Moral Polarism

Andrew Sepielli
University of Toronto

__Introduction__

Moral Polarism (first pass): The only considerations that are fundamentally relevant to the moral quality of an action are those that bear on the quality of the agent's exercise of agency or on the goodness and badness of what befalls a patient.

> Intuitive glosses: The agent and the patient are the two "poles", and nothing in between the poles matters; only the "takeoff" and "landing" matter morally.
>
> Opposing views: It matters fundamentally whether the agent's behaviour *caused* something to befall a patient; it matters whether an action falls into some commonsensical category: stealing, hitting, killing, lying…
>
> Polarism is not obviously opposed to any of the most well-known positions in moral philosophy — e.g. utilitarianism, deontology, contractualism, etc.; and it is not much of a guide to action. But I think it "cuts" a great deal of "ice" in fundamental moral philosophy, as I hope to show later.

__What is Polarism?__

It does not describe the bearers of moral significance in terms of the *metaphysics* of agency or patiency. It describes them in evaluative terms — i.e. in terms of what *bears on quality/goodness/badness*. So we are not required to delineate whether, e.g., the exercise of Zebulon's agency in his doing something to Yvette ends, and where the rest of this event begins.

An addendum: "If it doesn't matter out there, it doesn't matter in here": If, e.g., it is irrelevant according to Polarism whether the causal path from Zebulon's exercise of agency to what befalls Yvette is direct or circuitous, then Zebulon's thoughts about whether it is direct or circuitous are likewise irrelevant.

How to distinguish:

> "Z's exercise of agency in doing such-and-such to Y"
> "Z's doing such-and-such to Y"
> "It befalling Y that Z does such-and-such to her"
>
> One might ask: Isn't Z's exercise of agency just that she did such-and-such to Y? And isn't what befell Y just that X did such-and-such to her?

We have an intuitive sense, I think, if how to draw these distinctions, but we can also articulate the moral-theoretic role of each:

> The quality of Z’s exercise of agency in doing such-and-such to Y is the only determinant of the quality of Z’s doing such-and-such to Y that is fundamentally relevant to:
>
> Blameworthiness
> Punishment, on a reasonable retributivist view thereof
> Proper assessments of character
> The “fittingness” of remaining friends with someone
>
> The goodness or badness of what befalls Y is the only determinant of the quality of Z’s doing such-and-such to Y that is fundamentally relevant to:
>
> Compensation, “making whole”
> Matters of whose life has gone better/worse
> The “fittingness” of feeling sorry for someone (as opposed to feeling angry on their behalf)

___Arguments for Polarism___

___Argument #1: Mirroring___

The basic idea: Morality is about extending your concern for yourself outwards to others. Moral thinking addresses the matter of how to do so.

Suppose you care about or are concerned with good/acceptable things like having clean water to drink, or enrolling your child at a good school, or getting tickets to the Olivia Rodrigo concert. Then morality is about extending my care and concern to those things. *Ceteris paribus*, my moral thoughts are correct to the extent that they appropriately reflect care and concern about these things; my actions are right to the extent that they treat these as ends in the ways that I’d treat the objects of my own cares/concerns as ends. I thereby “mirror” your cares and concerns.

Suppose a third person thinks that you should not have clean water, or be able to enrol your child at a good school, or get tickets to the Olivia Rodrigo concert, and intends to thwart these aims of yours. Then he *anti-mirrors* your cares and concerns. And it is arguably good for me, then to *anti-mirror* his (bad) cares and concerns. Why?

> 1) Maybe because they have as their objects others not getting what they care about, are concerned with, and so mirroring them would take on the opposite valance;
> 2) Maybe because unless I anti-mirror his (bad) cares and concerns, I don’t “take your side”. But I should take your side.[1]
>
> But it’s also arguable that anti-mirroring has no *independent* value.

[1] On related ideas, see Hampton, “Correcting Harms versus Righting Wrongs” (1992).

In thinking that the goodness/badness of what befalls patients matters, we mirror their cares and concerns. In thinking that, e.g., agents' exercise of agency matters, we anti-mirror their bad intentions, insufficient concern for others, etc.

But in thinking that, e.g., the circuitousness of causal connection matters, or that "force transfer" matters, we don't mirror or anti-mirror anything. This stuff might matter aesthetically, but not morally.

___Argument #2: Debunking___

Two anti-Polarist views to debunk:

> 1) Causation matters morally — i.e. it matters that your action *caused* an effect (rather than just that the effect depended on your action such that you could be said to control it), and maybe it matters *how* your action caused an effect.
>
> 2) Form matters morally — i.e. it matters that your action instantiates some form like *killing*, *hitting*, stealing*,* *lying*, etc.[2]

On 1): We attribute causation, as opposed to mere counterfactual dependence, on the basis of features that are morally irrelevant:

> Maybe transfer of forces — Talmy, "Force Dynamics in Language and Cognition" (1988); Wolff, "Representing Causation" (2007)
>
> Maybe "portability": "the function of ascribing relations of actual causation is to locate dependence relations that are highly portable to other systems" (Hitchcock, "Portable Causal Dependence: a Tale of Consilience" (2012))
>
>> Portable dependence relations are ones between variables such that a small intervention on an independent variable allows one to make a substantial change in the specific way one wishes in the dependent variable in a variety of relevantly similar situations — such that the relations are said to be "portable" or "generalizable" to these either situations. E.g. lighting the match (relatum in portable relation w/ fire) vs. Changing amount of oxygen (relatum in non-portable relation w/ fire).[3]

[2] See Fodor et al., "Against Definitions" (1980) on causative verbs for arguments that 2) differs from 1).

[3] See Hitchcock (2012); Lombrozo, "Causal Explanatory Pluralism" (2010); Morris et. al., "Judgments of Actual Causation Approximate the Effectiveness of Interventions" (2018); Quillien, "When do we think that X caused Y?" (2020); Quillien and Lucas, "Counterfactuals and the Logic of Causal Selection" (2023).

On 2): We “chunk”[4] actions into forms on the basis of features and considerations that are, again, morally irrelevant, and then value gets assigned to those chunks *via* evolutionary and learning processes. Why do we chunk?

> Maybe for intentional understanding:
>
> > "a low-level skill for detecting meaningful structure in the behavior stream seems to be a prerequisite for infants' emerging sophistication in the realm of intentional understanding.” — Baird and Baldwin, “Making Sense of Human Behavior: Action Parsing and Intentional Inference” (2001)
> >
> > In other words, we parse/chunk actions in the way that we do because doing so is useful in assigning psychological states to other agents, and thus to predicting their future behaviour. “Zebulon hit/stole from Yvette” is useful in that regard; “Zebulon moved his arm swiftly to the right” is not. But why should the sorts of things to which fundamental moral significance is assigned (e.g. by a learning process) depend on such considerations of usefulness?
>
> Maybe for computational efficiency:
>
> > “What are the benefits of sequence-based action selection? As discussed in the previous section, expression of a sequence of actions is faster than selecting actions one by one, based on the action evaluation process. This can be for several reasons; e.g., identification of the current state by processing environmental stimuli can be time consuming; and the evaluation of actions using a model-based process is slower than having solely to select the next action from the sequence. Besides being faster, executing actions without going through the decision-making process makes it possible to perform a simultaneous task that requires decision-making resources.” — Dezfouli and Balleine, “Habits, Action Sequences, and Reinforcement Learning” (2012)
> >
> > “...chunking provides a mechanism for the acquisition and the expression of action repertoires that, without such information compression would be biologically unwieldy or difficult to implement.” — Graybiel, “The Basal Ganglia and the Chunking of Action Repertoires” (1998)
> >
> > But again: But why should the sorts of things to which fundamental moral significance is assigned depend on such considerations of computational efficiency or implementability?

[4] See Miller, “"The Magical Number Seven, Plus or Minus Two: Some Limits on Our Capacity for Processing Information” (1956).

__Implications of Polarism__

Causal structure doesn't matter. For example, Frances Kamm tries to explain why we shouldn't push the man off the footbridge, and why terrorism is *ceteris paribus* worse than conventional warfare (and much else besides) *via* something she calls the Doctrine of Productive Purity, the first part of which says:

> "If an evil cannot be at least initially sufficiently justified, it cannot be justified by the greater good that it is necessary (given our act) to causally produce. However, such an evil can be justified by the greater good whose component(s) cause it, even if the evil is causally necessary to help sustain the greater good or its components." — Kamm, "Toward the Essence of Nonconsequentialist Constraints on Harming" (2006)
>
> But if Polarism is right, then the order of goods and evils in a causal chain doesn't matter, nor does the difference between sustaining and causing.

Causation vs. counterfactual dependence doesn't matter. For example, Molly Gardner proposes the following principle:

> "A harmful action that causes greater benefits can sometimes be justified by those benefits, but a harmful action that does not cause greater benefits cannot be justified by any subsequent benefits that the action, itself, does not cause." Gardner, "When Good Things Happen to Harmed People" (2019)
>
> Gardner countenances ordinary cases of non-causal counterfactual dependence. But if Polarism is right, then the distinction between this and causation doesn't matter.

But if Polarism implies that causation is irrelevant, does it also imply that counterfactual dependence is irrelevant? Problems lurk either way:

> If we say <u>yes</u>, then then we seem to be left with an absurdity — that it doesn't matter how your action is materially connected to anything in the world.
>
> If we say <u>no</u>, then why not? This looks like special pleading…
>
> But consider: "Ought" seems to imply "can", and so that I ought to do something depends on whether, e.g., it helps people, but also on whether I, the agent, *can do it*. But it seems like these are very different sorts of things. I shouldn't care about what I can do in the same way I should care about people being helped. It doesn't go in the "pros" or "cons" column. I wouldn't admire you less for doing something helpful just because I couldn't do it myself! Ought-thoughts are playing two roles — evaluation, and the guidance of action.
>
> What I would say: Just as some actions are impossible, some agent-pole-event/ patient-pole-event pairs are impossible; and some are necessary in the sense that

> the patient-pole-event will happen regardless of whether the agent-pole-event does. These pairs are practically irrelevant, just as impossible actions, and necessary actions (if there are any) are practically irrelevant. To say that a PPE counterfactually depends on an APE is just to say that the pair is *not* practically irrelevant. So this is why counterfactual dependence matters to what you ought to do. But this same reasoning doesn't explain why causation as such, as distinct from counterfactual dependence matters. To think otherwise would be like thinking "It's plausible that 'ought' implies 'can', so it's plausible on similar grounds that ought implies 'can with your feet' or 'can without prosthetics'."

Implications for Jonathan Bennett's argument re: doing vs. allowing in *The Act Itself* (1998) and elsewhere:

> Bennett says that we do harm when we comport ourselves somehow, a harm occurs, and very few or the other ways of comporting ourselves are such that the harm would have happened when and how it did. We allow a harm to happen when most of the other ways are such that the harm would have happened. But very few vs. most is morally irrelevant, and so doing vs. allowing is, too, Bennett argues.
>
> Bennett allows that this may not be a perfect analysis of the concepts of doing and allowing harm, but writes:"When a purported analysis of some concept does not precisely capture the whole truth about when the concept-word would be the *mot juste*, the question is whether the omitted aspects matter for fundamental moral theory; and sometimes they do not."
>
> My thought is that Polarism might be way to rule out, in a principled way, certain "omitted aspects" as not mattering for fundamental moral theory.

Implications for Relationalism

> Relationalism: The fundamental moral facts are of the form "Zebulon has a duty to Yvette to do A (not to A) to her". These are thought to ground duties of compensation that Zebulon has to Yvette in the event that she is injured by his not doing A (doing A).[5]
>
> But if Polarism is true, then the fundamental objects of moral concern are the agent- and patient-polar facts; relations only matter for practical purposes, as explained above.
>
> This seems to lend support to a system where people pay in to a pool based on levels of fault, and then people get compensated (or I guess, just paid) based on levels of suffering.

[5] See, e.g., Thompson, "What Is It to Wrong Someone?: A Puzzle about Justice?" (2004), Zylberman, "Relational Primitivism" (2021).