UNITY IN THE THEORY OF REASONS

Michael Smith

1. Introduction

There are reasons for beliefs, desires, and actions, and reasons for various emotional reactions too. But what do all of these reasons have to do with each other? My conjecture is that there is unity in our theory of reasons about all of these, and that that unity derives from the fact that, at the most fundamental level, an account of what reasons for beliefs, desires, and actions are can be derived from a single source, namely, the concept of an *agent*. My view is thus a version of a *constitutivist theory of reasons,* the kind of theory defended by Christine Korsgaard and David Velleman. But as will emerge, my own version of the theory is very different from theirs. In what follows I provide some support for my conjecture about the unity in our theory of reasons. I will argue that though there are residual conflicts among reasons, these conflicts are all resolvable. Conflicts among reasons do not signal disunity in our theory of reasons.

2. Thomson on goodness-fixing kinds

Let's begin with a distinction made by Judy Thomson. Thomson distinguishes among kinds between those that are *goodness-fixing* kinds (eg *toaster, burglar*) and those that aren't (eg *pebble, smudge*). Goodness-fixing kinds are kinds with built-in standards of assessment. So long as these kinds do not have to be normatively characterized—more on this presently—goodness-fixing kinds thus provide us with standards of evaluation that are themselves non-normatively characterizable. If we aspire to provide a fully reductive theory of reasons, then a good place to start is with some appropriate goodness-fixing kind.

3. The kind agent *is a goodness-fixing kind*

What might that kind be? The kind *agent* is a goodness-fixing kind. A *minimal* agent is one who possesses and exercises the capacity to put his final desires together with his means-end beliefs with a view to realizing the objects of his final desires in the way characterized by the standard Humean story of action. An *ideal* agent is one who possesses and robustly exercises maximal versions of these two capacities: that is, the ideal agent has maximal capacities to *know the world in which he lives* and *realize his final desires in it*. It should already be clear why focussing on the concept of an agent might provide us with unity in our theory of reasons for beliefs, desires, and actions.

4. The two capacities possessed by an ideal agent

Note that this characterization of an ideal agent already traffics in the concepts of both reasons and rationality. An ideal agent is epistemically rational, where this is a matter of his having the capacity to believe for reasons, and he is also instrumentally rational, where this is a matter of his having the capacity to have instrumental desires that cohere in the right kind of way with his final desires and beliefs about means to ends. The two capacities possessed by an ideal agent are made for each other. The exercise of the capacity to realize desires would be impossible for a being that lacked the capacity to have knowledge of the world in which it lives, and a finite being's possession of the capacity to have knowledge of the world in which it lives is something for which there would no need at all if that being didn't have to come to a view about the way the world in which it lives is in order to satisfy its desires. A finite being's desires thus provide it with the appropriate target for its exercise of its capacity to know the world in which it lives. There is no requirement that it be ominiscient; knowledge is only required as and when needed in order to satisfy its desires.

5. The unity in the theory of reasons as so far characterized

Note that we already have a preliminary sense in which a focus on the concept of an agent may provide unity in our theory of reasons. We saw at the beginning that there are reasons for beliefs and desires, and what we have just seen is that reasons for instrumental desires turn out to be nothing over and above reasons for means-end beliefs. If, in the spirit of another of Judy Thomson's suggestions, reasons for beliefs in general can be reduced to evidence, or probability-raising, or truth-conduciveness, or entailment, and if these notions can be non-normatively characterized, and if instrumental coherence can be understood in terms of some function of the strengths of final desires and the degrees of means-end beliefs, and these too can be non-normatively characterized, then it will turn out that our theory of reasons for beliefs and desires is both unified and non-normative. Note, however, that one consequence of this theory would be that there are no reasons for final desires. It would also leave us without a theory of reasons for action.

6. The problem with the theory of reasons as so far characterized and the solution to that problem: certain self- and present-directed desires are partially constitutive of being an ideal agent

There is, however, a problem with the theory of reasons as so far characterized, because the two capacities possessed by an ideal agent do not fully cohere with each other. They do not cohere with each other in the sense that their joint exercise is not robustly compossible. If there is some psychological state an ideal agent could have that would make the joint exercise of its two capacities more robustly compossible, then it is plausible to suppose that an ideal agent would have to have that psychological state. This suggests that an ideal agent must have certain *coherence-inducing* desires. Specifically, it suggests that an ideal agent must have a dominant final desire to not now interfere with his current exercise of his capacity to know the world in which he lives. The argument for this is that the possession of such desires makes for a Pareto improvement in the extent to which an ideal agent is able to jointly exercise his capacities to know the world in which he lives and realize his desires in it.

7. Further self- and future-directed desires are constitutive of being an ideal agent

Given that agents exist over time, ideal agents must have further dominant final desires as well. Specifically, they must have a dominant final desire to not now interfere with their future exercise of their capacity to know the world in which they live, or their future exercise of their capacity to realize their other final desires, and they must also have a dominant final desire that they now do what they can to ensure that in the future they have the capacities to know the world in which they live and realize their final desires in it to exercise. These dominant final desires too are required because they make the joint exercise of their capacities to know the world in which they live and realize their desires in it more robustly compossible.

8. And yet further other-directed desires are constitutive of being an ideal agent

A variation on Parfit's argument in the 'Appeal to Full Relativity' section of *Reasons and Persons* suggests that there is no principled way to restrict the contents of these final desires of an ideal agent to that agent's present and future self. These dominant desires must therefore concern all beings whose exercise of their knowledge acquisition and final desire realization capacities depends on what the agent himself presently does. For short, let's put this by saying that every ideal agent must have dominant final desires to help and not interfere.

9. Why the potential for conflict in the desires constitutive of being an ideal agent is consistent with the joint exercise of an ideal agents two capacities being fully robustly compossible

Though an ideal agent's possession of such final desires is necessary because this is the only way in which to ensure that his exercise of his capacities to have knowledge of the world in which he lives and realize his desires in it is robustly compossible, note that there will cases in which these desires themselves come into conflict, and the desire that wins out will be the desire that precludes the ideal agent's having knowledge of the world in which he lives. Does this mean that the joint exercise of the two capacities is not fully robustly compossible? A point made earlier suggests not. An agent's desires provide him with the appropriate target for his exercise of his capacity to know the world in which he lives. There is no requirement that he be ominiscient. The conflict cases we are imagining are cases in which an agent does not need to have the knowledge that he will lack.

10. From an account of the dominant final desires of an ideal agent to an account of desirability as indexed to that agent

When we combine this account of the dominant final desires possessed by an ideal agent with the dispositional theory of value—this is the view that it is finally desirable$_a$ that p iff and because *a*'s ideal counterpart would finally desire that p—it turns out that for every agent, *x*, it is finally desirable$_x$ that *x* helps and does not interfere because *x*'s ideal counterpart has a dominant final desire that *x* helps and does not interfere, and indeed, it follows that this is more finally desirable$_x$ than acting in any other way available to *x*.

11. From an account of final desirability as indexed to an agent to reasons for final desires of that agent

We saw earlier that our initial account of what an ideal agent is like suggests that though there are reasons for beliefs, and reasons for instrumental desires, there are no reasons for final desires. However the account just provided of final desirability as indexed to an agent suggests that there are reasons for final desires after all. Yet another suggestion of Judy Thomson's is key. Thomson suggests that there are reasons for beliefs because of a feature of belief that belief shares with final desires. Belief is a mental state with a correctness condition, namely the truth of the proposition believed, and what makes a consideration a reason for being in a mental state with a correctness condition is the fact that that consideration is evidence for, or makes probable, the truth of the proposition that is that mental state's correctness condition. But final desire is also a mental state with a correctness condition. A final desire is correct just in case the object of the final desire is finally desirable. It follows that there are reasons for final desires too. A consideration is a reason for finally desiring that p just in case that consideration is evidence for, or makes probable, the truth of the proposition that p is finally desirable. The upshot is that reasons for instrumental desires and reasons for final desires both reduce to reasons for belief. If reasons for belief reduce to evidence, or probability-raising, or truth-conduciveness, or entailment, and if these notions can be non-normatively characterized, then it turns out that our theory of reasons for belief, instrumental desires, and final desires is both unified and non-normative.

12. From an account of desirability as indexed to an agent to an account of that agent's reasons for action

On the plausible assumption that what an agent has reason to do is fixed by which of his options would realize finally desirable states of affairs, the more finally desirable the stronger the reason, it further follows that every agent has a reason to help and not interfere, and that they also have a reason to do whatever they finally desire to do, so long as their so acting is consistent with their helping and not interfering. The upshot is thus not just that reasons for instrumental desires, reasons for final desires, and reasons for belief all reduce to the same thing. The upshot is also that reasons for action reduce to the elements out of which we constructed our account of desirability. Since these elements all emerged out of our analysis of the concept of an ideal agent, it follows that the concept of an agent provides us with the materials required for a unified and non-normative theory of reasons for belief, instrumental desire, final desire, and action.