

Why is Belief Involuntary?

Jonathan Bennett

This paper will present a negative result—an account of my failure to explain why belief is involuntary. When I announced my question a year or so ahead of time, I had a vague idea of how it might be answered, but I cannot make it work out. Necessity, this time, has not given birth to invention.

Still, my tussle with the question may contribute either towards getting it answered or showing that it cannot be answered because belief can be voluntary after all.

Most of the paper was written while I expected to get the question answered, and I have chosen not to hide that fact by revising the tone. I offer the paper as an essay in the ‘analytic’ manner in Descartes’s sense of that term. It is the manner of the *Meditations*—a presentation in the order of discovery or, in my case, of non-discovery.

1. Introduction

Could I get you to believe something by offering you an inducement to believe it? That is, could belief be produced by threats or bribes? Well, yes, in a couple of uninteresting ways; let us set them aside first.

(i) You don’t think I would ever threaten anybody, and you tell me this; I respond convincingly: ‘Believe that I am capable of threats! If you don’t, I’ll kick you on both shins.’

That might get you to acquire the belief, but only by giving you evidence for it, not by inducing you to acquire it. You would have acquired it just as easily and surely if you had heard me threaten someone else. My question is: Could I get you to believe something by offering you an inducement to believe it without thereby providing evidence for it? Or, for brevity, Could I induce you to believe something?

(ii) Pascal held that one’s chances of believing Christianity could be improved by keeping pious company, reading holy books, and praying. ‘That will dull your mind and make you believe,’ he charmingly said. Other kinds of mind-dulling can also produce credulity. We can make a person temporarily believe in UFOs or ESP (for example) by perceptually isolating him for several days and then barraging him with propaganda; and beliefs can be caused by post-hypnotic suggestion. There may be other such techniques, and some of them may be quite reliable. Clearly, I could induce you to start up one of those procedures leading to your acquiring a specified belief. My question is: Could I *immediately* induce you to believe something? That is, could an inducement offered by me bring it about that you believe something, where (i) I don’t provide evidence for the truth of the belief and (ii) this does not work by your being induced to do something that leads to your acquiring the belief?

I immediately induce you to raise your arm if I say ‘If you raise your right arm I’ll make it worth your while’ and you, greedy or curious, immediately raise the arm. This is an immediate inducement because you are not induced to do something that has as a consequence your raising your right arm. (There are indeed intermediate steps: you come to believe that I have offered the inducement, and you form an intention to raise your arm. But you are not induced to do either of those—they are constitutive parts of your being induced to raise your arm, not means you are induced to take to getting your arm up.) The inducement would be mediated if it induced you to arrange for someone else to bully you into raising your arm, or to get into a situation where you would have a nasty accident if you didn’t raise it, or to pull your right arm up with your left hand. I immediately induce you to do A if your response is to do A ‘just like that’, as Bernard Williams has said.

It will be important later on, but I had better say it now: What one is induced to do—or intends or plans or wants or aims or tries to do—is to bring about a state of affairs. When Norman hit the ball in my direction, I caught it, thereby producing a particular catch; but as I ran and extended my hand I was not trying to produce *that catch*, but rather to produce *a catch* or (better still) to make it the case that *I caught the ball*. Quite generally, inducements should be thought of as inducements to make it the case that P for various values of P. This can be defended by argument,¹ but in this paper I offer it as an axiom of the discussion.

A possible trap is worth a paragraph. When I offer you a reward if you will raise your arm and you raise it ‘just like that’, you don’t set yourself to do something which will lead to your arm’s going up; so the offer of a reward does not

relate in a *motivationally mediated* way to your arm’s going up. Still, your arm wouldn’t go up unless certain neural events occurred first; so the offer of a reward relates in an *ontologically mediated* way to your arm’s going up—the offer and the gesture lie on a causal chain with other items in between. It is clear that although the neural events mediate in the causal sequence, no motivating thought about them intervenes in the practical syllogism that takes you from the reward premiss to the action conclusion. If one doesn’t distinguish the two kinds of (im)mediacy one will think that belief is on a par with behaviour as regards voluntariness: each can be induced, and neither can be induced immediately.² But this masks what may be a real difference, namely that between what is motivationally (and thus also ontologically) mediated and what is merely ontologically mediated. It may be the case that behaviour can and belief cannot be induced in a motivationally immediate manner.

That is the question. Can I immediately induce you to acquire a belief? Can I, without giving you evidence for the belief and just by offering convincing bribes or threats, get you to acquire the belief ‘just like that’, immediately, without setting your thought on producing some intervening event that will lead to your getting the belief? The inducement need not be offered by another person—an impersonal part of the world would do as well. What matters is whether one can acquire a belief because one sees an advantage in having it, this being done as immediately as in raising one’s arm because one thinks it will be good for it to be up.

It is widely held that no-one could do this, and I think the usual view is that the impossibility is strong to the point of being conceptual or logical. Most of those who think about the question have found it incoherent or absurd to

¹ It is obliquely defended in J. Bennett, *Events and their Names* (Indianapolis: Hackett, 1988), ch. 14.

² Thus Margery Bedford Naylor, ‘Voluntary Belief’, *Philosophy and Phenomenological Research* 45 (1985) 427–436, especially p. 430.

suppose anyone should acquire a belief, just like that, *simply* because he wanted to, as though acquiring a belief were like raising one's arm. I agree. There is indeed something so chokingly unswallowable about the idea of someone's voluntarily coming to believe something that I have to suspect that this is ruled out at a deeper level than the contingent powers of our minds.

I want to get beyond the level of mere hunch or intuition, by explaining why voluntary belief is conceptually impossible. (When I say that belief cannot be voluntary, I mean that nobody could be immediately induced to believe something. For James Montmarquet, the involuntariness of belief consists in the fact that what one can believe is normally constrained by what theoretical reasons one has, which he compares with the fact that how one can act is constrained by what practical reasons one has. He infers that believing and acting are on a par so far as voluntariness is concerned.³ This is indeed a respect in which belief and behaviour are alike, but it does not make them both involuntary. Anyone who thinks it does must be taking voluntariness to require freedom from all constraints, total unconditionedness. In company with Locke, Leibniz, Hume and many others, I understand voluntariness as responsiveness to practical reasons. I take ϕ ing to be voluntary if one's ϕ ing depends upon inducements, that is, if one can ϕ or not depending on whether one thinks one has practical reasons to ϕ . Actions are voluntary in that sense, and beliefs seem not to be.)

³ James Montmarquet, 'The Voluntariness of Belief', *Analysis* 46 (1986) 49-53.

⁴ For helpful elaborations of the distinction, see Ronald B. de Sousa, 'How to Give a Piece of Your Mind: or, The Logic of Belief and Assent', *Review of Metaphysics* 25 (1971) 52-79; Mark Kaplan, 'Rational Acceptance', *Philosophical Studies* 40 (1981) 129-145; Patrick Maher, 'The Irrelevance of Belief to Rational Action', *Erkenntnis* 24 (1985) 363-384. I am indebted to Patrick Maher for my knowledge of the other two papers.

2. Acceptance vs. subjective probability

I am using 'believe that P' to mean 'regard P as highly likely'. My topic, then, is the impossibility of voluntarily raising the probability that one accords to a proposition. That is not the same as voluntarily accepting a proposition.⁴

It is not unreasonable to use 'belief' to stand for acceptance. The word is so used by those who have most strongly emphasized the difference between acceptance and high subjective probability. Here is a reason for using the term their way. Suppose the contrary: belief is a matter of high subjective probability; for example, to believe that P is to give it at least a 0.95 chance of being true. Then consider a fair lottery with a thousand entrants. What is your subjective probability for 'This entrant will not win the lottery' as applied to each entrant? It had better be 0.999. So, according to the present view, you believe, of each entrant, that he or she won't win; you also believe that these are all the entrants; yet it would be unreasonable for you to believe that no-one will win. This seems to be a case where it is utterly reasonable to believe that P, which obviously and elementarily entails Q, yet it is utterly unreasonable to believe that Q. That is incredible. A natural response to this is the thought: 'Do I really believe that Smith won't win? Perhaps I don't. I think he has a vanishingly small chance of winning, but his not winning isn't something I accept, adopt, make my own, stand behind.' That is the thought that acceptance should be distinguished from high subjective probability, and that 'believe' and its cognates should go with the former. Peter van Inwagen, who has helped me in my thinking about this,

suggests that subjective probabilities are just a species of acceptance: to regard P as likely is to accept the proposition that probably P.

I am sure that it is right to distinguish acceptance from high subjective probability, and it seems right to tie the word 'belief' to the former. I shall not do so, however. The thoughts I have to offer concern subjective probabilities, and I want 'belief' as a convenient word for them. If van Inwagen is right that the two relate as genus to species, they probably stand or fall together so far as (in)voluntariness is concerned, but this is not a topic I have explored.

One's acceptance of a proposition seems to be a state or condition that one can be in, but it can be associated with one or more kinds of intellectual act—affirming P in one's heart, turning one's thoughts away from P's chances of being false, throwing one's weight behind P in one way or another.

Now, many philosophers have seemed to say that belief can be voluntary, but I don't think they have been contradicting the presupposition of the present paper. It seems fairly clear that what they have been treating as voluntary is an intellectual act—often an inner speech act—that is associated with acceptance. Newman, for example:

Assent is an act of the mind, congenial to its nature; and it, as other acts, may be made both when it ought to be made, and when it ought not. It is a free act, a personal act for which the doer is responsible⁵. . .'

Similarly, Descartes says that we have 'a kind of freedom that enables us always to refrain from believing things which

are not completely certain', which might refer to acceptance or subjective probability; but he introduces that passage with the following: 'We have free will, enabling us to withhold our assent in doubtful matters. . .', which is clearly about some intellectual act rather than about belief.⁶ Chisholm writes:

'Belief' may. . . refer to a psychological act—to the fact that a man accepts a proposition, or takes it to be true. In this active sense of 'belief', a belief. . . may be called wrong and in much the same sense as that in which a criminal act may be called wrong.'⁷

I am not sure what it is that Chisholm here treats as voluntarily and potentially culpable, but I don't think it is the assigning of subjective probabilities, which is my topic in this paper.⁸

3. Williams's first reason

Bernard Williams asserts that 'it is not a contingent fact that I cannot bring it about, just like that, that I believe something, as it is a contingent fact that I cannot bring it about, just like that, that I'm blushing'; and he offers two reasons why this is so.⁹

The first of them, Williams says, 'is connected with the characteristic of beliefs that they aim at truth. If I could acquire a belief at will,' he continues, 'I could acquire it whether it was true or not. If in full consciousness I could will to acquire a "belief" irrespective of its truth, it is unclear

⁵ J. H. Newman, *An Essay in Aid of a Grammar of Assent* (Westminster, Md.: Christian Classics, 1973), p. 232.

⁶ René Descartes, *Principles of Philosophy* 1: 6.

⁷ Roderick Chisholm, 'Lewis's Ethics of Belief', in *The Philosophy of C. I. Lewis*, ed. P. A. Schilpp (La Salle, Ill.: Open Court), 223–242, at p. 223.

⁸ 'I found the Newman and Chisholm references in Louis P. Pojman, 'Believing and Willing', *Canadian Journal of Philosophy* 15 (1985) 37–55, at p. 38. Pojman also cites Thomas Aquinas, *Summa Theologica* Part II, Q4, Art. 2; William James, 'The Will to Believe', in *Essays in Pragmatism* (New York: Hafner, 1969); Jack Meiland, 'What we Ought to Believe or the Ethics of Belief Revisited', *American Philosophical Quarterly* 17 (1980).

⁹ Bernard Williams, 'Deciding to Believe', in his *Problems of the Self* (Cambridge University Press, 1973), 136–151, at pp. 148–51.

that before the event I could seriously think of it as a belief, i.e. as something purporting to represent reality.' If this were offered as a free-standing reason why beliefs cannot be immediately induced, it would fail. Before the event, one can think: 'I shall get myself to believe that P; the belief will be false, and I shall be deluded, but that is what I want.' This acknowledgment of falsity does not prevent me from thinking of the aimed-at state as one of belief.

Williams might agree, for he seems not really to trust this line of thought, and proceeds with a cautious variant on it:

At the very least, there must be a restriction on what is the case after the event; since I could not then, in full consciousness, regard this as a belief of mine, i.e. something I take to be true, and also know that I acquired it at will. With regard to no belief could I know—or, if this is to be done in full consciousness, even suspect—that I had acquired it at will. But if I can acquire beliefs at will, I must know that I am able to do this; and could I know that I was capable of this feat, if with regard to every feat of this kind which I had performed I necessarily had to believe that it had not taken place?

We are invited to answer No, but I answer Yes. Here is why.¹⁰ Credam is a community each of whose members can be immediately induced to acquire beliefs. It doesn't happen often, because they don't often think: 'I don't believe that P, but it would be good if I did'. Still, such thoughts come to them occasionally, and on some of those occasions the person succumbs to temptation and wills himself to have the desired belief. (Sometimes he merely wants to be the centre of attention and to amuse others. Someone who has no skill as an actor can instantly start to behave exactly as

though he believed that P, by coming to believe that P. It is fun to watch it happen.) When a Credamite gets a belief in this way, he forgets that this is how he came by it. The belief is always one that he has entertained and has thought to have some evidence in its favour; though in the past he has rated the counter-evidence more highly, he could sanely have inclined the other way. When he wills himself to believe, that is what happens: he wills himself to find the other side more probable. After succeeding, he forgets that he willed himself to do it.

The community have tried inducing paid volunteers to immediately acquire beliefs that have previously been deeply and radically at odds with their belief systems, and they have always failed. The trick cannot be worked if the protective forgetfulness would require that the rest of the person's beliefs be drastically rearranged.

After successfully willing himself to have a certain belief, a Credamite may later get evidence that that is what he has done; e.g. someone may tell him. Then he either rejects the evidence (e.g. disbelieves the informant) or else accepts that he has willed himself to have the belief, and usually loses the belief because of his knowledge of how he got it.¹¹ In the latter case, he still remembers having had the belief, and now remembers willing himself to acquire it. The belief itself suppressed the memory of how it was caused; when the belief went, the memory emerged. So each Credamite knows that he sometimes wills himself to believe something, even though it is never true that he *now* has a belief which he *now* remembers having willed himself to acquire.

There is my case for answering Yes to Williams's question. I do not believe that my story is possible because I do not think that one could immediately induce beliefs. But the

¹⁰ For a more detailed discussion of this argument of Williams's, see Barbara Winters, 'Willing to Believe', *Journal of Philosophy* 76 (1979) 243–256.

¹¹ Usually—but sometimes he retains the belief because since voluntarily acquiring it he has encountered evidence for it.

story satisfies Williams's demands; it is possible, for all he shows to the contrary; so his first reason does not explain why voluntary belief is impossible.

Furthermore, the supposed reason applies equally to any acquisition of belief, immediate or mediated, by means other than getting reasons that support it. When he says that I must not suspect that I have willed myself to believe something, his reason for this does not depend on my having acquired the belief 'just like that' but only on my having done it in some way other than acquiring evidence for it. But Williams holds that there is no conceptual obstacle to the mediated case, as in hypnotism or brainwashing: he objects to it morally, but thinks it coherent. Thus, his remarks imply the impossibility of something that he rightly thinks is possible.

4. Williams's second reason

Here is Williams's second reason:

Another reason stems from our considerations about perceptual belief: a very central idea with regard to empirical belief is that of coming to believe that P because it is so, that is, the relation between a man's perceptual environment, his perceptions, and the beliefs that result. Unless a concept satisfies the demands of that notion, namely that we can understand the idea that he comes to believe that P because it is so, and because his perceptual organs are working, it will not be the concept of empirical belief. . . . But a state that could be produced at will would not satisfy these demands, because there would be no regular connexion between the environment, the perceptions and what the man came out with, which is a necessary condition of a belief. . . .

That is the whole of what Williams has to say on this score. (The omitted portions strengthen what is claimed, but add nothing to the argument for it.) From here he goes on to discuss the limited possibility, and reprehensibility, of *arranging* to acquire beliefs by the use of drugs, hypnotists, and so on.

The main premiss of this argument needs editing. The thesis cannot be that every empirical belief is held because it is true, for there are false beliefs. So Williams must be saying only that it is a conceptual truth that empirical beliefs *can* be acquired in ways that are explained by their truth. The word 'empirical', taken strictly, reduces this to the trivial point that immediately induced beliefs, just because they do not arise out of sensory experience in the standard truth-tending way, should not be accounted 'empirical'. I think, though, that Williams is using 'empirical', as philosophers sometimes do, alas, to mean 'contingent': his thesis concerns the concept of *belief about how the actual world is*.

With the premiss thus construed, I agree with it. It is a conceptual truth that (i) sensory encounters with the world are the standard, typical source for beliefs about how the world is, and (ii) in such an encounter the belief that P can arise from sensory intake from a state of affairs in which P—which is to say that the person believes that P because it is so. I would defend (i) on the basis of the broad rightness of the functionalist analysis of the concept of belief, according to which the concepts of belief and desire have as their essential role the explanation of how a creature's behavioural outputs relate to its sensory inputs. No sensory inputs, no analysis, and so no work for the concept of belief. That does not give much centrality to (ii), but I don't think it is central in Williams's argument either, though it is perhaps covertly involved in his mention of a 'regular connexion' between 'the environment' and resultant belief.

What bearing does (i) have on whether beliefs can be immediately induced? That depends on what version of (i) we adopt. The version I would be willing to defend says that contingent beliefs must usually, centrally, and basically be acquired as a systematic result of sensory input from the environment. That makes no trouble for the possibility of immediately inducing belief. It does imply that immediately induced beliefs would have to be atypical or non-standard, but it does not rule them out. A state acquired in this way could count as a belief because of its relations—as exhibited in behaviour—with the creature’s desires and its other beliefs. This does not kick off a vicious regress: most of the beliefs count as such because of their relations not to other beliefs but to the sensory inputs that cause them. But some beliefs may be acquired in other ways.

Williams thinks otherwise. He relies on the stronger thesis that contingent beliefs must always be acquired in that way: a state that ‘could be produced at will’, he says, could not be a belief because ‘there would be no regular connexion’ between instances of it and the person’s sensory intake; and this has to mean that there would be no strictly, exceptionlessly regular connexion etc. That presupposes a version of (i) which does indeed conflict with the possibility of immediately induced belief, but I see no reason to accept (i) in that version of it. (All this applies equally to (ii), *mutatis mutandis*.)

My final point about Williams’s first reason also applies to his second. The latter relies on the claim that it is part of the conceptual essence of belief that beliefs are always rooted in perceptual evidence, and the fact that if beliefs could be acquired voluntarily those roots would sometimes be cut. But this threatens to prove too much: if it holds against beliefs acquired voluntarily ‘just like that’ then it holds against beliefs acquired in any way that doesn’t involve

gathering evidence, e.g. through hypnotism. If Williams is right in holding the latter to be possible, his second reason cannot explain why immediately induced belief is impossible.

5. Sharpening the problem

Each of Williams’s explanations of why one cannot immediately induce belief fails because it gets no help from ‘immediately’. If either of them did its intended job, it would also explain why one cannot induce belief in a mediated way; but there is no such fact to be explained.

Furthermore, there could be simpler, quicker, more reliable means for causing beliefs in people without giving them evidence. I passionately want to spend the evening in a state of confidence that the weather will be fine tomorrow (I have my practical reasons), so I give myself the thought of tomorrow’s weather being fine while snapping my fingers in a certain way, and sure enough I end up convinced that the weather will be fine tomorrow. We have no such fast, reliable techniques for producing belief without evidence, but they are not conceptually ruled out.

So we don’t leave the realm of abstract possibility if we think of people as being able to give themselves beliefs by snapping their fingers or downing a pill. If there is a conceptual impossibility, as Williams and I think there is, it comes between being immediately induced to *believe that P* and being induced to *do something as a means to believing that P*; the latter, however swift and sure the means-end link is, involves a mediated inducement to believe that P. The thinness of the line between those two makes my problem look harder than ever; but when the line was drawn in the wrong place the problem looked easier but was really insoluble.

6. The basic triangle

The sought-for explanation must have something to do with the basic triangle: what a creature (1) does depends on what it (2) thinks and (3) wants. A full grounding of the concepts of belief and desire require this triangle to be developed into a square: (2) beliefs and (3) desires mediate between (4) sensory inputs and (1) behavioural outputs. Item (4) briefly raised its head in section 4. above, but from now on I can set it aside and stay with the simpler triangular story.

Basically and initially it is a story about beliefs of the form: *Doing A is a way to get G*. Beliefs on other topics must be understood as adjuncts to, or means to, or upshots of, beliefs about means to ends. Whatever the grounds for saying that a creature somehow contains the information that P, it doesn't believe that P unless that information is manifested, through its beliefs about means to ends, in its behaviour.

There is a useful formula that was presented by Braithwaite in a seminal paper that reintroduced into twentieth century philosophy the insights of Alexander Bain's that had been neglected and forgotten.¹² The wording, but not the idea, is mine:

x believes that $P \equiv x$ is disposed to behave in ways which would (or will) satisfy x's desires if it were (or is) the case that P.

This, satisfactorily, covers not only beliefs about means to ends (Swimming is a way to get food) but also ones that have some bearing on the former (The prey went into the water). The corresponding formula for desire is this:

x desires that $P \equiv x$ is disposed to behave in ways that x thinks will bring it about that P.

Each analysed concept occurs in the analyses for the other. That is inevitable, which is why the functionalist literature proclaims that we must explain belief and desire together, through how they collaborate to explain behaviour.

The above formulae only approximate to analyses of the concepts of belief and desire; neither is true as it stands. To bring them closer to the truth we must qualify and complicate them in many ways—for example by playing off strength of desire against degrees of belief (levels of subjective probability). None of those details matter here, however. All I need is that the above two formulae capture the main lines of how the concepts of action, desire, and belief relate to one another. If you reject that, I have nothing positive to offer you concerning belief and the will.

7. Inducement and desire

The triangle prompts a question about whether desires are involuntary as beliefs are, and if so how that might be explained. I shall follow this up for a section, in the spirit of Hume's advice that when a search is proving difficult one should 'beat about the neighbouring fields'. I want the two searches to be as closely isomorphic as possible. Just as we have been examining the inducement of any change in the subjective probability one assigns to a proposition, so now I want to explore the inducement of any change in how strongly something is desired. Belief and desire are both matters of degree, and my topic is the will's relation changes in the degree to which something is believed or wanted.

¹² R. B. Braithwaite, 'The Nature of Believing', first published in 1932 and reprinted in A. P. Griffioths (ed.), *Knowledge and Belief* (Oxford University Press, 1967). Alexander Bain, *The Emotions and the Will*, first published in 1859 and re-issued by University Publications of America: Washington, DC, 1977. See especially pp. 568–573 in the reprint.

There is no problem about the mediated inducement of desires. In explaining why, I have to distinguish 'end desires' from 'means desires'; I have a means desire for something if I want it only as a means to something else that I want; I have an end desire for something if I want it for itself and not because of what I think it might lead to. I shall take the two sorts of desire in turn.

One's end desires can change—witnessing an execution turns a man against capital punishment, a life-threatening illness shifts his relative valuing of time and of money—and someone might undertake such a process in order to get an end desire that he doesn't now have.¹³ Lucrezia, for example, doesn't care about the welfare of others, but she comes to think that people who do care are more serene than she is, have richer friendships and like themselves better. She therefore sets out to become like them—actually to have the welfare of others at heart, really to want it as an end desire, nor merely to act as though she wanted it. If she achieves this, her individual good deeds will be motivated by altruistic concerns—her wish for the good of others will be an end desire—and yet she will have come to this point because of an inducement, a belief about means to ends. There is a certain piquancy about this situation, but no paradox. It is simply a case where an inducement works in a mediated way on someone's end desires, closely analogous to the mediated ways of getting someone to acquire a belief.¹⁴ Mediated inducements to acquire means desires are not only possible but in many cases easy. Suppose you were crazy enough to offer to reward me for *coming to want to put my hand into this bucket of mud* as means to something, I could easily

arrange to acquire this desire by rigging the situation so that putting my hand into the bucket is a means to something that I want.

So we come to the question of whether one could immediately induce desires, which I take to be the question of whether desire is voluntary. It seems intuitively right to say that it is not, and that there is something conceptually wrong with the idea of coming to desire that P, 'just like that', because one sees advantage in desiring that P. Of course I can immediately get you to want to raise your arm by convincingly offering you a reward for *raising* it but not by offering you a reward for *wanting to raise it*. Or so I say, expecting agreement, on the strength of my intuitive sense of how things stand. But here as with belief it would be good to have an argument for this conclusion, preferably in the form of an explanation of why the conclusion is true.

Here is an unsound argument for saying that an end desire could not be immediately induced:

- (i) If someone is induced to ϕ , then for some R he ϕ 's so as to make R obtain; that's what inducement is. It follows that if someone is induced to desire that P, then *for some R he desires that P so as to make R obtain*. (ii) If someone's desire that P is an end desire, then *there is no R such that he desires that P so as to make R obtain*; that's what an end desire is. The two are formally inconsistent. What better proof could there be of the absolute impossibility of inducing an end desire?

If this argument worked, it would work even against the mediated inducement of end desires; but the latter is possible,

¹³ 'The will can contribute indirectly, as though from a distance, to make something pleasing to us.' G. W. Leibniz, *New Essays on Human Understanding*, trans. and ed. P. Remnant and J. Bennett (Cambridge University Press, 1981), pp. 182–3.

¹⁴ Locke, in the *Essay*, gives a structurally similar example: knowing that smoking is good for their health, people doggedly train themselves to like tobacco! Any change in likes and tastes is a change in one's end desires.

so the argument must be defective. Its fault is that it trades on the ambiguity in 'He desires that P so as to make R obtain'. It could entail that he thinks (a) that his wanting P will lead to R, or (b) that P will lead to R. These are independent propositions; neither entails the other; they bring out the radical ambiguity in the sequence in question. In the above argument, (i) is true only in the sense of (a), whereas (ii) is true only in the sense of (b). So the argument fails.

The immediate inducing of means desires can be dismissed as a separate topic for investigation. Suppose that on the stroke of noon I come to want to put my hand into a bucket of muddy water so as to retrieve money from the bottom of it, this being a means desire that I didn't have at 11.59:59. This implies that at noon (i) I think that putting my hand into the bucket is a way to get money from it, and (ii) I want to get the money. To acquire that means desire, therefore, I must have acquired either (i) the causal belief or (ii) the end desire for the money. So the means desire could not be immediately induced unless beliefs or end desires can be immediately induced; so the immediate inducing of means desires has nothing new to offer us and can be dealt with automatically through the discussion of our other two topics.

First objection: 'Your new desire for the money might itself be a means desire.' True, but that doesn't block the overall line of thought if it is otherwise all right. Here is the revised version that shows how the objection is met: My new means desire must come either from (i) a new causal belief relating hand and bucket to money or from (ii) a new end desire for the money or (a new means desire for the money, arising either from (i') a new causal belief about money in relation to something else X or from (ii') a new end desire

for X or (a new means desire for X, arising either from (i'') a new causal belief about X in relation to something else Y or from (ii'') a new end desire for Y or (a new means desire for Y. . . .)). This has to terminate in either a new belief or a new end desire, so my argument finds its way through this labyrinth to the conclusion that I have drawn.

Second objection: 'It could be that all morning you wanted the money and thought you could get it by delving into the bucket, but not until noon did that belief and desire generate a downright desire to put your hand in the water. If that were so, the acquisition of the means desire would not involve acquiring either a belief or an end desire.' Yes it would. The case as presented is not intelligible unless before noon the positive case for reaching into the bucket is outweighed by other considerations (e.g. I have another project that demands clean hands, and I have no facilities for cleaning my hand after putting it into the bucket), and if at noon there was no change in my belief about hand-to-money or my desire for money, then there must have been a change in how I viewed the rival project; and the latter change must have involved a change of belief or a change of desire (or a change in my relation to some third project, with this arising from. . .). So the second objection fails. We really can set aside, as not a separate topic, the immediate inducement of means desires.

Coming back then to end desires: Could someone be immediately induced to want something? I have criticized one bad argument for answering No, and have offered no other. Still, I stand by my intuitively based conviction that that is the right answer, as it is to the corresponding question about the inducement of beliefs. It is reasonable to suspect, indeed, that the conceptual facts that generate one answer

¹⁵ They are more different than is commonly realized, I think, See Jonathan Bennett, 'Analysis Without Noise' (section 6), in Radu Bogdan (ed.), *Mind*

have a considerable overlap with those that generate the other. Enormously different as belief and desire are from one another,¹⁵ there are abstract structural similarities between them—and symmetries in how they relate to one another and to behaviour—that suggest that if they are both necessarily involuntary it is for similar reasons in the two cases.

8. Beliefs and desires as dispositions

The basic functionalist triangle confirms something that has been accepted by many philosophers, not all of them functionalists. H. Price, for example, wrote: ‘Believing a proposition is. . . a disposition and not an occurrence or “mental act”, though the disposition is not necessarily a very long-lived one. . . .’¹⁶ That seems right. Beliefs are not actions or events or processes; rather, they are sets of mind that are theoretically describable in statements about what would be the case about the believer if such and such came to be the case. A belief is a function from desires to actions. (Strictly: a total belief state is a function from desires to actions. A single belief is a function from desires and other beliefs to actions.)

Lately this view about beliefs has run into trouble on the grounds that beliefs cause behaviour whereas dispositions cannot be causes. I deny that beliefs cause anything, and suggest that the contrary opinion arises from a misunderstanding of the undeniable fact that attributions of beliefs help to explain behaviour. Are the explanations causal? That question is not determinate enough, just as it stands, for me to have an answer to it, but really it doesn’t matter. Nobody,

I presume, would deny that facts about a thing’s disposition can enter into causal explanations of its behaviour. I realize that I am being rather dogmatic about this matter; it’s a big topic, and I defend my position elsewhere.¹⁷

So I contend that to believe that P, for any P, is to have a certain complex behavioural disposition—more specifically, a function from desires to behaviour. And desires are dispositions too: a desire is a function from beliefs to actions. Many desires are accompanied by categorical processes, some of them all too painfully present to consciousness, but they are not of the essence of desire.

Is that a clue to the solution to my problem? A structural feature that is shared by beliefs and desires promises well as the source of the involuntariness of each, and the promise is confirmed by the fact that we have no trouble believing in the voluntariness of all the kinds of intellectual act that can be associated with acceptance (see section 2). When we try to make good on the promise, however, things are not easy.

Here is an argument starting from the premiss that dispositions supervene on categorical (= nondispositional) states. This implies that one can create a disposition only by altering a subvenient categorical state—making something fragile by altering its crystalline structure, making it edible by cooling it, and so on. To give someone a belief, then, one must alter some categorical state of him upon which his having of that belief supervenes. So the inducing of a belief cannot be immediate, and must go through the inducing of an action—a change of categorical state—upon which the change of belief supervenes. End of argument.

and Common Sense (New York: Cambridge University Press, 1990).

¹⁶ H. Price, ‘Belief and Will’, *Proceedings of the Aristotelian Society*, suppl. vol. 28 (1954), 1–26, at p. 15.

¹⁷ Jonathan Bennett, ‘Folk Psychological Explanations’, in John D. Greenwood (ed.), *The Future of Folk Psychology* (Cambridge University Press, 1991), pp. 176–95.

The argument is open to the objection that it neglects the distinction between motivational and merely ontological mediation which I introduced in section 1. I illustrated it through events that are causally required for some action to be done, whereas a change of categorical state doesn't cause the change of disposition that supervenes on it. Still, the relation between them is ontological—the world gets from the inducement to the change of belief through the subvenient categorical change—and it is a further question whether the person must aim to produce the subvenient change as a means to changing his beliefs. It has not been shown that he must. That is, it has not been shown that the categorical basis for a disposition yields not merely ontological but also motivational mediation. So the fact that inducements can change beliefs only in a motivationally mediated manner still waits to be explained. End of objection.

The objection itself overlooks a certain distinction. Let us compare

(i) the relation between an arm's being raised and the neural events that this involves

with

(ii) the relation between a change of disposition and the categorical change on which it supervenes.

I have mentioned one similarity, namely that both are ontological, and one unlikeness, namely that (i) is causal whereas (ii) isn't. Another difference is this: It is not absolutely necessary that there be any neural or other preliminaries to an arm's going up, whereas every disposition absolutely must have some categorical basis, though it is a contingent truth that a given disposition has the specific basis that it does have. The first half of that is familiar as implied by broadly Humean considerations that have been widely

accepted. The second half may need more comment.

I contend that although it could happen that two categorically indiscernible things were dropped in water and one dissolved while the other didn't, it couldn't be that the former was soluble while the other wasn't. This is because the statement that a thing *x* is soluble says something of the form

If *P* were the case, *x* would dissolve,

which is true only if at all the closest *P*-worlds *x* does dissolve. Just what closeness is, for purposes of the theory of these conditionals, is not yet quite settled; but there can be no doubt that closeness somehow depends upon categorical similarity. We have here a richly promising line of analysis according to which dispositional differences must rest upon categorical ones, and I confidently accept that conclusion.

It follows that there could not possibly be such an event as *just* acquiring a belief. There are worlds at which, when someone raises her arm, her arm goes up without any neurological or other antecedents, but there are no worlds at which someone acquires a belief without undergoing a subvenient change of state. So the statement that when Henry acquired a belief that is all that happened is impossible, very much as it is impossible that two unlike things should become alike without either of them changing categorically. I did not say 'change intrinsically', for there is the theoretical possibility that what someone believes might change because of changes not in his intrinsic nature but rather in how he is situated in the world.¹⁸ It is hard to imagine someone setting out to believe that *P* by so altering his situation that one of his present beliefs comes to qualify as the belief that *P*, but if such a thing did happen, under inducement, this inducement would obviously be working in a tremendously

¹⁸ In an earlier draft I did say 'change intrinsically', John Hawthorne warned me.

mediated way. So from now on I shall ignore that possibility and attend only to changes of disposition that supervene on changes of intrinsic state.

We have absolute necessity here, but not equivalence. If *x* and *y* became alike because *x* lost its radioactivity, the proposition that they became alike is not equivalent to the proposition that *x* lost its radioactivity, for the obvious reason that although the becoming-alike must supervene on some intrinsic change it need not have involved that one in particular. Similarly, attributing a disposition to a thing is not the same as attributing to it the disposition's categorical basis, because although the disposition had to have some categorical basis it didn't have to have that one.

Some may want to say that a particular disposition token is identical with its categorical basis. In my opinion most uses of 'token' in phrases like 'state token' and 'belief token' are vehicles for nonsense and covers for unthorough thinking. But I need not go into all that. My topic is what one can be induced to do, and thus what one can aim or try to intend or plan to do, and what one aims etc. to do is never to bring about a particular event but only to bring it about that there occurs an event of a certain kind.

So I stand by my claim that when someone is induced to acquire a belief, there must be something else that she is also induced to do. It is absolutely impossible for the belief to be acquired unless the person has a change of categorical state, but there is no such change *C* such that for her to be induced to believe that *P* is for her to undergo *C*. These are distinct states of affairs, yet one of them absolutely cannot obtain except as a consequence of the other or of some substitute for it.

9. The false answer

That, I thought for a while, is why belief cannot be immediately induced. The line of thought went as follows.

In a strictly analogous way, I cannot immediately induce you to make two dissimilar things alike. It is absolutely, conceptually impossible for you *just* to make them alike; you have to make them alike *by*—through the mediation of—making an intrinsic change in one or both of them. You may even be able to do this without knowing in what ways they are unlike. In such a case, you will not aim directly at the basis for their becoming alike—that is, the intrinsic changes that will make them alike—but rather at something that will cause those changes. For example, you might set about making two documents alike by implementing a software program that compares them and edits one of them to match the other.

Similarly with inducing beliefs. There is no reason in principle why some virtuoso of bio-feedback should not be able to focus his attention directly on the neural basis for acquiring the belief that *P*; but none of us can do this, for our own or other people's beliefs, and probably none of us ever will. What we have to do is to cause ourselves and others to have beliefs by aiming at something that will cause the belief by causing its neural basis: we subject the person to testimony or other evidence, or to brainwashing, hypnotism, drugs or the like, thus getting ourselves causally upstream from the neural basis and thus from the belief itself. In my account of the supervenience of beliefs upon intrinsic states, there is nothing to imply that to produce a belief in someone (myself or another) I must know anything at all about its neural basis. All that is implied is that I must know of something that is ontologically prior to the acquisition of the belief and must bend my efforts to the production of that.

Whether it is prior by *being* or by *causing* the neural basis for the acquisition does not matter. What does matter is that it must be one or other of those. I cannot bend my efforts directly onto the acquisition of the belief

This explanation of the involuntariness of belief has several *prima facie* virtues. (1) It is specifically about the *immediate* inducing of beliefs, and does not threaten to run wild and imply that beliefs cannot be acquired in any way except through the collection of evidence. (2) It applies to ‘the acquisition of beliefs’ when this is understood as meaning ‘changes in subjective probabilities’. It does not apply to any intellectual acts whatsoever, e.g. to acts of acceptance or inner affirmation, if such there be. (3) In section 5 I contended that there could be extremely fast, simple, reliable ways for mediately inducing beliefs, and commented on how thin the line is between that sort of mediation and downright immediacy, remarking that that made my problem look harder than ever. But that harder-seeming problem is just what my explanation solves. The explanation says nothing about the length, difficulty or uncertainty of any means for producing beliefs; its whole point concerns the difference between changes of disposition (which do need a basis) and categorical changes (which don’t). The line may be thin, but it is clear, sharp and absolute. (4) The explanation works equally well for the involuntariness of end desires. Indeed it can be generalized to argue that one could not be immediately induced to make *any* counterfactual conditional true, a result which I find intuitively acceptable.

10. Three objections to it

For all its seeming virtues, the explanation is a failure, as I now realize. Five of my friends have independently produced what amounts to the same objection.¹⁹ I now state it in my own terms.

Whenever someone is induced to make it the case than an E event occurs, the person’s thought must be in some degree abstract. I may, aim to raise my arm in some fairly specific manner, but I cannot aim to raise it ϕ ly where ϕ is a complete specification of an arm-raising; the eventual, actual raising of my arm must have details that were not represented in my intention. So, with any voluntary or immediately inducible kind of action there is the same supervenience pattern that I have been emphasizing. The arm’s going up supervenes, with absolute necessity, on the arm’s going up in some specific fashion; and yet ‘The arm goes up’ is not equivalent to ‘The arm goes up ϕ ly’ for any completely specifying ϕ , because the arm could go up in any one of countless different specific fashions. So I have committed myself to the conclusion that there cannot possibly be such an activity as *just* raising one’s arm, and the further conclusion that one cannot be immediately induced to raise one’s arm. This is offered as a *reductio ad absurdum* of my explanation, and as such it resoundingly succeeds

Perhaps my explanation was too parsimonious in its conceptual resources. Perhaps there is a sound explanation along those general lines—one that retains the *prima facie* virtues that I noted in section 9 but relies on some further fact about (a) how a disposition relates to its categorical basis. It would have to be something that distinguished (a) not only from (b) the relation of an arm-raising to its neural cause but also from (c) the relation of an arm’s being raised

¹⁹ John Hawthorne, Mark Lance, Mark Brown, Frances Howard, and Alastair Norcross.

to its being raised in some perfectly specific manner. There may be something to work on here: In the case of (b) and of (c) the subvenient and supervenient items are of the same metaphysical kind, whereas in (a) they are not. But I have no way of firming and clarifying that, nor any suggestions for how it might help to answer my question.

Furthermore, any attempt along those lines would have to steer clear of a further objection that has been brought against my proposed explanation.²⁰ The route from an inducement to a change of disposition must be ontologically mediated; that much is agreed. I have argued that the 'must' expresses absolute necessity, and my explanation infers that therefore the mediation is not just ontological but motivational. It seemed plausible at the time, but now that it has been challenged I can find no reason to believe it. Motivational immediacy has to do with the agent's thoughts, whereas the absolute necessity of a categorical basis for a disposition resides in a conceptual connection which need not constrain the agent's thoughts because he may be unaware of it or outright disbelieve it. The premiss that *everybody must realize that* dispositions must have a categorical basis would offer some hope of securing my conclusion, if only the premiss weren't patently false. This second objection seems to threaten not only the explanation that it was aimed at but also any variant of it which stays at the level of conceptual and metaphysical connections.

If we leave that level, we presumably walk out on the project of explaining my explanandum by appealing somehow to some relation between a disposition and its categorical basis. A reason has been offered to me for thinking that if we can't get the explanation out of that we can't get it at all.²¹ If something has to mediate between an inducement

and the acquisition of a belief, and if the mediating item is not the belief's categorical basis, then it must be or involve a mediating event. But it cannot be absolutely necessary that *any* event mediates between the receiving of the inducement and the acquisition of the belief. It is always theoretically on the cards that receiving the inducement should lead in an ontologically (and therefore motivationally) immediate way to the acquisition of the belief. Let S_1 be the state of wanting to believe that P because one has just been offered a reward for believing it, and let S_2 be the state of believing that P. It may be biologically impossible, but it cannot be absolutely, conceptually impossible that someone should be wired up in such a way that S_1 immediately causes S_2 . That is ontological immediacy, which entails motivational immediacy (or if it doesn't, I don't see why).

That threat to my explanandum can be countered. In explaining how, I am not mainly trying to protect my intuitive hunch that belief is essentially involuntary: the mere defence of intuitions is a trivial activity, and anyway this hunch of mine is not particularly strong. The importance of fending off the threat is that there is something to be learned from doing so.

I agree that the following is not conceptually ruled out:

A. Wanting to believe something could immediately cause one to believe it.

But I question whether that implies that beliefs could be acquired voluntarily. The view that it does follow presumably comes from the underlying view that

B. Doing something voluntarily is being immediately caused to do it by one's desire to do it,

or something close to that. Because I reject B, I can find no plausible way of getting from A to the voluntariness of

²⁰ By John Hawthorne, Frances Howard and Mark Lance.

²¹ This objection came from Mark Lance and Frances Howard.

belief. B has been popular through the centuries, but it seems pretty clearly to be mistaken. If it were right, the following wish would be unintelligible:

I wish that my arm would rise right now without my raising it, going up simply as an immediate consequence of my wanting it to go up.

This seems to me perfectly intelligible, and I offer that as one way of seeing that whatever we mean by doing something voluntarily it is more than, or different from, its happening

as an immediate consequence of wanting it to happen.

Still, the threat is valuable. It warns us of further complexities in our concept of voluntary conduct. The question of whether or why belief is essentially involuntary may be unanswerable until those further complexities are understood.²²

Syracuse University,
Syracuse, NY 13244-1170, USA.

²² My debt to friends who have helped me with this work goes well beyond what is indicated in the previous footnotes. All of them gave me other help also, enabling me to improve several sections of the paper. I am truly grateful for their assistance.