

# Reasons for Action: Agent-Neutral vs. Agent-Relative

*First published Thu Aug 11, 2005; substantive revision Thu Dec 1, 2011*

The agent-relative/agent-neutral distinction is widely and rightly regarded as a philosophically important one. Unfortunately, the distinction is often drawn in different and mutually incompatible ways. The agent-relative/agent-neutral distinction has historically been drawn three main ways: the ‘principle-based distinction’, the ‘reason-statement-based distinction’ and the ‘perspective-based distinction’. Each of these approaches has its own distinctive vices (Sections 1–3). However, a slightly modified version of the historically influential principle-based approach seems to avoid most if not all of these vices (Section 4). The distinction so understood differs from numerous other distinctions with which it might easily be confused (Section 5). Finally, the distinction so drawn is important to normative theorizing for a variety of reasons (Section 6).

- [1. The Principle-Based Conception](#)
- [2. The Reason-Statement-Based Conception](#)
- [3. The Perspective-Based Conception](#)
- [4. The Principle-Based Conception Revisited](#)
- [5. Related Distinctions](#)
- [6. Why the Distinction Matters](#)
- [7. Conclusion.](#)
- [Bibliography](#)
- [Academic Tools](#)
- [Other Internet Resources](#)
- [Related Entries](#)

## 1. The Principle-Based Conception

The principle-based version of the agent-relative/agent-neutral distinction actually predates the terminology ‘agent-relative’ and ‘agent-neutral’. Thomas Nagel instead uses the terms ‘subjective’ and ‘objective’ to mark a version of the principle-based distinction in his classic *The Possibility of Altruism* (Nagel 1970). The terms ‘agent-relative’ and ‘agent-neutral’ were later introduced by Derek Parfit (Parfit 1984) and Nagel himself then adopted Parfit's terminology (Nagel 1986). As background to Nagel's version of the distinction, we must first note that for Nagel, reasons are universal, in the sense that for every token reason there corresponds a predicate  $R$  which figures in a universally quantified proposition of the following form:

Every reason is a predicate  $R$  such that for all persons  $p$  and events  $A$ , if  $R$  is true of  $A$ , then  $p$  has prima facie reason to promote  $A$ . (Nagel 1970: 47)

With this conception of the universality of reasons in hand, Nagel articulates the distinction as follows:

Formally, a subjective reason is one whose defining predicate  $R$  contains a free occurrence of the variable  $p$ . (The free-agent variable will, of course, be free only *within*  $R$ ; it will be bound by the universal quantification over persons which governs the entire formula.) All universal reasons and principles expressible in terms of the basic formula either contain a free-agent variable or they do not. The former are subjective; the later will be called objective. (Nagel 1970: 91)

Drawn in such formal terms, the distinction can seem alien and difficult to grasp but the basic idea is actually not that complex. A few examples should help illustrate Nagel's idea. Suppose that there is reason for me to

phone a friend because doing so would make the friend happy. Now suppose that my reason is best expressed as follows—that phoning her would make someone happy. In that case, the fact that the person who is made happy is *my* friend is incidental. If phoning a stranger would have generated just as much happiness then I would have had just as much reason to call the stranger. This in turn suggests that the principle corresponding to this reason is of the form,

$(p, A)$  (If  $A$  will make someone happy then  $p$  has reason to promote  $A$ )

The use of ‘promote’ in Nagel's canonical formulation raises some interesting issues. Nagel holds that performing an action  $A$  is a trivial way of promoting  $A$ , so I can promote the calling of my friend by calling her. In one sense, Nagel builds teleology into his conception of a reason for action. For Nagel glosses ‘promotion’ in such a way that the fact that an action will produce the relevant sort of outcome is sufficient for there being a reason to perform that action (see Nagel 1970: 52). So every reason entails that if someone can promote the relevant outcome (which outcomes are relevant may depend on the agents involved if the reason is agent-relative) then it logically follows from Nagel's gloss of reasons in terms of promotion (in his sense) that there is reason to perform that action. In another sense, though, Nagel does not build teleology into his conception of a reason for action. For Nagel does not hold that causally producing (or even instantiating) are *necessary* for there to be a reason for the action. In his somewhat capacious sense of ‘promote’, it is sufficient for an action to count as promoting an outcome if the action is such that not performing it would bring about the non-occurrence of the relevant state of affairs (again, see Nagel 1970: 52). So actions which are necessary but not sufficient for a given outcome thereby count as promoting that outcome. So all reasons are not teleological in the sense of being reasons in virtue of their causally bringing about the relevant sort of outcome; simply preserving the possibility of that outcome is enough. However, all reasons are teleological in the sense that any reason for an agent  $A$  to  $X$  does entail that if a suitable agent (who counts as suitable depends on whether the reason is agent-relative; if it is agent-neutral then any agent will do) can promote  $A$ 's  $X$ -ing then there is reason for them to do so. Reasons are for Nagel teleological in the sense that they all entail reasons for people to perform actions in the right circumstances in virtue of what states of affairs they would causally bring about. This is not trivial. Some conceptions of reasons do not have this consequence, as in the case of those conceptions which gloss reasons for action in a way that does not advert to promotion at all.<sup>[1]</sup>

In any event, for Nagel, the preceding principle (and hence the corresponding reason) is agent-neutral because the antecedent contains no use of the free-agent variable ‘ $p$ ’. The reason is in this sense not relativized to the agent for whom it is a reason. However, we could instead hold that the fact that it is my friend who would be made happy is relevant to whether I have reason to call. In that case, the principle corresponding to the reason would be of the form,

$(p, A)$  (If  $A$  will make  $p$ 's friend happy then  $p$  has reason to promote  $A$ )

This principle is agent-relative because the free-agent variable ‘ $p$ ’ does appear in its antecedent. Basically, if the sufficient condition for the application of the reason predicate (the condition given by the antecedent of the principle corresponding to the reason) includes such a free-agent variable then the reason is agent-relative; otherwise it is agent-neutral. It is easy enough to see that on this conception, *ethical egoism* is an agent-relative theory (and hence concerns agent-relative reasons) while objective utilitarianism is an agent-neutral theory (and hence concerns agent-neutral reasons). For egoism holds that there is reason for a given agent to do something just in case his doing it would promote *his* welfare. Whereas objective *utilitarianism*, see the entry on [consequentialism](#), (on at least one version) holds that someone ought to do something just insofar as it promotes welfare, period (no matter whose it is). It is also important to be clear that the principles Nagel has in mind must be understood as the basic normative principles of a theory rather than the theory's account of what gives those principles their status as basic normative principles.<sup>[2]</sup> Otherwise various meta-ethical (see entry on [metaethics](#)) theories of what it is to be a reason (e.g., informed desire accounts) might be taken to imply agent-relativity when in fact they should be understood as neutral on this issue.

Nagel's version of the distinction is ‘principle-based’ in the fairly straightforward sense that one must first look to the principle corresponding to a given reason to determine whether it is agent-relative or agent-neutral. Nagel also makes this clear in his later work. In *The View From Nowhere* he holds that,

If a reason can be given a general form which does not include an essential reference to the person who has it, it is an *agent-neutral* reason... If on the other hand, the general form of a reason does include an essential reference to the person who has it then it is an *agent-relative* reason. (Nagel 1986: 152–153)

The context makes it relatively clear Nagel's reference to the 'general form' of a reason is a universally quantified principle corresponding to the reason of the sort he discussed in *The Possibility of Altruism*. Derek Parfit, who is the first one to introduce the terminology 'agent-relative' and 'agent-neutral' makes it even more clear that the distinction is in the first instance applied to normative theories. Having described a moral theory he calls *C*, he remarks that,

Since *C* gives to all agents common moral aims, I shall call *C* *agent-neutral*. Many moral theories do not take this form. These theories are *agent-relative*, giving to different agents, different aims (Parfit 1984: 27).

Parfit later explains how his terminology, when applied to reasons, maps onto Nagel's:

Nagel's *subjective* reasons are reasons only for the agent. I call these agent-relative... When I call some reason agent-relative, I am not claiming that this reason *cannot* be a reason for other agents. All that I am claiming is that it may not be. (Parfit 1984: 143)

This is slightly confusing. First, how does Parfit's distinction between agent-relative and agent-neutral theories map onto his distinction between agent-relative and agent-neutral reasons? The former is cast in terms of common aims while the second is cast in terms of whether a reason for one agent must also be a reason for anyone else. Second, how exactly does Parfit's distinction between agent-relative and agent-neutral reasons correspond to Nagel's version of the distinction? The former is cast in terms of whether a reason for one agent must also be a reason for anyone else while the latter is cast in terms of the occurrence of free agent variables in the sufficient condition for the application of the reason predicate corresponding to the reason. These are not obviously equivalent, yet Parfit takes his distinction to be the same as Nagel's.

The first of these two questions is slightly easier to answer than the second. A moral theory is agent-neutral if it gives us common aims, but if we have common aims then whenever there is a reason for you to promote an aim there will also be reason for me to promote that aim (if I can). So Parfit's agent-neutral moral theories provide agent-neutral reasons in his sense—in the sense of being such that a reason for one agent is guaranteed to be a reason for any agent situated to promote the end which figures in that reason. By contrast, a moral theory is agent-relative if it does not give every agent a common aim. However, if we do not have common aims then what is a reason for you may be no reason whatsoever for me *even if* I am in a position to promote the end which figures in that reason. So Parfit's agent-relative moral theories provide agent-relative reasons in his sense—reasons which are not such that a reason for one agent entails a reason for any agent in a position to promote the end which figures in that reason. So Parfit's agent-relative moral theories concern agent-relative reasons in his sense and his agent-neutral moral theories concern agent-neutral reasons in his sense. So the use of the same terminology is no coincidence, and presumably Parfit would agree that the agent-relativity of a reason is well understood in terms of the agent-relativity of the principle associated with that reason. Indeed, the primary reason we have included Parfit's version of the distinction in this discussion of the principle-based version of the distinction is the suspicion that his distinction between agent-relative and agent-neutral reasons is really just a corollary of his distinction between agent-relative and agent-neutral theories. The fact that Parfit clearly thinks that his distinction between agent-relative and agent-neutral reasons is just the same as Nagel's distinction also suggests that Parfit's distinction is well understood as principle-based, since Nagel's version of the distinction is a paradigmatic instance of a principle-based version of the distinction. The other (closely related) reason for including Parfit's version of the distinction here is that it also shares the chief vice of Nagel's version. As explained below, neither Nagel's nor Parfit's version of the distinction sits well with radical forms of moral particularism.

What about our second question about Parfit's distinctions, though? Why should Parfit assume his version of the distinction between agent-relative and agent-neutral reasons is the same as Nagel's even though they are not drawn in the same way? Here it is relevant that (in at least one sense; see above) Nagel builds teleology into the logical form of the reason predicate, so that a reason is always understood as a reason *to promote* something. Given this conception of reasons, it does seem that Parfit's conception maps fairly neatly onto

Nagel's. For if a reason is agent-relative in Nagel's sense, then whether there is a reason for me to promote the state of affairs for which it is a reason will depend on whether I am the agent who figures in the token reason under consideration. An agent-relative reason to promote *A*'s happiness will give me a reason only if I am *A* or suitably related to *A* (e.g., there can be agent-relative reasons to promote the welfare of *my* nearest and dearest). So reasons which are agent-relative in Nagel's sense are agent-relative in Parfit's sense. By contrast, if a reason is agent-neutral in Nagel's sense then it gives any agent who can promote the state of affairs it favors a reason to promote that state of affairs. In Nagel's colorful phrase, agent-neutral reasons in his sense transfer "across the gap between persons" (Nagel 1970: 79). So reasons that are agent-neutral in Nagel's sense are agent-neutral in Parfit's sense as well. Moreover, so long as we assume that all reasons are teleological, any reason which is agent-relative in Parfit's sense is agent-relative in Nagel's sense and any reason that is agent-neutral in Parfit's sense is agent-neutral in Nagel's sense. For plausibly the only way a reason for one person to promote *X* could fail to also provide someone else who could promote *X* a reason to do so would be if the reason was indexed to me in the way Nagel's agent-relative reasons are. So long as we hold onto the assumption that all reasons for action are teleological, Parfit's version of the distinction and Nagel's version of the distinction are at least extensionally equivalent in all possible worlds in spite of their rather different formulations of the distinction.

In different ways, both Nagel and Parfit formulate the agent-relative/agent-neutral distinction in terms of general principles and the distinctions they draw do indeed seem useful and important (more on why they are useful and important below). For a more explicit defense of drawing the distinction in terms of general principles (or 'rules' in their terminology), see also McNaughton and Rawling 1991.<sup>[3]</sup> However, one unfortunate consequence of these classical principle-based ways of drawing the distinction is that they are incompatible with radical forms of *particularism* of the sort recently defended by Jonathan Dancy and others. There are actually many forms of moral particularism (see Jonathan Dancy's entry on particularism in this Encyclopedia; see also McKeever and Ridge 2005a), but particularists are united in their embrace of holism in the theory of reasons. On a holistic conception of reasons, a consideration which functions as a reason in one context need not function as a reason at all in another context. For example, that my calling my friend would give him pleasure may be a reason here, but not a reason in another context (perhaps his pleasure would be purely sadistic in the second case, taking pleasure in some misfortune of mine). The atomist impulse is to build whatever is necessary for the consideration's status as a reason into the reason itself, so that whatever is a reason in one context is thereby guaranteed to be a reason in any other context as well. So in the preceding example, the atomist would insist that the second example just shows that we mischaracterized the reason in the first case. The real reason to call my friend in the first case is not merely that it will give him pleasure but rather that it will give him innocent pleasure, and that consideration is not present in the second case. Particularists argue at some length that we should resist this atomistic impulse and stick with the more natural characterization of the reason in the first case and accept a holistic conception of reasons. Particularists then argue that holism in the theory of reasons in turn supports the conclusion that morality is not well understood in terms of principles (for critical discussion of this move, see McKeever and Ridge 2005b).

Particularism is an interesting position and it would be a shame if some plausible version of the agent-relative/agent-neutral distinction were not compatible with the leading ideas behind particularism. However, it seems that both Nagel's and Parfit's versions of the distinction do not sit well with particularism. This is most clear in the case of Nagel. His doctrine of the universality of reasons directly contradicts holism about reasons yet plays an essential role in his account of what he calls the 'general form' of a reason which in turn plays an essential role in his version of the agent-relative/agent-neutral distinction. Parfit's distinction between agent-relative and agent-neutral theories of course presupposes that we should think about morality in terms of normative theories, which also contradicts many forms of particularism. Finally, Parfit's distinction between agent-relative and agent-neutral reasons holds that a reason is agent-neutral only if it is guaranteed to provide a reason for any agent who can promote the end which figures in the reason (again, we are assuming all reasons are teleological and hence always involve an end). Particularists will reject the idea that just because something is a reason for one agent it thereby must be a reason for any other agent who can promote the end which figures in the reason. For on a holistic conception of how reasons function, there may be some further aspect of another agent's situation which cancels the force of this consideration as a reason even though he could promote the relevant end. If this is right then for this sort of particularism it will follow trivially that all reasons are agent-relative and this makes the distinction useless for their purposes. So perhaps unsurprisingly, the principle-based version of the agent-relative/agent-neutral distinction does not seem like a promising one if we want to make sure that the distinction is useful to particularists as well as

their generalist opponents. We should therefore consider whether either the reason-statement-based version of the distinction or the perspective-based version of the distinction are more likely to be available to particularists as well while still doing roughly the same sort of important philosophical work that Nagel and Parfit had in mind.

## 2. The Reason-Statement-Based Conception

The second version of the distinction foregoes Nagel's appeal to the 'general form' of the reason, but instead is formulated in terms of the reason itself. We call this the 'reason-statement-based version' of the distinction because it holds that whether a reason is agent-relative depends on whether a full statement of the reason itself (forget about its "general form") involves pronominal back-reference to the agent for whom it is a reason. Philip Pettit puts forward a reason-statement-based version of the distinction:

An agent-relative reason is one that cannot be fully specified without pronominal back-reference to the person for whom it is a reason. It is the sort of reason provided for an agent by the observation that *he* promised to perform the action in prospect, or that the action is in *his* interest, or that it is to the advantage of *his* children. In each case, the motivating consideration involves essential reference to him or his.... An agent-neutral reason is one that can be fully specified without such an indexical device. (Pettit 1987: 75)

On Pettit's account, whether a reason is agent-relative depends on whether it (the reason) can be fully stated without pronominal back-reference to the agent for whom it is a reason. The use of anaphoric pronouns in Pettit's version of the distinction does not really mark much of a departure from Nagel's formulation. For Nagel's 'free agent variables' are really technical devices which function just like anaphoric pronouns by referring back to the agent for whom the consideration is a reason. The real difference between Nagel's formulation and Pettit's is that Nagel's formulation is cast in terms of the general form of a reason whereas Pettit's formulation is cast in terms of a full statement of the reason itself, which need not involve any mention of the general form of the reason. For Pettit, the full statement of a reason need not involve a universal quantifier at all. One of Pettit's illustrative examples of a full statement of a reason is "that the action is in *his* interest" and this statement involves no universal quantification whatsoever. So Pettit's version of the distinction, and more generally the reason-statement-based version of the distinction, seems unlikely to offend against particularist sensibilities.

Moreover, Pettit's distinction seems to do much the same philosophical work as Nagel's version of the distinction. For in both cases agent-relativity is characterized in terms of a certain sort of back-reference to the person for whom the consideration is a reason while agent-neutrality is in both cases characterized in terms of the absence of such back-reference. The only difference is that for Nagel the relevant back-reference is to be found in the general form of the reason (which turns out to be a universally quantified principle giving a sufficient condition for the application of the reason predicate) whereas for Pettit the relevant back-reference is to be found in a full statement of the reason itself. So far, it looks like we have a kind of dominance argument for the reason-statement-based version of the distinction. It seems likely to do much the same work as Nagel's distinction and mirrors the logical form of Nagel's distinction, but does so without making the distinction unavailable to particularists.

However, the reason-statement-based version of the distinction has problems of its own. Most significantly, this version of the distinction seems to presuppose a particular ontology of reasons, in which case we end up inviting particularists to the agent-relative/agent-neutral party only by rescinding the invitations of those who hold opposing ontological views about reasons. For suppose I think that reasons just are facts and also hold that there are no distinctively indexical facts. Rather, there are ordinary facts which can be characterized in indexical terms or characterized in non-indexical terms. So the fact that having a beer at 6:13 p.m. on January 17, 2005 would make MR happy is identical to the fact that having a beer now would make me happy. Here we have one fact which can be expressed in two different ways. Given this package of ontological views, the reason-statement-based version of the distinction looks unhelpful. For let us suppose that the fact that having a beer now would give me pleasure provides an agent-relative reason for me to have the beer—a sort of egoistic reason. On the reason-statement-based version of the distinction, the reason is agent-relative only if a full statement of the reason *must* involve pronominal reference to the agent for whom it is a reason. However, if reasons just are facts (this is my ontology of reasons) then it is simply not true that

a full statement of this reason (read: this fact) *must* involve pronominal back-reference to me. For instead of saying that having a beer would make him happy, you could say that my reason is that having a beer would make MR happy. On the ontology under consideration, this seems like a full statement (what is left out?) of the fact which is my reason to have the beer just as much as the anaphoric pronominal statement of my reason. Indeed, given these ontological views it is hard to see how any reason could be agent-relative since any fact can be given a full statement without the use of indexicals of any kind. Moreover, the problem here is not limited to this particular ontology. Someone who holds that reasons are true propositions or states of affairs and who also holds that there are no irreducibly indexical propositions or states of affairs (only indexical ‘modes of presentation’ of such propositions or states of affairs) will have trouble making use of the reason-statement version of the distinction for just this reason.

This is not to say that there are not ontologies which would provide a framework in which the reason-statement version of the distinction could be useful. Two ontologies spring to mind. First, those who hold that a reason is not just a fact but a fact plus a particular mode of presentation (a particular way of grasping the fact, as it were) might well be able to make good use of the reason-statement-based version of the distinction. Second, those who hold that reasons are just facts but also hold that there really are irreducibly indexical facts can also make sense of the reason-statement-based version of the distinction. Still, it seems like a considerable cost that this version of the distinction seems to exclude so many reasonable ontological views. So whereas Nagel's principle-based version of the distinction excludes particularists, Pettit's reason-statement-based version of the distinction excludes those who hold any of a wide range of ontological views about reasons. We should see if we can do better.

### 3. The Perspective-Based Conception

The third way in which the agent-relative/agent-neutral distinction has historically been drawn is in terms of the perspectives from which the reasons in question can be recognized as reasons. The basic idea is to specify a suitably objective perspective and hold that agent-neutral reasons can be appreciated as such from that perspective while agent-relative reasons cannot. Jonathan Dancy seems to read Nagel in this way:

Nagel holds that there are three kinds of reason. The first are stubbornly subjective reasons, such as those which are in play when we are choosing from a menu in a restaurant... But there are two classes of objective reasons. The first are agent-relative reasons... the second are agent-neutral reasons. Both are recognizable at some distance from here, as we shave peculiarities from our perspective in the move towards objectivity. The agent-relative ones are less objective, of course, though they can be recognized, and in some sense endorsed, from more objective points of view. However, what is recognized and endorsed is not the importance which the agent finds in (e.g.) his own life-time projects; this itself cannot be recognized from much further out. As we move away from the agent's own perspective, all that can be recognized is that he finds importance in them, which is quite a different matter. (Dancy 1993: 146)

This reading of Nagel is to some extent invited by Nagel's use of the metaphor of a ‘view from nowhere’. Moreover, it is true that in *The Possibility of Altruism* Nagel does argue that agent-relative reasons (which he then called ‘subjective’) cannot be recognized from a certain sort of impersonal perspective. However, this was precisely supposed to be the conclusion of a substantive and highly controversial argument which Nagel has since rejected (on the strength of an argument from Nicholas Sturgeon; see Sturgeon 1974). If agent-relative reasons cannot be appreciated from a suitably objective perspective then Nagel quite clearly thinks that this will need to be established by argument. It is *not* part of his definition of agent-relativity; otherwise he could have skipped most of the argument of *The Possibility of Altruism* as his main conclusion would have been established by linguistic fiat. As we have seen, he instead defines agent-relativity in terms of the general form of the reason.

Still, even if this is not what Nagel had in mind might this approach not be philosophically useful and important? It would, after all, have the virtue of being a distinction which is available to particularists. For the idea of more and less objective perspectives is not obviously incompatible with even very radical forms of particularism. Nor does it seem to exclude a variety of ontological views about reasons as the reason-statement-based version of the distinction did. However useful the distinction is, though, it is not one which is well-suited to doing the work traditionally associated with the agent-relative/agent-neutral distinction. For



a start, a plausible litmus test for any proposed version of the distinction is that it uncontroversially classifies egoistic reasons as agent-relative and utilitarian reasons (to maximize happiness, simpliciter) as agent-neutral. These are perhaps the most commonly cited paradigms of each sort of reason, after all. Also, the distinction is often thought to capture in a more abstract sense what Henry Sidgwick was discussing when he spoke of a ‘dualism of practical reason’ between reasons of self-interest and reasons of general benevolence (see Sidgwick 1907).

The crucial point is that it is far from obvious that the perspective-based distinction satisfies this litmus test. Nagel's attempt to prove that agent-relative reasons in his sense (including egoistic reasons) cannot be appreciated from a suitably objective perspective is widely regarded as a failure, as are many other attempts to refute egoism (and agent-relativism more generally) by showing it is incompatible with some suitable objective perspective. So perhaps egoistic reasons can, after all, be appreciated even from an ideally objective perspective on any of a wide range of conceptions of objectivity. Perhaps not, though; perhaps some clever argument will still show that agent-relativity is incompatible with some important and independent notion of objectivity. In any event, this remains at best a very controversial hunch rather than something which has been uncontroversially demonstrated. This alone should give us great pause about adopting the perspective-based approach to the agent-relative/agent-neutral distinction. Egoistic reasons are paradigmatic agent-relative reasons, and it should be trivial that they come out as such rather than a matter of great controversy. Moreover, the same point applies not only to egoistic reasons but to all of the typically invoked paradigms of agent-relativity. Reasons arising out of special relations to one's nearest and dearest are also paradigmatically agent-relative, but it is also far from clear that they will come out as such on the perspective-based approach. So even if we keep a single conception of objectivity as fixed there will be great controversy over whether seemingly paradigmatic instances of agent-relativity really are such. This is unfortunate. The agent-relative/agent-neutral distinction should be one which is useful for *framing* these debates at the outset (for more on this, see Dreier 1993) and not one which we can deploy with confidence only after the debates have been settled. However, there will also be great controversy over just what the right conception of objectivity is as well as controversy over the implications of any given conception thereof. This threatens to mire the agent-relative/agent-neutral distinction in such controversy from the outset as to make it virtually useless. Much better instead to define agent-relativity and agent-neutrality in terms of the logical form of the principle corresponding to the reason and then let it be a matter of substantive debate whether agent-relative reasons so understood can be appreciated from various objective perspectives.

## 4. The Principle-Based Conception Revisited

We seemed to have reached a dialectical cul de sac. The principle-based approach draws the distinction nicely but is inaccessible to particularists. The reason-statement-based approach to the distinction is tenable on certain ontological views about reasons but seems useless on a wide variety of other plausible ontological views about reasons. Finally, the perspective-based approach threatens to just change the subject or at least to mire the application of the distinction down in so much controversy as to make it useless as a tool for framing the debates which it seems so naturally suited to frame.

Fortunately, though, the principle-based approach can actually be understood in a way that is compatible with particularism after all, and this is how we propose to understand the distinction (here we go beyond the existing literature, but we do so in the spirit of a ‘friendly amendment’ to Nagel's account). The key move is to not build Nagel's conception of universality into the articulation of the distinction. So we might allow that a given consideration (or state of affairs, or feature of a situation; insert your favored ontology of reasons here) is a reason in one case but that very same consideration might be no reason at all (or even a reason with the opposite valence) in another situation due to some difference between the two situations. At least, our drawing of the distinction should not rule this out. In this way, we can make the distinction compatible with the particularist's favored holistic conception of reasons.

This very quickly leads to an awkward question, though. Nagel drew his distinction in terms of principles which incorporated his conception of universality. For Nagel's principles were universally quantified generalizations which held that *whenever* an agent can promote a certain sort of state of affairs there is reason to do so. So we need a different way of understanding ‘the general form of a reason’ or as one might prefer to put it, the principle corresponding to the reason. Fortunately, there is an alternative conception ready to hand. As Sean McKeever and Ridge argue elsewhere (see McKeever and Ridge 2006) there is a

species of hedged moral principle which is compatible with the particularist's holistic conception of reasons. We call these 'default principles', and they are well situated to make Nagel's principle-based distinction available even to as hardy a particularist as Jonathan Dancy. As background, recall that on the holistic conception of how reasons work, what is a reason here may be no reason at all or even a reason with the opposite valence elsewhere. So for example, the fact that it would give a father pleasure is a reason for his son to give him a kiss on the cheek but that very same consideration (that it would give him pleasure) may be no reason at all for him to watch a snuff film. Intuitively, the status of the fact that it would give him pleasure as a reason is 'defeated' in the latter case by the fact that the pleasure would be sadistic (or an expression of depravity, or whatever). Holists therefore call such facts 'defeaters'. We also can have what holists call 'enablers'—facts which are necessary for some other fact to function as a reason. With this machinery in hand we are in a position to see how default principles might be helpful in crafting a version of the agent-relative/agent-neutral distinction that everyone (or just about everyone) can live with.

Given holism, a true and non-trivial principle about reasons must accommodate the possibility of defeaters. We could do this in one of two ways. First, we could try to list all of the different possible defeaters in the antecedent and claim that none is present. However, a hardy particularist may insist that this is a fool's game, on the grounds that there is no saying in advance when we have a complete list of all possible defeaters. After all, part of the particularist idea is that morality is far too complex for any such principles ever to be within our ken. However, there is a second approach, and this is the approach which inspires our conception of default principles. Instead of trying to list all of the possible defeaters and countervailing reasons we could instead quantify over them. The proposal is most easily understood through illustrative examples. Returning to our example of pleasure and sadism, consider the following default principle:

- (P) For all possible agents ( $p$ ), and all possible actions ( $x$ ) and all facts ( $F$ ) If  $F$  is a fact to the effect that  $p$ 's  $x$ -ing would promote pleasure and no other feature of the situation explains why  $F$  is not a reason to  $x$  then  $F$  is a reason to  $x$ .

(P) is compatible with the particularist's holistic conception of reasons. For in those cases in which the status of a fact about pleasure as a reason is defeated by sadism (e.g.) the 'no other feature of the situation explains why  $F$  is not a reason...' clause is not satisfied. Moreover, as we argue elsewhere, it is hard to see how particularists could really object to principles as modest as (P). After all, (P) is compatible with the thesis that there are indefinitely many possible types of reasons and indefinitely many possible defeaters corresponding to each of those reasons. So the mere availability of default principles does not entail that the normative landscape could be finitely (much less manageably) codified in some short set of axioms like Ross's list of *prima facie* duties, for instance. Furthermore, the availability of such principles in logical space does not in itself entail that they are presupposed by the very possibility of moral thought and judgment. So the availability of such principles is compatible with Dancy's canonical formulation of particularism, according to which, "the possibility of moral thought and judgement does not depend on the provision of a suitable supply of moral principles" (Dancy 2004: 7).

What is the payoff of such principles in terms of the agent-relative/agent-neutral distinction? Default principles allow us to draw the agent-relative/agent-neutral distinction in much the same way that Nagel drew it but *without* thereby excluding particularists. An initial gloss of the distinction might go as follows. A given default principle will either include a free-agent variable in the statement of the consideration which is the reason or not. If it does then the reason is agent-relative; if not then the reason is agent-neutral. However, this is not quite right. For every statement of a reason will be a statement about an action which is a possible action for the agent for whom it is a reason. I cannot have reason to perform an action that only someone else could perform, after all. Since this sort of back-reference to the agent is entirely trivial, we must explicitly add to our definition that it is not sufficient to make a reason agent-relative. Otherwise, all reasons will implausibly come out as agent-relative for this trivial reason. Indeed, Derek Parfit already noticed this in his discussion, explaining that,

Even if you and I are trying to achieve some common aim, we may be in different causal situations. I may have reason to act in a way that promotes our common aim, but you may have no such reason since you may be unable to act in this way. Since even agent-neutral reasons are, in this sense, agent-relative, this sense is irrelevant to our discussion. (Parfit 1984: 143)



This suggests that we should slightly revise our statement of the agent-relative/agent-neutral distinction. The default principle corresponding to a given reason will either include a *non-trivial* free-agent-variable in the statement of the reason or not. If it does then the reason is agent-relative; otherwise it is agent-neutral. The idea is that the use of a free-agent variable to indicate that the action is one available to the agent for whom the fact is a reason is trivial in the sense that it must be included in the statement of any reason whatsoever. So the reasons associated with the default principle (P) [see above] are agent-neutral, as the only use of the free-agent variable 'p' is the trivial one that indicates that  $x$  is a possible action of  $p$ 's. By contrast, the reasons associated with the following principle would be agent-relative:

(P\*) For all possible agents ( $p$ ), and all possible actions ( $x$ ) and all facts ( $F$ ) If  $F$  is a fact to the effect that  $p$ 's  $x$ -ing would promote  $p$ 's pleasure and no other feature of the situation explains why  $F$  is not a reason to  $x$  then  $F$  is a reason to  $x$ .

Unlike (P), the statement of the reason given in (P\*) includes a non-trivial use of a free-agent variable in its insistence that the pleasure promoted must be  $p$ 's. So the reasons associated with (P\*) come out as agent-relative. It is not hard to see that the proposed reading of the distinction should easily pass the litmus test given above, classifying objective utilitarian reasons as agent-neutral (as with (P)) and egoistic reasons as agent-relative. Moreover, it is easy to see that the proposed reading of the distinction would sort other paradigmatic instances of agent-relativity and agent-neutrality in an intuitively satisfying way. Abandoning Nagel's assumption of universality in favor of our more modest default principles seems to allow us to make good sense of his distinction without excluding even fairly radical forms of particularism.

However, before we rest too easily with this conclusion, we must first address an important objection to default principles.<sup>[4]</sup> The challenge insists that they are all vacuously true simply in virtue of their logical form. Those sympathetic to particularism might think that even if default principles capture a kind of logic of reasons, they cannot be the locus of substantial moral knowledge because they are vacuously true. While serious, the vacuity objection is not sound, and seeing why will both clarify the content of our proposal and reveal why particularists should not be too eager to embrace default principles. Since default principles are universally quantified conditionals, they can be false only if they have an instantiation in which the antecedent is true and the consequent is false. The worry behind the vacuity objection is that the logical form of a default principle in conjunction with some very plausible assumptions about moral explanation entails that whenever the consequent of a default principle is false, its antecedent will be false as well, in which case the principle itself is vacuously true. Consider the following inane default principle:

LY: For all actions, ( $x$ ) (If (a)  $x$  would be done in a leap year and (b) no other feature of the situation explains why the fact that  $x$  would be done in a leap year is not a moral reason not to  $x$  and (c) the reasons in favor of  $x$  do not explain why  $x$  is not wrong in virtue of the fact that  $x$  would be done in a leap year then  $x$  is wrong in virtue of the fact that  $x$  would be done in a leap year).

LY is clearly absurd. For an action is wrong in virtue of a given fact or set of facts in the intended sense of 'in virtue of' only if the fact(s) in question is (or are) moral reason(s) not to perform the action which carry the day. We take it that the fact that an action would be done in a leap year could never be a moral reason to perform the action.<sup>[5]</sup> So LY had better come out as false. The vacuity objection, however, insists that LY turns out to be trivially true. Moreover, the objection continues, the reason LY is true carries over to any default principle.

LY is a universally quantified conditional ranging over possible actions and hence is false only if it has a possible instantiation in which its antecedent is true and its consequent is false. Presumably its consequent is necessarily always false—no action could be made wrong in the relevant sense by being done in a leap year—the fact that an action is done in a leap year could never in itself be a reason not to perform an action. So the crucial issue is whether any of its instantiations also has a true antecedent. The antecedent is itself a conjunction and hence will be true only if both conjuncts are true. However, the objection runs, the second of the three conjuncts [(b) above] is necessarily false or is false for all possible actions. We are taking it as given that the leap year fact is not a moral reason, but the objection insists on general grounds that it is implausible to suppose that there is no explanation of its failure to be a reason. Moral facts are not arbitrary, at least in the extremely minimal sense that whenever something is or is not a reason there will be *some* explanation of why this is so. Even someone who thought that morality was a direct function of the arbitrary will of God should admit this much, since we will on such an account always be able to explain moral

differences in terms of differences in God's will even if we cannot go on to explain why God willed one way rather than another. So moral differences can always be explained. Hence clause (b) in any default principle will always be false, which is enough to make the antecedent false and so enough to make the conditional trivially true.

An adequate understanding of why the vacuity objection is unsound requires careful attention to a further detail. The vacuity objection goes from the premise that every time a fact is not a moral reason there is some explanation for its failure to be a reason to the conclusion that the 'no further feature of the situation explains...' clause of the corresponding default principle will always be false whenever the fact in question is not a reason. However, this inference is valid only if it is valid to go from 'there is some explanation of  $p$ ' to 'some feature of the situation explains  $p$ ' and this inference is invalid. The inference would be valid only if we were to interpret 'feature of the situation' so broadly that any possible explanatory fact can count as a feature of the situation. This is not at all an intuitive reading of 'feature of the situation', nor is it the intended reading. Hence the vacuity objection is unsound.

For present purposes, the main point is just that a feature of a situation must be a contingent fact. Necessary facts apply equally to all possible situations and hence are never features of any particular situation in our sense. This is already enough to accommodate the plausible idea that every moral fact has an explanation of some kind while blocking the vacuity objection. For quite plausibly, the moral facts in question will be explained by some necessary fact. For example, the explanation of why the fact that a given action would be done in a leap year is not a moral reason not to perform that action might simply be that the fact that an action would be done in a leap year could never be any kind of reason for action. That explanation is perhaps not the most illuminating one, but it is a kind of explanation and it is an explanation given in terms of a necessary fact. An alternative (and more controversial) explanation might appeal to the putative fact that a given fact is a moral reason only if it is a fact about how the action bears in some way on welfare or treating people with respect. This is a very controversial explanation, but our point is simply that if the main premise of this explanation (a kind of pluralism involving utilitarian reasons and deontological reasons) is true then it very plausibly is a necessary truth. Once again, the fact in question does indeed have an explanation but it is not explained by a feature of the situation. So there is no need to reject the plausible suggestion that every moral fact has some explanation to refute the vacuity objection.

More importantly, no contingent feature of the situation could plausibly figure in an explanation of why the leap year fact is not a reason in a given case. Indeed, because the leap year fact could *never* be a reason its failure to be a reason here will in no way depend on any of the contingent features of this case. So none of those contingent features of the case at hand will figure in an explanation of why that fact is not a reason here. If the leap year fact were sometimes a reason then things would be very different. For in that case the contingent features that distinguish the situation in which it is a reason from those in which it is not a reason could intelligibly figure in an explanation of why the fact is not a reason here. Moreover, if asked to cite some particular contingent feature of the situation that explains why that fact is not a reason here any sane moral agent would simply be perplexed. Hence for any default principle citing a patently absurd candidate reason (and whose consequent is not necessarily true) there will be possible instantiations of that principle in which the antecedent is true and the consequent is false. This line of argument is perfectly general, so any principle citing a putative reason in its antecedent which could never actually be a reason (and whose consequent can be false; principles with tautologous consequents will of course be trivially true on anybody's account) will turn out to be false. The only reason to doubt that the antecedent would sometimes be true was the thought that the 'no feature of the situation explains...' clause is always trivially false. Since we have seen this rests on a failure to distinguish explanations in general from explanations cast in terms of contingent features, we can conclude that default principles are not all vacuously true simply in virtue of their logical form. So we can safely invoke default principles to articulate the agent-relative/agent-neutral distinction and thereby capture Nagel's basic idea without making the distinction useless to radical moral particularists like Dancy.

## 5. Related Distinctions

The agent-relative/agent-neutral distinction is a very useful and philosophically important one. However, as with all distinctions, its usefulness evaporates when it is confused with other related but different distinctions. This sort of confusion is depressingly commonplace, perhaps in virtue of an unfortunate

tendency for philosophers to use terms like ‘neutral’, ‘objective’, and ‘relative’ without always being completely explicit about what those terms are supposed to mean. To guard against such connotations, this section canvasses a number of distinctions with which one might easily confuse the agent-relative/agent-neutral distinction and explains how each differs from it. These distinctions are divided into six groups, where the distinctions are put into the same group insofar as they all have the same feature(s) in common with the agent-relative/agent-neutral distinction.

The first of these families of distinctions consists of those that are like the agent-relative/agent-neutral distinction in that they are drawn in terms of a relativization of the reason to the agent who has the reason, but in a different way from the way in which the agent-relative/agent-neutral distinction is. Only one widely employed distinction clearly falls into this family: Bernard Williams' distinction between internal and external reasons. On Williams' account, a reason for acting is internal just in case it counts as a reason in virtue of its connection to the agent's “motivational set” (desires, intentions, pro-attitudes, etc.); otherwise it is external (see Williams 1981b). It is not hard to see how this distinction might easily be confused with the agent-relative/agent-neutral distinction, for one might easily suppose that internal reasons just are agent-relative ones, while external reasons just are agent-neutral ones. However, the distinctions are not the same, for a reason may be external and still be agent-relative. Suppose, for example, that we accept a default principle according to which the fact that an agent's culture demands something is sometimes a reason to do it. Such a reason will be agent-relative in virtue of the use of a free-agent variable to indicate that it is the agent's own culture which determines what reasons she has. Or consider a default principle according to which the fact that an action would satisfy the agent's biological needs is sometimes a reason. Again, the relativization to the agent (here to the agent's needs) entails that such a reason is agent-relative. Each of the reasons grounded by either of these latter two principles will be both agent-relative *and* external, for an agent simply may not care about her culture's standards or her biological needs. Hence, reasons for acting also can be both external and agent-relative.

In a second family of distinctions, we find distinctions that are like the agent-relative/agent-neutral distinction in that they are also drawn in terms of the principles underwriting an agent's reasons for acting, but are unlike it in that they are not drawn in terms of their relativity to the agent who has the reasons. Two main distinctions fall into this family: universality/non-universality and generality/non-generality. Agent-neutrality is often confused with universality. A reason is universal insofar as any person, *A*, who judges that one agent, *B*, has a reason is thereby committed to making the same judgment of anyone else, *C*, whom they take to be in relevantly similar circumstances. So long as we assume reasons are associated with principles, this will mean that the principle associated with the reason will be universal in scope; that is to say, it will be of the form, “For all *x*, if *x* is an agent, then...” Note that this is much weaker than Nagel's conception of universality; even default principles employ universal quantifiers and are universal in this sense. Universality in this thin sense is rightly taken to be a highly uncontroversial, perhaps even trivial, feature of reasons. However, it would of course be a mistake to infer from this that agent-neutrality should be uncontroversial, for the concepts are quite different. Agent-relative reasons, as well as agent-neutral reasons, can satisfy universality in this sense. Nor, of course, is agent-neutrality the same as universality in Nagel's somewhat stronger sense according to which whatever is a reason in one case must be a reason anywhere. This is obvious once it is made explicit, but historical attempts to derive agent-neutrality from such forms of universality encourage the conflation (see, e.g., Hare 1963: 112–136).

Agent-neutrality is also easily confused with generality, where a reason is general just in case the principle that underwrites it contains no proper nouns or “rigidly designating” descriptions (see Kripke 1972). Like universality, generality is a function of the principles underwriting the agent's reasons. Upon reflection, though, it should be clear that the general/non-general distinction simply cuts across the agent-relative/agent-neutral distinction; agent-relative reasons could be either general or non-general, as could agent-neutral ones. For example, “The fact that God commands *X*-ing is a reason to *X* unless some other feature of the situation explains why it is not,” would be non-general (assuming that ‘God’ is a rigid designator) yet agent-neutral, whereas, “The fact that God commands each person to do as her conscience dictates is a reason for each person to do as her conscience dictates unless some other feature of the situation explains why it is not a reason” would be both non-general and agent-relative.

A third family of distinctions is like the agent-relative/agent-neutral one in that it is drawn in terms of a relativization to the agent for whom the consideration is a reason, but is *not* drawn in terms of the principle underwriting the reason. Here there are two very similar distinctions that merit discussion. In fact these two

distinctions are so similar that they might easily be confused with one another as well as with the agent-relative/agent-neutral one. The first of these two distinctions is the distinction between “deliberator relative” (DR) principles and “deliberator neutral” (DN) principles (see Postema 1998). Unlike the agent-relative/agent-neutral distinction, the DR/DN distinction is not one that is drawn in terms of the form of those principles themselves. Unlike agent-relativity or agent-neutrality, deliberator relativity and deliberator neutrality cannot simply be “read off” from an accurate statement of the principles themselves. For the DR/DN distinction concerns not the question of the form of the principles, but the question of the source of their authority, or if one prefers, their “force.” A principle's having force for a given agent is then glossed as being such that the agent must recognize the principle's validity *to avoid counting as irrational*. With this conception of force in play, the distinction is usefully characterized in the following way:

A principle is DN iff it has force for every possible agent, which is to say all rational agents must recognize the principle's validity to avoid counting as irrational.

A principle is DR iff (a) its force varies from one possible agent to the next, which is to say at least some possible rational agents might reject the principle's validity without thereby being irrational, or (b) it has force for no possible agents. This is just the denial of DN.

Once the DR/DN distinction is explicitly articulated and compared with the AR/AN distinction, it is clear that they are different distinctions, doing different theoretical work.

The other distinction falling into this third family is very similar to the DR/DN distinction. It also concerns the force, rather than the form, of a practical principle. The DR/DN distinction is in terms of which principles an agent must recognize as binding on her *to avoid counting as irrational*. A slightly different distinction is the between principles that really are binding on everyone (BN—“binding neutral”), *even if one might not accept their authority without thereby counting as irrational*, and those that are not (BR—“binding relative”). Drawing the distinction may commit us to the thesis that there is a fairly fundamental appearance/reality distinction to be drawn even with respect to principles of practical reason, so that even an ideally rational agent could, in principle, be mistaken at a given point in time about which principle(s) bind her *without thereby counting as irrational*. At any rate, it should be clear that this distinction, like the closely aligned DR/DN distinction is quite different from the AR/AN distinction.

A fourth family of distinctions with which the AR/AN distinction might easily be confused is of those distinctions that are like the agent-relative/agent-neutral distinction in that it is drawn in terms of relativization, but unlike it in that it is *not* in terms of a relativization to the agent who has the reason. There is perhaps only one important distinction that clearly falls into this category: what Nicholas Sturgeon has helpfully referred to as “appraiser-relativism” (Sturgeon 1994). Though this may not be essential to appraiser-relativism, it is worth noting that, unlike agent-relativism, appraiser-relativism is typically presented as a semantic thesis about terms of ordinary language (as opposed to technical terms like ‘agent-neutral’), to the effect that the truth value of one and the same ethical or practical judgment can vary from one appraiser to the next. On such views, my judgment that what Hitler did was wrong could in principle be true while another person making the very same judgment in a rather different context (but one in which Hitler's actions, their context and consequences are kept constant) could be false. By contrast, the judgment that there was agent-relative reason for Hitler to perform a given action will have a truth-value that is invariant across different appraisers. Agent-relativism is a substantive view about what kinds of reasons people have, and is distinct from the semantic thesis of appraiser-relativism. Appraiser-relativism involves a relativization to the person appraising an action, rather than a relativization to the agent who might perform the action.

A fifth family is of distinctions that in some sense divide reasons into categories that might usefully be thought of as the “private” and the “non-private.” In this respect, they are intuitively similar to the agent-relative/agent-neutral distinction, in that there is a recognizable sense in which agent-relative reasons are private—their status as reasons for an agent is irreducibly a function of the features of *that* agent as such. Correspondingly, agent-neutral reasons are public in the sense that this is, by definition, not true of them. At least two other distinctions are usefully thought of as dividing reasons into the categories of private and non-private. However, each of these two distinctions mark the private/non-private division in an importantly different sense from the one in which the agent-relative/agent-neutral distinction does.

The first of these two distinctions is one between reasons that are “essentially shared,” in a somewhat technical sense, and those that are not. This distinction is perhaps the one that is most frequently confused with the agent-relative/agent-neutral one, and the philosophical consequences are significant. As the distinction is typically drawn, a reason is supposed to be “essentially shared” just in case whenever the reason is a reason for one agent to perform an action it is equally a reason for *anyone* to promote his performance of that action; otherwise it is not. So, for example, if my reason to take a walk is essentially shared, it follows that it is equally a reason for anyone to promote my taking a walk. Put more quaintly, the question of whether reasons for acting are essentially shared is the question of whether a reason for me must provide everyone else with a corresponding reason to help me do as that reason recommends, insofar as they can. It is not hard to see how there is a sense in which this distinction divides reasons into the public and the non-public, as essentially shared reasons, unlike those that are not essentially shared, provide reasons for everyone who can promote the state of affairs in which people act in accordance with those reasons. That being “shared” has become a rather technical notion should be apparent, since in more ordinary terms you and I may both share a reason, that doing *X* would be pleasant say, where each of our reasons provides no corresponding reasons for the other.

It is also quite easy to see how this distinction might be confused with the agent-relative/agent-neutral one. Suppose one embraces the tempting view that all reasons for acting must be *teleological* in form, which is to say that any principle underwriting a reason for acting must individuate actions in terms of the states of affairs that they promote. However, the agent-relative/agent-neutral distinction itself should not be understood as incorporating such a highly controversial assumption, and the version of the distinction proposed in section IV as an improvement on Nagel and Parfit therefore does not incorporate it. Given the assumption of teleology, though, and given the rejection of holism in the theory of reasons (which Nagel's universality implicitly does reject) it would indeed follow that agent-neutrality and being essentially shared are necessarily co-extensive, since any universal (in Nagel's sense) agent-neutral principle will then have the following form:

- (A) The fact that *p*'s *X*ing would promote *N* [where *N* is some state of affairs which is specified with no non-trivial use of '*p*'] is a reason for *p* to *X*.

It follows trivially from the supposition that all practical principles have this form that whenever one agent has reason to promote a state of affairs that *any* agent who can promote *N* will have reason to do so.

However, the view that all reasons for acting are teleological, while tempting, is a substantive doctrine that one might reasonably reject. Teleology is certainly not a trivial feature of our understanding of practical reasons. To take just one example, T.M. Scanlon has recently argued at some length that not all reasons are teleological and that the assumption that they are badly distorts our conception of practical reason (Scanlon 1998: 79–107). So it would be a mistake to allow one's attraction to that substantive view to lead one to conflate the distinction between the agent-neutral/agent-relative distinction and the essentially shared/not essentially shared distinction. For insofar as one rejects that assumption, one can very well allow that there are reasons which are both agent-neutral and *not* essentially shared. Consider, for example, the following principle:

- (4) The fact that an action would be an instance of violently assaulting someone is a reason not to perform it unless some other feature of the situation explains why it is not.

The reasons generated by this principle are clearly agent-neutral, but they are *not* essentially shared, for it simply does not follow from this principle that I have any reason to perform actions that would reduce the total number of violent incidents (either by others or myself in the future). I may very well have such further reasons, but they would be associated with a different principle, and therefore might not be *as weighty* as the reasons associated with the above agent-neutral principle.

On a historical note, it is perhaps no surprise that these two distinctions are so often conflated. For we have seen that Nagel, the original and most influential proponent of the agent-neutral/agent-relative distinction, explicitly commits himself to the view that all reasons are teleological in form, and is driven by this view to conflate the two distinctions. Recall that Nagel embraces the following teleological conception of practical principles:

[E]very reason is a predicate  $R$  such that for all persons and events  $A$ , if  $R$  is true of  $A$ , then  $p$  has *prima facie* reason to promote  $A$ . (Nagel 1970: 47)

Nagel thinks this position is just an unproblematic simplification on the grounds that he treats the performance “of act  $B$  as a degenerate case of promoting the occurrence of act  $B$ ” (Nagel 1970: 47). The reason it is *not* merely an unproblematic simplification is that putting all practical principles into this form robs us of the ability to say that there is agent-neutral reason for an agent to but no reason for her to promote  $B$ ing, except in the admittedly degenerate sense in which  $B$ ing is a way of promoting  $B$ ing.

Of Nagel's critics, Christine Korsgaard has been most sensitive to the way in which he builds teleology into his view without argument from the very beginning. She notes, for example, that, “Nagel treats all reasons as reasons to *promote* something...Nagel is in danger of ending up with consequentialism because that is where he started” (Korsgaard 1996a: 300). In spite of this perceptive diagnosis of Nagel's error on this front, though, Korsgaard occasionally blurs the two distinctions, claiming that the criterion that the reason predicate not contain a “free agent variable” is simply a more formal way of saying that these reasons are “common property” and not “personal property,” which on her view amounts to the thesis that they are essentially shared, as we have explicated that thesis (see Korsgaard 1996b: 276). That a philosopher perceptive enough to note and so accurately diagnose Nagel's mistake could still blur the two distinctions is a testament to the depth of the confusion about these distinctions embedded throughout the current philosophical literature. For another instance of this mistake, see McNaughton and Rawling 1995a as well as Dreier 1993.

A second distinction falling into this larger family of distinctions is the distinction between reasons that are intersubjective, and those that are not, where intersubjectivity is cashed out in terms of the possibility of an agent's successfully communicating the force of the reason to other agents. Korsgaard has emphasized this distinction in a number of places and argued that all reasons for acting must be intersubjective (e.g., Korsgaard 1996b: 131–166). It is not hard to see how non-intersubjective reasons might plausibly be thought of as private and intersubjective reasons might be thought of as public; hence it is not too hard to see how that distinction might be confused with the agent-relative/agent-neutral one given the tendency to confuse that distinction with the private/public distinction. Still, once intersubjectivity/non-intersubjectivity is explicitly defined, it is relatively clear that it is different from the agent-relative/agent-neutral distinction. For the former distinction is drawn in terms of communicability and the latter distinction makes no reference whatsoever to communicability. Indeed, for all that has been said so far, a reason might be intersubjective and be either agent-relative or agent-neutral. Still, it is surprisingly easy to run together these two distinctions. One way this confusion might arise would happen in two stages. First, one might confuse the agent-relative/agent-neutral distinction with the distinction between reasons that are essentially *shared* and those that are not. Since intersubjective reasons are often characterized as reasons “we can *share*,” the stage is set for confusing both of those two distinctions with the further distinction between intersubjective and non-intersubjective reasons. For it would not be that difficult to confuse shareability with being essentially shared.

Finally, in the sixth family of distinctions, we find a distinction that might easily be confused with the agent-relative/agent-neutral distinction as a kind of historical artifact. Here we have in mind the distinction between “reasons for an agent to do something” and “reasons for something to happen.” Without actually confusing them himself, Nagel may have unintentionally encouraged this conflation in remarking, for example, that, “Ethics is concerned not only with what should happen, but also independently with what people should or may *do*. Neutral reasons underlie the former; but relative reasons can affect the latter” (Nagel 1986: 165). In fact, however, it should be clear that these two distinctions are different. The agent-relative/agent-neutral distinction concerns the form of a practical principle, whereas the other distinction concerns whether a reason is a reason for an agent to do something or a reason for something to happen. One way of making the separateness of these two distinctions especially vivid is to note that someone might embrace the former distinction and admit that there are both kinds of reasons, but reject the very notion of a “reason for something to happen” as resting on an obscure, confused notion of a reason that can float free of all possible agents. This is not to say that we should in fact reject that notion; Wilfrid Sellars argues that a distinction of this sort is an important one (see Sellars 1968: 175–229. See also Castaneda 1975, with a discussion of Sellars's distinction). Rather, it is simply to note that it is not obviously misguided to reject the very concept of a reason for something to happen, and we should leave room for someone to do this without supposing all reasons for acting are agent-relative, for that is simply another question. Even if I grant that all



reasons must be reasons for someone, in that a reason presupposes a possible agent, I can still hold that the principle underlying those reasons need not make any non-trivial, ineliminable pronominal back-reference to that agent. Hence the two distinctions are different.

In sum, then, the agent-relative/agent-neutral distinction shares a number of features with several other distinctions, and therefore is easily confused with those other distinctions. These other distinctions have been broken down into six families, where those families are divided in terms of what the distinctions in question have in common with the agent-relative/agent-neutral distinction. Having discussed all six of these families at some length, it may be useful to provide a concise review of the defining features of each family:

1. Distinctions that are like the agent-relative/agent-neutral distinction in that they are drawn in terms of a relativization to the agent who has the reason, but in a different way: *internal/external*.
2. Distinctions that are like the agent-relative/agent-neutral distinction in that they are drawn in terms of the form of principles underwriting reasons, but unlike the agent-relative/agent-neutral distinction are not drawn in terms of relativization at all: *universal/non-universal* and *general/non-general*.
3. Distinctions that, like the agent-relative/agent-neutral distinction involve relativization to the agent who has the reason, but unlike the agent-relative/agent-neutral distinction, are not drawn in terms of the form of the principles underwriting practical reasons but instead are drawn in terms of the source of the authority of those principles: *deliberator-relative/deliberator-neutral* and *bindingness-relative/bindingness-neutral*.
4. Distinctions that are like the agent-relative/agent-neutral distinction in that they involve relativization, but that are unlike it in that the relativization is to the appraiser of the reason, rather than to the agent who has the reason: *appraiser-relative/appraiser-neutral*.
5. Distinctions that are like the agent-relative/agent-neutral one in that there is a sense in which they divide reasons into private and non-private categories, but that do so in a different sense from the one in which the agent-relative/agent-neutral distinction does so: *intersubjective/non-intersubjective* and *essentially-shared/not-essentially-shared*.
6. Distinctions with which the agent-relative agent-neutral distinction might be confused primarily as a kind of historical artifact: *reasons-for-something-to-happen/reasons-to-do*.

## 6. Why the Distinction Matters

Having formulated the agent-relative/agent-neutral distinction and seen how it differs from other important distinctions, we are now in a position to consider why the distinction is an important one. The distinction has played a very useful role in framing certain interesting and important debates in normative philosophy.

For a start, the distinction helps frame a challenge to the traditional assumption that what separates so-called consequentialists and deontologists is that the former but not the latter are committed to the idea that all reasons for action are teleological. A deontological restriction forbids a certain sort of action (e.g., stealing) even when stealing here is the only way to prevent even more stealing in the long run. Consequentialists charge that such a restriction must be irrational, on the grounds that if stealing is forbidden then it must be bad but if it is bad then surely less stealing is better than more. The deontologist can respond in one of two ways. First, they could hold that deontological restrictions correspond to non-teleological reasons. The reason not to steal, on this account, is not that stealing is bad in the sense that it should be minimized but rather simply that stealing is forbidden no matter what the consequences (this is admittedly a stark form of deontology, but there are less stern versions as well). This is indeed one way of understanding the divide between consequentialists and deontologists, but the agent-relative/agent-neutral distinction, and in particular the idea of agent-relative reasons, brings to the fore an alternative conception. For arguably, we could instead understand deontological restrictions as corresponding to a species of reasons which are teleological after all so long as those reasons are agent-relative. If my reason not to steal is that I should minimize my stealing then the fact that my stealing here would prevent five other people from committing similar acts of theft does nothing to suggest that I ought to steal. To really have any chance of working, the reasons will probably need to be temporally relative as well as agent-relative. For otherwise the reason corresponding to a deontological restriction will give me reason to steal now if this is the only way to prevent me from stealing even more later.<sup>[6]</sup> If the reasons in play are agent-relative then perhaps the deontologist can do more to defuse the consequentialist's charge of paradox, though other problems now arise. The deontologist now can look overly self-indulgent, so obsessed with the purity of her own soul that she will not sacrifice her integrity for

the greater good (see Ridge 2001a). Another worry is that reasons which are both agent-relative and temporally relative are not really teleological after all in any interesting sense. For the only way to promote an action *right now* is simply by performing it. The broader and more standard conception of promoting an action by causing it simply has no foothold here, and if it did then the proposal would not correspond exactly with deontological intuitions. In spite of these worries, many philosophers have characterized the reasons corresponding to deontological restrictions as agent-relative. Indeed, the characterization of deontological restrictions as agent-relative (or agent-centered) is close to being an orthodoxy.<sup>[7]</sup>

If we can properly understand the reasons corresponding to deontological restrictions as agent-relative (and temporally relative) teleological reasons but teleological reasons all the same then in effect we can, as James Dreier puts it, ‘consequentialize’ deontology, surprisingly enough. The apparent success of deploying agent-relativity to ‘consequentialize’ deontology leads Dreier to defend the more bold hypothesis that *any* moral theory can be represented as a form of consequentialism so long as we are willing to allow that consequentialism comes in agent-relative as well as agent-neutral versions. The central idea behind consequentialism, on this way of thinking, is its teleology and commitment to maximizing, both of which seem compatible with agent-relativity about that which is maximized.<sup>[8]</sup> If Dreier is right about this then the agent-relative/agent-neutral distinction may be more important than the distinction between consequentialist theories and non-consequentialist theories.

Another advantage of the agent-relative/agent-neutral distinction is that it can bring to our attention important structural differences between what might otherwise seem like very similar normative theories. For example, a theory which holds that our overriding reason is always to maximize actual utility seems very similar in spirit to a theory which instead holds that our overriding reason is always to maximize *expected* utility. One would naturally have thought that both of these theories are agent-neutral so long as we assume that utility is understood in agent-neutrally (in terms of general happiness, e.g.). However, the reference to ‘expected’ utility in the second theory actually seems to entail that the reasons corresponding to that theory are agent-relative. For presumably the relevant expectations are the agent's, in which case we will need to mark this with a free agent variable and that free agent variable hardly seems trivial.<sup>[9]</sup> This is a surprising result, but there is no obvious way to block it on the conception of agent-relativity proposed here (or on the classical principle-based conceptions defended by Nagel, Parfit or even on the reason-statement-based version of the distinction defended by Pettit). This could itself be illuminating. Perhaps it suggests that we need to draw a fundamental distinction between value and reasons, contra T.M. Scanlon and others who see value claims as really just indicating the presence of reasons; see Scanlon's discussion of the ‘buck-passing account in Scanlon (1998). For such a distinction would allow us to say that while happiness is an agent-neutral good (thereby accommodating the intuition that there is something agent-neutral about the second theory) our reasons to promote that good are best understood in terms of the agent's expectations and hence agent-relative.

Nor should we forget that the first real use of the agent-relative/agent-neutral distinction was Nagel's in *The Possibility of Altruism*. There Nagel tried to prove that all reasons must be agent-neutral on pain of a kind of practical solipsism. Nagel eventually abandoned this argument in light of objections from Nicholas Sturgeon (see Sturgeon 1974), but the argument is ingenious and Nagel may have abandoned it prematurely. If any such argument could ever be made to work then we might be able to settle a wide range of difficult issues in normative philosophy without simply appealing to first-order intuitions about cases which so often seems to lead to philosophical stalemate. Moreover, if an argument like Nagel's could be made to work then its implications would be dramatic. Not only egoistic reasons but arguably deontological reasons and reasons arising out of special relations to one's nearest and dearest would stand refuted, as would what Nagel later referred to as ‘reasons of autonomy’ (see Nagel 1986: 165). Furthermore, Nagel is not the only one to have offered abstract considerations in favor of the thesis that all reasons are agent-neutral. For example, some of Derek Parfit's work on personal identity is supposed to undermine the importance of personal identity as such, and that in turn might undermine the tenability of agent-relativity (see Parfit 1984).

The agent-relative/agent-neutral distinction was also invaluable in James Dreier's exploration of the often neglected issue of how an expressivist might make sense of agent-relative norms (see Dreier 1996). Dreier's argument is subtle and complex, and we shall not try to reproduce it here. The point for present purposes is that his discussion highlights an important challenge for expressivists. Admittedly, a more narrow version of this challenge was previously seen by Brian Medlin (see Medlin 1957), whose work heavily influenced

Dreier. However, Medlin cast the challenge specifically in terms of egoistic reasons and that has important dialectical implications. As Dreier points out, Medlin's challenge applies to agent-centered norms more generally and this wider scope matters. For we might well be willing to abandon egoistic reasons, but if we also had to give up on the intelligibility of deontology then the costs of expressivism might well begin to seem too steep. Until Nagel and others drew the agent-relative/agent-neutral distinction, it was easy enough for philosophers like Medlin to not fully appreciate the scope and power of their own arguments.

Finally, the agent-relative/agent-neutral distinction can also provide a useful lens through which to examine some of the arguments of historical figures. Sidgwick's famous discussion of the 'dualism of practical reason' can now be seen as an instance of the more general tension between agent-relative and agent-neutral reasons. G.E. Moore's argument against ethical egoism would, if sound, refute agent-relative conceptions more generally (see Moore 1903: 96–105), since Moore's main objection is not to egoism in particular but (in effect) to agent-relative conceptions of the good more generally.

Also, there is an interesting debate about whether Kantian moral prohibitions must be understood as agent-relative even if we allow that all reasons are teleological (see Ridge 2009; compare Huckfeldt 2007). The strategy is to understand the end to be promoted in agent-neutral terms, but also in such a way that the agent at any given point in time can best promote that agent-neutral aim only by acting in accord with suitable deontological rules. One key idea behind this strategy is that the relevant end is "good willing" understood in a broadly Kantian way, combined with a robust theory of free will according to which one agent cannot ever fully control the will of another. The other key idea is that the agent-neutral teleological reasons in play are compatible with a non-maximizing theory of right action. In particular, the idea will be that the agent must always minimize the risk of the very worst of the available outcomes. If the very worst of the available outcomes is that everyone has a bad will, then one can ensure that the risk of that outcome is zero by preserving one's own good will. Given that one cannot in this way fully control the will of another, the needed self/other asymmetry is preserved without agent-relativity in the theory of reasons or value. A suitable theory of right action and free will can do the work that would otherwise require an agent-relative theory of reasons or value.

## 7. Conclusion.

The agent-relative/agent-neutral distinction is extremely important to a wide range of debates in normative philosophy. Yet the distinction is often drawn in very different ways, with the risk that philosophers are simply talking past one another. In this entry, different ways of drawing the distinction have been distinguished, and the virtues of a modified version of the principle-based approach has been defended. The distinction so drawn is different from a wide range of other distinctions with which it might easily be confused; these distinctions are laid out here to help guard against such confusions. Finally, the distinction so drawn is an important one in structuring central debates in normative theory, such as how to understand the divide between consequentialists and deontologists.

## Bibliography

- Broome, John, 1995. "Skorupski on agent-neutrality." *Utilitas*, 7: 315–17. [[Preprint available online](#)]
- Castaneda, H.-N., 1975. *Thinking and Doing*, Dordrecht, Holland/Boston, USA: D. Reidel Publishing Company.
- Cummiskey, D, 1996. *Kantian Consequentialism*, New York: Oxford University Press.
- D'Agostino, F. and Guas, G., 1998. *Public Reason*, Aldershot, England: Dartmouth Publishing Company.
- Dancy, J., 1993. *Moral Reasons*, Oxford: Blackwell Press.
- , 2004. *Ethics Without Principles*, Oxford: Oxford University Press.
- Dreier, James, 1993. "Structures of Normative Theories". *The Monist*, 76: 22–40. [[Preprint available online](#) (in PDF)]
- , 1996. "Accepting Agent Centred Norms." *Australasian Journal of Philosophy*, 74: 409–422. [[Preprint available online](#) (in PDF)]
- Hare, R.M., 1963. *Freedom and Reason*, Oxford: Oxford University Press.
- Herman, B., 1993a. *The Practice of Moral Judgment*, Cambridge, MA: Harvard University Press.
- , 1993b. "Leaving Deontology Behind. In Herman 1993a: 208–240.


- Huckfeldt, V., 2007. "Categorical and Agent-Neutral Reasons in Kantian Justifications of Morality," *Philosophia*, 35 (1): 23–41.
- Hurley, P., 1997. "Agent-Centered Restrictions: Clearing the Air of Paradox." *Ethics*, 108 (1): 120–146.
- Kagan, S., 1989. *The Limits of Morality*, Clarendon Press: Oxford.
- Korsgaard, C., 1996a. *Creating the Kingdom of Ends*, Cambridge: Cambridge University Press.
- , 1996b. *The Sources of Normativity*, Cambridge: Cambridge University Press.
- Kripke, S., 1972. *Naming and Necessity*, Cambridge, MA: Harvard University Press.
- Mack, Eric, 1989. "Against Agent-Neutral Value" *Reason Papers*, 14: 76–85. [[Preprint available online](#) (in PDF)]
- , 1998. "Deontic Restrictions are Not Agent-Relative Restrictions." *Social Philosophy and Policy*, 15: 61–83.
- McKeever, S. and Ridge, M., 2005a. "The Many Moral Particularisms." *Canadian Journal of Philosophy*, 35(1): 83–106.
- , 2005b. "What Does Holism Have to Do With Particularism?" *Ratio*, 18 (1): 93–103.
- , 2006. *Principled Ethics: Generalism as a Regulative Ideal*, New York: Oxford University Press.
- McNaughton, D., 1988. *Moral Vision*, Oxford: Blackwell.
- McNaughton, D. and Rawling, P., 1991. "Agent-Relativity and the Doing-Happening Distinction." *Philosophical Studies*, 63: 167–185.
- , 1995a. "Value and Agent-Relative Reasons." *Utilitas*, 7 (1): 31–47.
- , 1995b. "Agent-Relativity and Terminological Inexactitudes." *Utilitas*, 7 (2): 319–325.
- Medlin, B., 1957. "Ultimate Principles and Ethical Egoism." *Australasian Journal of Philosophy*, 35: 111–118.
- Moore, G.E., 1903. *Principia Ethica*, Cambridge: Cambridge University Press.
- Nagel, T., 1970. *The Possibility of Altruism*, Princeton: Princeton University Press.
- , 1986. *The View From Nowhere*, New York: Oxford University Press.
- Parfit, D., 1984. *Reasons and Persons*, Oxford: Clarendon Press.
- Pettit, P., 1987. "Universality Without Utilitarianism." *Mind*, 72: 74–82.
- Portmore, Douglas, 2001. "McNaughton and Rawling on the Agent-Relative/Agent-Neutral Distinction." *Utilitas*, 13 (3): 350–356.
- Postema, G., 1998. "Public Practical Reason: An Archeology." in D'Agostino and Gaus 1998, pp. 425–468.
- Ridge, M., 2001a. "Agent-Neutral Consequentialism From the Inside-Out: Concern For Integrity Without Self indulgence," *Utilitas* 13: 236–254. [[Preprint available online](#) (in PDF)]
- , 2001b. "Saving Scanlon: Contractualism and Agent-Relativity." *Journal of Political Philosophy*, 9: 472–481. [[Preprint available online](#) (in PDF)]
- , 2009. "Consequentialist Kantianism," *Philosophical Perspectives*, 23: 421–438.
- Scanlon, T.M., 1998. *What We Owe to Each Other*, Cambridge, MA: Harvard University Press.
- Scheffler, S., 1994. *The Rejection of Consequentialism*, Oxford: Clarendon Press.
- Sellars, W.F., 1968. *Science and Metaphysics*, London: Routledge & Kegan Paul.
- Sidgwick, H., 1907. *The Methods of Ethics*, 7th edition. Chicago: University of Chicago.
- Smith, Michael. 2009. Consequentialism and the Nearest and Dearest Objection, in *Minds, Ethics, and Conditionals: Themes from the Philosophy of Frank Jackson*, Ian Ravenscroft (ed.). Oxford: Blackwell.
- Sturgeon, N., 1974. "Altruism, Solipsism, and the Objectivity of Reasons." *Philosophical Review*, 83: 374–402.
- , 1994. "Moral Disagreement and Moral Relativism." *Social Philosophy and Policy*, 11 (1): 80–115.
- Williams, B., 1981a. *Moral Luck*, Cambridge: Cambridge University Press.
- , 1981b. "Internal and External Reasons." reprinted in Williams 1981a, 101–13.

## Academic Tools

 [How to cite this entry.](#)

 [Preview the PDF version of this entry](#) at the [Friends of the SEP Society](#).

 [Look up this entry topic](#) at the [Indiana Philosophy Ontology Project](#) (InPhO).

 [Enhanced bibliography for this entry](#) at [PhilPapers](#), with links to its database.

## Other Internet Resources

[Consequentializing: Part I](#) by Douglas Portmore (PEA Soup blog entry)

## Related Entries

[consequentialism](#) | [metaethics](#) | [moral particularism](#) | [obligations: special](#)

## Acknowledgments

Many thanks to David Copp and Walter Sinnott-Armstrong for helpful feedback on an earlier draft of this material.

[Copyright © 2011](#) by  
[Michael Ridge](#) <[mridge@staffmail.ed.ac.uk](mailto:mridge@staffmail.ed.ac.uk)>

Open access to the Encyclopedia has been made possible by a [world-wide funding initiative](#). See the list of [contributing institutions](#). If your institution is not on the list, please consider asking your librarians to contribute.

**Stanford** | Center for the Study of  
Language and Information

The Stanford Encyclopedia of Philosophy is [copyright © 2016](#) by [The Metaphysics Research Lab](#), Center for the Study of Language and Information (CSLI), Stanford University

Library of Congress Catalog Data: ISSN 1095-5054