Ideal Self In Non-Ideal Circumstances

Sušnik, Matej

10. Ideal Self In Non-Ideal Circumstances

MATEJ SUŠNIK

1. Introduction

Any plausible theory of reasons should say something about the role that reasons play in the explanation and justification of our actions. Practical reasons, in other words, are not only expected to exert a "motivational pull" on the agent, but they should be able to justify what the agent does as well. It might seem that the explanatory challenge could be simply met by arguing that that the existence of reasons is dependent on the existence of motivation. For if motivation explains action, and if there are no reasons unless motivation is present, then the relation between reasons and explanation ceases to be mysterious. But this could hardly be the whole story. Although such a move could elegantly account for the explanatory role of reasons, connecting reasons and motivation in this way would leave the second (justificatory) challenge unanswered because it would deprive reasons of their normative force. In light of these concerns, some philosophers argue that normative reasons do not depend on one's actual motivation, but rather on the motivation one would have if one were better epistemically placed. This view is also known as internalism about reasons, and it has been under numerous attacks ever since it was firstly developed by Bernard Williams (1981).

In this paper I will try to defend Williams's internalism from one such attack. Internalists are often interpreted as claiming that one's reasons are not dependent on the motivation of one's actual self, but rather on the motivation of one's better or ideal self. But as some philosophers point out (Johnson 1999; Sobel 2001), this view overlooks the possibility that the agent's reasons may be wholly determined by the fact that the agent is not ideally placed. There could be cases in which a person can have a reason to do something although his better self would not be motivated to do that thing if he were in his place. As a result of their attempt to avoid this difficulty, internalists revise their view, but then fail to account for the explanatory role of normative reasons. My aim is to consider this objection in more detail and try to see how internalists may respond.

2. Williams's Internalism

People engage in practical reasoning because they want to determine what to do; and once they determine their reasons for action, people often act on those reasons. But common as it may be, this apparently simple process raises some difficult and unresolved questions. While it is evident that practical reasoning often results in the agent acquiring the relevant motivation, one may wonder how this could ever be the case. The problem is well known, and its roots can be traced back to David Hume's picture of reason as motivationally inert. Hume famously argued that reason is not capable of generating motivation on its own, and that some additional help is needed for that to happen. Bernard Williams (1981) expands Hume's picture, and argues that the process of reasoning can generate the appropriate motivation only if there is some motivation already present; the whole process has to start from something that is capable of moving us to action, and since only desires are suitable for this job, it is postulated that they have to exist in the background. The upshot of Williams' view is that the truth of all reason claims depends on the agent acquiring the relevant motivation as a result of rational deliberation. Simply put, according to Williams, an agent has a reason to perform some action only if he could become motivated to perform that action through the process of reasoning. And whether he could reach the relevant motivation through the process of reasoning will largely depend on the agent's present motivation (i.e. on what he actually desires). Since reasons are derived from his desires, this internalist account nicely explains how reasons get their motivating power.

Williams is well aware that it would be a mistake to make the existence of practical reasons dependent on the agent's actual desires. An agent may desire something, for example, because he holds mistaken beliefs. Suppose that I desire to drink the content of the bottle in my car because I believe it contains fresh water. But the bottle in fact contains poison, so my belief is false. (Williams 1981: 102) What this shows, Williams argues, is that I do not have a reason to drink the content of the bottle. Although I actually desire to drink it, I would lose that desire if I knew there was poison inside. Hence, making practical reasons dependent on one's desires does not yet imply that there is no place for normativity. It is precisely because reasons are dependent on one's counterfactual desires – namely, the desires one would end up having if one engaged in the process of practical reasoning – that they can be considered normative. And the explanatory role of reasons is still preserved within this picture: unless it is true that a reasoning process may lead an agent to become motivated to perform some action, it is also not true that this agent has a reason to perform that action. Obviously, as many have noticed, the force of Williams' argument largely depends on

his conception of the process of deliberation. For the purposes of this paper, however, this question is not of crucial importance, so there is no need to provide a detailed answer about what this process involves. It is enough to say that it includes "at least correcting any errors of fact and reasoning involved in the agent's view of the matter." (Williams 1995: 36) To sum up, what an agent has reason to do in his specific circumstances, Williams holds, depends on what this agent would desire to do if he did not have false beliefs and if he did not make any mistakes in reasoning.

3. Stepping Into Someone's Shoes

An agent who does not make mistakes in reasoning and whose "view of the matter" is not based on factual errors is sometimes described as someone who is "fully rational" (Smith 2004a), or as someone "who has been properly brought up" like "Aristotle's *phronimos*." (McDowell 1995: 73; Williams 1995: 189) But no matter which description is used, the key point is that this view involves the process of idealization: what a person has reason to do is determined by the motivation he would have in those circumstances if he were idealized in relevant ways. Or, somewhat differently, what one has reason to do in his specific circumstances depends on what his ideal self would be motivated to do if he were in the shoes of his actual self. This characterization opens up some important questions.

The mental exercise of stepping into the shoes of someone else is a useful tool that we often employ in ethical thinking. Among other things, an agent in this way becomes able to see the situation from a different angle, gains a better understanding of how his actions might affect others, and ultimately becomes better equipped to reach the judgment about what to do. Similarly, we might want others to step into our shoes because we are seeking advice or because we want them to feel what we feel. At least this is what we hope to achieve when we ask questions such as "What would you do if you were in my place?" or "How would you feel if you were in my place?" But the talk of imagining yourself in the shoes of someone else brings some familiar puzzles. Richard Hare nicely describes the problem: "If I imagine myself in your shoes, do I imagine myself having the same likes and dislikes as I have now, only in your circumstances; or do I imagine myself with your likes and dislikes too? Are the likes and dislikes part of the shoes or not?" (Hare 2000/1963: 126) The difficulty arises because an accurate description of one's situation sometimes cannot be given without appealing to one's personal (physical or psychological) traits. If the agent were different in some way, he would not be in that exact situation. Now, this means that in some cases one can successfully imagine oneself being in the circumstances of someone else only if one imagines having some fea-

tures of that person. But if those features are completely different from the features of a person who does the imagining, then the question is whether that implies that one really needs to imagine being someone else. (compare Hare 1981: 119) Suppose someone asks an animal torturer how he would feel if he were in the shoes of a dog whose tail is set on fire. Does that mean that this person would need to imagine being a dog? That does not make much sense. As Charles Taylor writes when considering some cases of this sort, perhaps one way to avoid these difficulties would be to provide "a theory of personal identity which could allow that two men with completely different life histories and with distinct physical and psychological characteristics might yet be the same person." (Taylor 1965: 288) But as Taylor himself observes, it is hard to imagine what such a theory would look like.

But let us go back to internalism about reasons. Is this view vulnerable to a similar objection? True enough, when we think about the relation between one's ideal and one's actual self, it is implausible to say that there are "*two* men with *completely* different life histories and with distinct physical and psychological characteristics." After all, the theory says that the truth of reason claims depends on what an agent would desire if *he himself* were idealized. Nevertheless, the problem remains. Although my ideal self is in many respects just like my actual self, it could easily happen that the shoes of my actual self do not fit him. Namely, if the essential part of my circumstances is the fact that I am not ideal, then my ideal self cannot be in my circumstances without changing them *or* without ceasing to be my *ideal* self. And if our circumstances are different, then what I have reason to do may also be different from what my ideal self has reason to do.

In order to establish this point, Michael Smith (2004a: 19) gives the example of a squash player who, due to his inability to handle defeat, forms a strong desire to hit his opponent with a racket.[1] If this man were more rational and clearheaded, he would instead desire to congratulate his opponent and shake his hand. But what needs to be taken into consideration is that he is *not* more rational and clearheaded, and that his present emotional condition prevents him from reasoning correctly and forming such a desire. Then, although this man's ideal self would be motivated to shake his opponent's hand, this is not what he, as he is now, has reason to do. Given his present circumstances, what he has reason to do is to move away from his opponent and calm down. In a somewhat different example (Johnson 2003: 574; Markovits 2011: 150; Wiland 2000: 562-3), a man whose critical thinking skills are significantly diminished due to his tendency to make logical mistakes is offered to take a logic course as a way to improve those

[1] The example was originally presented by Gary Watson.

skills. And again, as internalists claim, whether this person has a reason to take a logic course depends on whether his ideally rational self would be motivated to do so. But his ideally rational self is someone who by definition does not make any mistakes in reasoning, so he would not be motivated to take a logic course. Why would anyone whose critical thinking skills are ideal be motivated to improve them? This result, however, does not seem to be satisfying. For someone whose capacities for thinking clearly are impoverished, and who has a chance to become a better thinker by taking a logic course surely has a reason to do so despite the fact that his ideally rational self would lack that desire. Their circumstances are not the same, and internalism about reasons – at least in its present formulation – fails to take that fact into account.

4. The Advice Model

One possible way to deal with the above counterexamples would be to modify the internalist theory. And this is what many internalists in fact do: since they believe that the counterexamples emerge only because the theory is not carefully formulated, they argue that the best way to block them is to clearly specify the relation between one's actual and ideal self. The best-known proponent of this strategy is Michael Smith, who adopts what he calls the "advice model" of internalism. (2004a: 18–20) Since it is obvious that there are cases in which one's shoes cannot fit his ideal self, there is no requirement, as this model would state, that they should fit him in the first place. The main point is *not* to make my ideal self step into my shoes, but rather to make him look after me. My ideal self now has a job to watch my back, so to speak, and he can accomplish this by giving me advice about what to do. So what I have reason to do in my particular circumstances depends on the advice my ideal self would give me. The proponents of this model are no longer interested in what one's ideal self desires to do for himself when he is in the shoes of one's actual self, but rather in what one's ideal self desires that his non-ideal self do in his own shoes.[2] According to this revised version of internalism, therefore, my ideal self is not someone who "sets an example" I should follow, but he is rather my advisor.[3]

[2] Also compare Peter Railton (2003: 11). It should be noted, however, that it is not irrelevant, according to Smith, whether one's ideal self *desires* or *advises* his actual self to act in a certain way: while one's *pro tanto* reasons are grounded in the former, one's overall reasons are grounded in the latter (see Smith 2004b).

[3] The phrase "sets an example" is used in order to indicate the contrast that Smith makes with what he calls the "example model" of internalism, namely the model which is not immune to counterexamples, and which has just been discussed.

Going back to the previously mentioned cases, although it is true that the agent's ideal self would not desire to hit his opponent after the defeat in a squash game, his ideal self is perfectly aware that his actual self has this desire. And when taking into account his negative emotions and lack of self control, the agent's ideal self would certainly not advise him to approach his opponent and congratulate him. Trying to keep his actual self from falling into temptation to do something he would later regret, he would rather advise him to stay away. So this is precisely what the agent, according to Smith, has most reason to do. Similarly, it is perfectly compatible with this version of internalism to say that I may have a reason to take a logic course despite the fact that my ideal self would not be motivated to do so. The fact that he himself would not be motivated to take that course is beside the point; what is relevant is that he would advice me that *I* take that course. So once again, this is what I have most reason to do.

Since the advice model can easily cope with the counterexamples, many agree that it represents an "improvement" over the example model. (Bedke 2010: 42; Sobel 2001: 229) However, as Robert Johnson observes (1999), the advice model faces another difficulty: it fails to show how reasons and motivation are related. And that reasons are indeed related to motivation has already been pointed out by Williams: "If it is true that A has a reason to φ, then it must be possible that he should φ for that reason; and if he does act for that reason, then that reason will be the explanation of his acting." (Williams 1995: 39) It seems, however, that the advice model blocks the possibility Williams is talking about. Johnson (1999: 61–71) considers a scenario in which A wakes up one morning with a belief that he is James Bond. While we may suppose that A's ideal self would advise A to seek medical attention, the problem is, Johnson argues, that A cannot become motivated to do so. A's ideal self, on the one hand, does not hold this belief, so there is no reason for him to seek medical help, and A, on the other hand, cannot recognize that something is wrong with him, so he does not think that this reason really applies in his case. Therefore, A will remain unmoved by that reason. And even if A somehow ends up in the hospital seeking medical attention, that will not be *because* his ideal self advised him to go there. There could be a number of different reasons for which A may go to the hospital, but none of these reasons will be connected to the desires of his ideal self. The reason A needs to go the hospital is grounded in the fact that his ideal self would advise him to go there, and Johnson's essential point is that A is unable to go there for that specific reason.

5. Advice From a Different Creature

The challenge is, therefore, to provide a version of internalism that can both deal with the counterexamples and preserve the connection between reasons and motivation. The puzzle arises because it is not entirely clear how to understand the link between one's actual and ideal self. Since the circumstances someone faces could be largely determined by one's personal traits, it is no surprise that the situation I am in could be completely different than the situation my non-ideal self is in. In those cases, it appears, the result of the idealization process becomes unimportant for the actual person. For whatever my ideal self might desire, it does not need to have any effect on my particular situation. As David Sobel remarks: "[t]he idealization process […] turns us into such different creatures that it would be surprising if the well being of one's informed self and one's ordinary self consisted in the same things." (Sobel 2001: 228)

Interestingly enough, Williams anticipated this difficulty in his discussion with John McDowell (1995). Williams does not picture one's improved self as fully rational, fully knowledgeable or ideal in any other way. On the contrary, he explicitly claims that if practical reasons are analyzed in terms of the desires of some "ideal type," then reason statements can sometimes fail to be "distinctively" about some particular agent and his particular circumstances. And this is why he finds it important that such statements (such as the statement "A has a reason to φ") "say something special about A," and that they relate "more closely to the actual nature of A." (Williams 1995: 190) Reason claims should not be analyzed in terms of the desires of some *ideal* agent, but rather in terms of the desires of that *same* (actual) agent. Unless this condition is met, the counterexamples are always possible: it could easily happen that the actual agent and his ideal self have different reasons. In Williams' words, "…problems of this type can always in principle arise, until the distance between the actual and the imaginary improved agent has been reduced to zero… ." (Williams 2001: 94)

The reasons of the two selves are different because their circumstances are different, but also their circumstances are different because *they themselves* are different. The trouble is, as Sobel notes above, that the actual person and his ideal self are "different creatures." And this again raises some interesting questions. If me and my ideal self are "different creatures," then is there any relevant sense in which my ideal self is *my* ideal self? Moreover, even if he is *my* ideal self, it seems that – taking into account the circumstances I am actually in – this fact does not play any important role for me. Perhaps he is my ideal self after all, but he is just so distant from my actual self that it is irrelevant that he is *my* ideal self.

Let me briefly turn to the advice model in order to clarify what I have in mind. According to the advice model, as previously stated, what an agent has reason to do in his situation depends on what his ideal self would advise him to do in that exact situation. But when seeking advice, why should one turn to his ideal self rather than someone else who is also ideally placed? One would think that as long as the advisor is ideally placed (i.e. as long as he is in the ideal epistemic situation and makes no mistakes in reasoning), his identity is not important. Since I can receive advice from anyone who is ideally placed, my advisor could also be some neutral and impartial observer. So why is it necessary that this ideally placed advisor is *my* ideal self? What difference does it make? This remark shows that one could be doubtful whether this model falls under the rubric of internalism at all. For if my reasons do not really stem from my desires (actual or counterfactual), but rather from the advice of an ideally placed agent – an agent who could, at least in principle, be someone other than me – then in what sense are these reasons really internal?[4]

In summary, if internalists claim that one's ideal self is required to step into the shoes of his actual self, then one's ideal self cannot meet that requirement without changing the circumstances that his actual self is in, but if one's ideal self is not required to step into the shoes of his actual self and is rather pictured as his advisor, then it is no longer clear that this view deserves the internalist label. Can an internalist resolve this difficulty? I think the answer is yes, and in order to see that we need to return to Williams's version of this view.

6. *The Route Out*

Perhaps it would be useful to start this last section by giving the exact formulation of internalism, as understood by Williams. In his last paper dedicated to this topic, he formulates this view in the following way: "*A* has a reason to *φ* only if there is *a sound deliberative route* from *A*'s subjective motivational set (which I label "*S*," as in the original article) to *A*'s *φ*-ing. (Williams 2001: 91) Notice that Williams is only saying that such a deliberative route needs to *exist* for an agent to have a reason, and leaves it open whether an agent could in fact take that route from where he currently is. For, as he also notes, "[p]erhaps some unconscious obstacle, for instance,

[4] Johnson makes a similar point, but on different grounds. He is not concerned with the identity of one's ideal self, but rather with the fact that reasons on the advice model cannot find their path to motivation. He writes: "[I]t is misleading to present [the advice model] as a model of the internalism requirement. The latter connects reasons to motivation, but the advice model does not; it connects reasons to advice." (Johnson 1997: 621)

would have to be removed before [one] could arrive at the motivation to φ" (Williams 1995: 188). These obstacles can take various forms. If the agent is, for example, deceived – as is the case with the person deceived about the content of the bottle in his car – then the obstacle takes the form of a false belief. And this false belief, then, needs to be removed before the agent takes the deliberative route.[5]

Now it remains to be seen how Williams's version of internalism accounts for the above mentioned difficulties. Let us go back once again to Smith's example of a squash player. According to Smith's understanding of this case, as we may recall, although it is true that the agent's ideal self would be motivated to congratulate his opponent, this is not what the actual agent, due to his highly intense emotional state, has reason to do. The actual agent has reason to leave the place of the squash match. But this is not the correct description of what is going on in this case. The case of a squash player only shows that the agent presently cannot *act* on that reason, not that he does not *have* that reason. Under the assumption that there is a deliberative route from his *S* to his congratulating his opponent, then congratulating his opponent is what he has reason to do. The fact that his intense emotions present an obstacle for him to do so does not imply that his reason is removed, but rather that the obstacle itself should be removed. Perhaps it could be objected that if the agent is *not capable* of acting on some reason, then there is a sense in which he does not have that reason at all. But I do not think that it is even true to say that he is not capable of acting on that reason. It is not as if he does not have a capacity to deliberate; it is rather that he fails to exercise that capacity. And his failure to exercise that capacity is the obstacle that needs to be removed. The obstacle, in other words, comes in the form of him failing to deliberate.

What about the example of a person whose bad deliberative skills provide him with a reason to take a logic course? That example is somewhat different, but it also does not refute Williams's version of internalism. First, it could be undermined simply by attacking its basic premise – namely, that one's ideal self would not be motivated to improve his competence in reasoning because there would be no reason for him to do so. As it has already been pointed out, Williams does not think that one's reasons depend on the desires of one's *ideal* self, but rather on the existence of a sound deliberative route leading from one's *S* to him being motivated to act. If one does not deliberate ideally, then there is always room for improvement. Second, there is no inconsistency in saying that an agent could become motivated

[5] This interpretation of Williams is also favoured by Steven Arkonovich. He writes: "It is the possibility of deliberation *given* correct beliefs that grounds the agent's reasons, according to Williams." (Arkonovich 2011: 411, n.9)

through correct deliberation to improve his deliberation skills. For if one deliberates correctly in the practical domain, that does not mean that he also deliberates correctly in the theoretical domain. Perhaps it is even true that once the agent takes the deliberative route, he cannot reach the end of that route without being changed or improved in some relevant respects. But the internalist can give an account of that change or improvement. He may, in Williams's words, "impose … some constraints on what counts as 'deliberating correctly'." (Williams 1995: 188)

The "James Bond" example threatens to undermine any version of internalism, not only the advice model. Its basic message is that there could be reasons with no explanatory role – reasons that could motivate neither one's actual nor ideal self. An agent who suddenly forms the belief that he is James Bond has a reason to seek medical attention, but there is no way for him to get moved by that reason. For he can be moved by that reason only if he drops that belief, but as long as he holds that belief, that reason will not be capable of motivating him. Once again, it seems to me that the advocate of Williams's internalism can find the way out of this difficulty. According to my understanding of this position, whether there exists a deliberative route leading from the agent's *S* to him being motivated accordingly depends on whether it is possible for the agent to actually take that route. If there is an obstacle blocking the route, for example, and the agent cannot possibly remove it, then the route does not actually exist for him. Let us then assume that the obstacle blocking the route cannot possibly be removed by the agent. More specifically, assume that it is not possible for the agent to realize that his "James Bond" belief is false. In that case, it seems to me, we should say that the agent has *no* reason to visit a psychiatrist. Since he cannot possibly remove the obstacle, this means that there is no deliberative route leading from the agent's *present S* to him being motivated to visit a psychiatrist. Simply put, if he sincerely believes he is James Bond, and if he has no way of finding out that something is wrong with him, then *he*, as he is now, indeed has no reason to visit a psychiatrist.

But why suppose that it is not possible for him to start doubting that he is James Bond? What if the agent falsely believes that he is James Bond, but he also holds many other true beliefs from his past life? In that case it could be possible for him to remove the obstacle. Perhaps once he is told that he is *not* James Bond, he might be capable of realizing, through deliberation, that all his beliefs do not cohere well with the belief that he *is* James Bond, so he might in the end understand that he suffers from delusional disorder.[6]

[6] Although he does not directly defend Williams's model, this is in effect the proposal developed by Mark Van Roojen (2000). Van Roojen suggests that the problem disappears if one's better self is pictured as less than ideal.

So, if it is possible for him to start doubting the truth of his "James Bond" belief, then there might exist a deliberative route leading from his present *S* to him being motivated to visit a psychiatrist. In that case, the internalist should say, the agent indeed would have a reason to visit a psychiatrist. The advocate of William's internalism, therefore, should simply argue that whether this agent has a reason to seek medical attention will exclusively depend on the possibility of him realizing that he is in a delusional state.

Williams's internalism, then, not only that it has resources to cope with different counterexamples, but it also brings on the surface the importance of the process of idealization. Just as we learn something about other people when we imagine ourselves in their shoes, we also learn something about ourselves when we engage in the process of idealization. We gain a better understanding of what we really desire, what we plan to do, and what is the best way for us to proceed in given circumstances. Likewise, just as there is no point in imagining oneself in the shoes of someone else if that process implies that the agent needs to *become* someone else, there is equally no point in idealizing if somewhere along that process the agent completely ceases to be himself.

REFERENCES

Arkonovich, S. (2011). "Advisors and Deliberation." *Journal of Ethics* 15 (4): 405–424.

Bedke, M. S. (2010). "Rationalist Restrictions and External Reasons." *Philosophical Studies* 151: 39–57.

Hare, R. M. 2000. (1963). *Freedom and Reason*. Oxford University Press.

Hare, R. M. (1981). *Moral Thinking: Its Levels, Method and Point.* Oxford University Press.

Johnson, R. N. (1997). "Reasons and Advice for the Practically Rational." *Philosophy and Phenomenological Research* LVII/3: 619–625.

Johnson, R. N. (1999). "Internal Reasons and the Conditional Fallacy." *The Philosophical Quarterly* 49: 53–71.

Johnson, R. N. (2003). "Internal Reasons: Reply to Brady, Van Roojen and Gert." *The Philosophical Quarterly* 53: 573–580.

Markovits, J. (2011). "Internal Reasons and the Motivating Intuition." In Brady M. (ed.) *New Waves in Metaethics.* Basingstoke: Palgrave Macmillan: 141–165.

McDowell, J. (1995). "Might There Be External Reasons?" In Altham J. E. J. and Harrison R. (eds.) *World, Mind and Ethics: Essays on the ethical philosophy of Bernard Williams*. Cambridge University Press: 68–85.

Railton, P. (2003). "Moral Realism." In Railton P. *Facts, Values, and Norms: Essays toward a Morality of Consequence*. Cambridge University Press: 3–42.

Smith, M. (2004a). "Internal Reasons." In Smith M. *Ethics and the A Priori*. Cambridge University Press: 17–42.

Smith, M. (2004)b. "The Incoherence Argument: Reply to Shafer-Landau." In M. Smith *Ethics and the A Priori*. Cambridge University Press: 43–55.

Sobel, D. (2001). "Explanation, Internalism, and Reasons for Action." *Social Philosophy and Policy* 18 (02): 218–235.

Taylor, C. C. W. (1965). "Freedom and Reason by R. M. Hare." *Mind* 74 (294): 280–298.

Van Roojen, M. (2000). "Motivational Internalism: A Somewhat Less Idealized Account." *The Philosophical Quarterly* 50: 233-241.

Wiland, E. (2000). "Good Advice and Rational Action." *Philosophy and Phenomenological Research* 60 (3): 561–569.

Williams, B. (1981). "Internal and External Reasons." In Williams, *Moral Luck*. London: Cambridge University Press: 101–113.

Williams, B. (1995). "Replies." In J. E. J. Altham and R. Harrison (eds.) *World, Mind and Ethics: Essays on the ethical philosophy of Bernard Williams*. Cambridge University Press: 185–224.

Williams, B. (2001). "Postscript: Some Further Notes on Internal an External Reasons." In E. Millgram (ed.) *Varieties of Practical Reasoning.* MIT Press: 91–97.