THE JOURNAL OF PHILOSOPHY

VOLUME XCIII, NO. 9, SEPTEMBER 1996

FREEDOM IN BELIEF AND DESIRE*

People ordinarily suppose that there are certain things they ought to believe and certain things they ought not to believe. In supposing this to be so, they make corresponding assumptions about their belief-forming capacities. They assume that they are generally responsive to what they think they ought to believe in the things they actually come to believe. In much the same sense, people ordinarily suppose that there are certain things they ought to desire and do and they make corresponding assumptions about their capacities to form desires and act on them. We chart these assumptions and argue that they entail that people are responsible and free on two fronts: they are free and responsible believers, and free and responsible desirers.

In the first section, we characterize some assumptions that people make about one another and about themselves within what we call *the conversational stance.* Drawing on this characterization, we go on in the second section to describe people's specific assumptions about their capacities as believers. In the third section, we extend the characterization to encompass the assumptions people make about their capacities as desirers. In the fourth section, we show that these assumptions about the formation of belief and desire commit people to a belief in their responsibility. And then in the fifth and final section, we connect this belief in responsibility with a belief in freedom. We argue that to be responsible in desire is to hold your desires

* We would like to thank John O'Leary-Hawthorne, Galen Strawson, and Susan Wolf for their very helpful comments. We are also grateful for comments received when the paper was presented to the Australasian Association of Philosophy, University of New England (1995), and at the Human Action and Causality Conference, University of Utrecht (1996), and, as well, when it was read to staff seminars at the Australian National University, the University of Hong Kong, and the University of Caen.

0022-362X/96/9309/429–49

freely—if you like, to enjoy free will—and that to be responsible in belief is to hold your beliefs freely: to enjoy free thought.

The ideals of free will and free thought are not usually connected and we are conscious of offering a relatively novel view of their relationship. As we see things, philosophers have taken a unified perspective on freedom to be unavailable because of an unfortunate tendency to overstate the differences between belief and desire. The paper is part of a more general project of undermining this unfortunate tendency.[1]

I. THE CONVERSATIONAL STANCE

One of the most striking things about human beings is conversation, in particular conversation conducted to intellectual effect. People do not set out just to form their own intellectual beliefs and then inform others of them. They listen to one another in the course of belief formation and they invest one another's responses with potential importance. They are prepared often to change their own minds in the light of what they hear from others and, if they are not, then they usually feel obliged to make clear why they are not and why indeed the others should alter their views instead.

It is true, of course, that most human exchange is not primarily intellectual in character. Conversation is the means whereby we recognize others and seek recognition from them. It is the forum in which we tell our jokes, confess our antipathies and form our friendships, coax and persuade and flatter in the furthering of our ends, and, in general, ring changes on the basic themes that engage the human sensibility. But for all that conversation achieves in these respects, it is also in some part a forum in which we put our beliefs on the line and expose them to the reality test that others represent for us. It often assumes an intellectual character.

Conversation of an intellectual kind is such a common feature of everyday life that it is easily taken for granted, but such conversation involves assumptions that are actually rather remarkable.[2] First, people assume that they each form beliefs or judgments and that these beliefs bear on common questions. This appears in the fact that people balk at any perceived discrepancy between their respective attitudes: they take the discrepancy to signal that someone is in the wrong. Second, people assume that they are each authorities worth listening to, even if the likelihood of error varies from individual to

[1] See our "Backgrounding Desire," *Philosophical Review,* XCIX (1990): 569-92, and "Practical Unreason," *Mind,* CII (1993): 53-79.

[2] Pettit, *The Common Mind: An Essay on Psychology, Society and Politics* (New York: Oxford, 1993; 2nd ed., with new postscript, 1996), ch. 4.

individual. No one commands or expects to command universal deference, and no one gets dismissed out of hand or expects to get dismissed out of hand. And third, people assume that when they differ in their judgments, a review of the evidence commonly available can usually reveal who is in the wrong and thereby establish agreement; they assume that good evidence, if there is good evidence available, will serve to put the mistaken ones right.

This recourse to evidence will not always be successful in resolving differences. But even then it is striking that people do not happily acquiesce in the existence of the discrepancy. They make the auxiliary assumption that there is probably a certain sort of explanation available. They judge that one or the other does not have access to all the evidence—that the evidence is not equally available on all sides—and try to put that right. Or they judge that the available evidence, or even all the evidence possible, leaves the difference between them unresolved—it is not good enough to constrain belief uniquely—and that, within certain limits, no one is blocked from going his own way. Or they judge that those who dissent are misled by something like inattention or illogic or just laziness of mind. By taking one or another of these views, people are saved from having to conclude that those who dissent are out of their minds and not worthy of attention: that they are not even presumptive authorities. They may be driven to the out-of-their-minds conclusion as a last resort but the default position is more optimistic.

A final feature of conversation is that not only do people make all these assumptions—the three basic assumptions together with this auxiliary assumption—they apparently each accept that the assumptions are a matter of common belief. Each person believes them, each person believes that everyone else believes them as well—or at least no one disbelieves that everyone else believes them[5]—and so on. That everyone believes them shows up, as indicated, in their responding appropriately: they balk at every discrepancy, look for a resolution, and try to explain any failure to achieve it. That everyone believes that everyone believes the assumptions shows up in the fact that no one is surprised at anyone's responding in that way. That everyone believes that everyone believes that everyone believes them shows up in the fact that no one is surprised that no one is surprised at anyone's responding in that way. And so on.

[5] David Lewis, "Languages and Language," reprinted in his *Philosophical Papers, Volume 1*, (New York: Oxford, 1983), p. 166.

Do many conversations have the intellectual character that engages these assumptions? It is true that people often only go through the motions of pretending that there is a common subject matter under intellectual discussion with others, or that others are really worth listening to on that subject matter, or that they have anything to say. But we take it that the very fact of such pretense is itself testimony to the effect that everyone has a notion of what properly intellectual conversation is and that they all assume that it at least occasionally occurs. People do genuinely, if only intermittently, reach out toward others and seek a meeting of minds with them: they authorize their interlocutors and in turn assume authorization by them.

Conversation in the sense characterized need not involve different people in exchange at or over the same time. As someone makes up his mind about what to believe on some matter, conscious that he will return to the topic again, or as someone reflects on what he came to believe earlier, assessing the worth of the reasons that moved him, he enters into a sort of conversation with himself. He takes it, whether at the earlier or later moment, that there is a common content at issue; that neither self can spurn the voice of the other; that any discrepancy ought to be subject to resolution; that if it is not, then that is probably due to limited evidence or a local failure on one or the other side; and that these are matters of common belief between his different, interlocuting selves.

This characterization of intrapersonal conversation is borne out, not by appeal to introspection, but by reflection on the assumptions implicit in the ways people conduct their thought. An earlier self will always balk at the prospect of a later discrepancy, as of course the later self will balk at the experience of such discrepancy with the past: the prospect of later discrepancy may even give an earlier self reason now to rethink commitments.[4] We see in evidence here the assumption on each side that there is a common content addressed by the two selves: or addressed, if you prefer, by the same person at different times. Again, in face of perceived discrepancy, neither the earlier nor the later self defers to the other, or dismisses the other out of hand; and so there is also an assumption of shared authority at work. Moreover, to go to the assumptions relating to resolution, the earlier self will try in anticipation, or the later self in recollection, to come to a common mind in the light of common evidence or, fail-

[4] Bas C. van Fraassen, "Belief and the Will," this JOURNAL, LXXXI, 5 (May 1984): 235-56; and Richard Holton, "Deciding to Trust, Coming to Believe," *Australasian Journal of Philosophy*, LXXII (1994): 63-76.

ing that, will look for some contingent explanation of the expected or actual lack of consensus. And, finally, that these various assumptions are endorsed by earlier and later selves is a matter of common belief between the two: they chime with their expectations, and their expectations about expectations, in the required manner.

II. BELIEF

It only makes sense to adopt the conversational stance in relation to someone—yourself or another person—if three conditions are satisfied: first, there are norms relevant to the issue of what she ought to believe; second, she is capable of recognizing this to be so; and third, she is capable of responding appropriately to the norms: that is, capable of believing in the way she should.

Consider the first condition. There are belief-relevant norms that apply, by everyone's lights, to any conversational interlocutor. When you ascribe beliefs to someone you identify certain propositions as the contents of those beliefs. You say that what she believes is that eucalypts are evergreen or that Princeton is not in England or that ripe tomatoes are red or whatever. So far as certain propositions are the contents of the subject's beliefs in this way, it is right for her to maintain those beliefs in certain circumstances, or in certain apparent circumstances, wrong for her to do so in others. This follows, at least under plausible assumptions, from the fact that beliefs are representations of the subject's environment. Furthermore, it is right for the subject to form certain new beliefs, wrong for her to form certain alternatives, depending on what follows from what: it is right for her to believe that this eucalypt is evergreen, that this English town is not Princeton, that this green tomato is unripe, and so on.

The general lesson is that when you ascribe beliefs you assume, in effect, that certain belief-relevant norms apply to the performance of the believer. There are norms that govern what the subject ought to believe in the presence of certain facts. And there are norms, therefore, that govern what the subject ought to believe in the light of the corresponding evidence—the apparent facts—so far as that evidence is determinate. Are the norms that you postulate in this way objectively valid? Yes, in at least one sense. Such norms are not like the conventional norms of behavior that might be recognized in a given club or circle: they are not norms, such that subjects can be imagined deciding whether to embrace them or not, depending on their contingent attitudes or alignments. To be a thinker who believes in certain determinate contents is to be subject to norms like: believe that *p* if and only if *p*; believe that *p* if and only if all the evidence points to *p*; and believe that *p* when *p* is entailed by some of the things you be-

lieve, and is not inconsistent with anything else you believe! The relevance of such norms—such evidential norms, as we will call them—is *inescapable.*

Consistently with acknowledging this you can, of course, admit the possibility that your own view of the relevant norms is inadequate or distorted. For all anyone can guarantee, there may be reasons for casting evidential norms quite differently from the way they are habitually cast. There may be reasons, for example, for revising the received view on what follows from what, or what supports what: standard views on deductive and inductive logic may be mistaken. To think that there are objectively valid norms of evidence is not necessarily to pretend to be infallible on the matter of what those norms are. It is only to hold that believers fall willy nilly under norms of evidence, and that the norms in question apply across the different groups to which the subjects belong. They do not segregate on lines of culture or class or gender.

Consider now the second condition. Does the conversational stance suppose that a conversational interlocutor recognizes evidential norms? Someone will recognize such norms if she has beliefs with contents of the form: it is true that *p* or it is false that *p*; the evidence supports the hypothesis that *p* or is against the hypothesis that *p*; *q*, and the fact that *q* implies or entails that *p*; and so on. To believe that a certain proposition—a certain potential belief content—is true or is supported by the evidence or is entailed by something that is itself accepted is to believe, in effect, that it is right to believe the proposition, wrong to disbelieve it: it is to believe that there are norms that require the attitude, at least when other things are equal.

We think that in order to attract and sustain authorization, the conversational interlocutor must manifest beliefs involving notions like truth and support and entailment and that in this sense she must recognize certain belief-relevant norms. When you authorize someone in conversation, you hold her to the expectation that she will balk at discrepancies between the two of you, and do so in a way that invites the ascription of beliefs like the following: that you each have different belief attitudes toward the same content, that the evidence available may rule out one of those beliefs as unsupported or false, and that attention to the evidence may reveal which, if either, of the beliefs should be given up. If an interlocutor failed to live up to that expectation, if she failed to manifest any notion of there being a common content of belief or a common fund of evidence, for example, then you would have no reason to take her attitudes seriously; you would have no reason to invest her responses with any au-

thority. But this is to say, then, that in authorizing someone in conversation, you treat her as recognizing certain familiar norms for beliefs. You take her to see that for any content believed there may be evidence for or against the proposition, and that the state of this evidence may make it right or wrong to entertain the belief.

The third condition is that people not only be disposed to recognize the demands of evidential norms in their own case but that they also be disposed to respond to those registerings: that they be disposed to maintain beliefs that comply with the norms and to reject beliefs that fail to do so. What would happen if, on being challenged about a certain belief—say, in virtue of a discrepancy between the two of you—an interlocutor was disposed to examine the belief for its compliance with evidential norms but was not disposed to maintain or reject the belief, depending on what the examination revealed? Suppose the person was unrevisably committed to her beliefs, for example, or that examination of her beliefs occasioned an arbitrary pattern of maintaining or rejecting the belief in question. What would happen then?

In such a case you would, once again, have no reason to invest such a person with the authority of a conversational interlocutor. You might use them, in the way you might use a clock, as a prima facie check on your beliefs. But you could not assign to her the sort of role that you must expect a conversational interlocutor to fulfill. You could not treat her as a subject such that it may well be possible to achieve the resolution of any discrepancy that appears between you. You would not treat her as a subject such that the failure to achieve such resolution is a serious challenge for each of you to face. There will be no possibility of resolution, if she is not disposed to respond to what she registers in her own case as the demands of evidential norms.

This is not to say that the subjects we invest with conversational authority must be cognitive saints or god-like creatures. As already noticed, there are all sorts of obstacles, by the light of our conversational assumptions, that may stop our partners from responding on some particular occasion in the way they know and we know they should. Some of these obstacles are seen as disabling, others as not disabling. We may think that though someone failed to respond on a given occasion—say, through inattentiveness—further conversational pressure would have brought her around. Such nondisabling obstacles are readily countenanced. But we may also think that when other sorts of obstacles are in place, conversational pressure can do no good; on the topic affected—say, by bias—or over the period af-

fected—say, by passion—the person is disabled as a conversational partner. Such obstacles are difficult to countenance since they force us to regard the interlocutor as unworthy of being taken seriously on the topic, or at the time, in question. But still they can be countenanced, provided they are sufficiently insulated from the person's general performance.

To sum up, then, authorizing a subject as a conversational interlocutor makes sense only if there are certain norms governing what that subject ought to believe, the subject is disposed to recognize those norms, and she is disposed to respond in the way required. Whether it be yourself or someone else, authorization involves postulating that the subject has a variety of sophisticated belief-forming capacities.

III. DESIRE

We turn now to the assumptions that people make about their capacities, not in forming beliefs, but rather in forming and acting on desires. Our argument is that these capacities are assumed to be on a par with belief-forming capacities.

Desire-forming capacities will be on a par with belief-forming capacities so far as they fulfill three conditions: first, there are norms governing what agents should desire and do at any moment; second, agents are capable of recognizing these demands; and third, agents prove generally responsive in their desires and actions to the impact of the norms they recognize. Insofar as people converse, not just about more or less theoretical questions, but also about practical matters—about what it is right or wrong, good or bad, rational or irrational, sensible or stupid, to do in a given situation—we think that they are more or less bound to treat one another—and, of course, themselves—as satisfying the three conditions mentioned.

The first condition is that there are norms governing what the agent should desire and do. Why does conversing with someone about practical matters—why does conversationally authorizing him in this role—presuppose that he satisfy this condition? Imagine a difference with someone about whether it would be rational or irrational for someone—say, the interlocutor—to act in a certain way. Such a difference often has the very same characteristics as a disagreement about any ordinary matter of fact. The pair of you balk at the conversational discrepancy and seek out ways in which you might resolve your evaluative difference. You assume that both of you cannot be right, and you assume further that a careful weighing of your reasons for your different judgments will reveal which of you is right, which of you is in error.

When you take yourselves to be concerned in this way with facts that support or undermine the claim that an act is rational or irrational, you treat it as a matter of fact—a matter of agreed fact—that there are norms available to govern the things that the agent should desire and do. While the norms countenanced may leave certain choices open—while they may be permissive of various differences—you are agreed that this is not one of those cases. And by your lights, the business of seeing what the norms require in such a case—the business of deciding whether it is rational or irrational for the agent to act in a certain way—is a serious enterprise. It is not one of just sorting out your respective feelings, for example, nor one of just manufacturing a mutual accommodation, nor anything of the kind. As you practice it, practical evaluation amounts to a world-directed enterprise of sifting out fact from fiction; a matter of trying to determine what, in light of the facts, the agent is required to desire and do.

Of course, it is one thing to argue that practical evaluations make claims that can be justified by appropriate reasons, quite another to give an account of the metaphysical underpinnings of this idea. Should we be antirealists, or quasirealists, or realists with regard to the contents of our practical evaluations?[25] In the present context, we can afford to be ecumenical. Different theorists are free to give the idea of a practical evaluation their preferred treatment. All that our argument requires is that the accounts they give be consistent with the truism that practical evaluations are indeed conversationally interrogable. We return to the issue of metaphysical underpinnings later, but for the moment we may let it pass.

So much for the claim that people presuppose that there really are norms governing desire and action. The other two claims we need to defend are that you have to presuppose with any interlocutor in a conversation about what he ought to do—yourself included, of course—that he is disposed to recognize the demands of those norms and that he is disposed to respond appropriately. These claims do not have to hold true for absolutely every desire. Just as otherwise perceptive and responsive believers may be subject to certain disabling obstacles, so otherwise perceptive and responsive desirers may be subject to similar constraints: they may be the victims

[5] For quasirealist and antirealist accounts, see Simon Blackburn, *Spreading the Word* (New York: Oxford, 1984); and Allan Gibbard, *Wise Choices, Apt Feelings* (Cambridge: Harvard, 1990). For realist accounts, see Smith, *The Moral Problem* (Cambridge: Blackwell, 1995); Frank Jackson and Pettit, "Moral Functionalism and Moral Motivation," *Philosophical Quarterly*, XLV (1995): 20-40.

of certain restricted fetishes and obsessions, for example, and they may be susceptible to certain disabling moods or passions. And even where they are not so disabled, they may be subject to obstacles that give way only with efforts at self-control or under conversational challenge.

Once we recognize these escape clauses, we can have little hesitation in agreeing that when you authorize someone in conversation about matters of practical evaluation, and in particular on matters that bear on what he ought to do himself, you must treat him as being disposed both to register those demands and to respond appropriately. You will expect him to change his evaluations in light of the evidence, and his desires in light of his evaluations.

Were you to think that your interlocutor lacked the dispositions to register and respond to the demands of the norms governing evaluations that you both countenance, and lacked them even in the provisoed measure allowed, you would either have to put his evaluative understanding or commitment in serious question or you would have to regard him as something close to a zombie or psychopath. How could the interlocutor agree that doing such and such is irrational, so you will ask, but not see that the prescription applies to him? Or, if he does admit that it applies to him, how could he fail to adjust his desires and actions accordingly? In particular, how could he fail to do these things, when the failure is not to be explained by reference to familiar obstacles? The only answer available would seem to be that he is not seriously or sincerely involved in the business of practical evaluation, or that if he is, then he is not reliably attuned to the practical values in question. In either case, you lose solid grounds for authorizing him as a conversational interlocutor. You must cease to see any point in conducting a conversation that is supposed to bear on how he should behave.[6]

The upshot is that to take someone as a serious conversational partner on questions of value, and in particular on questions of what

[6] The argument of this paragraph requires us to assume that there is an internal connection between evaluative judgment and the will. Externalists will not allow this assumption. Unfortunately, this is not the place to provide a full-scale defense of internalism (but see Smith, ch. 3, and "The Argument for Internalism: Reply to Miller," forthcoming in *Analysis*). Suffice it to say that, since the account of freedom in the sphere of desire and action which we go on to characterize assumes internalism, externalists will be unable to accept the account. They are therefore bound to see a huge gulf between freedom in the sphere of belief, and freedom in the sphere of desire and action, where we see a continuity. Whether this provides yet another argument against externalism presumably depends on the plausibility of the externalist's account of freedom in the two spheres.

he ought to do, is to take him to recognize norms governing his desires and actions and to be disposed to register and respond to those norms in his own case. But we cannot stress sufficiently that this is not necessarily to regard him as a paragon of insight and virtue. You can authorize him fully—authorize him fully in domains where there are no disabling obstacles—and yet see him as seriously mistaken on a wide variety of matters: seriously mistaken in some of his evaluations. Here, we recall the similar point made in the last section. And equally you can authorize him fully and recognize that he may often be slow to see the implications of his evaluations for his own behavior, or that when he does see the implications, he may be slow to bring his own desires and actions into line. You may see your interlocutor, for example, as someone who can only manage to bring his behavior into line by resort to direct or indirect methods of self-control: methods of coping with weakness of will and the like.[7] The important point is simply that you cannot fully authorize a conversational interlocutor on matters of value, and see him as someone with whom to discuss what he ought to do, while expecting to find him only randomly sensitive to the demands of the values. You must be able to see him as having a reliable disposition, however limited and unspontaneous, to track the values in his evaluations, and to link his evaluations to his self-prescriptions and so to his choices and actions.

An example of Gary Watson's[8] illustrates the point. Watson describes a man who suffers an ignominious defeat in a game of squash. As a result, he finds himself wanting very much to smash his opponent in the face with his racquet. Let us stipulate, with Watson, that by the man's lights, there is nothing whatsoever to be said in favor of acting on this desire. Even the satisfaction he would feel if he were to smash his opponent in the face counts for nothing with him, given that the satisfaction would have been obtained by inflicting undeserved harm. He wants to hurt his opponent, but can provide no justification for doing so. To the extent that you authorize such a person in conversation, and see some point in discussing what he ought to do with him, you must take him to be capable of recogniz-

[7] See our "Practical Unreason," *Mind*, CII (1993): 53-79; Jeanette Kennett, "Mixed Motives," *Australasian Journal of Philosophy*, LXXI (1993): 256-67; Kennett and Smith, "Philosophy and Commonsense: The Case of Weakness of Will," in Michaelis Michael and O'Leary Hawthorne, eds, *Philosophy in Mind* (Dordrecht: Kluwer, 1994), pp. 141-57; and our "Brandt on Self-control," in Brad Hooker, ed., *Rationality, Rules and Utility* (Boulder: Westview, 1994), pp. 33-50.

[8] "Free Agency," this JOURNAL, LXXII, 8 (April 24, 1975): 205–20, reprinted in Watson, ed., *Free Will* (New York: Oxford, 1982), pp. 96-110.

ing the demands his own values make upon him: to be capable, that is, of seeing that smashing his opponent in the face with his racquet would be a completely unjustifiable thing to do. And you have to assume further that, having recognized that fact, he is capable of responding appropriately in his desires and actions. He may find it impossible to resist smashing his opponent in the face if he stays on the court, but you must suppose him to be capable of removing himself from the scene and settling himself down, or something of that sort. Were you to reject either of these assumptions then you would weaken the grounds for taking him seriously as a conversational interlocutor on relevant matters of value. Certainly, you would remove any point from conversing with him about what he in particular ought to do. In the old phrase, you might as well be talking to the wall.[9]

To sum up, then, authorizing someone as an interlocutor on certain questions of practical evaluation makes sense only under three conditions: first, there are relevant norms governing desires and actions; second, the interlocutor has the capacity to recognize the demands of these norms; and third, the interlocutor has the capacity to respond appropriately to the demands he recognizes. In other words, the person's desire-forming capacities satisfy conditions that parallel the conditions that you take his belief-forming capacities to satisfy.

IV. RESPONSIBILITY

We have argued in the last two sections that human beings treat themselves as possessed of belief-forming and desire-forming capacities that satisfy three conditions. There are norms governing what

[9] The fact that there are norms governing what people should desire and do, norms that they have the capacity to recognize and respond to, is thus implicit in what we have elsewhere called the *deliberative* perspective on human agency: see our "Backgrounding Desire"; "Practical Unreason"; Smith, "Valuing: Desiring or Believing?" in David Charles and Kathleen Lennon, eds., *Reduction, Explanation and Realism* (New York: Oxford, 1992), pp. 323-60; *The Moral Problem.* Human beings do not just house desires to do this or that, desires that come and go without invitation or welcome. They can and do often deliberate about which action to choose and, in the course of their deliberations, they invariably make evaluations of this or that object of desire. They ask whether the course of action desired is right or wrong, good or bad, rational or irrational, sensible or stupid, or whatever, and, insofar as they function properly as rational deliberators, their answers to these questions have an impact upon what they desire and do. They do not just *observe* that acting in a certain way would be rational or sensible or whatever, they at least sometimes do what they do *because* they judge it to be the rational or sensible thing to do. Thus, to suppose that the squash player in Watson's example did not have the capacity to recognize and respond to his own values would be to imagine someone who is incapable of effective deliberation.

they should believe and desire; they are capable of recognizing the demands of those norms in their own cases; and by and large they are capable of responding appropriately, if only via stratagems of self-control, to those demands.

The message in each case can be restated in the language of responsibility. People who engage in conversation suppose that it is appropriate to address a certain 'ought' to their interlocutor, on three grounds. There is a relevant norm in play, as the first condition has it; the interlocutor can see what the norm requires, as the second condition makes clear; and the interlocutor can do what the norm prescribes, as the third condition spells out. The message is that as you see someone in the role of a suitable conversational interlocutor, you cannot help but see her as the addressee of certain 'oughts' and, by the same token, as the realizer of certain 'cans'. You must ascribe the 'can' of perception involved in the second condition, and the 'can' of performance associated with the third: you must see the interlocutor as capable both of recognizing and of responding to the norms.

To see someone in this way is to see her as a responsible subject in the relevant domain. The interlocutor whose beliefs are engaged is depicted as someone who can be made to answer to the norms of evidence governing what is the case. And the interlocutor whose desires and actions are in question is depicted as someone who can be made to answer to the sorts of reasons that can be offered for and against evaluative claims. The one person can be held responsible, as we say, for what she believes; the other person can be held responsible for what she desires and does.

This picture of responsible believing and responsible desiring is very different, it should be noted, from the standard picture associated with talk of what Daniel Dennett[10] calls *the intentional stance*. Under the standard image, believers and desirers may be extremely rational, being well-attuned to demands of evidence, demands of consistency, and the like. They may realize almost perfectly, for example, the Bayesian model of theoretical and practical coherence. But under that image, people can remain passive or mechanical subjects who harmonize and update their beliefs and desires in a more or less autonomic way. Such adjustments as are involved may happen within them without any recognition of why they should happen and without any efforts on their behalf to help them happen.

[10] *The Intentional Stance* (Cambridge: MIT, 1987).

The picture of responsible believing and desiring that we associate with the conversational stance suggests a very different style of attitude formation.[11] The subject may or may not be particularly rational, may or may not be particularly in tune with the demands of evidence, consistency, and the like. But the subject is certainly not a mere passive or mechanical system. She does not just revise her beliefs and desires autonomically, or at any rate, not when they operate beyond the reach of the occasional disabling obstacles that get in her way. She revises them under the spur of recognizing what the relevant norms require of her. She revises them, in particular, when that spur is applied in the interactive business of conversation, whether the conversation be with one another or with herself.

We have argued elsewhere that it may be useful to give up on the ideal of autonomy or self-rule in favor of the ideal of "orthonomy" or right rule.[12] Responsible believers and desirers are orthonomous subjects, in the sense that they recognize certain yardsticks of right belief and right desire and can respond to the demands of the right in their own case. They may vary among themselves in how far they actually conform their beliefs and desires to those yardsticks; they may be more or less thoroughly ruled by the right. But outside the domain of disabling obstacles, they are all equally orthonomous in at least this sense: they are all able to answer the call that the right makes upon them.

Orthonomy will be taken to be of lesser or greater significance, whether in belief or desire, to the extent that the nomos or rule to which people are said to be responsible is cast as being more or less objective in its standing. The point is worth remarking because, while the norms that govern evidence for ordinary matters of fact are generally taken to enjoy a robust objectivity, there are different theories about the metaphysical underpinning of values, and these theories have a differential impact on the objectivity of the norms governing desire and action. Depending on which of these theories is adopted, orthonomy in desire will be taken to be of lesser or greater significance.

The force of this observation appears when we consider the importance attaching to the notion of autonomy which Harry Frankfurt[13] defends: the notion that agents are autonomous to the extent that they act only on first-order desires by which they have a second-

[11] See also Pettit, chs. 2 and 3.

[12] Our "Backgrounding Desire" and "Practical Unreason."

[13] "Freedom of the Will and the Concept of a Person," this JOURNAL, LXVIII, 1 (January 14, 1971): 5-20; reprinted in *Free Will*, pp. 96-110.

order desire to be moved; they are not the victims, as it were, of wanton first-order desires, first-order desires that fail to attract second-order support. One objection to this view is that second-order desires need not be anything special, being inherited perhaps from childhood conditioning.[14] The idea is that Frankfurt's condition stipulates only that one sort of desire should have a controlling influence over others, and that while this condition may be realized by some agents and not by others—while second- or higher-order desires may be more effective in some than in others—this does not mark any significant difference between them. It may serve only to distinguish those with relatively powerful hang-ups from those with hang-ups that they manage often to defeat. It may serve to distinguish one pattern of internal conflict resolution—one internal state of harmony—from a different but no less eligible pattern.

To be orthonomous, as distinct from autonomous, an agent's evaluations and desires have to be sensitive to his recognition of normative requirements: reasons that may be offered in support of evaluative claims. To the extent that there are normative requirements to be satisfied, the achievement of orthonomy will therefore represent something distinct from any sort of internal harmonization; it will represent a way of coming into line with something outside the realm of desire: with the reasons in favor of the relevant evaluative claims. But some accounts of the metaphysical underpinnings of the norms governing our desires and actions may fare better than others in substantiating our image of ourselves as orthonomous subjects. Go for an antirationalist, reductive, psychological account of what values are—go, say, for a crude subjectivist or emotivist account—and the significance of orthonomy will be diminished. Coming into line with what the norms require may be only barely distinguishable from internal harmonization. Go for a more rationalist or realist account and the significance of orthonomy will be magnified. Coming into line with the norms will require either a sensitivity to rationally binding reasons or attunement with the world.

Our own preference is for a more rationalist or realist account of practical evaluations, though we differ in the details of the accounts we prefer.[15] Our own disposition is therefore to take practical orthonomy as enjoying great importance: the sort of importance, in-

[14] See Watson.

[15] Smith, *The Moral Problem*; Pettit and Jackson, "Moral Functionalism and Moral Motivation."

deed, that we see in theoretical orthonomy. But others may take a different view on this matter, consistently with sharing our belief in practical orthonomy, and we do not want to contest the issue here. For them, the sort of responsibility people have for their desires and actions will lack the significance of the responsibility they have for their beliefs. For us, the two are on a par.

V. FREE WILL AND FREE THOUGHT

We want to show why the account of responsible believing and desiring defended in the previous section is, precisely, an account of how responsible believers and desirers can enjoy freedom in the matter of what they believe and of what they desire and do. Responsibility or orthonomy in belief means that people enjoy free thought. Responsibility or orthonomy in desire means that people enjoy free will.

Freedom in the sense associated with free will is traditionally defined in terms of the ability of the agent, for anything they do, always to have done otherwise.[16] A believer or desirer would be free in this sense to the extent that no matter what he believes or desires, he is such that he could always have believed and he could always have desired otherwise. Freedom in such an unqualified sense—if, indeed, it deserves to be called "freedom" at all—would not be particularly attractive from our point of view. If an agent believes or desires rightly according to the evidence and the values, then there will be nothing attractive in itself about being such that he could have believed or desired otherwise. Believing or desiring otherwise will simply be a matter of his getting it wrong, and so doing much worse than he actually did. The ability to have believed or desired otherwise will be something inherently attractive from our point of view only so far as it is the person's ability for anything that is not rightly believed or desired always to have believed or desired otherwise. We argue that responsible believers and desirers are free in the sense of having this ability.[17]

Before looking at the argument, we have one short comment on the conception of freedom just mentioned. While freedom is a person's ability, in the event of getting things wrong, to get them right, the question of whether someone believes or desires freely arises in the case where he get things right as well as in the case where he gets

[16] O'Leary-Hawthorne and Pettit, "Strategies for Free-will Compatibilists," forthcoming in *Analysis*.

[17] We therefore find much to agree with in Wolf, *Freedom within Reason* (New York: Oxford, 1990).

things wrong. Suppose a person believes or desires rightly: the norms require, or, due to underdetermination, allow that belief or desire. He may do so out of brute luck, in the sense that he would have had that belief or desire no matter how the world was or seemed to be. In that case, he does not hold the belief or desire freely even though he believes or desires rightly. Alternatively, he may believe or desire rightly, not out of pure luck, but in such a way that, were the belief or desire wrong, still he has the ability in such an event to get it right. In such a case he does hold his belief or desire—the belief or desire that is actually right but might have been wrong—freely. To hold a belief or desire freely is to hold it in the presence of an ability, should the belief or desire be wrong, to get it right. The question of whether someone believes or desires freely thus arises both for the case where he gets things right and for the case where he gets things wrong.[18]

Now let us turn to the argument that responsible believers and desirers possess the ability associated with freedom. Suppose that you conversationally authorize someone in a given domain, taking him to be a responsible believer in regard to those topics. The fact of conversationally authorizing the person gives you certain expectations; indeed, these expectations are part of what it is to authorize him. You expect that, whether or not he in fact believed in accordance with the evidence in the domain in question, were he to be challenged about the demands of the evidence then he certainly would do so. Perhaps he would not do so immediately, because of some surmountable obstacle, but he would do so under sustained conversational pressure. The expectation means that, if you find that your interlocutor does not believe in accordance with the evidence, as you see things, then you must believe that were you to challenge him with that evidence, he would come to adjust his beliefs appropri-

[18] Wolf claims that freedom is asymmetrical: "being psychologically determined to perform good actions is compatible with deserving praise for them, but being psychologically determined to perform bad actions is not compatible with deserving blame" (*op. cit.*, p. 79). Our position in the text is that people only deserve praise for getting things right if their getting them right is governed by an ability which in the nature of things will not be displayed in the actual world, where they get things right, but only in the possible world in which they get things wrong. Presumably someone attracted to the asymmetry view of these matters (a view in the spirit of Wolf's remark) will say that what is important is not that that ability is present in the actual world but that it would come to be present in such a possible world. We think that this position gives praise where something else is due: something like congratulations on the good fortune that these people enjoy. But someone who sticks with the asymmetry view can still go along, of course, with the rest of our argument.

ately. In entertaining this expectation, so we claim, you display the assumption that the interlocutor has the ability for anything wrongly believed—or at least for anything wrongly believed by available lights—to believe otherwise.

When will an interlocutor have the ability, for anything wrongly believed, to believe otherwise? Following a well-established tradition, we postulate that two conditions are going to be necessary and sufficient: first, the interlocutor would believe otherwise in the event of it being impressed on him that he believes wrongly, and second, the possibility that this is impressed on him is suitably accessible: it is not a possibility that could only be realized, for example, via a total transformation of his nature.[19] You are bound to think that your interlocutor meets both of these conditions. So far as you continue to authorize the person in conversation, you have to think that he would come around to the right belief in the event of your pressing him with the demands of the evidence. And so far as you regard yourself as an agent who can realize at will the options you discern, in particular the option of challenging your interlocutor with the evidence, you have to think that the possibility of pressing the demands of the evidence on him is suitably accessible. Thus, you have to think that for anything that the interlocutor wrongly believes, or at least wrongly believes by available lights, he is capable of believing otherwise.

This means, in our sense, that you have to believe of the interlocutor that in the domain in question, and at least so far as available lights go, he is a free believer or a free thinker. His beliefs do not just come and go in a natural procession of events. His beliefs are subject to an ability on his part that is characteristic of being free. The beliefs come and go in a manner that is consistent with the person's being able to get anything right that he happens to get wrong: or at least that he gets wrong by lights that are available to him in conversation.

Indeed, not only do you have to believe of any interlocutor that you authorize, including yourself, that the person has the sort of ability described: the ability to adjust to evidence that you are in a position to produce. Since you recognize the interlocutor as himself capable of challenging you with the demands of evidence, you must also believe that he has a *self-starting* version of the ability in question. You must suppose that he has the ability to adjust, not just to evi-

[19] See, for example, Roderick Chisholm, "Human Freedom and the Self," reprinted in *Free Will*, pp. 24-35.

dence that you are in a position to produce, but also to evidence that he will often be in a position to produce himself. The interlocutor does not necessarily depend on you, or on any other, for enjoying the freedom of thought that he is presumed to have insofar as you adopt the conversational stance. He may have it by grace of the conversational community he establishes with himself.

We have argued that to acknowledge someone as a responsible or orthonomous believer, as we do in conversationally authorizing a person, is to see him as possessed of free thought. An analogous argument will show that to acknowledge someone as a responsible or orthonomous desirer, as we do in conversing with him about what he ought to do, is to see him as possessed of free will. You cannot hold out any values as matters to which your interlocutor is answerable without thinking of him in the image of a subject who can get any desires right that by available lights he actually gets wrong. You have to envisage that in the suitably accessible event of your drawing his attention to the demands of the values, he will come to adjust what he desires and does. And you have to envisage, indeed, that he will often be in a position himself to play the role here allotted to you: he will often be in a position to require and promote such an adjustment in himself.

The fact that you cannot think of an interlocutor, or indeed of yourself, as a responsible believer or desirer without postulating freedom of thought and will does not in itself establish the fact of such abilities. Perhaps you cannot fail to think in this way, and yet this way of thinking is mistaken. We acknowledge that this possibility is logically open. We acknowledge that people may be in massive error when they take one another to be responsible believers and responsible desirers and so to be possessed of free will and free thought. But if people were to embrace that possibility they would have to adopt a wild and self-defeating stance on one another and on themselves. They would have to discount everything they must assume in order to practice conversation, and relate more broadly in an interpersonal fashion.[20] Indeed, since thinking itself is a kind of intrapersonal conversation, as we saw earlier, they would have to discount everything they must assume in order to practice conversation with themselves: everything they must assume in order to think.

We are therefore happy to embrace the conclusion that people rightly treat one another, and they rightly treat themselves, both as

[20] Peter Strawson, "Freedom and Resentment," reprinted in *Free Will*, pp. 59-80.

responsible and free believers and as responsible and free desirers. The conclusion is inscribed in habits of thought that we can scarcely imagine anyone being prepared to give up. The conclusion, moreover, looks capable of being reconciled with what the sciences have taught us about the natural world. We see no reason, in principle, why we might not each be purely physical, even deterministic, systems and yet it be true that we are individually capable of responding to the call of the right. Freedom of will and thought, as we have characterized and supported it, looks capable of being realized in quite nonspooky subjects.[21]

We would like to discuss one more issue, in conclusion. Free thought is a matter of just as much interest from our perspective as free will. It is a matter for congratulation not just that we are free in the formation of desire and the performance of action, where indeed we are free, but also that we are free in the formation of belief. But the prevailing orthodoxy is that the sort of freedom we enjoy, when we enjoy freedom of thought, has nothing whatsoever to do with the sort of freedom we enjoy when we enjoy freedom of the will. How can it be that the tradition has left us with such divergent understandings of free thought and free will? The answer we propose is this. On at least many occasions where a person fails to exercise the ability associated with free will, that failure will be manifest to them, but nothing of the same kind holds for failures to exercise the ability associated with free thought.

Imagine that your beliefs run counter to what evidence and fact require. In such a case, your beliefs will not allow those requirements to remain visible because the offending beliefs themselves give you your sense of what is and your sense of what appears to be. You are therefore denied an experience whose content is that you are believing such-and-such in defiance of the requirements of fact and evidence. This is why, as G. E. Moore[22] observed, you cannot simultaneously think that while you believe that *p*, yet it is not the case that *p*.

[21] What are we to say about the nature of free choice? What we have said so far is that a subject's beliefs and desires are free to the extent that they are the product of an ability, in the event of his being wrong, to get them right. The natural position for us to take is therefore that a subject's choices are free to the extent that they are the product—product, no doubt, "in the right way"—of beliefs and desires that are themselves free. Note that the claim that beliefs and desires are free in our sense is not equivalent to the claim that they are freely chosen, on pain of an infinite regress familiar in discussions of free will.

[22] In P.A. Schilpp, ed., *The Philosophy of G.E. Moore* (Evanston: Northwestern, 1942; 3rd ed., Open Court, 1968), p. 543.

Now, the same is true on the evaluative side, of course, to the following extent. Imagine that you are firmly committed to a particular evaluative claim even though the reasons available to you favor an alternative. As in the belief case, you are once again denied an experience whose content is that you have a particular evaluative commitment in defiance of the reasons available to you. But nonetheless the evaluative case is different in a further crucial respect. You may be denied any experience of evaluating contrary to reasons, but you are by no means denied the experience of desiring contrary to evaluation. On the contrary, it is an all too common experience that your evaluative commitments lead you on one path but that you go nonetheless another: the spirit is willing but the flesh is weak. Agents are aware in such an experience of what the right requires them to desire and do, at least in the light of their evaluative commitments, despite the fact that their actual desires and actions do not conform to that requirement.

Failures to exercise free will—say, through weakness or compulsion or whimsy—are matters of everyday experience, but failures to exercise free thought—say, through being careless or conditioned or subject to group pressure—are not: they are essentially elusive. We believe that this asymmetry may explain why people are keenly aware of free will, being conscious of how their will may fail, but are more or less oblivious of free thought. Not being conscious of the ways in which thought may fail, they do not recognize the ideal of thinking freely as one that parallels the ideal of free will. They do not see that as caution and self-control are necessary for achieving freedom of their will, for example, so vigilance and self-criticism are needed for attaining freedom of thought. They do not see that in this respect, as in others, freedom is one and undivided.

PHILIP PETTIT
MICHAEL SMITH

Australian National University