

The *analysanda* we are considering are of the form 'the disposition to manifest *m* under stimulus *s*'. On the approach just mentioned, to satisfy the conditional analysis in its application to some actual disposition, every possible antidote must in principle be detailed and held to be absent within the specification of the disposition's stimulus. There are all sorts of potential antidotes to a disposition, maybe limitlessly many. And so for an actual disposition it will be impossible to write down what *s* is. Even if the schematic conditional analysis is correct we cannot give literally true instances of it. This in turn raises a different question of analysis: how are we to analyse concepts like 'poisonous', 'fragile', 'soluble'? If we were to try 'x is (water-) soluble iff x has the disposition to dissolve when placed in water', as I have just indicated, there are no simple dispositions of that sort, since the stimulus clause ('when placed in water') does not mention the absence of antidotes which would prevent, e.g., a sugar cube dissolving in water. If that analysis of 'soluble' were correct, then nothing would be correctly called soluble. If things are correctly to be called soluble, as we think they are, then we must have 'x is (water-) soluble iff x has the disposition to dissolve when *s**', where the specification of *s** includes being placed in water and the negation of all possible antidotes. Is something which is soluble in this qualified way really soluble? Lewis tends to think that it is, but regards the question as essentially idle. Even if the simpler analysis were correct, it would be misguided pedantry to deny that a sugar cube is soluble. There is imprecision in our talk, and the degree to which it is permissible is the task of pragmatics to assess. This, I suspect, is a key difference between supporters of the conditional analysis and its opponents. While I take the foregoing to suggest that an implicit reference to 'normal circumstances' or some other indexical element is an ineliminable part of our characterization of dispositional concepts, Lewis regards these matters as belonging to the general pragmatic topic of what qualifications we may permissibly leave unmentioned.

University of Edinburgh

MOTIVATIONAL INTERNALISM: A SOMEWHAT LESS IDEALIZED ACCOUNT

BY MARK VAN ROOJEN

In the contemporary debate over motivational internalism about moral values, internalists have shouldered the burden of making their view defensible despite certain seeming counter-examples. While it seems plausible that there must be some

connection between values¹ on the one hand and motives on the other, we know that people are sometimes not motivated by values, even by values that they share.² Thus contemporary internalists postulate a defeasible yet necessary connection between values and motives. Typically they idealize the conditions under which an agent must be motivated by values, claiming for example that motivation must be present in all rational persons who accept the moral judgement or for whom the latter is true.³ Thus these versions of internalism will be conditional, stating that an option is valuable for an agent only if that agent would be motivated to perform it under appropriately idealized conditions.

Robert Johnson has convincingly argued that this strategy leads to its own problems. He suggests that these versions of internalism may have trouble avoiding the 'conditional fallacy'. This fallacy involves overlooking ways in which the conditions in the antecedent of the conditional expressing the analysis are incompatible with the claim under analysis. A relevant example of this sort would be where the idealization involved in setting up appropriate conditions for being motivated is incompatible with the agent's being in the conditions which ground the reasons. We could then construct a counter-example to the internalist claim by describing a situation in which the agent has a reason, where the reason depends on the agent's not being in the ideal conditions for eliciting motivation.

Johnson's argument does not stop at this point. We might be able to construct a version of internalism that avoids the conditional fallacy. But, he argues, it is more difficult to avoid the fallacy while retaining the attractions of internalism. In service of this claim he very nicely sets out (p. 59) two main philosophical merits of internalism: internalism allows reasons or values to figure in both the intentional explanation of an agent's acting and in the justification of that action.

I think that Johnson's arguments work against the internalist targets he discusses, and I think he has captured the attractions of internalism. In this paper I want to piggy-back on Johnson's arguments a somewhat different proposal for defining the central internalist claims. My proposal involves changing the specification of the conditions in which motivation must be manifest so that it is less idealized. With proper care we can specify conditions which are ideal enough to ensure motivation but which are not so ideal as to be incompatible with the grounds of an agent's reasons. Since Johnson himself suggests (p. 54) that internalism can yet be vindicated, my thesis in no way contradicts his line of argument. In fact, I shall argue that the new proposal should be preferred to those considered so far, precisely because it meets the two desiderata Johnson sets out, while avoiding the sorts of counter-examples that plague the more idealized internalist theses. Since his discussion lays out the difficulties clearly, I can re-summarize his points rather briefly.

¹ This paper is a response to R. Johnson, 'Internal Reasons and the Conditional Fallacy', *The Philosophical Quarterly*, 49 (1999), pp. 53–71. I follow Johnson in using the term 'value' here as a general term to cover, among other things, judgements about rightness and goodness.

² See D. Brink, *Moral Realism and the Foundations of Ethics* (Cambridge UP, 1989), pp. 45–60; and M. Stocker, 'Desiring the Bad: an Essay in Moral Psychology', *Journal of Philosophy*, 76 (1979), pp. 738–53, esp. p. 744.

³ M. Smith, *The Moral Problem* (Oxford: Blackwell, 1994), pp. 60ff.; and C. Korsgaard, 'Skepticism about Practical Reason', *Journal of Philosophy*, 83 (1986), pp. 1–25.

I. JOHNSON'S ARGUMENT

Johnson focuses his discussion on a schematic statement of the conditional version of motivational internalism about practical reasons. In the schema, 'A' refers to the agent, 'A+' to the agent in ideal conditions, 'ϕ' to an action, and 'C' to the conditions in which the agent should do the action:

IR. There is a reason for A to ϕ in C → A+ would want to ϕ in C.

Motivational internalism about practical reasons is a view distinct from motivational internalism about moral judgements or values, but it is often part of the justification for these other views. (IR), plus the rationalist idea that if an action is morally right then there is reason to do it, will yield

MI. It is morally right for A to ϕ in C → A+ would want to ϕ in C.

And that is a schematic statement of motivational internalism about the moral value *rightness*. Meta-ethical rationalists, who hold that moral rightness entails a reason to act,⁴ often accept (MI) precisely because they accept (IR). Any difficulties for (IR) will (if rationalism is correct) be inherited by (MI). I shall follow Johnson in focusing my discussion on motivational internalism about practical reasons of the sort represented by (IR).

Johnson (p. 59) summarizes the appeal of internalism about practical reasons: 'any account of what reasons are must make plain how, roughly, *the* reason there is for A to ϕ could be A's *reason* for ϕing. And something could be A's *reason* for ϕing only if it could feature in both (i) the intentional explanation of his ϕing, and (ii) his rational justification for ϕing.' The internalist constraint ensures that the first desideratum is met by requiring that in some ideal possible circumstances the agent would have a motive which would in turn partly explain acting on the reason. In addition it ensures that the second desideratum is met by requiring that those circumstances must be rationally optimal. If it is rational for me to be motivated to do something, then I am justified in doing it. One somewhat oversimplified way to look at this is that the connection with motivation makes reasons apt for explanation, while the idealized nature of that connection makes them apt for justification.

Troubles arise because the grounds of some practical reasons are incompatible with the conditions specified in the idealization. Sometimes we have reasons to act precisely because we are not fully informed and fully rational. So we would not have those reasons if we were fully informed and rational. Johnson provides two examples. I might now have a reason to investigate some matter precisely because I am not fully informed. If I were fully informed I would not have that reason and would be unlikely to be motivated to investigate. I might now have a reason to see a psychiatrist because I am subject to irrational delusions which cause me to think I

⁴ Relativists such as Gilbert Harman in 'What is Moral Relativism?', in A. Goldman and J. Kim (eds), *Values and Morals* (Dordrecht: Reidel, 1978), pp. 143–61, count as rationalists in the intended sense. So of course do non-relativists such as Christine Korsgaard.

am James Bond. But if I were fully rational I would not have those delusions, and hence would have no reason and no motive to see a psychiatrist. The problem in both examples is due to the nature of practical reasons themselves. Reasons are not bare facts. Rather we have reasons to act because grounds exist for acting in that way. One sort of ground we may have is to do with our being less than ideal in some way. Thus the idealization is in conflict with supposing that the grounds for the reason in question obtain.

But we cannot do without the idealization. Without it, an internalist about practical reasons runs up against examples similar to those employed by Brink against motivational internalism about moral values or moral judgements. If we require that an agent has a reason to ϕ in some circumstances only if the agent would want to ϕ in those circumstances, we do allow the conditions under which motivation is relevant to co-exist with the grounds for the reason. Thus in many cases the account can explain how a person might be motivated by just those grounds. Yet the restriction to actual conditions is incompatible with agents' having reasons in conditions where they apparently do have reasons. For example, the irrational agent has a reason to see a psychiatrist, even if the irrationality manifests itself partly by causing the agent to fail to see the need to see a psychiatrist and hence not to be motivated. Furthermore, it seems we lose the connection with justification that internalism was designed to capture. How can it justify an action that it satisfies an irrational desire? As Johnson puts it, 'this doctrine "trims reason to the size of individual motives", and by so doing, sacrifices their capacity to be a justification for acting'.³

The remaining strategy, then, is to vary features of the idealization so as to make them compatible with the grounds of the reasons in question. Johnson considers several ways of doing this. One is to require that reasons must be connected to second-order desires of our most rational selves. Formally this is expressed as

R₁. There is a reason for A to ϕ in $C \rightarrow A+$ would want A to want to ϕ in C .

The problem with this reformulation is that it leaves unexplained the justificatory status of reasons. On the original internalist accounts, ϕ ing was justified by the fact that it was what one would want to do if completely rational. But that is no longer true here. And we cannot in general assume that where one has a reason to desire to do something one also has a reason to do it. There are many cases where the reasons to desire an end are due not to the desirability of the end, but instead to the effects of having the desire.

Other attempts at reformulation run into trouble with explaining action. For example, a reformulation might require that a reason be necessarily connected with a desire of one's more ideal self to do the action in question if in circumstances such as one's own. Formally this idea could be expressed as

R₂. There is a reason for A to ϕ in $C \rightarrow A+$ would want to ϕ if he were A in C .

³ Johnson p. 62, partly quoting W. Frankena, 'Obligation and Motivation in Recent Moral Philosophy', in A.I. Melden (ed.), *Essays on Moral Philosophy* (Univ. of Washington Press, 1958), pp. 40–81, at p. 80.

This formulation would avoid the conditional fallacy, since an agent can have a desire to do something conditional on being in certain circumstances, even though those circumstances would be incompatible with retaining that desire. Yet the requirement would not be useful in explaining why A might ϕ . For were A in circumstances appropriate for satisfying that desire by ϕ ing, A would no longer have the desire.

Johnson deploys the same objection against a similar formulation proposed by Michael Smith,⁶ before considering whether some of the explanatory ability of the requirement could be rescued by coupling it with the claim that moral judgements themselves necessarily motivate. For reasons too complex to consider in detail here, Johnson argues that the attempt will not succeed in forging a general connection between reasons and motivation. Thus, he concludes, internalists have an unfinished task – to specify an account of the connection between reasons and motives which avoids the conditional fallacy and at the same time allows practical reasons both to explain actions and to justify them.

II. A SOMEWHAT MORE MINIMAL PROPOSAL.

As I indicated earlier, I agree with Johnson's argument against the internalist candidates he surveys. In fact, because I believe he has rightly characterized the appeal of internalism, I think his argument suggests that we should renew our efforts to construct an internalist constraint that avoids the problems he highlights. A strategy is suggested by the internalist candidates Johnson himself has surveyed. They occupied two extremes with respect to the extent of their idealization. One set altogether avoided idealizing the conditions in which motivation was required. These had trouble showing how reasons could justify actions. At the other extreme, the remaining proposals required full rationality and information. I suggest we try to aim between these two extremes. We might aim to idealize less, but still idealize enough to provide a justification for acting.

How could we do that? We should start by reminding ourselves that rationality comes in degrees. The people we know are more or less rational, but none is completely rational. And rationality is structured, by which I mean that it has many components. Different theories will offer different accounts of those components, but a rough approximation would include component norms governing the formation of beliefs in the light of evidence, norms for the pursuit and gathering of evidence, norms governing the choice of ends, and norms governing the choice of means to those ends. People can be in greater or lesser conformity to these norms, and some sorts of breaches may be more irrational than others. Again, different substantive accounts of rationality will put differing weights on various violations of rational norms. Be that as it may, a full theory of rationality would allow us to order people in terms of rationality from less rational to more rational. More importantly for our purposes, it would allow us to rank possible states of a single person based on

⁶ M. Smith, 'Internal Reasons', *Philosophy and Phenomenological Research*, 55 (1995), pp. 109–31.

whatever features the correct account of rationality employs. The relevant information necessary to do so would include at least information about agents' wants, goals and desires, about their beliefs, about their evidential states, about their intentions and actions, and about how all of these interact over time; but exactly how this would work would depend on the substantive account.

Given this complete ordering of states of persons, we can compare different total states which share certain features for their degree of rationality. For example, we can fix the evidence available to people, as well as certain of their beliefs, goals and desires, and ask which of the as yet unspecified features would make these people out to be most rational. With that background, we can now formulate an internalist principle that is not subject to the conditional fallacy. An agent has a practical reason to do an action on certain grounds in certain conditions, only if, *of the complete descriptions of that agent that include those grounds and circumstances*, the ones that make the agent out to be most rational and relevantly informed include a motive to do that action. In other words, we are defining the conditions that fill in the nature of *A+* for purposes of (IR) as those in which *A* has the features (beliefs, desires, goals, intentions) that ground the reason for acting, and is otherwise as rational and relevantly informed as possible consistently with that requirement and with holding *C* fixed.

This enables us to avoid the problems with the more idealized versions of (IR). For we are not requiring *A+* to be completely rational or completely informed; it was these requirements which made (IR) incompatible with the features of *A* that grounded the reason in question. Our agent who thinks he is James Bond may now have a reason to go to the psychiatrist. For, holding fixed his actions to this point, his evidence, his beliefs and his desires, it would be more rational to recognize his delusion and form the intention to see the psychiatrist than not to do so.

The proposal captures both the justificatory and explanatory desiderata that Johnson emphasizes. First, justification: according to the account, one would be motivated by one's practical reasons if one were as well informed and rational as possible consistently with the grounds of those reasons still obtaining. Thus a process of information-gathering, rational reflection and reasoning in the light of that information would yield a motive to act in a rational agent. I am assuming here that more rational agents will if possible reflect on and reason from their evidence, given sufficient opportunities. This assumption may be somewhat problematic. Perhaps a rational agent will not have time to deliberate and hence will not go through the reasoning that should lead to motivation. To rule out that possibility, it may be that we should build a requirement of adequate opportunity for reflection right into the account of the conditions under which motivation necessarily follows upon having a reason. But in any case, this internalist principle will ensure that there is a 'sound deliberative route' from the agent's current psychological set to motivation.⁷ An agent motivated after such deliberation will be justified in acting on the relevant practical reason.

⁷ The phrase comes from Bernard Williams, 'Internal Reasons and the Obscurity of Blame', in his *Making Sense of Humanity* (Cambridge UP, 1995), pp. 35–45, at p. 36, quoted by Johnson on p. 58 to emphasize the justificatory dimension of the internalist requirement.

As for explanation, that will depend, as it always does with intentional explanations, on the assumption that the agent is rational. What the account assures is that if an agent is made aware of relevant information and rationally reflects on it, he will be motivated to act as the reason would indicate in appropriate circumstances. It will be possible in those conditions for the reason to explain the agent's action. If the agent is sufficiently irrational all bets are off. But that is the fate of all intentional explanations in such circumstances. Such explanations work only against a background assumption that the agent is relatively rational.

III. SOME OBJECTIONS AND REPLIES

Perhaps I should consider some objections. A first objection might strike at the ability of reasons such as this to justify. For the account entails that a reason based on false belief is still a reason. Holding fixed an agent's false beliefs and asking what the agent would do if otherwise fully informed and rational, we often find that the agent would be motivated precisely because he has a false belief. Yet someone like Bernard Williams might object that if a person mistakenly thinks that a glass of petrol is gin, and then mixes it with tonic to satisfy a desire for a gin and tonic, it is not true that the person has a reason to drink the petrol in the glass.⁸ That is because someone in that position is not justified in drinking what is in the glass. Motivation stemming from false beliefs does not justify, and hence does not give one a reason to act. I think Williams would be wrong here. While there may be a sense of 'reason' which makes it sensible to say that the drinker thinks he has a reason but really has not, I do not think that sense is the usual or primary sense of 'reason'. An agent who satisfies the revised internalist requirement is in a state which is both explanatory and justificatory. That was what the internalist constraint on reasons was supposed to capture. The particular example of the misinformed drinker does not show that such agents have no justification for acting – that is, for drinking the petrol. Relative to this agent's beliefs there is a fairly good justification. We might note here that the drinker need not be irrational to believe that a glass of petrol is a glass of gin. He might simply lack a sense of smell. But that is not the crucial point. What is crucial is that it would be more irrational (given *otherwise* full information and time to assess it) not to be motivated to drink what is in the glass. Hence there is a perfectly good sense in which the drinker has a reason to drink what is in the glass.

Another objection might stem from doubts that this sort of internalism really avoids the conditional fallacy. There is a certain kind of case where the less idealized internalism seems to imply that an irrational agent does not have a reason to act based on what seem like relevant grounds. The case I have in mind is one in which the agent is truly incapable of being the least bit rational. My revision of the internalist account entails that this sort of agent would then not have a reason grounded in this extreme irrationality to see a psychiatrist. The purported ground for the

⁸ The example is from Williams, 'Internal and External Reasons', in his *Moral Luck* (Cambridge UP, 1980), pp. 101–13, at p. 102. Williams claims that the drinker has no internal reason to drink the petrol.

reason to act, that is, this agent's extreme irrationality, is incompatible with being motivated, because there is no possible state of the agent in which he has just this sort of irrationality and is motivated by that irrationality to seek help. Thus according even to the more modestly idealized internalism I am proposing, this sort of irrationality cannot ground a reason to act.

This upshot may seem implausible. Should not such extreme irrationality give an agent even more reason to seek help than a similar person who could be just a bit more rational, rational enough to recognize the delusion and make an appointment? Should we not reject any account that denies that this agent, because of more severe irrationality, has an extra reason to get help over and above those had by a somewhat less irrational counterpart?

Here I bite the bullet, and endorse what the modestly idealized internalism suggests: *such hyper-irrationality* gives an agent *no more* reason to seek help than a person who is just a bit more rational. But I do not think that this really causes problems for the account. Modestly idealized internalism need not deny that this person has a reason to see a psychiatrist. The extremely irrational agent has grounds of the same sort to make an appointment as a slightly more rational counterpart has, based on the fact that each shares other sorts of irrationality and on the fact that, holding these other grounds fixed, the most rational thing to do would be to make the appointment. The view is only committed to denying that alleged grounds for action which are strictly speaking incompatible with motivation really do give such agents a reason to act. Any sorts of irrationality that are compatible with being motivated can ground a reason to act. In this way, denying the hyper-irrational agent extra reasons to get help hardly seems a problem for the theory.

Suppose instead, the objector might continue, that a person has a rational defect which is in principle inseparable from its other manifestations and which in principle makes it impossible to be more rational. I do not actually think that a concrete example could be made out, because I do not see what could underwrite necessary connections between distinct manifestations of irrationality. But if such a case could arise, I think it a fair response to deny that the agent would have a reason to act based on that sort of irrationality. Practical reasons are supposed to be able to guide actions, and here, in principle, they could not. It is also worth reminding ourselves at this point of the connection between internalism about practical reasons and internalism about moral judgements. We were exploring the motivational implications of practical reasons in part to get at the motivational commitments of judgements regarding moral values. The version of internalism about values we were considering had to do with rightness. A case like this would be just the type of case where one might be tempted to say that while it might be good if the agent was motivated to act, it is not right or obligatory. If that is the right thing to say at the level of value, it would indicate that it is also the right thing to say about motivation, at least for the variety of rationalist who accepts motivational internalism about moral judgements because of a commitment to motivational internalism about reasons.

Doubtless there is more to say here, but I think that the general strategy for capturing the internalist desiderata will be robust in the face of further objections. Even though we shall no doubt need to tweak the account to capture exactly the

conditions in which to expect motivation, the general strategy of recognizing that rational and epistemic privilege comes in degrees, and of requiring motivation of agents who are as privileged as possible consistently with the ground of the reasons, is on the right track.

I have not discussed another sort of example here, one which might cause internalist principles to fall foul of the conditional fallacy for a different reason. Those examples involve goals that can only be achieved through indirection. They would require us to supplement the present account. However, I think that the strategy I have suggested does avoid the problem Johnson highlights, that the conditions of full rationality and information can be incompatible with our reasons to act. The revised more modestly idealized internalist requirement captures the two desiderata that Johnson has clearly highlighted, while avoiding the specific pitfalls he rightly cautions us against.⁹

University of Nebraska Lincoln

⁹ I am grateful to Jonathan Dancy, Stephen Darwall, Robert Johnson and Sigrún Svavarsdóttir for discussion of this topic at Johnson's presentation of his paper at the Eastern Division meetings of the American Philosophical Association in December 1998, and to Johnson for correspondence. I also owe thanks to Harry Ide for discussion of the problem and of my proposal for handling it, and to Jennifer Haley for comments on an earlier draft of this paper. Finally I would like to thank anonymous referees of *The Philosophical Quarterly* for some very helpful suggestions about the paper and the topic under discussion.