Agent-Relative vs. Agent-Neutral

Word Count (main text only): 3,746

The agent-relative/agent-neutral distinction has been drawn with respect to many things, including values, normative theories, and reasons for action (*see* REASONS; REASONS FOR ACTION, MORALITY AND). It is widely recognized as being one of the most important distinctions in practical philosophy, especially in distinguishing between theories that can and theories that cannot account for certain basic elements of commonsense morality, such as associative duties, agent-centered options, and agent-centered restrictions. (*See* ASSOCIATE DUTIES; AGENT-CENTERED OPTIONS; AGENT-CENTERED RESTRICTIONS.)

Reasons for Action

Thomas Nagel offers the following suggestion for how to draw the distinction between agent-relative and agent-neutral reasons:

> 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. (1986: 152-153)

By the "general form of a reason," Nagel has in mind some universally quantified principle such as:

(x)(the fact that p constitutes a reason for x to ϕ).

This abbreviates: 'For all agents, x, the fact that p constitutes a reason for x to ϕ', where 'p' refers to some statement of fact and 'ϕ' refers to some action.

Nagel's idea, then, is that if p contains an essential reference to x, it's an agent-relative reason. Otherwise, it's an agent-neutral reason. Thus, the fact that *my* friend is in need of some good cheer is an agent-relative reason for me to throw her a surprise party, for, when we put this reason into the above general form, 'p' would stand for 'x's friend is in need of some good cheer', which contains an essential reference to the agent, x. By contrast, the fact that Jane is in need of some good cheer is an agent-neutral reason for me to throw her a surprise party, for, when we put this reason into the above general form, 'p' would stand for 'Jane is in need of some good cheer', which contains no reference to the agent, x.

There is, however, a problem with drawing the distinction in this way. To see this, note that the fact that Jane is in need of some good cheer couldn't possibly be a full and accurate description of my reason for throwing her a surprise party—assuming that I even have such a reason. After all, the fact that Jane is need of some good cheer wouldn't be a reason for me to throw her a surprise party unless my throwing her a surprise party would bring her some good cheer. What if Jane hates surprise parties, or what if she

would hate *my* throwing her a surprise party, because, say, she would interpret my doing so as a sign that I intend to stalk her? If either is the case, then the fact that Jane is in need of some good cheer is no reason for me to throw her a surprise party. Thus, it seems that, in those instances in which my throwing Jane a surprise party is something that would bring her some good cheer, the reason that I have to throw her a surprise party, if fully and accurately described, is that *my* throwing her a surprise party would bring her some good cheer. But if that's right, then this is an agent-relative reason, for, when we put this reason into the above general form, we find that '*p*' stands for '*x*'s throwing Jane a surprise party would ensure that Jane experiences some good cheer', which does contain an essential reference to the agent, *x*. Yet, this was supposed to be a paradigm instance of an agent-neutral reason. It seems, then, that this way of drawing the distinction gets the wrong result once we fully and accurately describe the reason. What's worse, it seems that all reasons, when fully and accurately described, will turn out to be agent-relative on this account, for the simple fact that any reason for an agent to perform an act must appeal to some fact about the agent's performing that act—e.g., that the agent's performing that act would ensure that Jane experiences some good cheer.

This worry is only compounded by the realization that every fact that constitutes a reason for x to ϕ must make some reference to x if only to make reference to the fact that x can ϕ. For instance, the fact that my throwing Jane a surprise party would bring her some good cheer is a reason for me to throw her a surprise party only if throwing her a surprise party is something that I *can* do. Michael Ridge makes the point thusly:

> [E]very 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. (Ridge 2008: 20)

Interestingly, some philosophers embrace this implausible result, arguing that all reasons are, on purely formal grounds, agent-relative reasons. Rønnow-Rasmussen, for instance, argues: "*Since all reasons apparently are reasons for someone and a reason is only a reason for someone, if it somehow involves or refers to this someone, it follows that all reasons are by their very form reasons that refer to the person who has the reason to* ϕ" (2009: 231). But, like Ridge, I think that we should draw the agent-relative/agent-neutral distinction in a way that ensures that not all reasons come out agent-relative, thereby preserving the philosophical importance that the distinction is widely recognized as having. One way to do this—and this only a suggestion—is as follows.

Drawing inspiration from Rønnow-Rasmussen 2009 and McNaughton and Rawlings 1991, we can let '$x\phi c\mathrm{E}p$' stand for 'x's ϕ-ing in circumstances c would ensure that p' and formulate the distinction as follows:

(x)(ϕ)(c)(p)(if the fact that $x\phi c$Ep constitutes a reason for x to ϕ in c, then this fact constitutes an agent-relative reason for x to ϕ in c if and only if p contains an essential reference to x, and any reason for x to ϕ in c that doesn't constitute an agent-relative reason for x to ϕ in c constitutes an agent-neutral reason for x to ϕ in c).

The statement p contains an essential reference to x if and only if there is no non-x-referring statement q such that, for all x, the world in which x ensures that q is identical to the world in which x ensures that p. And note that I'm assuming that there is some determinate fact as to what the world would be like were x to ϕ in c. This assumption is sometimes called *counterfactual determinism* (Bykvist 2003: 30). If counterfactual determinism is false, then I'll need to let '$x\phi c$Ep' stand instead for 'x's ϕ-ing in c would make it probable that p'.

To see how this works, let's look at the following facts, which each, we'll suppose, constitute a reason for x to ϕ in c:

AR_1 $x\phi c$E(x abides by the Categorical Imperative)

AN_1 $x\phi c$E(Smith abides by the Categorical Imperative)

AR_2 $x\phi c$E(Jones feeds x's child)

AN_2 $x\phi c$E(Jones feeds Jones's child)

AR_3 $x\phi c$E(x's promise-breakings are minimized)

AN_3 $x\phi c$E(promise-breakings are minimized)

Whereas AR_1, AR_2, and AR_3 each constitute an agent-relative reason for x to ϕ in c, AN_1, AN_2, and AN_3 each constitute an agent-neutral reason for x to ϕ in c, for whereas each of AR_1, AR_2, and AR_3 contain an essential reference to x in the parentheses, none of AN_1, AN_2, and AN_3 do. Here and elsewhere, I've substituted the contents within the parentheses for p.

Note, however, that the following two facts each constitute an agent-neutral reason for x to ϕ in c despite the fact that the second one contains a reference to x in the parentheses:

AN_4 $x\phi c$E(utility is maximized)

AN_5 $x\phi c$E(x maximizes utility)

They each constitute an agent-neutral reason for x to ϕ in c, because neither includes an *essential* reference to x in the parenthesis. The reference to x inside the parenthesis of AN_5 is superfluous, because, for all x, the world in which x ensures that utility is maximized is identical to the one in which x ensures that x maximizes utility.

The foregoing account of the distinction between agent-relative and agent-neutral reasons captures everything that we could want from such an account. First, it allows us to account for what Rønnow-Rasmussen calls the *personalizability-feature* of reasons for action: that is, the feature according to which any fact constituting a reason (agent-relative or agent-neutral) for x to ϕ in c must in some way refer to the agent for whom it

is a reason so as to account for why that fact constitutes a reason *for x* in particular (2009: 230). Since, on the above account, reason-constituting facts all take the form $x\phi c\mathrm{E}p$, they all do refer to the agent (viz., x) for which it is a reason.

Second, this account remains neutral with regard to most, if not all, substantive issues about reasons for action. Although, on this account, all reasons for action are facts about what agents can ensure, this does not commit advocates of the account to a teleological conception of reasons, where reasons for action are all a function of what it would be good for agents to bring about. On this account, the fact that x's ϕ-ing in c would ensure that x abides by the Categorical Imperative (*see* CATEGORICAL IMPERATIVE) can constitute a reason to x to ϕ in c even if x's abiding by the Categorical Imperative would produce bad consequences. And this account allows that any sort of consideration, be it past-regarding or forward-looking, can count as a reason. For instance, the fact that x's ϕ-ing in c would ensure that x has fulfilled her past promise to ϕ can count as reason for to x to ϕ in c.

What's more, this account doesn't commit us to denying particularism about reasons for action (*see* PARTICULARISM). One could accept this account and hold, as particularists do, that a consideration that counts as a reason in one context doesn't count as a reason in some other context. For instance, it could be that the fact that x's ϕ-ing would ensure that Jones takes pleasure in some state of affairs is a reason for x to ϕ in those circumstances in which Jones would be taking pleasure in someone else's pleasure, but not in those circumstances in which Jones would be taking pleasure in someone else's pain.

Third, the account correctly identifies paradigmatic instances of reasons of both kinds. For instance, AR_1, AR_2, and AR_3 are all paradigm instances of agent-relative reasons, and this account correctly identifies them as such. And AN_1, AN_2, AN_3, and AN_4 are all paradigm instances of agent-neutral reasons, and this account correctly identifies them as such.

Fourth and last, this account of the distinction preserves the philosophical importance that it is widely regarded as having. It does so by ensuring that not all reasons are agent-relative reasons, and it does so irrespective of what substantive views we accept about reasons for action (cf. Rønnow-Rasmussen 2009).

Normative Theories

We can distinguish not only between agent-relative and agent-neutral reasons for action, but also between agent-relative and agent-neutral normative theories. As many philosophers draw the distinction, a normative theory is agent-neutral if and only if it gives every agent the exact same set of substantive aims. Otherwise, it's agent-relative. (See, for instance, Parfit 1984: 27 and Dreier 1993: 22.) But this talk of aims is at best metaphorical. Take one class of normative theories: moral theories. Moral theories provide criteria for determining whether an act has a certain deontic status (e.g., permissible, impermissible, obligatory, or supererogatory), but they don't explicitly state which aims agents ought to have (*see* SUPEREROGATION). And it's not clear how we are to move from talk about an act's having a certain deontic status to there being a certain aim that its agent ought to have. Suppose, for instance, that a moral theory holds

that Smith's giving at least ten percent of his income to charity is obligatory and that Smith's giving more than ten percent is supererogatory. Does such a moral theory give Smith the aim of giving ten percent, giving more than ten percent, or something else? It's not clear.

Perhaps, though, we can understand talk of aims in terms of talk of reasons for action. And, perhaps, we can just stipulate that the following relationship holds between deontic status and reasons for action:

> If, on a normative theory *T*, *x*'s ϕ-ing in *c* is either obligatory or supererogatory in virtue of the fact that $x\phi c\mathrm{E}p$, then, on *T*, the fact that $x\phi c\mathrm{E}p$ constitutes a reason for *x* to ϕ in *c*.

Once we make this stipulation, we can distinguish agent-relative theories from agent-neutral theories as follows:

> A normative theory *T* is an agent-relative theory if and only if, on *T*, there exist some possible agent *x*, some possible act ϕ, some possible circumstances *c*, and some possible statement *p* such that the fact that $x\phi c\mathrm{E}p$ constitutes an agent-relative reason for *x* to ϕ in *c*. Otherwise, it's agent-neutral.

To illustrate, consider act-utilitarianism (*see* UTILITARIANISM). It holds that, for all *x*, *x*'s ϕ-ing in *c* is obligatory if and only if ϕ-ing in *c* is the only way for *x* to ensure that

utility is maximized. And, on act-utilitarianism, acts are never supererogatory. So, on act-utilitarianism, agents only have reason to ensure that utility is maximized, and since this is an agent-neutral reason, act-utilitarianism is an agent-neutral theory. Universalizable egoism (*see* EGOISM), by contrast, is an agent-relative theory. It holds that, for all x, x's ϕ-ing in c is obligatory if and only if ϕ-ing in c is the only way for x to ensure that x's utility is maximized. Since, on egoism, each agent has an agent-relative reason to ensure that *her* utility is maximized, egoism is an agent-relative theory.

To illustrate how supererogation can come into play, consider what I'll call *super-act-utilitarianism*. Like act-utilitarianism, super-act-utilitarianism holds that, for all x, x's ϕ-ing in c is obligatory if and only if ϕ-ing in c is the only way for x to ensure that utility is maximized. Unlike act-utilitarianism, though, super-act-utilitarianism holds additionally that, for all x, x's ϕ-ing in c is supererogatory if and only if x's ϕ-ing in c would ensure that x performs a permissible act that produces less utility for x than some other permissible alternative would. Suppose, for instance, I must choose whether or not to push a certain button. Let's assume that, if I push the button, I'll receive only ten units of utility, whereas others will in aggregate receive ninety units of utility. And let's assume that, if I refrain from pushing the button, I'll receive ninety units of utility, whereas others will in aggregate receive only ten units of utility. Either way, utility will be maximized. Thus, I'm permitted to do either. Nevertheless, on super-act-utilitarianism, my pushing the button is supererogatory, for in doing so I ensure that I perform a permissible act that produces less utility for me than some other permissible alternative would. Since, on super-act-utilitarianism, each agent has an agent-relative

reason to perform those permissible acts that produce less utility for *herself*, this is an agent-relative theory.

Values

The agent-relative/agent-neutral distinction can also be drawn with respect to values. Those who postulate agent-relative values tend to subscribe to the view that there is some bi-conditional relationship between *A*'s being better than *B* and its being fitting to prefer *A* to *B*. Those who accept what's known as the buck-passing or fitting-attitude account of value (*see* BUCK-PASSING ACCOUNT; VALUE, FITTTING-ATTITUDE ACCOUNT OF) certainly think that there is such a bi-conditional relationship, but their Moorean rivals often endorse such a bi-conditional relationship as well. Although the buck-passer and the Moorean must disagree about which side of the bi-conditional has explanatory priority, they can agree on the bi-conditional itself.

In any case, if we accept that there is such a bi-conditional relationship, we can distinguish between agent-relative and agent-neutral values as follows:

> For all states of affairs *p* and *q*, it is better-relative-to-*x* that *p* is the case than that *q* is the case if and only if it would fitting for *x* to prefer *p* to *q*. And for all states of affairs *p* and *q*, it is better (agent-neutrally speaking) that *p* is the case than that *q* is the case if and only if it would be fitting for an impartial spectator to prefer *p* to *q*. (Suikkanen 2009: 6)

Note that an impartial spectator must be both impartial (*see* IMPARTIALITY) and a mere spectator. To ensure that she is impartial, we must assume that the impartial spectator has no personal relations with anyone involved. And to ensure that she is a mere spectator, we must assume that she is not involved either in bringing it about that p or in bringing about that q.

To illustrate the distinction between agent-relative and agent-neutral values, suppose that I can actualize only one of two possible worlds: either the one in which I save my own daughter (call this w_1) or the one in which I ensure that some stranger, named Smith, saves his slightly more cheerful daughter (call this w_2). Let's assume that a world with more pleasure is, other things being equal, better than a world with less pleasure. And let me just stipulate that w_2 contains slightly more pleasure than w_1, for Smith's slightly more cheerful daughter would experience slightly more pleasure than my daughter would. Assume, though, that everything else is equal. Clearly, given these stipulations, w_2 is better than w_1 (agent-neutrally speaking), for it seems fitting for an impartial spectator to prefer w_2 to w_1. Arguably, though, w_1 is better-relative-to-me than w_2. Given the special relationship that I have with my daughter, it seems fitting for me to prefer w_1 to w_2. Indeed, if I didn't care more about the preservation of my daughter's life than I did about such small increases in the aggregate pleasure, we would think that there is something wrong with me. Fathers ought to care more about the preservation of their children's lives than they do about such small increases in the impersonal good.

So whereas one possible world is better than another if and only if it would be fitting for an impartial spectator to prefer the one to the other, one possible world is

better-relative-to-*x* than another if and only if it would be fitting for *x* to prefer the one to the other. And what it is fitting for an impartial spectator to prefer can differ from what it is fitting for *x* to prefer, because *x*'s agential involvement in actualizing one of the two worlds and/or *x*'s personal relationships with those who would be better off in one of the two worlds can affect which it is fitting for *x* to prefer.

It should be noted, though, that the existence of agent-relative values is quite controversial. Mark Schroeder (2007), for instance, has argued that the better-relative-to-*x* relation is just a theoretical posit and that there is no theory-neutral way to draw the distinction between agent-relative and agent-neutral values. Furthermore, he has argued that there is no way to explain what the better-relative-to-*x* relation means using ordinary language terms such as 'better', 'better for', or 'better from a particular point of view' (*see* GOOD AND BETTER; GOODNESS, VARIETIES OF). Whether the above account succeeds where previous attempts have failed is something that I leave for the reader to judge.

The Importance of the Distinction

Central to commonsense morality are various agent-relative (or agent-centered) elements, such as associative duties, agent-centered options, and agent-centered restrictions. Theories that fail to accommodate such agent-relative elements will, as a result, have implications that conflict with our commonsense moral intuitions. Take agent-centered restrictions, for instance. Agent-centered restrictions prohibit agents from performing certain act-types (such as, murder) even in some circumstances in

which performing that act-type is the only way to minimize comparable commissions of that act-type. Any moral theory that lacks agent-centered restrictions will seem too permissive, allowing agents to commit heinous acts such as murder whenever doing so will minimize comparable commissions of that act-type overall.

Agent-neutral theories are unable to accommodate such agent-relative elements. To illustrate, consider traditional act-consequentialism (hereafter, simply consequentialism) and its inability to accommodate agent-centered restrictions (*see* CONSEQUENTIALISM). Consequentialism holds that, for all x, x's ϕ-ing in c is obligatory if and only if ϕ-ing in c is the only way for x to ensure that the (agent-neutral) good is maximized. And, on consequentialism, no act is supererogatory. Thus, consequentialism is an agent-neutral theory, as defined above. Since, on consequentialism, any act that maximizes the good is permissible, consequentialism cannot accommodate any restrictions on maximizing the good. Although the consequentialist can hold that murdering one is worse than allowing two innocent others to die from natural causes and thus hold that it is impermissible to commit murder in order to prevent two deaths by natural causes, the consequentialist cannot hold that it is impermissible to commit murder in order to prevent two comparable murders, for on any agent-neutral account of the good on which murder is bad, two murders will, other things being equal, be worse than one murder.

Although no agent-neutral theory can accommodate agent-centered restrictions, it doesn't follow that no teleological theory can accommodate agent-centered restrictions. It is a mistake to equate the agent-relative/agent-neutral distinction with the

deontological/teleological distinction (*see* DEONTOLOGY). Consider, for instance, agent-relative teleology (ART). It holds that, for all x, x's ϕ-ing in c is obligatory if and only if ϕ-ing in c is the only way for x to ensure that what is good-relative-to-x is maximized (Portmore 2005; Schroeder 2007). This is an agent-relative theory, and, thus, the ARTist can accommodate agent-centered restrictions. For instance, the ARTist can accommodate an agent-centered restriction against the commission of murder by claiming that, other things being equal, the state of affairs in which x commits one murder is worse-relative-to-x than the states of affairs in which x allows two others to commit comparable instances of murder. Thus, the important distinction to make in delineating those theories that can and those theories that cannot account for various agent-relative elements (such as agent-centered restrictions) is the agent-relative/agent-neutral distinction, not the teleological/deontological distinction.

Cross-references

AGENT-CENTERED OPTIONS; AGENT-CENTERED RESTRICTIONS; ASSOCIATE DUTIES; BUCK-PASSING ACCOUNT; CATEGORICAL IMPERATIVE; CONSEQUENTIALISM; DEONTOLOGY; EGOISM; GOOD AND BETTER; GOODNESS, VARIETIES OF; IMPARTIALITY; PARTICULARISM; REASONS; REASONS FOR ACTION, MORALITY AND; SUPEREROGATION; UTILITARIANISM; and VALUE, FITTTING-ATTITUDE ACCOUNT OF.

References

Bykvist, Krister 2003. "Normative Supervenience and Consequentialism." *Utilitas*, 15: pp. 27-49.

Dreier, James 1993. "Structures of Normative Theories." *The Monist*, 76: pp. 22-40.

McNaughton, David and Rawlings, Piers 1991. "Agent-Relativity and the Doing-Happening Distinction." *Philosophical Studies*, 63: pp. 167-185.

Nagel, Thomas 1970. *The Possibility of Altruism.* Princeton: Princeton University Press.

Nagel, Thomas 1986. *The View from Nowhere.* New York: Oxford University Press.

Parfit, Derek 1984. *Reasons and Persons*. Oxford: Oxford University Press.

Portmore, Douglas 2005. "Combining Teleological Ethics with Evaluator Relativism: A Promising Result." *The Pacific Philosophical Quarterly* 86: pp. 95-113.

Schroeder, Mark 2007. "Teleology, Agent-Relative Value, and the 'Good'." *Ethics*, 117: pp. 265-295.

Ridge, Michael 2008. "Reasons for Action: Agent-Neutral vs. Agent-Relative." *The Stanford Encyclopedia of Philosophy (Fall Edition)*, Edward N. Zalta (ed.), URL = <http://plato.stanford.edu/archives/fall2008/entries/reasons-agent/>.

Rønnow-Rasmussen, Toni 2009. "Normative Reasons and the Agent-Neutral/Relative Dichotomy." *Philosophia*, 37: pp. 227-243.

Suikkanen, Jussi 2009. "Consequentialism, Constraints and the Good-Relative-to: A Reply to Mark Schroeder." *Journal of Ethics & Social Philosophy*, www.jesp.org, pp. 1-8.

Suggested Readings

Portmore, Douglas 2009. "Consequentializing." *Philosophy Compass* 4: pp. 329-347.

Schroeder, Mark 2008. "Value Theory." *The Stanford Encyclopedia of Philosophy (Fall Edition)*, Edward N. Zalta (ed.), URL = <http://plato.stanford.edu/archives/fall2008/entries/value-theory/>.

Smith, Michael 2003. “Neutral and Relative Value after Moore.” *Ethics* 113: pp. 576–98.