



The Doctrine of Internal Reasons

HALLVARD LILLEHAMMER

Faculty of Philosophy, University of Cambridge, Cambridge, CB3 9DA, UK

1. The Doctrine of Internal Reasons

According to advocates of internalism about reasons for action, there is an interesting connection between an agent's reasons and the agent's present desires. On the simplest version of this view, an agent has a reason to act a certain way at some time if and only if acting that way would promote his present desires. Let us call this the sub-Humean model.¹ The sub-Humean model is widely regarded as too simple on the grounds that there are adverse conditions, such as massive confusion, in which desires are irrationally possessed or acquired, thereby failing to provide reasons for action.²

In light of this, philosophers have modified internalism by supplementing the sub-Humean model with the requirement of a rational link between the present desires and rational motivation of an agent. On this conception, an agent has a reason to act a certain way at some time if and only if acting that way would promote some desire he might have after the modification of his present desires by a process of rationally sound deliberation. Let us call this the doctrine of internal reasons.³ Different defenders of the doctrine would have us construe the notion of rational deliberation in different ways. For some, the deliberation is undertaken in the agent's actual circumstances. For others, it is undertaken in some counterfactual circumstance where the agent has undergone a purge of false beliefs and psychological anomalies like depression. Furthermore, different defenders of the doctrine fix on different desires as determinants of the reasons of agents. For some, the desires in question are the counterfactual desires of agents regarding their counterfactual circumstances after deliberation. For others, the desires are the counterfactual desires of agents regarding their actual circumstances antecedent to deliberation. Finally, different defenders of the doctrine would have us construe the reasons themselves in different ways. For some, the reasons are mental states. For others, they are facts about the counterfactual desires of agents. What makes all these positions internalist is their commitment to the possibility, for any agent at any time, of a rationally sound deliberative route the outcome of which determines which reasons exist for him at that time, and which is con-

strained, either directly or via the purge of false beliefs and psychological anomalies, by the contents of the present desires of the agent. This is true even of internalists who wish to underplay the dependence of an agent's reasons on his present desires. Michael Smith, for example, requires there to be non-arbitrary convergence among the desires of agents in reflective equilibrium in order for there to be moral reasons for action.⁴ For this claim to be interesting it must be possible to imagine the desires of agents not converging after deliberation. But since the nature of the rational process envisaged by Smith is the same for all agents, and since it includes becoming aware of all relevant facts, the only potential for divergence must lie in the content of the initial psychological states of the agents, including their desires.

Internalism is attractive for philosophers who are dissatisfied with the sub-Humean model but who fail to detect standards of rational agency which obtain independently of the desires of agents. If there are constraints on rational agency beyond what agents presently desire, where can they originate if not in the motivations arising from sound practical reasoning on the basis of those desires? Even if reasoning on the basis of present desires need not take present desires as part of its content, its results might still be felt to be somehow constrained by those contents. According to an internalist, if you could not be motivated in accordance with utilitarian principles at the end of a process of sound practical reasoning, then you have no reason to be a utilitarian. Your neighbor might be motivated in accordance with utilitarianism at the end of a process of sound practical reasoning. He would then have reasons to act accordingly. Someone who thinks agents have reasons to act in accordance with utilitarian principles regardless of a development of their present desires by sound practical reasoning would be mistaken on this view. Philosophers who claim that there are such reasons are generally known as external-reasons theorists.⁵ On its weakest reading, their claim is that reasons which derive from a rationally sound development of the present desires of an agent are not the only reasons for action there could be.⁶ Agents might have reasons which derive from the desires of an agent at different times, from some the desires of a different deliberator, or perhaps even reasons which exist independently of practical reasoning as such.⁷

Most arguments against the doctrine of internal reasons depend on one or both of two questionable strategies. Some external-reasons theorists give arguments which depend entirely on the claim that internalism conflicts with common externalist intuitions about particular cases.⁸ The problem with this strategy is that it fails to convince anyone who fails to share the relevant intuitions. Other external-reasons theorists argue that some particular argument for internalism is invalid.⁹ The problem with this strategy is that it fails to show what is wrong with the internalist conclusion. It is less common to reject the doctrine of internal reasons while remaining neutral on the issue of whether there are external reasons. Such a neutral strategy would be success-

ful if it could be shown that the doctrine of internal reasons fails on its own terms as a direct consequence of its commitment to the notion of a rationally sound deliberative route. If the existence of a rationally sound deliberative route entails the existence of rational requirements on action which are not constrained by the contents of the present desires of agents, it follows that reasons which depend on the upshot of a rationally sound deliberative route entail the existence of external reasons, thereby contradicting the doctrine of internal reasons.¹⁰

2. The Notion of a Sound Deliberative Route

According to internalists, what you have reasons to do is basically what you want to do, subject to correction by deliberation and a rational purge of your present desires. The notion of a rationally sound deliberative route stands as a filter and generator of motivation between what you presently desire and what you might desire if practically rational. It stands as a filter when desires are undermined because of reliance on systematic confusion. It stands as a generator when desires are produced by making your desires systematically justifiable. What you might desire if you were to rationally deliberate provides you with reasons for action. But which forms of deliberation does practical reason demand? What makes for a rationally sound deliberative route? An internalist owes us an answer to this question.

Internalists construe the notion of a rationally sound deliberative route in different ways. Gilbert Harman, for example, has construed the notion of rationally sound deliberation as the process of creating maximal unity and coherence among the existing beliefs and desires of an agent.¹¹ Bernard Williams does not include a coherence constraint on rationally sound deliberation, but claims that the reasons of an agent are determined by desires which rest on no relevant factual ignorance and the existence of which is compatible with extensive imaginative awareness regarding their objects.¹² More recently, Smith has supplemented the account defended by Williams with a demand for unity and coherence among attitudes, while omitting the requirement of imaginative awareness.¹³ How can we choose between the different internalist conceptions of rationally sound deliberation? Williams argues that practical reasoning is essentially heuristic and indeterminate. He might therefore say that there is no real distinction between different internalist conceptions of a rationally sound deliberative route. But even he should agree that if internalism is to have determinate content, it must rule out some developments of an agent's desires as rationally unsound. If so, there is a real distinction between different conceptions of rationally sound deliberation. For example, we can imagine someone who by undertaking the route specified by Williams would remain unmotivated by some consideration, but who by undertaking the route speci-

fied by Smith would be motivated by it. According to the conception defended by Williams, the person could not be rationally faulted. According to the conception defended by Smith, he would be. We can also imagine someone who includes being moved by good family rhetoric as a rationally sound deliberative route, something which Williams would deny. An internalist owes us an account of what makes a deliberative route rationally sound, and which retains an interesting link between reasons and the present desires of an agent. We are entitled to ask what such an account would look like. In particular, we are entitled to ask whether what counts as rationally sound deliberation is itself constrained by the content of the present desires of an agent.

An internalist has two options. Either what counts as a rationally sound deliberative route for an agent is relative to the content of his present desires, or it is not. Let us call the first option process-relative and the second option outcome-relative. For to process-relative internalism, the reasons of an agent are determined by the upshots of a rationally sound deliberative route, where both the outcome of the route and the nature of the route are constrained by the present desires of the agent. For outcome-relative internalism, the reasons of an agent are constrained by the upshots of a rationally sound deliberative route, but the nature of the route itself is not constrained by the content of the present desires of the agent. Neither the outcome-relative nor process-relative interpretation of internalism is coherent. On the process-relative interpretation, internalism is intrinsically implausible. On the outcome-relative interpretation, internalism collapses into externalism.

3. Process-Relative Internal Reasons

There is more than one way for an internalist to construe the notion of a sound deliberative route as process relative. First, the internalist can define a rationally sound deliberative route in terms of a process which an agent presently desires to engage in, where this entails that the agent has a desire with a content to the effect that he will reason this way. Call this conception principles as ends. Second, he can define a rationally sound deliberative route in terms of a process the employment of which would promote the agent's present desires. Call this conception principles as means. Third, he can define a rationally sound deliberative route in terms of a process the agent would either desire to engage in, or which would promote the satisfaction of his desires, if he were practically rational. Since an internalist defines practical rationality in terms of the upshots of a rationally sound deliberative route, this strategy boils down to the claim that what counts as rationally sound deliberation is itself determined by the upshot of rationally sound deliberation. None of the three options are promising.

The conception of principles as ends is ill-founded. This view entails that whatever an agent endorses as a principle of deliberation thereby constitutes a rationally sound principle of deliberation. The view therefore fails to rule out principles which cannot be consistently applied. You may wish to apply what you take to be a consistent principle of deliberation where that principle is inconsistent. A principle may be straightforwardly inconsistent in virtue of its content. You cannot consistently act so as to maximise and minimise happiness at the same time. A principle may also be incoherent in the sense that no agent could actually implement it. No agent can coherently apply the principle that a person should always act only on false beliefs, since in order to apply this principle he would need true beliefs about his own beliefs in order to make sure that they are false. This fact indicates that the soundness of principles of deliberation is not a simple function of which principles an agent is disposed to accept. Maybe some inconsistent principles can be usefully applied if the inconsistency in question is unlikely to turn up in practice. But some principles cannot be applied at all. There are no grounds to accept that they are nevertheless rationally sound principles simply in virtue of being endorsed as such.

The conception of principles as means is equally ill-founded. First, to construe rationally sound deliberation in terms of means to the satisfaction of the present desires of an agent entails a collapse of internalism into the sub-Humean model which the doctrine of internal reasons is designed to avoid. Second, what counts as a rationally sound deliberative route cannot be construed as wholly relative to what will satisfy the desires of an agent since the desires of an agent may be satisfiable without recourse to any deliberation. Third, even if we help ourselves to the notion of deliberation, not every deliberative route which happens to promote the present desires of an agent can be rationally sound. An agent might satisfy a desire of his either by misapplying some deliberative principle, or by attempting to apply a principle which cannot be consistently applied. In order for a deliberative route to count as rationally sound, therefore, it must satisfy some constraints beyond the fact that it counts as a means to the satisfaction of a desire of an agent. It follows that internalism cannot be process relative in this sense.

It is no more promising to define a rationally sound deliberative route in terms of the desires of an agent after a process of sound deliberation. This attempt at definition fails to illuminate, since it merely reinvokes the notion of sound deliberation we are seeking to explain the basis of. The crucial question of what makes a deliberative route rationally sound remains unanswered, thereby threatening the internalist position with emptiness, regress, or circularity. Perhaps there is no way to make sense of the notion of rationally sound deliberation without engaging in deliberation. It does not follow that antecedent to deliberation any attempt to deliberate is as good as any other. First, we are entitled to some antecedently given constraints on how a deliberative

process should take place in order to count it as rationally sound. Second, we are also entitled to some antecedently given constraints on what counts as a process of deliberation.

There will no doubt be other process-relative views waiting in the wings. But no grounds have been provided by internalists for thinking that any such view would be coherent. We therefore have no reason to think of rationally sound deliberation as process relative.

4. Outcome Relative Internal Reasons

Outcome relativity entails that the soundness of a deliberative route is an objective matter to the extent that its soundness is determined independently of the content of the present desires of an agent. But the present desires of the agent might still constrain her reasons for action in their capacity as inputs to the deliberative process, and thereby determine the outcome of deliberation jointly with the deliberative route itself, or with the deliberative route plus a purge of false beliefs and psychological anomalies. Christine Korsgaard has argued that if the constraints on rationally sound deliberation are sufficiently demanding, the nature of the input will be irrelevant to the output.¹⁴ Such a predetermined convergence of deliberative outputs would certainly remove any interesting connection between reasons and the present desires of an agent, and thereby give practical reason a distinctly externalist flavor. But even if deliberative outputs are not predetermined to converge in this way, we still want to know what makes a deliberative route rationally sound in complete independence from the contents of the present desires of an agent. The external reasons theorist has a simple answer to this question. She is already committed to the existence of rational requirements on action which obtain independently of the present desires of agents. To postulate the existence of independent rational requirements on deliberation adds no further mystery to the externalist view. This is not to say that there is no mystery about how there can be such independently rational requirements. The point is rather that given the existence of such requirements on action, there is no further mystery about how there could be such requirements on deliberation. Deliberation is just a species of action. The case is different for an internalist. An internalist needs to ground her notion of rationally sound deliberation without admitting the existence of rational requirements on action which obtain independently of the present desires of agents.

There is good reason to think that this internalist task cannot be consistently carried out. Consider the case of a bad deliberator. Suppose that rationally sound deliberation includes a constraint of imaginative awareness. Suppose the bad deliberator is imaginatively insensitive to the effects of his acts upon others. He is never be able to develop “a concrete sense of what

would be involved” in acting the way he does, as he is unable to sufficiently exercise his imagination.¹⁵ It can still be true that were the bad deliberator to reach a state of imaginative awareness, he would be differently motivated. If so, he might have a reason to be motivated differently. But for bad deliberators a non-rational process like therapy is the only way to reach a state of rational action. Given their deliberative incapacity, they certainly cannot soundly deliberate themselves into such a state. Any version of internalism must allow for the possibility of bad deliberators. If there could be no bad deliberators, there could be no good deliberators, and thus no rationally sound deliberative routes. But the very possibility of bad deliberators undermines any version of internalism which entails that reasons exist conditionally on the possibility of a rationally sound deliberative route which agents can undertake in their actual circumstances in order to reach a state of rational motivation.¹⁶ For in the case of bad deliberators there is, by hypothesis, no such route.

No internalist should therefore claim that a rationally sound deliberative route must be possible for an agent to undertake in his actual circumstances. A wiser internalist would say that an agent’s reasons are determined by the outcome of a rationally sound deliberative route undertaken in some counterfactual circumstance where he can deliberate soundly. Suppose that rationally sound deliberation requires the presence of true beliefs plus a capacity for extensive imaginative awareness. An internalist might then argue that the reasons for action of an agent are determined by the outcome of a rationally sound deliberative route applied to the desires the agent would have after a purge of false beliefs and the acquisition of the capacity to gain extensive imaginative awareness. The desires he would have regarding his actual circumstances in these counterfactual circumstances could then supply his actual self with reasons for action. Smith has convincingly argued that a counterfactualised advice model along these lines provides a more plausible version of the doctrine of internal reasons.¹⁷

The possibility of an internalist advice model undermines a counterexample to internalism recently developed by Elijah Millgram, who plausibly argues that an agent can have reasons which exist conditionally on his incapacity to deliberate soundly.¹⁸ The case developed by Millgram fails to undermine the internalist advice model, since the fully rational counterpart of an agent is not barred from grasping reasons which are necessarily unavailable to his deliberatively incapacitated actual self. But the advice model of internalism also undermines the idea that there is any interesting connection between the reasons of agents and their present desires. To see this, consider the counterfactual verdicts of an agent regarding his actual circumstances after a rational development of his desires. The verdicts are reached at the end of a process which takes the present desires of the agent as inputs and produces fully rational counterparts of his desires as outputs. The challenge to an internalist is to specify what constrains this process, if not the adjustment of the present

desires of an agent to rational requirements on action which obtain independently of his present desires. There is no consistent internalist response to this challenge. Consider an agent who chooses to employ some deliberative route consequent to a purge of false beliefs. For the route in question to be rationally sound it must be undertaken in some ways rather than others. By outcome relativity, the way it is to be undertaken is determined independently of the content of the present desires of the agent, whether consequent or antecedent to the purge. This fact entails the existence of external reasons. For to deliberate in some ways rather than others is a way the agent can act. If the agent has reasons to deliberate in some ways rather than others which obtain independently of the contents of his present desires, he has external reasons to act in some ways rather than others. Furthermore, he has these external reasons both consequent and antecedent to the purge. His external reasons consequent to the purge include his reasons to deliberate a certain way. His external reasons antecedent to the purge include his reasons to act whichever way his fully rational counterpart would desire that he act. It follows that if there are any rationally sound deliberative routes available to the rationally purged counterparts of agents, their existence entails the existence of external reasons for actual agents, rationally purged or not. This contradicts the doctrine of internal reasons.

The same argument cannot be applied against the internalist conception of a rational purge. For undergoing a rational purge is not a way the agent can act. Still, the purge has to take some determinate form in order to count as rational. Once again, it is implausible that the nature of a rational purge is determined by the contents of the present desires of an agent. If it is not so determined, it must be determined by constraints which obtain independently of the contents of the present desires of the agent. While this does not directly entail the existence of external reasons for action, it does entail that the contents of the agent's desires after the rational purge are in part determined by rational constraints which obtain regardless of the contents of his present desires. It is a moot point whether or not an internalist can rest comfortably with this conclusion.

The upshot of this argument is somewhat puzzling. If we look at the way the doctrine of internal reasons has been motivated by such recent proponents as Harman and Williams, we find that its main target is the postulation of objectively valid requirements of practical reason which apply to agents regardless of the content of their desires. What we find on closer inspection is that there are reasons for action in the outcome-relative sense only if there are such objectively valid requirements of practical reason. It follows that the outcome-relative version of internalism is self-defeating, since it entails the postulation of rational requirements on action which it is of its nature to deny.

5. Reasons for Action and Reasons for Belief

The doctrine of internal reasons is a thesis about the nature of reasons for action, or what is known as practical reason. It might be thought that reasons for action exist in mutual independence of reasons for belief, or what is known as theoretical reason. On this view, practical reason includes rational requirements on action, while theoretical reason includes reasons to deliberate in determinate ways in order to proportion beliefs to the strength of available evidence. If so, it might also be thought that an agent can have theoretical reasons to deliberate in determinate ways, even if she has no practical reasons to deliberate that way. If the standards of theoretical reason pick up the slack where the standards of practical reason run out, perhaps an internalist can explain after all why agents have reasons to deliberate in some ways rather than others even if there are no external reasons for action.

But the standards of theoretical reason cannot pick up the slack left by the standards of practical reason on an internalist account. Given that actions which require the exercise of theoretical rationality are a subset of all actions open to an agent at any time, and given that an agent may have no reasons to engage in such actions at that time, the agent may have no reasons to observe the standards of theoretical rationality at that time. An agent who has no reasons to observe the standards of theoretical rationality is an agent who has no reasons to deliberate in order to proportion her beliefs to accord with available evidence. If she has no practical reason to be theoretically rational, she plausibly has no theoretical reason to deliberate in one way rather than another. But an internalist is committed to the possibility that an agent might have no practical reason to observe the standards of theoretical rationality, since it is a contingent matter whether she would be motivated to do so after sound deliberation. It follows that an internalist cannot make independent appeal to standards of theoretical reason in order to defend the claim that agents have reasons to deliberate in some ways rather than others, independently of their present desires.

6. Conclusion

We have found no coherent interpretation of internalism on which reasons are interestingly constrained by the present desires of agents. If the sub-Humean model is too simple, we have no grounds to believe in the existence of internal reasons. It does not follow that we have grounds to believe in the existence of external reasons. To acquire such grounds we need an account of what makes a process of deliberation rationally sound. Only when we have such an account can we begin to determine whether agents have external reasons for action.¹⁹

Notes

1. See Bernard Williams, "Internal and External Reasons," in *Moral Luck* (Cambridge, England: Cambridge University Press, 1981), pp. 101–113, and "Internal Reasons and the Obscurity of Blame," in *Making Sense of Humanity* (Cambridge, England: Cambridge University Press, 1995), pp. 35–45.
2. See Michael Smith, "Internal Reasons," *Philosophy and Phenomenological Research* 55 (1995), and Derek Parfit, "Reasons and Motivation," *Proceedings of the Aristotelian Society, Supplementary Volume* 71 (1997).
3. See Williams *op. cit.*, Smith, *op. cit.*, Gilbert Harman, "Moral Relativism Defended," *Philosophical Review* 85 (1975), and Harman, "Practical Reasoning," *Review of Metaphysics* 29 (1976).
4. Michael Smith, *The Moral Problem* (Oxford: Basil Blackwell, 1994).
5. See Williams, *op. cit.*
6. See E.J. Bond, *Reason and Value* (Cambridge, England: Cambridge University Press, 1983).
7. See Thomas Nagel, *The Possibility of Altruism* (Oxford University Press, 1970), and John McDowell, "Might there be External Reasons?" in J.E.J. Altham and R. Harrison, eds. *World, Mind and Ethics* (Cambridge, England: Cambridge University Press, 1995).
8. See Bond, *op. cit.*, Rachel Cohon "Are External Reasons Impossible?," *Ethics* 96 (1986), and J. Robertson, "Internalism about Moral Reasons," *Pacific Philosophical Quarterly* 67 (1986), and Elijah Millgram, "Williams's Argument against External Reasons" *Nous* 30 (1996).
9. See Brad Hooker, "Williams' Argument against External Reasons" *Analysis* 47 (1987), and Jonathan Dancy, *Moral Reasons*. (Oxford: Basil Blackwell, 1993), and McDowell *op. cit.*
10. See Jean E. Hampton, *The Authority of Reason* (Cambridge, England: Cambridge University Press, 1998).
11. See Harman, *op. cit.*
12. See Williams, *op. cit.*
13. See Smith, *op. cit.*
14. See Christine Korsgaard, "Skepticism about Practical Reason," *Journal of Philosophy* 83 (1986).
15. See Williams, *op. cit.*, pp. 104–105.
16. See Williams, *op. cit.* and Harman *op. cit.*
17. See Smith, *op. cit.*
18. See Millgram, *op. cit.*
19. I am grateful to Jimmy Altham, Jonathan Dancy, Ross Harrison, Jane Heal, Alex Oliver, and especially Philip Lakelin for helpful criticisms.