changed desire or belief; so why do we not do so, even in cases where the intention was expressly formed in anticipation of such changes? Providing an answer to that question will be an ongoing theme of this book. But let me say a little here about the kind of structure that resolutions might have.

Clearly, if it is to work, a resolution has to be something that holds firm against temptation. At one extreme we could think of them simply as intentions with a specially high degree of stability. But that doesn't seem to get it right. It is no part of the nature of a resolution that it *will* be effective; the point is rather that it is *meant* to be. At the most intellectual level, resolutions can be seen as involving both an intention to engage in a certain action, and a further intention not to let that intention be deflected. Understood in this way they involve a conjunction of two simpler intentions, one first-order and one second-order (i.e. an intention about an intention). So, when I resolve to give up smoking, I form an intention to give up, and along with it I form a second-order intention not to let that intention be deflected.²⁰

There is plenty of reason to think that we can and do form second-order intentions. For instance, on forming a goal intention to visit the Taj Mahal, I might form the further intention to make some relevant implementation intentions at a later time; that is a second-order intention. But it might seem that understanding resolutions as containing them is to make them rather too complicated. Can't children form resolutions without forming second-order intentions? Some clarifications here might help. To form a second-order intentions children would not need the idea of a second-order intention; nor would they need to have the idea of a resolution. What they would need is the idea of a first-order intention, for only then could they form intentions about intentions; but this idea need not be terribly complex or well-articulated. They certainly would not need to have any grip on the sort of account proposed here, the idea of intentions as states that are stable and committing (how many adults have a grip on that?). All they would need is the idea of what they are going to do, where this is distinct from what they want to do. And even this need not be a conscious idea; it would be good enough if their behaviour indicated that they could grasp it unconsciously.

¹⁹(Bratman, 1987) pp. 87-91. If these 'one more won't hurt' argument are not to be simply irrational, they will need to involve a degree of preference shift. See below, Chapter Six.

²⁰ I am grateful to Alison Gopnik for discussion here.

Still it would certainly be evidence against the second-order construal of resolutions if it looked as though children could form resolutions before they came to understand the idea of an intention whether consciously or unconsciously; and it would be some evidence in favour of the construal if the ability to form resolutions came around the same time as, or subsequent to, the grasp of the idea of intention. We do not have overwhelming evidence on this point. But we do have evidence that children get much better at delaying gratification—that is, at resisting the temptation to take a small benefit when by waiting they can gain a larger one—at around the age of five: that is, at about the time that they come to understand the idea of intention.²¹ Now, of course, there may be ways that one could delay gratification without forming a resolution to resist: one might simply distract oneself, as many of the children seemed to do. Nevertheless, it does appear that some of the children were forming resolutions, something that they will go on to do with increasing frequency as they get older. And the evidence is that that ability comes along with an understanding of the nature of intentions, just as a higher-order theory of resolutions would suggest.

PRESENT-DIRECTED INTENTIONS

The picture of intentions sketched so far is primarily a picture of future-directed intentions: intentions to do something at some future time. But we might wonder whether there couldn't be present-directed intentions: intentions to perform an action *now*.²² And if there are, then it might seem that the account offered will not cover them. For some of the main advantages that were held to accrue from having intentions—curtailing deliberation, aiding inter- and intra-personal coordination, enabling resistance to temptation—do not seem to apply to present-directed intentions.²³

A first thing to say is that if intentions are thought of as enduring states, then, for almost all intentions, there will come a point at which they are to be implemented. In that sense then, they will be intentions to perform an action *now*. A future-directed intention will simply turn into a present-directed intention with the passage of time. So the issue is not really whether there can *be* present-directed intentions. The issue is rather whether, on the account proposed here, there can be reason for an agent to *form* a present-directed intention.

There can. It is simply not right to say that none of the advantages discussed so far apply to the forming of present-directed intentions. Indifference and incommensurability are problems that can affect present-directed action just as much as future-directed (I can be indifferent about what colour to paint the door *now*); and by hypothesis these are cases in which desires and beliefs are not sufficient to decide one way rather than another. So an agent faced with either of them would seem to have reason to form a present-directed intention to break the impasse.

²¹ I discuss the delayed gratification literature in Chapters Five and Six. For evidence of when children grasp the idea of an intention see (Schult 2002).

²² See, for instance, the notion of proximal intentions in (Mele, 1992), Part II.

²³ The criticism is made by David Velleman in his review of Bratman's *Intention, Plans and Practical Reason* (Velleman 1990) and again in (Velleman 2007).

Moreover, even if it were held that this on its own is too slight a reason to admit the existence of present-directed intentions we might challenge the whole thrust of the argument that says that we should only concede their existence if we can find specific advantages to be had from forming them. It could be that they are there because of the advantages they bring in the future-directed case, and that then they get co-opted to work. If intentions are real psychological kinds, identified in the kind of empirical work that has been discussed, then it is an interesting empirical question whether present-directed action works, very quickly, using them; or whether one can decide to perform an action and go on to perform it directly, without need of an intervening intention. I don't know of any empirical work that addresses this question, so for now I think that we should leave it open. Indifference and incommensurability give us some reason for admitting present-directed intentions, but not enough to be clinching.

Suppose it turned out that direct action were possible, without an intervening intention of the kind we have been discussing. Would that show that the account of intention I have offered is partial? It only would if we thought that there were some other kind of intention that obtained in cases of present-directed action, a kind that needed a further account. But why think that? The only reason I can see is that such actions would be naturally described as 'intentional'; and where there is intentionality, it might be thought, there must be intentions. That brings us to the complex issue of the relation between an action being intentional, and it being done on the basis of an intention.

INTENTIONS AND ACTING INTENTIONALLY

There is a simple and straightforward theory that insists that an agent performs an action intentionally if and only if they act on an intention to perform that action. Some have proposed such an account because it provides the basis of a reduction: if acting on an intention is equivalent to acting intentionally, and acting intentionally can be reduced to acting on the basis of a certain belief and desire, then intentions can be reduced to beliefs and desires.²⁴ Others have simply assumed that two such closely related linguistic constructions as 'intention' and 'intentionally' should be equivalent in this way, or that one should be derived from the other.²⁵

However plausible this might initially seem, there are good reasons for rejecting the equivalence. It has occasioned a great deal of discussion that I shall not repeat here.²⁶ Let me just point to two considerations. First, anyone struck by the linguistic proximity of 'intention' and 'intentionally' in English would do well to look to the complexity of translating those notions into other languages. In German or Spanish, for instance, one finds a number of different terms, and choosing the right one is a delicate business. If we started with those languages we might be more likely to think of a set of related but different notions, rather than

²⁴ See (Davidson 1963); in later work he identifies intentions with all-out evaluations, which, rather oddly, he identifies with the actions themselves. See (Davidson, 1969) and (Davidson, 1978). For discussion see (Bratman 1985) at pp. 22-3

²⁵ Anscombe insists that the locutions should be treated together (Anscombe 1963) §1, and is often seen as someone who wants to privilege talk of intentional action over that of intention. In fact though her position is rather complex. Her focus on intentional action is to some degree procedural (§§ 3-4); and she argues, via a series of rather strange thought experiments, that talk of intentional action cannot be understood in the absence of talk of intentions (§20).

²⁶ For discussion see (Bratman 1987) pp. ??; for a useful summary of much recent work see (Mele 1997).

the single notion that English idiom suggests.²⁷ Indeed, even in English the patterns of usage are more complex than one might expect if the simple equivalence held. Doing something *on purpose* is very close to doing something intentionally; but there doesn't seem to be a use of the noun 'purpose' that is quite the same as that of 'intention'.

Second, and more compelling, is evidence from Joshua Knobe that appears to show that 'intentionally' is a normative term in the way that 'intention' is not. Knobe asked one group of subjects to consider the following passage:

The vice-president of a company went to the chairman of the board and said, 'We are thinking of starting a new program. It will help us increase profits, but it will also harm the environment.'

The chairman of the board answered, 'I don't care at all about harming the environment. I just want to make as much profit as I can. Let's start the new program.'

They started the new program. Sure enough, the environment was harmed.

He asked whether the chairman had intentionally harmed the environment; 82% thought that he had. Knobe then asked a second group to consider an almost identical passage, except that the programme was said to *help* the environment. But here the judgment about whether the help was intentional was radically different. 77% *denied* that the chairman had intentionally helped the environment.²⁸

Other experiments have brought similar results. Ask people about an intentional action that has a *bad* consequences as a foreseen side effect, and Knobe found that most of them will say that that side effect was brought about intentionally; ask them about a parallel case in which the side effect is *good*, and a majority will say that it was not brought about intentionally. In contrast, in both cases a majority will say that there was no intention to bring about the side effect.²⁹

So normally people are more likely to see an action as intentional if its consequences are bad than in if its consequences are good. It turns out that the full picture is more complicated than this suggests. There are a few cases in which attributions of intentionality are made even when the act is judged good, and working out quite why is a tricky business.³⁰ And it can be seen that the moral status of the side effect has an impact on some people's judgments not just of whether an act is performed intentionally, but of whether it is performed on the basis of an

²⁷ Joshua Knobe and Arundra Burra, who develop this line, produce interesting data showing that the Hindi term corresponding to 'intentionally' is morphologically related to the term for knowledge rather than the term for intention; see (Knobe and Burra, 2006)

²⁸ (Knobe, 2003)

²⁹ For summary of Knobe's results, and some others, see (Knobe 2006).

³⁰ See (Knobe, 2007). Knobe suggests that the full account requires a distinction between spontaneous unconscious attributions of goodness or badness, and those that are more measured; it is the former rather than the latter that are operative in attributions of intentionality. I suspect rather that it turns on a difference between actions that involve the violation of a requirement and those that don't; but I shall not develop this here.

intention. A minority apparently use the term ‘intention’ much as the majority use the term ‘intentionally’. Nonetheless, there are very good grounds here for thinking that there are two different notions in play: one, normally denoted by the noun ‘intention’, that is a psychological notion; and one, normally denoted by the adverb ‘intentionally’ that combines the psychological with the normative in some rather complicated way. To say this is not to say that the terms are utterly distinct: it is not to say that ‘intention’ only occurs in ‘intentional’ as, say, ‘bus’ occurs in ‘business’. Of course they come from the same root, and have a great deal in common. It is just to say that, by a process of lexicalization, a process that is familiar for words that have been in the language for a very long time, they have come to have different meanings in standard English.³¹ We see that with many other terms: consider, for instance, the connections and differences between ‘awful’, ‘awesome’ and ‘awe-inspiring’. My concern in this book is with intentions, not with intentional actions.

REDUCTION

I mentioned at the beginning that the third element of Bratman’s account was an argument that the role of intention cannot be played by beliefs or desires. I hope that what I have said so far shows that intentions have an important role to play; we cannot just give them up and talk in terms of beliefs and intentions instead. But that leaves open the possibility that intentions can be analyzed in terms of beliefs and desires, so that ultimately those two can do all of the work.

Certainly the intention to ϕ does not look like either the simple belief that one will ϕ or the simple desire to ϕ . One can believe that one will succumb to a temptation without intending to do so; and equally one can want to perform some forbidden act without thereby intending to perform it. Moreover, cases of indifference are ones where, by hypothesis, one lacks the desire to perform either action, and the belief that one will. So if intention is to be reduced to belief and desire some more complex analysis must be given. Several authors have tried to give one, and they do indeed become very complex. For instance, David Velleman argues that the intentions consist in ‘self-fulfilling expectations that are motivated by a desire for their fulfillment, and that represent themselves as such’³². And on the desire side, with even greater complexity, Michael Ridge contends that

A intends to ϕ if and only if (a) A has a desire to ϕ , (b) A does not believe that ϕ -ing is beyond her control, (c) A’s desire to ϕ is a *predominant* one, which is just to say that there is no desire ψ , such that A does not believe ψ -ing is beyond her control, she desires to ψ as much as or more than she desires to ϕ , and she believes that a necessary means to her ϕ -ing is that she refrain from ψ -ing, (d) A has a desire not to deliberate any more about whether to ϕ unless new, relevant information comes to light.³³

³¹ The evolution of the terms is actually quite complex. The *OED* gives a number of other meanings for ‘intention’ that are now obsolete, including some notion of mental application or effort, as in Locke’s *Essay*: ‘When the Mind with great Earnestness, and of Choice, fixes its view on any Idea . . . it is that we call Intention or Study’ (Locke 1690) II xix §1

³² (Velleman, 1989) Chapter 4.

³³ (Ridge 1998)

I doubt that either of these accounts work, but I will not say much to defend this doubt.³⁴ I want to raise a more fundamental question: why should we want to reduce in the first place? Ridge contends that a reduction will allow us to maintain the “attractive Humean view that a full, rationalizing, explanation of an agent’s intentional actions always can be had without appealing to anything other than beliefs and desires”. In so far as the Humean account is supposed to provide a contrastive explanation of an agent’s actions, so that it explains why they performed one action *rather than* another, I think that there are other reasons for rejecting it, reasons that will be discussed in Chapter Four. But, independently of that, why should we prefer an account that offers explanations of agents’ actions in terms of beliefs and desires rather than one that makes use of beliefs, desires and intentions?

An obvious response is the Ockhamist one: an account with fewer ontological commitments is, *ceteris paribus*, better than one with more. But if this is to be plausible the *ceteris paribus* clause will need to do a lot of work. Reducing our ontological commitments by reduction should not be an end in itself. To take an example from Kripke: when describing a population of married people, we could quantify only over the women, reducing all that we wanted to say about their husbands to statements about the wives. Of course the predicates would become more complicated—each woman would now have both a wife-weight and a husband-weight, and so on—but we would have halved our ontological commitments. Still, clearly the approach is absurd. Reduction is good if it brings greater explanation, and this does not. The women are different from the men, and reducing the latter to the former only brings confusion.³⁵

I suspect that something similar is true of intentions. A reduction of intentions to beliefs and desires only confuses things. Certainly the accounts that we have seen from Velleman and Ridge bring a great deal of complexity. If they are to justify this, they need to bring some explanatory advantage beyond a mere claim of ontological reduction.

Velleman does indeed argue that his account brings such an advantage. Treating intention as a form of belief explains some otherwise puzzling features that it has. In particular it explains (i) why it is that intending to perform a certain action entails a belief that one will perform that action; and (ii) why it is that intention is governed by a consistency requirement that makes it irrational to intend to do two things that one believes to be inconsistent.

That would be all to the good if intention had these features. But they are controversial. I contend that the first is simply illusory: intention does not entail belief. The second is more complicated. Whilst there are consistency requirements on intention, I contend that they are quite weak, and give no support to the thesis that intention is a form of belief. These issues are the topic of the next chapter. I start with question of whether intention entails belief.

³⁴ In brief: Against Velleman I doubt that we can form entirely self-fulfilling beliefs prior to having reasons to believe that they are true. Compare: I tell you that there is a handkerchief inside the box before you, and that, by some complex process, it is whatever colour you believe it to be. Knowing this, and having no independent reason for thinking it is any particular colour, can you just come to believe that it is, say, red? I think not. For similar worries, though stressing the normative inappropriateness of forming self-fulfilling beliefs as well as their impossibility see (Langton 2004). Against Ridge I take it as evident, his protestations notwithstanding, that if I desire to do something but intend to resist this desire, I do not intend to do it. And the general recipe for creating Humean analyses that he proposes seems to me to beg the question.

³⁵ The example was given in a seminar in Princeton in the late 1980s.