Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Agent-Relativity

João da Rocha Rodrigues

id11394@uminho.pt

University of Minho

Research Article

Keywords: Agent-relative, Agent-neutral, Parfit, Nagel, reasons, consequentialism

Posted Date: November 17th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11340401/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Title: Agent-Relativity

Abstract: This paper examines the agent-neutral/agent-relative distinction, a major but contested achievement in moral philosophy. It addresses three questions: the proper application of agent-relativity, whether the distinction is trivial, and whether it has been adequately formalized.

It will reject the application of agent-relativity to values and ought-propositions but permit it in relation to reasons, theories, and rules. It will also defend the distinction's substance against one objection which claims that all reasons are agent–relative. It will do this by distinguishing between real agent-relativity and a merely pseudo-agent-relativity.

The existing attempts at formalizing the distinction as it applies to reasons are shown to be misguided. They do not fulfill the original motivation behind the distinction, due to what we will call the "derivability problem": agent-relative reason-statements can be transformed into agent-neutral ones while retaining the same ethical content, and agent-neutral theories can generate derivative agent-relative reasons.

The paper's main thesis is then proposed: we should rethink agent-relativity as primarily a feature of comprehensive moral outlooks, not isolated reason-statements. A reason is agent-relative or agent-neutral only in relation to such an outlook. The essence of agent-relativity lies in the possibility of conflict between the injunctions and normative reasons of a given moral outlook. Consequently, the agent-relativity of a reason is secondary, to be understood only in relation to a moral outlook that renders it a true normative reason and determines its scope across agents. This new framework bypasses the derivability problem and restores the distinction's philosophical significance.

Keywords: Agent-relative; Agent-neutral; Parfit; Nagel; reasons; consequentialism.

Introduction

The distinction between agent-neutrality and agent-relativity has been regarded as one of the major achievements of contemporary moral philosophy (Hurka, 2003). At the same time, however, it has been the target of persistent criticism and dismissed as either unfounded, trivial, or even meaningless (see, e.g., Bykvist, 2018; Schroeder, 2007).

Even among those who accept the distinction and regard it as philosophically significant, there remains substantial disagreement about how it should be understood in formal terms (see, e.g., McNaughton & Rawling, 1995a, pp. 319–325). No proposed formalization has proved entirely uncontroversial and free of difficulty.

This paper will focus on the following three essential questions concerning the idea of agent-relativity. What is agent-relativity properly applied to? Is the distinction trivial or meaningless? Has it been adequately formalized and if not, why?

Section 1 is concerned with the first question and will consider a wide range of candidates, such as agent-relative reasons, values, rules, moral theories, and even ought-propositions.

Section 2 will tackle the second question and look at an objection according to which the distinction becomes meaningless once it is formulated in terms of reasons. Since all reasons are reasons for someone, so the objection goes, all reasons are agent-relative, and the contrast with agent-neutrality therefore collapses.

In answering the third question, sections 3 and 4 are the most substantive parts of the paper and contain its main thesis: agent-relativity is first and foremost a feature of comprehensive moral outlooks, and only secondarily a feature of reasons. A reason should only be understood as agent-relative or agent-neutral in relation to such an outlook. The several attempts to draw the distinction in formal terms have all been misguided. This is because they look to classify a reason as agent-relative or agent-neutral in isolation, typically on the basis of some linguistic feature of the corresponding reason-statement. This approach is vulnerable to what I will call "the derivability problem".

1. Agent-Relativity of What?

1.1 The Original Distinction

The distinction between the agent-relative and agent-neutral was coined by Parfit and originally applied to moral theories. A moral theory is agent-neutral if it assigns the same aim to all agents, and agent-relative if it assigns different aims to different agents (Parfit, 1984, pp. 26, 142). Standard direct consequentialism is an unproblematic example of an agent-neutral theory, since it gives every agent the same aim: maximizing welfare. By contrast, a form of commonsense morality was presented as an example of an agent-relative theory since, according to Parfit, it gives each agent the moral aim of complying with a set of duties, such as not killing or not lying. In such cases, each agent has the aim that he himself not kill or not lie, and so each agent has a different aim.

Thomas Nagel adopted this terminology[1] and set the agenda for many of the questions that would subsequently become central to the debate. In his account, the focus does not go from theories to reasons; rather, it begins with reasons themselves. Agent-neutral reasons are those reasons that can be stated without an essential

[1] Although he could also be said to be the true originator of this dichotomy with his treatment of objective and subjective reasons in The Possibility of Altruism (Nagel, 1970), which are largely similar to agent-neutral and agent-relative reasons.

reference to the person who has them, whereas agent-relative reasons are those that cannot be stated without such a reference.[2]

Nagel distinguishes three types of agent-relative reasons that resist reduction to agent-neutral ones. The first consists of the reasons a person may have to pursue her own life projects, even at the expense of the greater good. The second concerns reasons arising from specific attachments, for example, to friends, family, or others with whom one stands in a special relationship. The third concerns deontic constraints: reasons not to act in certain ways, even when doing so would produce a better overall outcome.

This is only a rough summary, and the distinction could be drawn in somewhat different terms. Still, agent-relative reasons are seen to arise in those areas that have posed persistent difficulties for fully agent-neutral theories such as consequentialism, which has struggled to account for the apparently irreducibly personal value of one's projects, commitments, and close relationships. These difficulties were famously highlighted in Bernard Williams's influential critique of utilitarianism (Williams, 1973), where he grouped such considerations under the value of integrity. Deontic constraints, although they too could be framed in terms of integrity, are more directly a restatement of the longstanding tension between consequentialism and deontology, or of the conflict between consequentialism and the demands of justice that would prevent us, for example, from mistreating people in certain ways even if doing so would produce a better overall outcome. It is therefore unsurprising that agent-relative reasons have been interpreted as being constitutive of deontology.

Nagel also points to a central motivation behind the distinction, which Jonathan Dancy describes as the possibility of coherently maintaining that something matters immensely "for me" even though it is not especially important in general, or objectively (Dancy, 1993, p. 233). This description is sufficiently close for present purposes. In every one of these cases, there is a contrast between what might be regarded as objectively or impartially better on the one hand, and the demands imposed on a specific agent on the other, in virtue of some feature that is his own. This feature can be a project, a relationship, or his own conduct as a moral agent. For a reason to be agent-relative, it must single out that particular agent (or a subset of agents), rather than apply equally to everyone.

1.2 Rules as an Attempt at Formalization

Agent-relativity has also been associated with rules or principles, values, and even ought-statements. We will look at each of these in turn.

In an effort to formalize the agent-relative/agent-neutral distinction, some theorists have argued that agent-relativity is best expressed in terms of rules. One of the most influential attempts at this sort of formalization comes from McNaughton and Rawling. On their view, moral rules are best understood as "universally quantified directives, governed by a pro tanto operator. Where 'xS' abbreviates, "x should ensure, to the best of x's abilities, and insofar as there is not a conflicting duty of greater weight, that [...]". All rules have the following form: (x)(xS [...])." (McNaughton & Rawling, 1995b, pp. 34–35)

According to this view, a rule is agent-relative if and only if two conditions are met. First, there must be an occurrence of x inside the square brackets that is bound by the initial universal quantifier. Second, the rule must not be equivalent in ethical content to an agent-neutral rule.

[2] Nagel (1986, pp. 152–153). Pettit (1987, p. 75) also defines an agent-relative reason in a similar way, as one that "cannot be fully specified without pronominal backreference for the person for whom it is a reason".

Thus, a rule such as (x)(xS [(x does not kill the innocent)]) is agent-relative, since there is an x within the brackets bound by the initial universal quantifier. In contrast, a rule such as (x)(xS [(y)(y does not kill the innocent)]) is agent-neutral, since there is no x within the brackets. This particular example captures a familiar difference between an agent-neutral rule (or reason) to maximize a state of affairs in which innocents are not killed and an agent-relative rule or reason not to kill innocents oneself.

This is a serious and illuminating attempt at formalizing the core of Nagel's original distinction. I will turn to its difficulties in the next section. Right now, what's important to keep in mind is that McNaughton and Rawling treat the distinction at the level of rules as equivalent to the original distinction at the level of reasons: one has an agent-relative reason when one is to act on an agent-relative rule, and likewise in the agent-neutral case (McNaughton & Rawling, 1995b, p. 36).

1.3 Why Value is the Wrong Locus

A second way of thinking about agent-relativity is through its relation to value. Nagel himself appears to allow for this, as do others (e.g., Dancy, 1993; Mack, 1993; Stewart, 1993). Just as reasons may be agent-relative, so too, it is said, may value be agent-relative. One's relationships may have agent-relative value for the person involved in them, or one's personal projects may have agent-relative value for the person whose projects they are.

Following Mark Schroeder, we will reject the agent-relativity of value. Schroeder criticizes a family of theories that accord such value a central place, what he calls "agent-relative teleology" (Schroeder, 2007, p. 265). Like the distinction between agent-neutrality and agent-relativity, these theories emerged as an attempt to preserve what is attractive about consequentialism while resolving some of its most familiar difficulties. Broadly speaking, they retain the consequentialist requirement to always maximize good states of affairs, but relativize the value of those states of affairs to individual agents. Thus, for an agent A, a state of affairs in which he, for example, commits a murder, would have less value in relation to A; but that very same state of affairs could have more value in relation to another agent, B. What is to be maximized by each agent is now the agent-relative value of states of affairs in relation to that agent. In this way, things like deontic constraints can supposedly be justified while retaining an overall consequentialist structure, since A would be barred from committing murder because of the agent-relative disvalue of a state of affairs in which he himself commits it.

Schroeder correctly identifies the main difficulty of this approach in the dubiousness of the notion of "good-relative-to", and hence of agent-relative value more generally, especially when compared with the more familiar and less problematic notions of "good" and "good-for".

It is clear that some things, including states of affairs, can be good; and it is equally clear that some things can be good for certain people without thereby being good *simpliciter*. To use Schroeder's example, a tax code can be good without being good for Dick Cheney's pals, and it can be good for Dick Cheney's pals without thereby being good (Schroeder, 2007, pp. 277–278). But where exactly does "good-relative-to" enter here? It seems that no coherent sense can be made of it as a distinct notion that is neither "good" nor "good-for" someone.

Goodness and value are usually understood as the properties of objects (such as states of affairs) considered either in themselves or in relation to someone. Reasons, in contrast, are always relational in the sense that they apply to some set of agents, which may, of course, be the set of all agents. The agent-relativity of reasons is therefore simply the claim that certain reasons do not apply to all agents. That claim is perfectly intelligible. But because goodness and value are properties of objects, it makes little sense to say, in addition to the possibilities of

being good simpliciter and good for someone, that there is a third possibility in which an object possesses the independent goodness of being good simpliciter, while that same independent goodness is somehow relative to different agents. If, on the other hand, the object is not good or valuable independently, then we are simply dealing with the familiar notion of it being good for some set of agents, and there is thus no distinction between this and the notion of it being "good-relative-to" that set of agents.

There remains, however, one sense in which talk of agent-relative value may be acceptable. Namely, if one adopts the so-called "buck-passing" account of value (Scanlon, 1998). On this view, value always cashes out in terms of reasons, which become primary. To say that something has value is equivalent to saying that it has properties that provide reasons in favor of its attainment or realization. In this sense, if there are agent-relative reasons bearing on something, it can be said that this thing has agent-relative value. But this is a deflationary account of value, since it reduces claims about value to claims about the existence of certain kinds of reasons. Under such a conception, there is no distinct notion of agent-relative value over and above the notion of an agent-relative reason.

1.4 Why Ought-Propositions are the Wrong Locus

Although it remains a minority position, some philosophers have associated agent-relativity with ought-statements or ought-propositions (Humberstone, 1991; see also Lee, 2024).

Humberstone, for example, discusses two forms of agent-relativity in deontic logic and in relation to ought-propositions. The first is what he calls "onus agent-relativity". It may be the case not merely that something ought to happen, but that it ought to happen through a particular agent. His example is that of a drug deal that ought not to occur. In many legal jurisdictions the onus of preventing it from happening is placed on the seller rather than on the buyer of the drugs (Humberstone, 1991, pp. 146–151).

This kind of agent-relativity appears to be largely redundant, since the onus can be built into the very event to which the ought is attached. That A ought to do X, or that X ought to come about through A, always differs from the case in which X comes about through B or C. In Humberstone's legal example, what is really at issue is an attribution of responsibility after the fact, not a difference in the moral injunction itself. Neither the dealer nor the user ought to bring about the drug deal, but only the dealer is held legally responsible in the relevant way. And in any case, the drug deal's occurring through the user and its occurring through the dealer are simply different events, and therefore correspond to different oughts rather than to the same ought qualified in different ways.

Humberstone then introduces a second sort of agent-relativity in relation to ought statements. This is an "ought agent-relativity" based on the "variability of circumstances" between different perspectives (Humberstone, 1991, pp. 152–154). His example is that of a Nazi doctor about to perform an experiment on a concentration camp inmate without anaesthesia. A nurse—who opposes the experiment but cannot prevent it—has the opportunity to administer the anaesthesia without the doctor's knowledge. Ought she to do so?

In such a case, what ought to happen is that the experiment does not take place at all. If the experiment ought not to occur, it may seem odd to say that the nurse ought to administer the anaesthesia, since that prescription presupposes the occurrence of the experiment itself. Hence, the true proposition "she ought to administer the anaesthesia" is classified as agent-relative to her perspective, given her particular circumstances and the limitations on what she can do, especially the fact that she cannot prevent the experiment.

This points to an important truth about agent-relativity more generally. What we ought to do, or what we have reason to do, is often shaped by the circumstances in which we find ourselves, and by the power and

opportunities available to us in those circumstances. There is nothing incoherent about this kind of relativity, though when it is cast in terms of ought- propositions it can generate apparent dilemmas (Humberstone, 1991, pp. 154–155).

Be that as it may, this kind of relativity does not align with the original motivation for the distinction. It amounts merely to the claim that our obligations vary according to the power or opportunity available to us in the circumstances in which we act.

To conclude, we have considered the different kinds of things to which agent-relativity has been thought to apply. I have argued that it can be intelligibly applied to theories, reasons, and rules, the latter being one influential attempt to formalize the original distinction at the level of reasons. By contrast, I have rejected its application to values and to ought-propositions.

2. Pseudo-Agent-Relativity

As we saw at the end of the previous section, there is a kind of trivial agent-relativity that may be applied both to ought-statements and to reasons, simply due to the fact that different people find themselves in different circumstances and therefore have different powers, opportunities, and options available to them. This point may appear obvious, but it is central to an objection against the distinction between the agent-relative and the agent-neutral that I now wish to examine.

If what I've just described happens for almost all reasons—which it does since all moral agents (except maybe God) are limited to some capacity—then all of them would be agent-relative, and the distinction with agent-neutrality would be pointless.

Simply put: the objection is that the distinction is trivial since all reasons are reasons for someone, and who this "someone" they apply to is, largely depends on the circumstances; therefore all reasons are agent-relative (Rønnow-Rasmussen, 2009, pp. 230–231). Supporting this objection is the further thought that, for someone to have a reason to do something, it must be possible for that person to do it. This is a version of "ought implies can" transposed into the realm of reasons: something like "having a reason implies can" (Rønnow-Rasmussen, 2009, p. 236). If this is so, then pretty much all reasons an agent could have are agent-relative, since their application to *that* agent (or lack thereof) always depends on the circumstances he finds himself in, and this necessarily prevents those same reasons from applying to (at least) some other agents. I will only have reasons to help *these* people who are close enough to me that I can help, or to fulfill this or that obligation that it's possible for me to fulfill at this time. There will always be others in different circumstances who cannot do this, and so none of the reasons that apply to me can apply to everyone. All reasons are limited in this way, and so all reasons are agent-relative, so the objection goes.

All reasons are indeed agent-relative in this way, but from now on we will call this sort of agent-relativity "pseudo-agent-relativity." A reason is pseudo-agent-relative when its restricted scope is solely due to the circumstantial limitations on the possibilities available to the agents it doesn't apply to, rather than it being due to any characteristic of those agents as such—like, for instance, their place in a web of special relationships, or the fact that they occupy a certain institutional role.

Suppose that someone is drowning on the other side of the world. Do I have a reason, however weak, to save that person, or do I have no reason at all, given that it's presently impossible for me to do so? Someone like Nagel might say that everyone has such a reason, since he holds that the fact that someone is in pain gives everyone

else a reason to bring that pain to an end (Nagel, 1986, pp. 157–161), whether or not they are in a position to do so. If we accept this, then we have preserved the class of such reasons as agent-relative in the fullest sense, but seemingly at the cost of rejecting the "having a reason implies can" principle.

Luckily we are not forced to choose between the rejection of this principle and the loss of an acceptable agent-relative/agent-neutral distinction. Whatever view one takes, an intelligible version of the distinction remains available. The reasons whose scope gets cut down solely due to the variability of circumstances may still be classified as agent-neutral, even if they are also pseudo-agent-relative.

Scanlon, for example, defines a reason as a relation holding between a fact F, an action or attitude A, an agent X, and a set of circumstances C. For him, the full statement of a reason is a statement of the form R(F, A, X, C), which says that F is a reason for X to do A in circumstances C. (Scanlon, 2014, p. 31)

According to this conception, the application of a reason may vary with the circumstances, while the reason itself remains agent-relative or agent-neutral depending on the agent-parameter X. If X ranges over all agents, then the reason applies to everyone in those circumstances, and is therefore agent-neutral, even though it can still be pseudo-agent-relative. If X ranges over only a subset of all agents within a fixed C parameter, then the reason is genuinely agent-relative.

With this in mind, we can reject the earlier objection. Even if all reasons are reasons for someone, and even if circumstances and limitations necessarily exclude at least some agents from this class, there still remains an intelligible and philosophically useful notion of agent-relativity that is distinct from mere pseudo-agent-relativity. It is therefore still possible to classify some reasons as agent-neutral, and the distinction between the two types of reasons is preserved.

3. The Derivability Problem

As we saw in the first section, attempts have been made to formalize the agent-relative/agent-neutral distinction. However, none of them has been uncontroversially successful. Different reasons have been given for this[3], but we will now argue that there is a greater difficulty with the general way these attempts have been made so far.

For a formalization to be successful, it must find a principled rationale for distinguishing between the two types of reasons; but it must also draw the distinction so as not to deny many of the standard associations between agent-relativity, agent-neutrality, and other moral considerations. For example, many think that a theory like consequentialism must be agent-neutral, and that theories such as egoism or some forms of deontology have to be agent-relative, at least to some degree.[4] A successful formalization should also classify deontic constraints as agent-relative, together with the reasons related to our personal projects or our close attachments. It might be possible to find a plausible formalization for the distinction that does not fulfill these desiderata. But in that case we would probably have arrived at a different distinction and not at the one that has organically emerged in the philosophical literature, which is tied to a specific set of issues and the tensions that result in their interaction with the functioning of paradigmatic moral theories.

[3] Dancy and Schroeder have been critical of Nagel. McNaughton and Rawling's formalization has been criticized in Hammerton (2018). Others like Pettit and Skorupski have been criticized in Bykvist (2018) and Buckland (2019), respectively.
[4] Although there is less consensus in the case of deontology.

Many have sought to formalize the agent-relative/agent-neutral distinction by focusing on reason-statements, and on whether these statements contain a reference to the agent to whom the reason applies. According to this approach, a reason is agent-relative if the statement of that reason contains a reference to the agent for whom it is a reason. Otherwise, it is agent-neutral.

Nagel understood this reference to consist in the reason-statement containing what he called a free agent-variable. (1970, p. 90) In the rule formulation of McNaughton and Rawling the reference happens through the variable X occurring inside the universal quantifier at the start of all reason-statements. Skorupski takes a similar path and draws the distinction as follows:

> (It's being the case) that Rφ gives x reason to φ.
>
> Where 'x' is replaceable by terms denoting agents and 'φ' by terms denoting actions open to x. [...]
>
> Now we define agent-neutral and agent-relative reasons as follows:
>
> (i) If 'R' contains a free occurrence of 'x' then it is an agent-relative predicate. If it does not, it is an agent-neutral predicate.
>
> (ii) A reason for action which is expressible by an agent-neutral predicate is agent-neutral. A reason for action which is not so expressible is agent-relative (Skorupski, 2011, pp. 63–64).

Pettit (1987, p. 75) resorts to the device of pronominal back-referencing. If a reason-statement contains a pronominal back-reference to the agent for whom it is a reason, then that reason is agent-relative (and agent-neutral if otherwise). A reason-statement like "because she is my daughter" contains the pronoun "my" which references the agent for whom this is a reason, which therefore makes it agent-relative.

In all these cases, the agent-relativity of a reason consists in there being a reference in the reason-statement to the agent for whom it is a reason.

The main difficulty with finding a successful formalization along these lines is what we will now call "the derivability problem".

There are two dimensions to it. The first is that many reasons, some of them agent-relative, can be derived from other, agent-neutral reasons. The most obvious example can be found in standard consequentialism. It seems to be a clear case of an agent-neutral theory, where the only true reason to do anything is that it maximizes welfare. This reason applies to all agents, makes no reference to any specific set of agents, and is therefore agent-neutral.

However, if agent-relativity is all about the occurrence of a reference in reason-statements, then even in consequentialism, agent-relative reasons remain.

For example, if taking my brother to see a movie will increase his welfare, and this is what I can currently do to maximize the overall welfare, then it can be said that I have a reason to take my brother to see a movie. Its reason-statement would reference me since it would have to contain a pronominal back-reference like "my brother", or a free agent-variable to indicate that it is *me* who must take him to the movies in order for welfare to be maximized. Expressed in this way, this reason would be agent-relative, even though it is clearly derived from the agent-neutral reason to maximize welfare that, under consequentialism, everyone has.

If this is correct, then even consequentialism would be forced to accept agent-relative reasons, and any attempt to formalize the distinction would be doomed to failure even in this simple goal of classifying it as an agent-neutral theory.

The second dimension of the derivability problem is that many agent-relative reason-statements—related to the questions of personal projects, special relationships, deontic constraints, and so on—can seemingly be transformed into agent-neutral ones and, for all relevant purposes, retain practically the same ethical content.

Take deontic constraints, for example; if a person X has a deontic constraint not to murder innocent others, this could be expressed in a number of ways. It could be said, going by McNaughton and Rawling's formalization, that he has a reason of the sort: (x)(xS [(x does not kill the innocent)]). In a more general sense it could be said that the reason "X shouldn't kill the innocent" has the free-agent variable X. But this same deontic constraint could be expressed by saying "there is a reason for anyone not to murder the innocent."[5] In the rules-based formalization, what's inside the brackets could be replaced by "nobody kills the innocent". This reason-statement has no occurrence of X within the brackets and can be expressed without any free-agent variable, so it would be classed as agent-neutral. Yet it is expressing the same sort of deontic constraint that should be agent-relative.

Egoism can be described as the theory where each agent has a reason to benefit himself. This reason could, in turn, be expressed as agent-relative: the fact that A would benefit X is a reason for X to do it. This contains a clear reference to X. However, this same reason can be expressed by saying: the fact that one would benefit from doing A is a reason for one to do it. If such a reason were expressed in Skorupski's formalization, for example, it would result in something like:

> (x, φ) (It's being the case) that doing φ has a propensity to benefit an agent who does φ gives x reason to φ.

Which would be agent-neutral.[6]

A similar transformation can be done in the case of reasons to pursue our personal projects. They can be expressed in an agent-neutral form like:

> (x, φ) (It's being the case) that doing φ furthers the personal project of the agent who does φ gives x reason to φ.

In the case of special relationships, one could even imagine a world where all the reasons to favor those close to us had an agent-neutral form, without thereby losing any of their normative justificatory power.

Imagine a world where every person has a unique name that is determined by the application of an algorithm on the names of that person's parents, who are also unique. In this world, if I tell you someone's name, you can automatically know who that person's parents are.[7]

The people of this world do not use words like "son" or "daughter", and they also don't use indexicals in expressions such as "my offspring" or "my child". In this world, it would not be possible for reason-statements to contain pronominal back-references to the agents for whom they are reasons. Nonetheless, the people of this world believe in roughly the same moral demands regarding one's family that we do.

So a father could offer a reason for saving his son from a group of shipwrecks as "because he is XYZ2151 [the son's name]". Could this not be his reason?[8] Now, since the names of children are unique and determined by the names of the parents, it is somewhat implied that XYZ2151 is necessarily his child. But this link is a feature

---

[5] This was Schroeder's example (Schroeder, 2007, pp. 276–277).

[6] This was John Broome's point (Broome, 1995, pp. 315–317), also described in Buckland (2019).

[7] For the algorithm to determine the name of a person, it would also have to take as an input the previous number of children of the parent couple. It would require this in order to give each child a unique name.

[8] In response, someone like Pettit might argue that such reason statements must be fully specified, and that the one in question isn't. But the problem is that there are no clear standards for what counts as a fully-specified reason-statement. In this case, the reason-statement requires no further specification to function properly as a reason that someone might give to justify their action.

of his world, not of the reason-statement. The reason-statement doesn't have a pronominal back-reference to the father, and so, according to a formalization like that of Pettit, it would be agent-neutral, even though it concerns the sort of special relationships partiality that should make it agent-relative.

The tension is the same across all of these issues. Reason-statements are fickle bits of language, and they can seemingly be altered from agent-relative to agent-neutral with slight changes that leave their ethical content largely intact. Furthermore, reason-statements are not reasons, understood as the underlying normative reality that the statements are describing. This means, among other things, that not every statement corresponds to a distinct reason, and that in fact wildly different reason-statements seem to refer to the same normative reality even if they are linguistically different. In the case of consequentialism, the reason to take my brother to the movies and the reason to maximize welfare are not distinct reasons; they are the same reason expressed by two different reason-statements. The same is true in the examples of personal projects, special relationships, and deontic constraints. Although the corresponding reason-statements are manipulated from agent-relativity to agent-neutrality, the underlying normative reality is the same, and both reason-statements seem to justify the same thing in the same way, so much so that it seems improper to speak of two reasons, one relative and the other neutral, as if they were distinct things.

If this is so, why insist on defining the agent-relativity of reasons through some grammatical or linguistic property of a reason-statement? A proper conception of agent-relativity should point to some underlying moral property about reasons that is not just about the language used to describe them.

4. Looking Back to Parfit on Reasons and Theories

So how can we know whether a reason is agent-relative without resorting to the linguistic properties of isolated reason-statements?

To get a hint on how we might solve this problem, let us put the question regarding normative reasons to the side and consider a new question: can explanatory reasons be agent-relative or agent-neutral? We will argue that they can't, and that the distinction only makes sense in relation to normative reasons. This will show us that the key to finding the proper conception of agent-relativity lies in the normative aspect reasons, and not just in their structure as reasons described by a certain kind of reason-statement. So let us briefly discuss the ontology of different types of reasons.[9]

For theorists like Scanlon and Raz, reasons are essentially facts with a certain relation to actions or attitudes (Scanlon, 2014; Raz, 2010, Section 5).

And what is this relation? This depends on what type of reason we are talking about. In this respect there are two main types: explanatory reasons and normative reasons.

Explanatory reasons are the reasons why things happen. Regarding agents, they could also be called motivating reasons. If a person feels hungry and this leads him to eat, then his feeling of hunger was the explanatory reason for the action of eating. These reasons enter into causal explanations of actions without necessarily justifying them. In this respect, asking why someone did something is no different than asking why it rained yesterday or why a building collapsed. In both cases, we are asking for the facts (the reasons) that played

[9] We are here in stark opposition to McNaughton and Rawling (1995b, p. 33), who believed that the distinction should be seen as purely structural and independent of more substantive ontological claims.

a causal role in what ended up happening. Before an action happens, to have a reason in this sense would just be to have a motivation or some other consideration that exerts causal influence on some agent.

Scanlon and Raz, on the other hand, are mainly concerned with normative reasons. Such reasons have to do with justification and moral authority (Scanlon, 2014, p. 10). Raz saw the normative reasons for an action as facts in the case for that action, as well as its value properties (Raz, 2010, Section 5). These are the sorts of reasons we use to morally justify our actions, and when we ask others to justify theirs, we expect them to appeal to normative reasons and not just to explanatory ones. Reasons are facts, and the activity of reasoning, including practical reasoning, consists in identifying and responding to these facts (Raz, 2010, Section 2).

Let's now propose this question. Can there be agent-relative and agent-neutral explanatory reasons? If we look at our reason-giving practices, it becomes difficult to see how this could be so.

Let's say someone punches his colleague. When asked why he did this, this person replies, "Because he was being annoying." Now, this statement of a reason doesn't make any reference to the agent, but it doesn't seem to us agent-neutral on that account. And what about the more general distinction of agent-relativity/neutrality as being about whether or not the reason applies to all agents or only to some? Well, what could this mean in relation to explanatory reasons? It could mean that someone being annoying would always cause his colleague to punch him, whoever he was, and that, in this case, the reason would therefore be agent-neutral. In this case, this is certainly false, so would the reason therefore be agent-relative? We could maybe call it that, but this would be vacuous, since there is no fact that would cause a specific action in a uniform way across all agents, even if one assumes similarity of circumstances. To deny this is not to say anything substantive about the explanatory reason described by the statement "Because he was being annoying" in relation to the act of punching a colleague.

Could the distinction depend on other beliefs or attitudes that the agent holds? Perhaps if the agent thinks that everyone would act like he did (even if they don't) then the reason would be agent-neutral, and agent-relative if otherwise. The problem with this is that the same reason, "because he was being annoying," could be given by the agent if he holds these beliefs, if he holds the opposite beliefs, or if he holds no beliefs in either way, or if he simply hasn't thought of the question.

The point is that, when it comes to explanatory reasons, the distinction is vacuous. Explanatory reasons are meant to provide a particular causal link between an agent and an action or attitude. They fully function as explanatory reasons independently of the agent's beliefs and independently of whether the facts of the reason would cause the same action in others. They may or may not, but a particular explanatory reason always refers to a particular causal link. To demand something more, like a "full specification" of such reasons that would include either beliefs of the agent or a declaration of whether it would always cause the action in question, is already to artificially restrict the sorts of things we accept as explanatory reasons in our reason-giving practices.

This hints at the fact that we must look for the key to agent-relativity in the normative aspect of reasons, in whatever it is that makes them true normative reasons.

This final part will be a tentative exploration of how this can be done, and of how the agent-relative/neutral distinction should be understood once we reject the attempts to formalize it through the linguistic properties of isolated reason-statements.

The main claim is that we should go back to Parfit and revive the essential link between the agent-relativity of moral theories, or what we will call moral "outlooks", and the agent-relativity of the reasons associated with

them. Reasons are either agent-relative or agent-neutral only in relation to some moral outlook, in relation to which those reasons are true normative reasons.

An outlook is similar to, yet broader than, a theory. A moral theory can be understood as a collection of principles for regulating the conduct of moral agents, institutions, or what have you. In the case of comprehensive moral theories, these principles usually have an underlying, unified rationale, and they tend to be totalizing in the sense of establishing how agents should act in either all circumstances or at least in a wide variety of them. In contrast, what we're calling an "outlook" doesn't require a unified rationale, and it doesn't have to be totalizing in the slightest. An outlook could contain only a single injunction like "Do not kill innocent people", or multiple such injunctions without a unified justification for all of them.

Of course, the true or correct moral outlook would probably require this. What's crucial to note for our purpose of giving content to the agent-relative/neutral distinction is that an outlook is assumed to be true and then taken as a fixed point for the consideration of specific reasons. It is only in relation to a moral outlook, assumed to be true, that reasons will be either agent-relative or agent-neutral.

In what follows, standard moral theories such as consequentialism or egoism will be used as examples of moral outlooks, as well as political theories such as rights-libertarianism. The specific versions of these outlooks won't be specified in depth, but this shouldn't be necessary for our present purpose.

For Parfit, the agent-relativity of theories consisted in whether or not they gave different aims to different agents. He was criticized for supposedly importing too much teleology into this conception of agent-relativity (Löschke, 2021, Section 4). There is truth to these criticisms, since a moral theory or outlook may not give agents aims in such a strict sense, and its injunctions may instead be about refraining from acting in certain ways towards others.

However, we shouldn't abandon Parfit's key insight. All moral theories either tell agents how to act or how to be, or restrict their actions in some ways while demarcating a space where they may act freely or according to their own opinions of what's best. A moral outlook like rights-libertarianism might not establish aims in a strict sense, but it does limit our conduct towards others in certain ways. Such a theory does make claims like "you shouldn't coerce others", and therefore gives agents normative reasons not to do so, even if this doesn't establish aims or goals in the teleological sense. Also, outlooks of this sort do not claim a monopoly on all moral injunctions, and they allow agents to follow their own personal moralities as long as they don't infringe on the rights of others. An outlook like Rawls' Justice as Fairness, a form of liberal egalitarianism, is similar in this regard. It works against a backdrop of ethical pluralism by establishing principles of political justice and a space where each person is allowed to follow their own comprehensive moral doctrine. So it's true that not all moral outlooks set up aims in the teleological sense, but the essence of agent-relativity can still be found in the interplay of the different reasons and injunctions that a moral outlook lays out.

With this in mind, we can finally define agent-relativity, first in outlooks and then in reasons. A moral outlook is agent-relative if the normative reasons it either demands or allows agents to adopt for themselves can come into conflict. If they can't, then all of them can be said to harmonize, and the outlook is agent-neutral.

So, an outlook like rights libertarianism doesn't give agents aims in a strict sense, but it does allow agents to adopt different normative reasons in a specified sphere, and these won't be harmonizable. Even if no one is infringing on the rights of others, this is not the only way that the pursuit of one's normative reasons by one person can thwart that same pursuit in someone else. For example, if I build a monument to my religion on a piece of

land, a person with a different religion can no longer build his monument on that piece of land. The first person's pursuit of the normative reasons he adopted for himself is not harmonious with the second person's pursuit of the reasons *he* adopted for himself, even if no one's rights are being infringed. Since the outlook permits both of these pursuits, it is agent-relative. In contrast, an outlook like standard consequentialism is agent-neutral because the reasons that follow from its injunction to maximize welfare always harmonize. One agent's pursuit of general welfare maximization never conflicts with the same pursuit in another agent, since it is always the same welfare being maximized. Of course, in practice the two pursuits can conflict, even if both agents believe in consequentialism. But the point is that this would have to happen due to mistakes, lack of information, or weakness of will. If all agents had perfect information and could perfectly adhere to all that consequentialism demands, then their pursuits could not conflict, and this makes consequentialism an agent-neutral moral outlook.

In relation to an outlook, a reason described is agent-relative if the outlook is agent-relative, agent-neutral if the outlook is agent-neutral, or the reason is not a true normative reason at all (in relation to that outlook) and is therefore neither agent-relative nor agent-neutral (even though it can be in relation to a different outlook). It is only in relation to an outlook that a reason can be classed as agent-relative or agent-neutral. This can't be done by considering the reason in isolation, much less by pointing to a linguistic property of the reason-statement that describes it.

There is a special caveat. There can be agent-neutral reasons in relation to agent-relative outlooks, meaning reasons that apply to all agents even if they are in relation to an agent-relative outlook, but only if this is explicitly defined as such in the outlook. For example, we can think of a possible outlook[10] that is mostly egoistic but makes an exception for an injunction to have slavery abolished that does apply to everyone and supersedes every other reason as long as slavery still exists. This specific pursuit would be harmonious amongst all agents since, according to the outlook, there can't be others that are required or permitted to follow other injunctions or reasons at the same time.[11] This would be an agent-relative outlook, but in relation to it, the reason to have slavery abolished would be agent-neutral. For such cases to happen, the outlook must make clear that, whenever the injunction to abolish slavery is operative, all other reasons cease to apply. If the demand to abolish slavery did not supersede egoistic reasons and co-existed with them in the same timeframe, then the reason to abolish slavery would not be agent-neutral in relation to the outlook we are considering.

There can't, however, be agent-relative reasons in relation to agent-neutral outlooks. If an agent-neutral outlook explicitly defined an agent-relative reason within itself, or if it followed from one of its injunctions, it would cease to be agent-neutral.

With this in mind, we can go back to the sorts of reason-statements that were discussed over the course of this paper.

---

[10] In the sense of logically possible, even though there most likely isn't any justification or convincing argument that can be made for it.

[11] This preserves Nagel's association of agent-neutrality with a lack of conflict in the promotion of states of affairs, while avoiding the controversial move of making all reasons into reasons for the promotion of states of affairs (Schroeder, 2007, p. 277). In these specific cases where an agent-relative outlook contains an agent-neutral reason, it must be explicit that the reason applies to everyone. But this is not enough. A reason "to grab the last slice of pizza because it is tasty" can apply to everyone in the sense of everyone having the same reason, but it is not therefore agent-neutral. This is because the pursuit of this reason is not harmonizable amongst those who have it, since one person getting the last slice will prevent others from doing so. My framework doesn't require all reasons to be reasons to promote something, but for a reason to be agent-neutral in this special case there must be harmony in all its different pursuits under the moral outlook.

An interesting consequence of our view is that, in relation to the moral outlook of consequentialism, even the reasons picked out by statements like: "I should take my brother to the cinema because it would increase his welfare", turn out to be agent-neutral, when they would be indisputably agent-relative if we were still focused on references to the agent present in the reason-statement.

This might seem strange, but it is only strange if we are still in the mindset of looking for agent-relativity in isolated reason-statements, unconnected to a moral outlook. There is nothing strange about thinking in this more holistic way, since whatever it is that makes a reason truly normative will also be what determines its scope of application to other agents, and whether or not its pursuit is in harmony with other normative demands. In this case, in relation to consequentialism, all these reasons are agent-neutral because they are derived from the one true normative reason to maximize welfare. It is thus no problem if these sorts of reasons are admitted into consequentialism, as they are themselves agent-neutral in relation to it.

Also in the case of personal projects, special relationships, or deontic constraints, the derivability problem is not to be feared. It doesn't matter if we can derive equivalent reason-statements that do not contain either a free agent-variable or a pronominal backreference to the agent for whom they are reasons. The reasons such statements pick out don't therefore become agent-neutral. They are still agent-relative in relation to any outlook that accepts the moral rightness of personal projects, special relationships, and deontic constraints. Any outlook that does this will necessarily be agent-relative, since the different pursuits of these things are not harmonizable—and so the reasons themselves, no matter how they are described, will also be agent-relative.

The reason for X not to murder the innocent is the same as the reason for anyone not to murder the innocent. The reason for a person to further his project is the same reason for X to further his project. The reason for the father to save his son is the same reason for the father to save XYZ2151. All of them are agent-relative even without an essential reference to the agent in their reason-statements, as long as they are understood in relation to an agent-relative moral outlook, which they must be if they are to be truly normative reasons at all.

These same reasons, in relation to a different moral outlook, might not be true normative reasons at all. In fact, in relation to an outlook like egoism, many of the reasons we are most confident in, like, for example, "the fact that a man is drowning is a reason to save him", are not true normative reasons, and are therefore neither agent-relative nor agent-neutral. Interestingly, that same reason-statement doesn't point to a true normative reason in relation to egoism, but it can point to an agent-relative reason in relation to a form of commonsense morality, or to an agent-neutral reason in relation to consequentialism (if saving the drowning man is the best way to maximize welfare).[12]

This too might seem strange: the fact that a reason's status can change like this. But here it is important to clarify that the actual normative reality doesn't change according to the moral outlook that we choose to consider. The true morality and the truly normative reasons, as well as their status as agent-relative or agent-neutral, do not change. But if we are trying to know whether an individual reason is agent-relative or agent-neutral, we must assume as true a wider normative reality–what I've been calling a moral outlook–and it is the relation of that reason to the other reasons, injunctions, and the general structure of the outlook, that will allow us to know whether the reason under consideration is agent-relative or agent-neutral. If the reason we are considering simply isn't part

[12] Just like different reason-statements can refer to the same reason, understood as the underlying normative reality, so can the same reason-statement refer to multiple different underlying moral realities.

of the outlook, then, cut off from its normative fabric, it can't even be considered a true normative reason and is neither agent-relative nor agent-neutral, just like an explanatory reason.[13]

The essence of agent-relativity is the conflictual nature of certain moral outlooks, the fact that they pull, or allow agents to be pulled, in different directions, so to speak, which makes it impossible to harmonize all their different pursuits. Parfit understood this along the lines of a theory giving different aims to different moral agents, and while we do not have to think strictly in terms of aims to be achieved, his key insight, that the agent-relativity of theories and reasons must be understood together, should never have been abandoned.

I think this is why McNaughton and Rawling's formalization, considered by some to be the best so far (Hammerton, 2018), must turn every reason into a rule that instructs agents to act to the best of their abilities to fulfill it, as long as there are no explicitly conflicting rules. This formalization takes the fixedness of a theory or outlook and turns it into a feature of the rule itself. Every rule becomes a kind of mini outlook, and to have a reason is just to be under a specific rule.

Even beyond the problems we've looked at so far, I think this approach is mistaken in the same way that Nagel's or even Scanlon's approaches are. They all claim that normative reasons must be expressed in a certain form: either as rules with a universal quantifier, as reasons to promote events, or as functions with variables that determine the circumstances they apply in and the agents they apply to. The problem is that we accept a variety of different things as reasons in our reason-giving practices, and there's no privileged form of expression that they all must abide by. Normative reasons are intimately tied to our practices of morally justifying ourselves to others, and various kinds of reason-statements fulfill this role without making any claims about rules, events, or whatever else. Claims that such reasons must be "properly expressed" or "fully specified" according to a specific form are of no help. The standards for what counts as a proper expression depend on context, and the justificatory role of normative reasons is very often fulfilled without their full specification, or even if they contain a degree of ambiguity.

Thus we shouldn't try to shoehorn the chaotic world of reasons into one neat form of expression. To do so is to unacceptably cut it up into shapes that we can manage, instead of facing it as it truly is. This resulted in the mistake of identifying agent-relativity with a property of this privileged form of expression—a form that is our own creation and therefore ephemeral—instead of with something substantive like the conflictual nature of a moral fabric in relation to which reasons can be truly normative and have the scope of their application determined.

Conclusion

In section 1, we considered the different applications of the notion of agent-relativity and rejected it in connection to values and ought-statements, but accepted it in connection to reasons, theories, and rules.

In section 2 we tackled the claim that the distinction between agent-relative and agent-neutral reasons is vacuous since all reasons are reasons for some set of agents, and they are therefore all agent-relative. We rejected this claim since, even if we accept that a reason's scope of application is always limited by the power and opportunities available to specific agents, a substantive notion of agent-relativity remains.

[13] This also does not imply that we must first know the correct moral outlook before we can justifiably claim something to be a true normative reason. This gets into questions of moral epistemology that are outside of our present scope, but our beliefs about true normative reasons influence our beliefs about the true moral outlook and vice versa, in a way where both are up for revision.

Once we see that the distinction is not vacuous, section 3 looks at different attempts to formalize it. It concludes that such attempts are not likely to succeed as long as they focus on the linguistic properties of isolated reason-statements. This is due to what we have called the "derivability problem", which is the fact that many agent-relative reason-statements can be transformed into agent-neutral ones while retaining the same ethical content, together with the fact that, even in agent-neutral theories like consequentialism, agent-relative reasons can be derived from agent-neutral ones. This makes it very difficult for any formalization to fulfill the original motivational impulse behind the distinction. On such a view, agent-neutral theories like consequentialism would turn out to be agent-relative, uncontroversially agent-relative theories like egoism would turn out to be agent-neutral, and the reasons associated with personal projects, special relationships, and deontic constraints could all be stated in an agent-neutral form.

In section 4 we argue for a different way of conceptualizing the distinction. We re-establish Parfit's original link between the agent-relativity of moral theories–that we now call "outlooks"–and that of reasons. In this framework, a reason is agent-relative or agent-neutral only in relation to a moral outlook that is assumed to be true. The essence of agent-relativity lies first and foremost in the fact that the normative reasons and injunctions of a given moral outlook are not harmonizable. The agent-relativity of reasons is secondary to this, and should be understood in relation to a moral outlook that both makes them true normative reasons and determines their scope of application across different agents.

References

Broome, J. (1995). Skorupski on agent-neutrality. *Utilitas*, 7, 315–317.

Buckland, J. (2019). Skorupski and Broome on the agent-neutral/relative distinction. *Utilitas*, 31(1), 59–82.

Bykvist, K. (2018). Agent-relative and agent-neutral reasons. In D. Star (Ed.), *The Oxford handbook of reasons and normativity*. Oxford: Oxford University Press.

Dancy, J. (1993). Agent-relativity—the very idea. In R. G. Frey & C. W. Morris (Eds.), *Value, welfare, and morality* (pp. 229–247). Cambridge: Cambridge University Press.

Hammerton, M. (2018). Distinguishing agent-relativity from agent-neutrality. *Australasian Journal of Philosophy*, 96(4), 661–675.

Humberstone, I. L. (1991). Two kinds of agent-relativity. *The Philosophical Quarterly*, 41(163), 144–166.

Hurka, T. (2003). Moore in the middle. *Ethics*, 113, 599–628.

Lee, J. (2024). On the agent-relativity of 'ought'. *Analysis*, 84(3), 483–492.

Löschke, J. (2021). Agent-relative reasons and normative force. *Philosophia*, 49, 359–372.

Mack, E. (1993). Agent-relativity of value, deontic restraints, and self-ownership. In R. G. Frey & C. W. Morris (Eds.), *Value, welfare, and morality* (pp. 209–232). Cambridge: Cambridge University Press.

McNaughton, D., & Rawling, P. (1995a). Agent-relativity and terminological inexactitudes. *Utilitas*, 7(2), 319–325.

McNaughton, D., & Rawling, P. (1995b). Value and agent-relative reasons. *Utilitas*, 7(1), 31–47.

Nagel, T. (1970). *The possibility of altruism*. Oxford: Clarendon Press.

Nagel, T. (1986). *The view from nowhere*. Oxford: Oxford University Press.

Parfit, D. (1984). *Reasons and persons*. Oxford: Clarendon Press.

Pettit, P. (1987). Universalizability without utilitarianism. *Mind*, 96(381), 74–94.

Raz, J. (2010). Reason, reasons, and normativity. In R. Shafer-Landau (Ed.), *Oxford studies in metaethics* (Vol. 5). Oxford: Oxford University Press.

Ronnow-Rasmussen, T. (2009). Normative reasons and the agent-neutral/relative dichotomy. *Philosophia*, 37, 213–238.

Scanlon, T. M. (1998). *What we owe to each other*. Cambridge, MA: Harvard University Press.

Scanlon, T. M. (2014). *Being realistic about reasons*. Oxford: Oxford University Press.

Schroeder, M. (2007). Teleology, agent-relative value, and 'good'. *Ethics*, 117(2), 265–300.

Skorupski, J. (2011). *The domain of reasons*. Oxford: Oxford University Press.

Stewart, R. (1993). Agent-relativity, reason, and value. *The Monist*, 76(1), 66–80.

Williams, B. (1973). *Utilitarianism: For and against*. Cambridge: Cambridge University Press.