

The restrictions of consequentialism

Jan Gertken, Humboldt-Universität zu Berlin

1. Introduction

It seems plausible to many that there are *deontic restrictions*, meaning that performing a morally problematic act is not justified by the fact that someone else will perform an act of this type if one does not do it.

It has been argued that in order to integrate restrictions into a consequentialist outlook, one has to make suitable assumptions about so-called *agent-relative value*. I will argue that it is not our understanding of value, but rather our understanding of outcomes that is crucial for a consequentialist account of restrictions.

2. Conceptual groundwork

Consequentialism: an act is morally right if and only if, and because, its outcome realises at least as much value (of a certain sort) as all alternative outcomes that the agent can bring about.

I use “action” in a wide sense that includes everything that is a possible object of an intention. By *an outcome* of an action, I mean a state of affairs that is brought about by that action. I use “the outcome of an action” to refer to the collection of *all* states of affairs brought about by an action.

Agent-neutral value is the type of value ascribed by unqualified judgments about what is good, valuable or desirable *simpliciter* (and their comparative forms).

Agent-relative value is the type of value ascribed by judgments to the effect that certain things, events or states of affairs are good, valuable etc. *relative to some particular agent* (and their comparative forms).

Agent-neutral consequentialism comprises all consequentialist theories that entirely rely on ascriptions of agent-neutral value in their axiology.

Agent-relative consequentialism comprises all consequentialist theories that involve an axiology that is at least partly formulated in terms of agent-relative value.

There is a *deontic restriction against ϕ -ing* if and only if: (i) acts of this type are morally problematic, and (ii) one’s own ϕ -ing is not justified by the fact that someone else will ϕ unless one ϕ -s.

If there is an *absolute restriction* against ϕ -ing, then is *never permissible* to ϕ , no matter what the consequences of not ϕ -ing might be. The existence of a *threshold restriction* against ϕ -ing is compatible with the idea that ϕ -ing is permissible in some situations, e.g. when very bad consequences or catastrophic events would otherwise occur (cf. Kagan 1989: 189–191).

There is a *minimal restriction* against ϕ -ing if, and only if, for all agents A: absent further relevant facts, it is wrong for A to ϕ , even in situations where someone else will ϕ unless A ϕ -s.

The minimal restriction against killing innocents against their will (=MRK):

(MRK) For all agents A: absent further relevant facts, it is wrong for A to kill an innocent person against her will, even in situations where someone else will kill an innocent person against her will unless A does so.

A moral theory T *consequentialises* MRK if and only if T is a consistent consequentialist theory that implies MRK. T *fully accommodates* MRK if and only if T consequentialises MRK without incurring controversial commitments over and above those that are involved in MRK itself.

Deontic restrictions are sometimes (cf. Portmore 2011: 97–98) defined such that they imply a *constraint on maximising the good*, meaning that some actions may not be done in order to maximize the good (for the relevant notion of constraint, cf. Nozick 1974: 28–35, Kagan 1989: 4, and Schroeder 2007). This is not how I understand “restriction” in this talk.

3. The argument against agent-neutral consequentialist accounts of restrictions

There is a well-known argument for the view that no agent-neutral consequentialist theory can consequentialise a restriction such as MRK (cf. Schroeder 2007). Consider the following two schematic descriptions of situations that are covered by MRK:

Table 1

Case #1: B will kill D unless A kills C	Case #2: A will kill D unless B kills C
Outcome of A’s killing (=o ₁): < A kills C, B does not kill D, C is dead ... >	Outcome of B’s killing (=o ₃): < B kills C, A does not kill D, C is dead, ... >
Outcome of A’s refusal to kill (=o ₂): < B kills D, A does not kill C, D is dead, A lets D die, ... >	Outcome of B’s refusal to kill (=o ₄): < A kills D, B does not kill C, D is dead, B lets D die, ... >

Assume that there is an agent-neutral consequentialist theory T that implies MRK. In order to do so, T needs to imply both that it is wrong for A to kill in case #1, and that it is wrong for B to kill in case #2. Hence, T needs to imply both that *o₂ is better than o₁* and that *o₄ is better than o₃*.

Abstracting from all possible contingent differences that might obtain, the relevant evaluative asymmetries between o₂ and o₁ and o₄ and o₃ must somehow result from the different negative contribution made by *A’s killing* and *B’s killing*. If there is a relevant difference between these outcomes, it cannot concern *what* is done, but only *who* does it.

So in order to imply both that o₂ is better than o₁ and that o₄ is better than o₃, T needs to assume both that *A’s killing is worse than B’s killing*, and that *B’s killing is worse than A’s killing*.

These assumptions are inconsistent. Hence, there is no agent-neutral consequentialist theory that consequentialises MRK.

4. Two possible responses

Many advocates of agent-relative consequentialism argue that we need an agent-relative axiology if we want to accommodate deontic restrictions into a consequentialist framework (cf. Sen 1983, Dreier 1993, Portmore 2005, and Smith 2009). It is usually assumed that “A’s killings are worse *relative to A* than B’s killings” and “B’s killings are worse *relative to B* than A’s killings” are jointly consistent.

Friends of agent-neutral consequentialism can question a hidden assumption on which the argument from part 3 relies. Assume that a correct description of the outcomes of A’s killing and of A’s refusal to kill is such that if A kills in case #1, the outcome of his action includes *A’s killing of C*, whereas if A does not kill, the outcome of his action does not include *B’s killing of D*, but merely *D’s death* and *A’s act of letting D die*. Assume that similar modifications are made with regard to the outcomes of B’s actions in case #2. This gives us:

Table 2

Case#1: B will kill D unless A kills C.	Case #2: A will kill D unless B kills C.
Outcome of A’s killing (=o ₁): < A kills C, B does not kill D, C is dead, ... >	Outcome of B’s killing (=o ₃): < B kills C, A does not kill D, C is dead, ... >
Outcome of A’s refusal to kill (=o ₂): < A does not kill C, D is dead, A lets D die, ... >	Outcome of B’s refusal to kill (=o ₄): < B does not kill C, D is dead, B lets D die, ... >

Given this description of the relevant outcomes, any axiology that takes killings of innocents to be *agent-neutrally* worse than mere *deaths* of innocents and *agent-neutrally* worse than acts of *letting innocents die* will imply both that o₂ is better than o₁ and that o₄ is better than o₃ without any inconsistency.

A conception of outcomes is *agent-centred* if, and only if, it implies: for all agents A, there are at least some types of actions ϕ such that the only person who can bring about outcomes including instances of ϕ -ing that are *done by A* is *A*. A conception of outcomes is *agent-indifferent* if and only if it is not agent-centred.

Given an agent-centred conception of outcomes, both agent-neutral and agent-relative axiologies can be used to consequentialise restrictions.

What is more: unless consequentialists accept an agent-centred conception of outcome, even an agent-relative axiology will not allow them to *fully accommodate* deontic restrictions, such as MRK.

5. Intermezzo: What does it mean to talk of agent-relative value?

How are ascriptions of agent-relative value to be understood, and how does agent-relative value differ from agent-neutral value?

“Good relative to an agent” cannot be understood as referring to what is *good for an agent*, to *what is good from an agent’s perspective*, and it also cannot be treated as a primitive technical term (cf. Schroeder 2007). The most promising approach to understanding agent-relative value seems to be a fitting attitude analysis.

Two possibilities:

- (i) p is better than q *relative to A* if and only if, and because, it is *fitting for A* to prefer p to q; p is better than q if and only if, and because, it is *fitting for everyone* to prefer p to q (cf. Schroeder 2007).
- (ii) p is better than q *relative to A* if and only if, and because, it is *fitting for A* to prefer p to q; p is better than q if and only if, and because, it is *fitting for an impartial observer* to prefer p to q (cf. Suikkanen 2009).

6. No full accommodation of MRK without an agent-centred conception of outcomes

Given an agent-indifferent conception of outcomes, an agent-relative version of consequentialism needs to assume the *evaluative asymmetry hypothesis* (= EAH) in order to consequentialise MRK:

- (EAH) For all agents A and B: other things being equal, A’s acts of killings innocents against their will are worse *relative to A* than B’s acts of killings innocents against their will.

It is not only plausible to assume that A’s killings are bad *relative to A*. It is also plausible to assume that B’s killings are bad *relative to A*. EAH furthermore cannot be postulated as a basic assumption. If x and y both have negative (or positive) value, then x cannot have more negative (positive) value than y without there being something that makes it the case that x has more negative (positive) value than y.

Consequentialists who accept EAH in order to consequentialise MRK need to give an account of *why* (other things being equal) A’s killings are worse relative to A than B’s killings, even though both acts have negative value relative to A.

The only differences between A’s killings and B’s killings to which consequentialists could appeal for this purpose must have something to do with the fact that A’s and B’s killings are performed by a *different agents*. This leads us to the *Identity as Intensifier Hypothesis* (=IIH):

- (IIH) If ϕ has negative value relative to A, and if ϕ is performed by A, then the fact that A is identical to the agent performing ϕ *intensifies* the negative value that ϕ has relative to A.

IIH undermines the consequentialisation of MRK, given the framework of an agent-indifferent conception of outcomes. (For objections to IIH itself, cf. Kamm 1996: 240–253 and Emet 2010).

The premises of my argument:

- (1) In order to consequentialise restrictions, one must accept a notion of outcome that implies: if p is part of the outcome of A 's ϕ -ing (and hence part of what A brings about by ϕ -ing), then A 's *bringing about of p* is also part of that outcome (cf. Sosa 1993: 112).
- (2) Not only acts of killing have negative value (agent-relative or agent-neutral), acts of bringing about acts of killing have negative value as well.
- (3) For all agents A : A 's act of killing cannot differ in value (either agent-neutral or agent-relative) from A 's act of bringing about his act of killing, at least in those cases where A brings about his act of killing *by killing*.

From (1), together with the assumption that the relevant conception of outcomes is agent-indifferent, it follows that:

- (4) The outcomes of A 's killing ($=o_1$) and of A 's refusal to kill ($=o_2$) in a case covered by MRK will at least involve the following:

A kills ($=o_1$): $\langle A$ kills, B does not kill, A brings it about that A kills, ... \rangle

A refuses to kill ($=o_2$): $\langle A$ does not kill, B kills, A brings it about that B kills, ... \rangle

IIIH, together with (2), implies that:

- (5) A 's bringing about of B 's killing is worse relative to A than B 's bringing about of B 's killing.

From (3), it follows that:

- (6) B 's killing has the same negative value relative to A as B 's bringing about of B 's killing.

From (5) and (6), it now follows that:

- (7) A 's bringing about of B 's killing is worse relative to A than B 's killing.

As was pointed out before, IIIH implies that:

- (8) A 's killing is worse relative to A than B 's killing.

Hence, we can conclude that:

- (9) Outcome o_1 contains A 's *killing*, which is worse relative to A than B 's *killing*. Outcome o_2 contains A 's *act of bringing about B 's killing*, which is also worse relative to A than B 's *act of killing*.

Remember: in order to consequentialise MRK, an agent-relative consequentialist theory must imply that *the whole outcome of A 's killing* is worse relative to A than *the whole outcome of A 's refusal to kill*. Given (9), it is hard to see how this could be achieved.

For o_1 to be *overall worse* relative to A than o_2 , it is not enough that A's killing (part of o_1) is worse relative to A than B's killing (part of o_2). It also needs to be the case that A's killing (part of o_1) is worse relative to A than A's bringing about of B's killing (part of o_2).

The gap: IIIH does not allow us to do conclude that A's killing is worse relative to A than A's bringing about of B's killing. A's killing and A's bringing about of B's killing are not *actions of the same type* performed by *different persons*, but rather *different types of actions* performed by *the same person*.

Evidence against MRK: According to IIIH, both A's bringing about of B's killing and A's killing are worse relative to A than B's killing due to the fact that they *involve A*. So IIIH offers strong support for the view that the outcome of A's killing ($=o_1$) is *not* worse relative to A than the outcome of A's refusal to kill ($=o_2$).

Given the framework of an agent-indifferent conception of outcomes, agent-relative consequentialist theories that aim to consequentialise MRK face a dilemma. Either they postulate an ungrounded evaluative asymmetry between one's own killings and the killings done by others, or they accept IIIH, in which case their account fails to consequentialise MRK. In either way, they do not fully accommodate MRK.

7. Conclusions

Consequentialists who want to fully accommodate deontic restrictions should accept an agent-centred conception of outcomes. In this case, both an agent-relative and an agent-neutral axiology will allow them to consequentialise deontic restrictions, such as MRK. This approach should be taken seriously, as least from a consequentialist perspective. It might not lead to a successful consequentialist account of deontic restrictions – but it might nevertheless be the consequentialist's best option.

Department of Philosophy, Lund University, 27 October 2016
jan.gertken@hu-berlin.de

References

- Dreier, J. (1993). "Structures of Normative Theories." *The Monist* 76: 22–40.
- Emet, S. (2010). "Agent-Relative Value and Agent-Relative Restrictions." *Journal of Ethics and Social Philosophy* 4: 1–13.
- Kagan, S. (1989). *The Limits of Morality*. Oxford: Clarendon Press.
- Kamm, F. (1996). *Morality, Mortality*. Volume II. Oxford: Oxford University Press.
- Nozick, R. (1974). *Anarchy, State, and Utopia*. New York: Basic Books.
- Portmore, D. (2005). "Combining Teleological Ethics with Evaluator Relativism: A Promising Result." *Pacific Philosophical Quarterly* 86: 95–113.
- Portmore, D. (2007). "Consequentializing Moral Theories." *Pacific Philosophical Quarterly* 88: 39–73.
- Portmore, D. (2011). *Common Sense Consequentialism*. Oxford: Oxford University Press.
- Schroeder, M. (2007). "Teleology, Agent-Relative Value, and 'Good'." *Ethics* 117: 265–295.
- Sen, A. (1983). "Evaluator Relativity and Consequential Evaluation." *Philosophy and Public Affairs* 12: 113–132.
- Smith, M. (2009). "Two Kinds of Consequentialism." *Philosophical Issues* 19: 257–272.
- Sosa, D. (1993). "Consequences of Consequentialism." *Mind* 102: 101–122.
- Suikkanen, J. (2009). "Consequentialism, Constraints and The Good-Relative-to: A Reply to Mark Schroeder." *Journal of Ethics and Social Philosophy*: 1–8.