THE PITFALLS OF 'REASONS'

Ralph Wedgwood

University of Southern California

These days, many philosophers are tempted by an approach that can be epitomized by the slogan "Reasons First". According to this approach, there is a crucial notion of a "normative reason" for an action or attitude, which is the most central of all normative concepts that appear in the parts of our thinking that are concerned with normative questions. As Joseph Raz (1999, 67) says, "The normativity of all that is normative consists in the way it is, or provides, or is otherwise related to reasons." In a similar vein, T. M. Scanlon (2014, 2) says that he is "inclined to believe" that "reasons are fundamental" in the "sense of being the only fundamental elements of the normative domain, other normative notions such as *good* and *ought* being analysable in terms of reasons." Besides Raz and Scanlon, many other contemporary philosophers are drawn to a similar approach – including John Skorupski (2010) and Mark Schroeder (2007 and 2010), among others.

In this essay, I shall argue against the basic presupposition of this approach – the presupposition that there is a single central concept of a "normative reason". On the contrary, I shall argue that there is a plethora of different concepts expressed by these philosophers' talk of "normative reasons", and none of these concepts is any more central than any other. I shall also propose an interpretation of the way in which the word 'reason' functions in English. According to this interpretation, *none* of the concepts expressed by the term 'reason' is fundamental: on the contrary, the concepts expressed by the relevant uses of 'reason' can all be defined in terms of other normative concepts – such as the concepts that are expressed by deontic modals like 'ought' and 'should', and those that stand for various different kinds of value.

1. "Reasons": Some preliminaries

It has been all but universally recognized that the term 'reason' is at least to some extent ambiguous or polysemous. In particular, almost all philosophers who have discussed "reasons" distinguish between *motivating reasons* and *normative reasons*.[1]

A person's motivating reasons for acting or thinking in a certain way are the reasons *for which* the person acts or thinks in that way (or the person's reasons *for* acting or thinking in that way). These motivating reasons provide a certain sort of distinctive psychological explanation of the person's acting or thinking in this way: for example, Martin's reason for flying to Ireland is at least one of the reasons *why* Martin is flying to Ireland. In this way, these motivating reasons appear to be a special case of *explanatory* reasons (such as *the reason why the bridge collapsed*).

[1] See for example Schroeder (2007, 10–15) and Grice (2001, 31).

By contrast, the normative reasons for an agent to perform a certain action or to have a certain attitude need not in any way motivate or sway the agent to perform the action or to have the attitude. They simply count *in favour of* that action or attitude; in other words, they are considerations that *support* or go at least some way towards *justifying* that action or attitude. My focus throughout this discussion will be on normative reasons, not on motivating reasons.

Some philosophers distinguish between speaking of the reasons that *there are* for an agent to perform an action or to have an attitude, on the one hand, and the reasons that the agent *has* to perform the action or to have the attitude, on the other. According to these philosophers, a reason that an agent has to perform a certain action is a reason that there is for the agent to perform the action of a special kind – namely, a reason that the agent is in some way *aware of* (or has a certain kind of access to).[2]

It does not seem plausible to me that this distinction is really marked by the difference between these two constructions in ordinary English; there is nothing infelicitous about saying 'It turns out that we had good reason to be cautious – although we couldn't have known it at the time'. The real difference seems to be one that has been highlighted by John Broome (2013, 65). The sentence 'There is a reason for Alex to get a severe punishment' could be true in a given context even if the sentence 'Alex has a reason to get a severe punishment' is not true in that context. I shall explain what this difference amounts to later on; but this difference will not matter for most of what follows.

There is a further difference between uses of the term 'reason' that philosophers have occasionally wondered about.[3] Sometimes, we use the term 'reason' as a mass noun (as when we say, 'What have we most reason to do?' or 'He saw little reason to accept the invitation'); on other occasions, we use the term as a count noun (as when we say, 'There are many reasons for you to doubt his trustworthiness'). Does this difference mark an important distinction?

In fact, however, I doubt that this difference will turn out to have great importance for our purposes. Other terms like 'explanation' and 'benefit' can also be used in both ways. We can say both 'This requires little explanation' and 'Which strategy will provide most benefit?' (using the terms as mass nouns) and also 'What is the best explanation?' and 'This option will provide many benefits' (using them as count nouns). The general pattern is that the count noun refers to items that are instances or species or sources of what the corresponding mass noun refers to; it seems that a similar pattern holds with the term 'reason'. At all events, I shall not worry about these differences here. For most of what follows, I shall write as though 'There is a reason for you to φ', 'There is reason for you to φ', 'You have a reason to φ' and 'You have reason to φ' are all equivalent to each other.

---

[2] See for example Schroeder (2011).

[3] This phenomenon has been noticed by Grice (2001, 31) and discussed by Fogal (forthcoming).

Scanlon (1998, 17) seeks to clarify his talk of normative reasons by saying that the distinguishing feature of a normative reason for an action or an attitude is that it "counts in favour" of that action or attitude. However, in his view, this does not count as any sort of definition or analysis of the concept of a normative reason, since the relevant uses of the phrase "counts in favour of" are in fact simply synonymous with – that is, express the very same concept as – the corresponding uses of the phrase 'is a reason for'. It is plausible that in many contexts, the phrases 'is a reason for' and 'counts in favour of' can be used in such a way that they express the same concept. However, this point does not establish that there is a unique concept that these phrases express in all of these contexts; as I shall argue later in this essay, these phrases can in fact be interpreted in a large number of importantly non-equivalent ways.

Most philosophers who adhere to the "Reasons First" approach – including both Raz (2011, 18) and Scanlon (1998, 17) – claim that the concept of a normative reason is utterly primitive and incapable of being defined by means of other concepts. Moreover, Scanlon (2014, 2) also claims that it is impossible to give any metaphysical reduction of facts about reasons to strictly naturalistic facts. Neither of these claims is a necessary component of the "Reasons First" approach as I shall understand it here. This is particularly clear in the case of the second claim, since some proponents of the "Reason First" approach – most notably, Schroeder (2007, chap. 4) – advocate precisely the kind of naturalistic reduction of the normative that Scanlon rejects.

2. Alternatives to "Reasons First": Two ways of defining reasons

In assuming that there is a unique notion of a "normative reason", which is the most central of all normative concepts, the adherents of the "Reasons First" approach reject any attempt to define the notion of a "normative reason" by means of other normative concepts. But other philosophers have offered definitions of just this kind. Broadly speaking, there are two main varieties of such normative definitions of "normative reasons" that we need to consider.

First, there is the definition of normative reasons that has been developed by John Broome (2004). According to this definition, the reasons that there are for an agent to act in a certain situation are facts that play a certain sort of role in *explaining* the truth about how the agent *ought* to act in that situation. As Broome puts it, a reason for an agent to φ is a fact that plays the "pro φ-ing" role in a *weighing explanation* of how the agent *ought* to act in the situation in question. The key point of this definition of normative reasons is that it interprets normative reasons as key elements in an *explanation* of a *normative* fact.

The general idea then is this: Together with the facts about which options are available, the reasons that one has in favour of the available options, and the reasons that one has against the available options, determine what one has *most reason* to do, all things considered; and if one has most reason to do something, all things considered, then it is also what one *ought to do*.

In general, if the reasons that one has in favour of and against the available options determine what one has most reason to do, each of these reasons must have something like a *weight*, which determines the effect that this reason has when it is weighed up with the other reasons. There are various conceptions of how these reasons determine what one has most reason to do, all things considered. I shall such not be able here to investigate either what determines the weight of reasons, or how these weights combine to determine what one ought to do.

It is clear why, on this definition, it is reasonable to describe a reason for a certain action or attitude as "counting in favour of" that action or attitude; this reason counts in favour of that action or attitude because it goes some way towards making it the case that one has *most* reason for that action or attitude – that is, it goes some way towards making it the case that one *ought* to perform this action or have this attitude.

A second approach to defining "normative reasons" takes a very different approach. On definitions of this second variety, reasons are the *starting points* for processes of sound or rational *reasoning* or *deliberation*. For example, one philosopher who has given a definition of reasons of this second kind is Kieran Setiya. According to Setiya (2007, 12), "The fact that *p* is a reason for *A* to φ just in case *A* has a collection of psychological states, *C*, such that the disposition to be moved to φ by *C*-and-the-belief-that-*p* is a good disposition of practical thought, and *C* contains no false beliefs." This definition is similar to Bernard Williams' (1995, 35) idea that there is a reason for you to φ just in case there is a "sound deliberative route" from your current state of mind to your being moved to φ. What Setiya adds to Williams' idea is a certain conception of what counts as a "sound deliberative route": specifically, for Setiya, a deliberative route counts as "sound" just in case it is the manifestation of a "good disposition of practical thought" and it does not involve being "moved" by any "false beliefs".

Many other versions of this second variety of definition are possible. For example, consider the definition of reasons that is given by Stephen Kearns and Daniel Star (2009). According to Kearns and Star's definition, a fact counts as a reason for an agent to φ if and only if the fact is evidence that the agent ought to φ. This may sound quite different from the definitions of Broome and Setiya that we have just considered. In fact, however, it is fairly plausible that if a fact is evidence that you ought to φ, then there is a "sound deliberative route" that leads from your considering that fact to your being inclined or moved towards φ-ing. So the definition of Kearns and Star seems to be fundamentally akin to the kind of definition that is given by Setiya.

Again, it is intelligible how a definition of this second variety would make it reasonable to describe a normative reason for a course of action or an attitude as "counting in favour" of that course of action or attitude: according to this definition, the normative reason counts in favour of that course of action or attitude because a suitably rational or well-informed agent would *respond* to this reason by being *inclined* or *moved* to take that course of action or have that attitude.

On the face of it, however, these two varieties of definitions are importantly different from each other. The first definition associates reasons with a *justificatory* story – that is, with a story that *explains* the truth about which action or attitude one has, all things considered, most reason to do. According to this definition, normative reasons are what provide explanations of normative facts. If there is anything that satisfies this definition, then, just to have a label, we may call the reasons that satisfy this definition the "normative-explanation reasons".

According to definitions of the second variety, normative reasons are tied to an *ideal deliberative or motivational procedure*. In effect, definitions of this second variety interpret normative reasons as idealized possible motivating reasons: they are, very roughly, what *would* be our motivating reasons if we were suitably well informed and rational. If there is anything that satisfies this definition, then, just to have a label, we may call the reasons that satisfy this definition the "ideal-motivation reasons".

On the face of it, it seems plausible that something like each of these definitions is satisfied by some items or other; if that is right, then both the "normative-explanation reasons" and the "ideal-motivation reasons" exist. But are the normative-explanation reasons the same as the ideal-motivation reasons? Or are they different?

Before addressing this question, however, I wish to raise a question about whether the definitions of the "normative-explanation reasons" and the "ideal-motivation reasons" that I have just given are completely univocal. Both definitions involve normative terms, like 'ought', or 'good disposition of practical thought' (or 'sound' and 'rational'), and the like. But now we need to ask, Are these terms completely univocal, or do they express different concepts in different contexts?

If there are many different kinds of 'ought', then there will presumably be correspondingly many kinds of normative-explanation reasons – the reasons that explain the facts that can be articulated using each of these kinds of 'ought'. In particular, consider the idea that there are both *objective* and *subjective* kinds of 'ought'.[4] For example, suppose that you are on top of a tower tracking someone who is making his way through a maze on the ground. You might say, 'He has no way of knowing it, but he ought to turn Left at this point'. But you might also say, 'Since all the evidence that he has had so far supports going Right, he ought to turn Right (and not Left) at this point.' Both your statements seem perfectly true, when taken in their intended sense. But it surely can't be true in this case that he ought both to turn Left and not to turn Left. So, it seems, we must distinguish between "objective" and "subjective" senses of 'ought'.

If 'ought' is polysemous in this way, then some reasons will be facts of the sort that explain what one *objectively* ought to do, while other reasons will be facts of the sort that explain what one

---

[4] I have argued in favour of the conclusion that we must distinguish between these different kinds of 'ought' elsewhere; see Wedgwood (2007, Section 5.1).

subjectively ought to do. Presumably, the reasons that explain what one objectively ought to do may include facts that one does not know, and perhaps even facts that one is not in a position to know – whereas the reasons that explain what one *subjectively* ought to do will be limited to facts that are in some way reflected in one's perspective or in the information that is available to one at the relevant time.

To illustrate this point, suppose that your evidence is misleading, in the sense that some of the propositions that given your evidence you are rationally required to believe are in fact false. For example, suppose that given your evidence, you are rationally required to believe that the man approaching you is an enemy soldier who will kill both you and your children unless you shoot him. In fact, however, the man is entirely innocent and poses no threat of any kind. Do you "have a reason" to shoot him?

The best way to handle such cases, it seems to me, is to say that in one "subjective" sense, you ought to shoot the man, while in another "objective" sense, you ought not to. So in one sense, you have "normative reason" to shoot (there is something that explains why you in this subjective sense "ought" to shoot him). But arguably, in another equally legitimate sense, you do not have "normative reason" for shooting him: the fact that you are rationally required to believe shooting him to be necessary in order to prevent him from killing you and your children does not seem to be a reason in relation to the objective 'ought' in the same way – for in the case where you have a true belief about your situation, it seems typically to be the fact that makes the belief true, and not the belief itself, or the fact that the evidence supports the belief, that counts as a reason in relation to the objective 'ought'. (It would be a strange sort of double-counting to include both the fact that makes the belief true, and the fact that the evidence supports the belief, as distinct reasons in favour of the same act.) If that is right, it would be an unilluminating pseudo-problem to worry about whether you *really* have reason to shoot him: in one sense you have a reason, and in another sense, you have not.

This approach can also help us to understand the contexts in which 'There is a reason for Alex to get a severe punishment' is true while 'Alex has a reason to get a severe punishment is not'. Some occurrences of 'ought' are in a way indexed to practical situation that Alex is in at a particular time; as Broome (2013, 12–15) would say, the concept expressed by these occurrences of 'ought' is "owned" by Alex. The sentence 'Alex has a reason to φ' is naturally heard as concerned with reasons that can explain the truth about what Alex ought to do – in a sense of 'ought' that is indexed to Alex's practical situation at a contextually salient time; the sentence 'There is a reason for Alex to φ' can more easily be heard as concerned with reasons that can explain the truth about what ought to be the case with Alex – in a sense of 'ought' that may not be indexed to Alex's practical situation at any time.

In this way, the terms that were used in the definition of the normative-explanation reasons were not completely univocal. There is not just one kind of normative-explanation reasons, but many different kinds of such reasons.

Moreover, a parallel point seems to hold about the terms that appeared in the definition of the ideal-motivation reasons as well. On the face of it, there are several different ways in which a process of motivation or deliberation could be "ideal". It could be ideally rational; it could lack all false beliefs (or at least all false beliefs about a certain range of subject-matters); it could be ideally well informed about *all* the empirical facts (or at least about a certain range of empirical facts); or it could be ideally well informed about all facts whatsoever, including normative and ethical facts; and so on.

In fact, different philosophers who define reasons in this second way seem to have appealed to quite different kinds of idealization in their definitions. For example, as we have seen, according Kieran Setiya's definition of what it is for the fact that *p* to count as a reason for an agent to φ, the ideal motivational process must manifest a "a good disposition practical of thought", and it must set out from a possible collection of mental states that includes no false beliefs, and includes the belief that *p*, but otherwise is as similar as possible to the agent's actual mental states. This is quite different from the kind of the kind of idealization that is appealed to by Michael Smith (1994, 155–161), according to whom the relevant ideal agent has a fully coherent set of desires, and has all relevant true beliefs and no false beliefs whatsoever. On the face of it, each of these different sorts of idealization could define a different concept of "ideal-motivation reasons", and on the face of it, it is unclear why we should be more interested in any one of these concepts rather than in any other.

In this way, then, many different precise versions of each of these two varieties of definitions could be formulated; and each of these formulations would define a different kind of normative reason, without giving any encouragement to the view that any of these kinds of reasons is any more fundamental than any other. In short, the most plausible alternative to the "Reasons First" approach is one on which there are innumerable different concepts of "normative reasons", none of them any more central than any other.

3. Arguing against "Reasons First": The strategy

Even though the "Reasons First" theorists do not give any definition of "reasons" in normative terms, the overwhelming majority of these theorists hold that normative reasons play the kinds of roles that these other theorists have invoked in their definitions of reasons. Thus, virtually all the "Reasons First" theorists hold that reasons play both a normative-explanation role and an ideal-motivation role.

Thus, for example, Joseph Raz (2011, 23–6) holds that the normative reasons in favour of and against all the available options determine what "one has conclusive reason to do"; and if it is true that one has conclusive reason to φ, then it will also be true (in at least one sense of the term 'ought') that one ought to φ. In this way, he thinks that normative reasons play a version of the normative-explanation role. But he also holds that normative reasons must be capable of playing a version of the ideal-motivation role. As Raz (2011, 27) says, "normative reasons must be capable of providing an explanation of an action: If that R is a reason to φ then it must be possible that people φ for the reason that R and when they do, that explains (is part of an explanation of) their action". This is a constitutive feature of reasons according to Raz (2011, 86): "Reason does not make reasons into reasons …. But they are reasons because rational creatures can recognize and respond to them with the use of Reason."

In a similar way, Schroeder (2007, 130) explicitly argues that the totality of an agent's reasons along with their weights explain what the agent ought to do. In addition, Schroeder (2007, 26) also claims that reasons are subject to what he calls the "Deliberative Constraint": "when Ryan is reasoning well, the kinds of thing about which he should be thinking are his reasons." In this way, he also accepts that reasons play both the normative-explanation role and the ideal-motivation role.

Like Raz and Schroeder, Scanlon (2014, 108) also thinks that the normative reasons in favour of and against the available options determine what the agent has "sufficient reason" and what she has "compelling reason" to do, and that what one has "compelling reason" to do is also what one ought to do. In this way, he accepts that reasons play the normative-explanation role. Unlike Raz, Scanlon does not exactly hold that all normative reasons as such must be capable of playing some version of the ideal-motivation role. But he does hold that insofar as an agent is rational, her *beliefs* about her reasons will play a motivating role. As Scanlon (2014, 54) says, "if a rational agent believes that *p* is a reason to do *a*, she will generally do *a*, and do it *for this reason*."

To give one final example, it is one of the central principles of Dancy's (2000, 2) view of reasons that one and the same kind of thing is both the agent's normative reason, of the kind that could potentially explain what the agent ought to do, and also the agent's motivating reason, of the kind for which the agent might act. So Dancy also thinks that normative reasons play versions of both the normative-explanation role and the ideal-motivation role.

Clearly, there are a few differences between these philosophers' claims, especially in their formulation of the "ideal motivation role" that reasons must be capable of playing. In fact, however, these differences will not matter for my purposes. All that matters is that these "Reasons First" philosophers all claim that normative reasons play some version or other of both of these two roles. To simplify matters, however, I shall not try to find a precise formulation of the ideal-motivation role that all of these philosophers would agree to. I shall simply assume that they all accept that for every fact or proposition *p*, if *p* is a normative reason for an agent *x* to φ, it must be possible for the agent, if she is sufficiently rational and well-informed, to have an attitude, such as a belief or the like, towards this proposition *p*, and for this attitude to incline the agent to φ.

In Sections 4–6, I shall argue that it is basic mistake to assume that *anything* plays both of these two roles. The conclusion that we should draw is that the items that play the normative-explanation role – the normative explanation reasons – and the items that play the ideal-motivation role – the ideal-motivation reasons – are two different kinds of normative reasons. The basic presupposition of the "Reasons First" approach, that there is a single central concept of a normative reason, turns out to be false. It is this false presupposition that explains why the "Reasons First" theorists made the mistake of supposing that normative reasons must play both of these two roles.

Then, in Section 7, I shall propose an interpretation of the meaning of the term 'reason' in the relevant contexts. According to this interpretation, there is a further mistake in the "Reasons First" approach: not only is there no single central concept of a "normative reason", but none of the concepts expressed by 'reason' in these contexts is fundamental – on the contrary, they can all be defined in terms of other more basic normative concepts.

4. Criteria of rightness *vs.* ideal decision procedures

On the face of it, the assumption that the very same items – the normative reasons – play both the normative explanation role and the ideal-motivation role should seem dubious in the light of contemporary ethical theory. Famously, a number of ethical theorists – most notably, theorists in the consequentialist tradition, like Peter Railton (1991) – have insisted that we need to distinguish between a *criterion of rightness* and a *decision procedure*. A criterion of rightness is a principle that gives the ultimate explanation of the truth about which acts are right, and which are not. By contrast, a decision procedure – even an ideal decision procedure – is an actual mental process by means of which agents might make their decisions about what to do.

According to these ethical theorists, the question of which decision procedure is ideal – whatever exactly we mean by speaking of a decision procedure's being "ideal" – is not settled simply by determining what the correct criterion of right action is. It could well be that the ideal decision procedure will not always or even usually involve thinking consciously about the ultimate criterion of right action at all. Instead, the ideal decision procedure may involve the agent's simply manifesting certain ingrained motivational dispositions or habits of mind, which correspond to reliable rules of thumb – that is, rules that typically, in normal circumstances, lead to the agent's making the right decision, even if they need not do so infallibly in every case.

The "Reasons first" theorists who assume that one and the same kind of reasons play both the normative explanation role and the ideal motivation role may be tempted to think that this Railton-inspired objection does not apply to their approach. According to this assumption, after all, the ideally rational agent need not think about the *ultimate principle* according to which the normative-explanation reasons explain what the agent has most reason all things considered to do. These assumptions imply only that the ideal rational agent must think about, and deliberate from, these reasons themselves – which are normally thought of as contingent facts that play only a certain special sort of role in explaining the truth about what the agent has most reason to do.

Nonetheless, as I shall argue, there are parallel problems with the assumption that the very same kind of reasons plays both of these roles. It is not plausible that the justificatory story, which explains the truth about what one has most reason to do, or about what attitude one has most reason to have, has such a tight connection with the story about the ideal deliberative or motivational process

The first problem concerns the inevitable limits to the agent's knowledge. Unless the agent is that extraordinary genius, the brilliant moral philosopher of the future who will discover the whole ultimate truth of ethics, it seems overwhelmingly likely that the ultimate explanation of what the agent has most reason to do will not be known – at least not in full detail – by the agent herself.[5]

Few philosophers have explicitly recognized that the various different normative or evaluative truths differ with respect to how easy it is to know them. But on reflection it seems clear that normative and evaluative truths do differ in this respect. Some normative and evaluative truths are relatively easy to know, while others may be wholly unknowable. It is easy to know that the atrocity of September 11, 2001, was a wrongful act. On the other hand, it may be that – for reasons that have been spelled out Timothy Williamson (2000, chap. 4) – it is impossible for any moral thinker to pinpoint exactly where the threshold lies between the amount of altruistically helpful behaviour that is strictly morally required and the amount that is supererogatory.

In cases where the normative truth is hard or impossible to know, the ideal agent will take account of her epistemic limitations. In the case of altruistically helpful behaviour, for example, an ideally virtuous agent will typically do *more* than is strictly required of her – given that the chances of her exactly hitting the threshold between what is required and what is supererogatory are so low.

Why is this relevant to our discussion? Let us take a simple example. There is usually a reason against killing a person. But suppose that a person asks you to kill him – where this person has an excruciating degenerative disease, which will kill him in a few weeks' time if you do not kill him now, and knows that if you kill him now, you will thereby save the lives of many innocent people. It is not clear that in this case there is any reason against killing this person at all.

I am tempted to say that what this shows is that even in an ordinary case, the mere fact that an act is a killing is not the reason that explains why it should not be done; it is a more complicated

[5] This point is rightly stressed by Star (2015).

fact – such as the fact that the act (i) causes serious irreparable harm to an innocent victim, (ii) is done without the victim's consent, and (iii) accomplishes no good results sufficient to justify this non-consensual harming of an innocent. So, in most ordinary cases of acts of killing, the true "normative-explanation reason" against the act – that is, the reason that gives the ultimate explanation of the act's normative status – will be a complicated feature of the act of this sort. At all events, the bare fact that the act is a killing will not be a "normative-explanation reason" against the act.

Ordinary rational agents cannot, I believe, be expected always to be able to identify the true normative-explanation reason. Instead, ordinary agents will be directly motivated by simpler facts about their situation. The bare fact that an act is a killing will trigger an aversion that they have to killing, which will normally motivate them to refrain from killing. So, the bare fact that the act is a killing does seem to be an "ideal-motivation reason" against the act.

One might object that the ideal motivation assumption focuses on *ideal* motivational processes, and so abstracts away from the agent's epistemic limitations. In fact, however, as we have seen, there are many different versions of the ideal motivation assumption – only some of which abstract away from *all* of the agent's epistemic limitations, including the agent's ignorance of ultimate normative or evaluative truths as well as empirical truths. Anyway, even if we focus on agents who are ideally well informed about all relevant truths whatsoever, it is still not clear why their knowledge of what the ultimate normative-explanation reasons are should inform their actual processes of deliberation or motivation. They might still continue to respond with horror to the mere thought of killing, even if they know that the bare fact that an act is a killing is not a normative-explanation reason against the act. So, even if we focus on ideal agents of this sort, it is unclear whether the gap between the normative-explanation reasons and the ideal-motivation reasons has been closed.

5. Overestimating the importance of belief

There are further problems with the assumption that the very same reasons play both the normative-explanation role and the ideal-motivation role. Perhaps the most serious problem is that it leads to a distorted picture of rational deliberation and motivation – specifically, a picture that in a systematic way overestimates the centrality of *outright belief*.

This problem arises because it is plausible to assume that to say that one is rationally required to think in a certain way is to say – at least in a certain sense – that one *ought* to think in that way. If reasons play the normative-explanation role, there must be reasons that (perhaps together with the facts about which ways of thinking are available) explain how one ought or is rationally required to think. These reasons must be *facts* or *true propositions*. This is because the explanation of any fact must itself consist of facts; a sentence of the form '*p* because *q*' is true only if both *p* and *q* themselves are true.

If these reasons play not only the normative-explanation role, but also the ideal-motivation role, then it must be possible for the rational agent to respond to or deliberate from these reasons. More precisely, according to the formulation of the ideal-motivation role that I have given above, it must be possible for the rational agent to have an attitude like *belief* towards the true proposition that constitutes the reason, and to be appropriately motivated by that belief.

Indeed, some theorists – most notably, John Hyman (1999) – restrict the notion of "responding to reasons" still more narrowly, and insist that if a given fact is a reason for one to act or think in a certain way, the only way in which one can respond appropriately to this reason is by *knowing* this fact, and then reacting appropriately to one's knowledge. Even those theorists who do not interpret responding appropriately to reasons as responding to known facts in this way, the kind of belief that one must have in the facts that constitute the reasons that one is responding to is typically thought of, not as a partial belief or a level of confidence or credence that may fall short of certainty, but as an *outright belief*. To have an outright belief in a proposition is to be disposed simply to treat the proposition as true – as some philosophers would say, to treat the proposition as a "premise in practical reasoning".[6]

A few adherents of the "Reasons First" approach – most notably, Mark Schroeder (2011) – have relaxed this approach to the extent of supposing that the reasons that one can respond to are propositions that are the objects of a range of mental states which includes not only outright belief, but other types of mental state – such as episodic memories or perceptual experiences – as well. Schroeder suggests that the relevant range of mental states consists of what he calls "*presentational* attitudes"; but it is important to him that this range of mental states does not include all kinds of mental states, but is restricted to these "presentational" attitudes.

The picture that emerges is this: whenever you are rationally required to think in a certain way, the reasons that explain why you are rationally required to think in that way must all be facts that you are capable of believing (or having a presentational attitude towards) and rationally responding to by thinking in the way that is rationally required. The simplest explanation of this picture is that thinking rationally is simply a matter of in this sense "responding appropriately" to the reasons that explain the way in which we are rationally required to think.[7]

The problem with this account is that it implies that if the agent is thinking rationally in this way, then the way in which she is rationally required to think is explained purely by the outright beliefs (or presentational attitudes) that she has in true propositions. This account totally ignores all the *other* kinds of mental states that agents may have, apart from outright beliefs and other

---

[6] For further discussion of the distinction between partial belief and outright belief, see Wedgwood (2012).

[7] For example, one of the most distinguished "Reasons First" theorists, Raz (2011, 86) identifies "Reason" with "our reflective capacity to recognize and respond to reasons", and defines rationality as follows: "Reason, i.e. the rational powers or capacities, is involved in activities such as choosing, deciding, reasoning, These activities, and therefore their results, are rational so long as the rational powers guiding them function properly" (93).

presentational attitudes: in particular, it ignores the agents' partial beliefs or levels of confidence or credence; their desires, emotions, and preferences; their plans and intentions; and so on.

In cases of uncertainty, rational agents will typically be guided, not only by their outright beliefs, but also by their partial beliefs. For example, you might be guided by the fact that you have a 0.5 degree of belief in *p* and a 0.5 degree of belief in '¬*p*'. Since these propositions *p* and '¬*p*' are contradictories, only one of them is true; and so, since normative-explanation reasons must be true propositions, only one of them can be a normative-explanation reason of the way in which you are rationally required to think. Nonetheless, in being guided by your degrees of belief in these two propositions, you are no more responding to or being guided by the true proposition than by the false proposition, since your relationship to these two propositions is entirely symmetrical.

Moreover, the outright beliefs that a rational agent has do not determine how the agent is rationally required to think. It seems perfectly possible for there to be two agents who are thinking in an equally rational way, and have exactly the same outright beliefs, but differ in their partial degrees of beliefs – that is, in their levels of confidence or credence.[8] Within a Bayesian framework, so long as two agents had different "prior probabilities", their now having exactly the same outright beliefs – or the same "evidence", as many epistemologists would say – is quite compatible with their also having quite different probability assignments or partial beliefs.

In such cases, rationality may well require that these two agents should think in different ways. This result is certain to follow if what rationality requires is that our response to our cognitive situation must maximize the expectation of some sort of value (it does not matter for our purposes whether this value is utility or a value of some other kind), so long as the relevant "expectation" is to be calculated using the rational agent's partial degrees of belief in various hypotheses about the value of the various available responses to the situation.

It will not help to suggest that in these cases, the agent can form a true outright belief *about* her partial degrees of belief or credences, and respond appropriately to this outright belief. Not all rational agents have numerically precise concepts of the different degrees of belief (arguably, such concepts were not developed until the seventeenth century at the earliest); and so such agents could not form such outright beliefs about their own partial degrees of belief at all.

Moreover, even if a rational agent does have such outright beliefs about her own partial degrees of belief, and her partial degrees of belief make it rational for her to choose to buy a certain lottery ticket, it is not clear that it is rational (or "appropriate") for the agent to choose to buy the

[8] This would be denied by philosophers like Timothy Williamson (2000, chap. 10), who believe that there is a single special privileged probability function P such that every believer at every time rationally should proportion their credences to the result of conditionalizing P on the facts that constitute the reasons that they have at that time. But virtually all formal epistemologists would reject the presupposition that there is any single special privileged probability function of this sort.

lottery ticket in direct response to this outright belief about her own partial degrees of belief. It seems, rather, that if the agent is to make this choice rationally, the agent will not be directly responding to any such higher-order beliefs about her own degrees of belief; she will be directly responding to her degrees of belief themselves. After all, it is her rational degrees of belief that make it the case that the choice is rational, while her outright beliefs about these degrees of belief could in principle be false – in which case the choice might not be rational at all.

In short, it seems quite possible for a true proposition *q* to belong to the set of truths that explains why the agent is rationally required to make a certain choice, even if it is not possible for the agent to have an attitude like a belief (or other presentational attitude) towards this proposition *q*, and to rationally respond to this attitude by making the choice in question.

The "Reasons first" theorists cannot accept this point without significantly weakening the assumption that reasons play both the normative-explanation role and the ideal-motivation role. The simplest way to weaken this assumption would be by replacing it with the weaker claim that it must be possible for the agent to believe (or have a presentational attitude towards) and rationally respond to at least *some* of the reasons that determine how she is rationally required to think. However, there will be cases where one is so deeply uncertain about one's situation that the only true propositions that one rationally has an outright belief in are propositions about which options are available – in effect, propositions of the form 'I can φ' and the like. It seems bizarre to say that this proposition is a reason for φ-ing, and yet there may be no other true proposition that one is capable of having an outright belief in and rationally responding to in this case. So even weakening the assumption that the normative reasons play both the normative-explanation role and the ideal-motivation role in this way is not enough to save the "Reasons First" theorists from giving a badly distorted picture of rational reasoning in these cases involving partial belief and uncertainty.

6. Level confusions

At this point, it is tempting to object: surely there must be *some* link between the justificatory story, which appeals to the normative-explanation reasons, and the story of ideal motivation or deliberation, which appeals to the ideal-motivation reasons? Surely there must be some connection between the two?

It is clear, I think, that there is indeed a connection here. Once we get clear about the nature of this connection, however, the deepest error in the identification of normative-explanation reasons and ideal-motivation reasons will come to light. It is, fundamentally, a *level confusion* – that is, a confusion between (a) what must be known or grasped by the *theorist* who is giving an account of a certain sort of agent, and (b) what must be known or grasped by the *agent herself*.[9]

[9] See the seminal discussion of such level confusions that was given by Alston (1980).

Suppose that a certain process of deliberation or motivation is indeed ideal. To say that it is ideal is to make a normative or evaluative claim about it. In effect, this process has a certain normative or evaluative property – the property of being ideal in the relevant way. We can now ask: What *explains why* it is ideal in this way? The explanation that answers this question will identify the reasons why this process is ideal. These reasons that explain why the process is ideal can clearly in a sense be thought of as reasons in favour of that process: each of these reasons contributes towards explaining why one in some sense ought to engage in that process, or at least why it is in some way good for one to engage in the process.

Clearly, this explanatory story about why this process is ideal must be true *of* the agent. That is, the agent must exemplify or instantiate this story. Equally clearly, however, there is no reason why the agent herself needs to know or believe or think of or in any way be aware of this story. The agent can be rational without knowing why she is rational; indeed, in my view, an agent can be rational without knowing that she is rational or even possessing the concept of rationality.

In general, the rational agent and the virtuous agent do not themselves have to be theorists of rationality or virtue. The virtuous agent and the rational agent just need *dispositions* that reliably lead them to reason and to act in the ways that rationality and the virtues require – they do not need to have a theoretical understanding of the nature of these dispositions. So long as these agents are manifesting dispositions of this kind, it will be no accident that their actions and reasoning conform to the requirements of rationality and the virtues, and this seems to be enough to ensure that they are not just doing the right thing, but doing it "for the right reason" (as we say).

While the "Reasons First" theorists typically accept that the ideal agent need not know the whole explanation of why the acts that she performs are right, or why the reasoning that she is engaged in is rational, this theorist typically still insists that the ideal agent must respond to, or deliberate from – and so presumably must believe or have presentational attitude towards – some of the facts that figure in this explanation. As I have argued, there are serious grounds for doubting this. It seems that one of the mistakes that lie behind these theorists' position is a simple level confusion of this sort.

The problems that we have canvassed in Sections 4–6 seem to show that it is a mistake to identify the normative-explanation reasons with the ideal-motivation reasons. But it seems plausible that both kinds of reasons exist: we can, after all, perfectly well make sense both (a) of the facts that explain the truth about how the agent ought to act or to think (in various different senses of 'ought'), and (b) of the considerations that the agent is responding to, insofar as she is reasoning in a rational or ideal manner (in various different senses of 'rational' and 'ideal'). So, we should acknowledge that both kinds of reasons exist. In short, the normative-explanation reasons and the ideal-motivation reasons are two distinct kinds of normative reasons. In this way, one of the basic presuppositions of the "Reasons First" approach fails: there is not in fact a unique central notion of a normative reason; there are many different kinds of normative reasons, and nothing that makes any one of these kinds of reason any more central than any other.

7. The language of 'reasons'

If there are so many different concepts that can be expressed by talking about "normative reasons", how is it that all these concepts can be expressed by means of the same term? What meaning does the term 'reason' have in ordinary English that permits it to express all these different concepts?

In fact, the English language is somewhat unusual in the way in which it accommodates talking of reasons. It is particularly easy in English to combine the noun 'reason', not just with an infinitive of the form 'to φ', but with an infinitival phrase like 'for *x* to φ'. In many other languages, such constructions would be at least more awkward and less common, if they are even possible at all.

Indeed, there seem to be many languages that lack any word that corresponds to the English word 'reason' as it is used to refer to normative reasons. It is particularly striking that this is the case with one of the canonical languages of Western philosophy and Christianity – namely, ancient Greek. In ancient Greek, there does not seem to be any term that in any of its normal senses coincides in meaning with these uses of the English word 'reason'.[10] Greek certainly contains a rich array of words that correspond to 'good', 'right', 'ought', and the like. In effect, however, the easiest way to translate talk about normative reasons into ancient Greek would be by using explicitly explanatory terms (of which ancient Greek has many), and talking about *why* a certain action or choice is right or fine, or about *why* a virtuous agent would act as she should.

Historically, our word 'reason' derives, through the French term 'raison', from the Latin word 'ratio' – which comes from a verb that simply refers to *thinking* or *calculating*. As a historical matter, then, the origins of our word 'reason' lie in a word that referred to *reasoning*. In other languages, the word most closely corresponding to 'reason' has a quite different origin. For example, in German, the word most naturally used to translate the relevant uses of 'reason' is 'Grund', which literally means *ground*. In German, the word 'Grund' is very commonly used in an explanatory sense: what "grounds" something is what explains or causes it.

In general, I propose that this explanatory meaning is crucial. In Latin and the Romance languages, the meaning of the Latin 'ratio', the Italian 'ragione', and the French 'raison', was transferred from *reasoning*, to *right* or *correct* or *proper reasoning*, and from there to what is articulated by such correct or proper reasoning – namely, the *correct explanation* for something. In short, the word 'reason' functions as the nominalization of an explanation.

[10] For a compelling argument for the conclusion that Aristotle's use of the phrase 'right reason' (ὁ ὀρθὸς λόγος) refers to *the correct explanatory account*, see Moss (2014).

If the root meaning of the relevant uses of ‘reason’ in contemporary English is explanatory, then we can see why in some contexts it comes to refer to motivating reasons – since motivating reasons provide a certain kind of explanation for the action or attitude that is in question. When the context indicates that we have a normative concern with evaluating possible actions or attitudes, there are then two ways in which we could “normativize” this essentially explanatory meaning: either by focusing on the explanation of a normative fact, or by focusing on an idealization of possible motivating reasons.

In short, the occurrences of the word ‘reason’ that “Reasons first” theorists categorize as expressing the concept of a “normative reason” in fact express concepts of two fundamentally different kinds:

a. Concepts of *normative-explanation reasons* – where every such concept stands for a fact that contributes towards *explaining* a *normative fact* of some contextually salient kind.
b. Concepts of *ideal-motivation reasons* – where every such concept stands for a fact that could explain a possible response on the part of the relevant agent that would count as in the contextually salient way a good or ideal response.

In this way, if this account is correct, it would explain how the term ‘reason’ can end up expressing concepts of these fundamentally different kinds; it would also help to explain how the “Reasons First” theorists made the mistake of assuming that the term ‘reason’ expresses a single especially central normative concept.

In fact, however, there is a great plethora of concepts that can be expressed by the relevant uses of ‘reason’. As we have seen, some of these concepts fall into the two categories of normative-explanation ‘reason’-concepts and ideal-motivation ‘reason’-concepts. But it is also plausible that each of these two categories contains many concepts.

The concepts in the category of normative-explanation ‘reason’-concepts differ from each other depending on kind of normative fact that the items falling under the ‘reason’-concept contribute towards explaining. For example, they may contribute towards explaining “objective” normative facts, or towards explaining “subjective” or “information-relative” normative facts; they may contribute towards explaining normative facts that are “owned” by the agent who is mentioned in the sentence, or towards explaining normative facts that are not “owned” by that agent (as with the example ‘There is a reason for Alex to get a severe punishment’); and so on.

The concepts in the category of ideal-motivation ‘reason’-concepts differ from each other depending on the kind of idealization or goodness that would be exemplified by the possible response that the item falling under the ‘reason’-concept would explain. As we have seen, there are many different kinds of idealization or goodness of thinking that could factor into these different ‘reason’-concepts here.

8. Conclusion

As I said above, it seems to be a presupposition of the “Reasons first” program that there is one central concept of a normative reason that is more basic and central than all other normative concepts. If the arguments of this paper are correct, then we cannot identify any such concept just by pointing to the language that is standardly taken to express normative reasons: there is no unique concept expressed by those uses of language, but a big family of concepts instead.

There are two final manoeuvres that the “Reasons first” theorists might attempt at this point. First, they might try to identify one member of this big family of ‘reason’-concepts, and claim that that concept is the basic central concept in terms of which all normative phenomena are to be explained. But in fact all these concepts seem to be broadly parallel and analogous to each other, and so it seems at best arbitrary to claim that any one of these concepts is basic, and that all the rest are derivative. Indeed, discriminating invidiously in this way among the concepts that can be expressed by the term ‘reason’ seems worse than merely arbitrary: it seems positively implausible. Since all these concepts are broadly parallel and analogous concepts, it is just implausible to claim that any one of these concepts is any more basic or central than any other.

A second manoeuvre that the “Reasons first” theorists might attempt at this point might be to claim that this whole family of ‘reason’-concepts is as a whole conceptually more basic than all other normative concepts. But the explanation of ‘reason’-concepts that I have given above casts doubt on this manoeuvre too. To distinguish between these different ‘reason’-concepts, we needed to refer to different concepts that can be expressed by ‘ought’, or to different concepts of what is “ideal” or “good”. So the claim that the notion of a “reason” is more basic than those expressed by ‘ought’ and ‘good’ also looks doubtful.

In this way, if my linguistic account of the meaning of the term ‘reason’ is correct, then the “Reasons First” theorists are mistaken in a second way as well. If this account is correct, then it seems that none of the concepts that can be expressed by these uses of ‘reason’ is a primitive indefinable concept. On the contrary, it seems that all of these concepts can be defined by means of combining *explanatory* notions with normative notions that can be expressed by means of *other* terms (like ‘ought’, ‘right’, ‘good’, ‘rational’, and so on).

In this way, the “Reasons First” theorists have utterly misunderstood the layout of this region of conceptual space. The concepts that are expressible by the term ‘reason’ are not among the most fundamental normative concepts. They are among the very *least* fundamental normative concepts; it is presumably for this reason that so many natural human languages, like ancient Greek, get by perfectly well without having any term for this concept – whereas no human language could get by without having terms corresponding to ‘good’ and ‘ought’ and the like.

This is not to say that reasons of various kinds do not come first in any sense. Indeed, since the normative-explanation reasons are what explain normative facts, such reasons – by definition – come first in the order of normative explanation. But it is the *facts that constitute the reasons* that come first in the order of normative explanation. The *reason-relation* between these facts and the action or attitude for which they are reasons does not come first in the order of explanation at all. On the contrary, this reason-relation can be analysed in terms of the holding of a relevant explanatory relationship between (a) the fact that constitutes the reason and (b) some normative fact concerning the action or attitude for which the fact in question is a reason.

However, the “Reasons First” theorists do not claim merely that the facts that constitute the reasons come first in the order of normative explanation. They claim that the notion of a “reason” comes first in the order of conceptual analysis. This claim, as I have argued here, rests on a profound misunderstanding of the concepts involved.[11]

References

Alston, W. P. (1980). “Level-Confusions in Epistemology”, *Midwest Studies in Philosophy* 5: 135–150.

Broome, John (2004). “Reasons”, in *Reason and Value: Essays on the Moral Philosophy of Joseph Raz*, ed. R. J. Wallace, Michael Smith, Samuel Scheffler, and Philip Pettit (Oxford: Oxford University Press).

Broome, John (2013). *Rationality through Reasoning* (Wiley-Blackwell).

Dancy, Jonathan (2000). *Practical Reality* (Oxford: Oxford University Press).

Fogal, Daniel (forthcoming). “Reasons and Reason: Count and Mass”, in *Weighing Reasons*, ed. Errol Lord and Barry Maguire (Oxford: Oxford University Press).

Grice, H. P. (2001). *Aspects of Reason* (Oxford: Oxford University Press).

Hyman, John (1999). “How knowledge works”, *Philosophical Quarterly* 50 (197): 433-451.

Kearns, Stephen, and Star, Daniel (2009). “Reasons as Evidence”, *Oxford Studies in Metaethics* 4: 215-42.

Moss, Jessica (2014). “Right Reason in Plato and Aristotle: On the Meaning of *Logos*”, *Phronesis* 59, no. 3: 181-230.

Railton, Peter (1984). “Alienation, consequentialism, and the demands of morality”, *Philosophy and Public Affairs* 13 (2): 134-171.

[11] Earlier versions of this paper were presented in 2013 at a conference at the University of St Andrews, in 2014 at New York University, and in 2015 at the Massachusetts Institute of Technology. I am grateful to all those audiences for helpful comments.

Raz, Joseph (1999). *Engaging Reason* (Oxford: Oxford University Press).

Raz, Joseph (2011). *From Normativity to Responsibility* (Oxford: Oxford University Press).

Scanlon, T. M. (1998). *What We Owe to Each Other* (Cambridge, MA: Harvard University Press).

Scanlon, T. M. (2014). *Being Realistic about Reasons* (Oxford: Oxford University Press).

Schroeder, Mark (2007). *Slaves of the Passions* (Oxford: Oxford University Press).

Schroeder, Mark (2010). "Value and the right kind of reason", *Oxford Studies in Metaethics* 5: 25-55.

Schroeder, Mark (2011). "What does it take to 'have' a reason?" in Andrew Reisner and Asbjørn Steglich-Petersen, eds., *Reasons for Belief* (Cambridge: Cambridge University Press): 201-22.

Setiya, Kieran (2007). *Reasons without Rationalism* (Princeton, NJ: Princeton University Press).

Skorupski, John (2010). *The Domain of Reasons* (Oxford: Oxford University Press).

Smith, Michael (1994). *The Moral Problem* (Oxford: Blackwell).

Star, Daniel (2015). *Knowing Better* (Oxford: Oxford University Press).

Wedgwood, Ralph (2007). *The Nature of Normativity* (Oxford: Oxford University Press).

Wedgwood, Ralph (2012). "Outright Belief", *dialectica* 66, no. 3, Special Issue on Belief and Degrees of Belief, ed. Philip Ebert and Martin Smith: 309–329.

Williamson, Timothy (2000). *Knowledge and Its Limits* (Oxford: Oxford University Press).