Free & Equal | A Journal of Ethics and Public Affairs

Chappell, Richard Yetter 2025 Preference and Prevention: A New Paradox of Deontology. *Free & Equal: A Journal of Ethics and Public Affairs*, 1(1): 175–201.DOI:https://doi.org/10.16995/fe.18062

Open Library of Humanities

Preference and Prevention: A New Paradox of Deontology

Richard Yetter Chappell, Philosophy, University of Miami, r.chappell@gmail.com

It's commonly thought that we can reasonably oppose serious wrongdoing. For example, deontologist bystanders may prefer that an agent allows the killing of five rather than wrongly killing one as a means to saving the five. But this preference turns out to conflict with caring sufficiently strongly, after the one is killed, that the remaining entirely gratuitous killings are successfully prevented. This surprising incompatibility suggests that, whatever view we accept for ourselves, we cannot want others to abide by deontology.

Preference and Prevention: A New Paradox of Deontology

Richard Yetter Chappell

I. INTRODUCTION

The consequentialism/deontology debate is arguably the most central, fundamental dispute in ethical theory. Although it has seemingly stalled in recent years, new progress may be made by approaching the problem from a new angle. Whereas most past discussion has focused directly on the deontic status of a focal action, less attention has been paid to the question of what morally-motivated bystanders should *hope* is done. It's natural to expect that deontologists could go either way on this question.[1] In this paper, I show that the more appealing of the two paths must be closed. We cannot robustly endorse deontic constraints, as this turns out to be surprisingly incompatible with sufficient concern to avoid entirely gratuitous killings. As a result, whether we end up accepting consequentialism or deontology for ourselves, when we have no personal interests at stake, we must all prefer that *others* act as consequentialism recommends.

Deontic constraints prohibit agents from committing certain evils, even to prevent a greater number of like evils. Deontology—understood as the theoretical (not merely pragmatic)[2] endorsement of deontic constraints—is widely accepted.[3] Despite its prominence, surprisingly few philosophical objections have been mustered against

[1] Personal experience confirms: deontologists I've asked tend to split about evenly.

[2] Note that two-level consequentialists like R. M. Hare, Moral Thinking: Its Levels, Method, and Point (New York: Oxford University Press, 1981) may endorse deontic constraints as part of a value-promoting *decision procedure* for fallible agents, without thinking that these constraints really yield decisive objective normative reasons at a more fundamental level. For such consequentialists, constraints may constitute part of instrumental rationality (for non-ideal agents): part of how we can best hope to achieve goals that themselves make no essential reference to such constraints. For deontologists, by contrast, the moral significance of constraints is non-instrumental and essential to ethics, such that they would guide infallible angels as well as fallible humans.

[3] "Consequentialism" secured support from less than 1/3 of philosophers in the 2020 PhilPapers Survey. "Deontology" was supported by about 1/3, and "virtue ethics" by slightly over 1/3. David Bourget and David Chalmers, "Philosophers on Philosophy: The 2020 PhilPapers Survey," *Philosopher's Imprint* 23, no. 1 (2023): 1–53. I would guess, but do not know for sure, that most of those who selected "virtue ethics" would also endorse deontic constraints, making "deontology" in my sense the majority view.

deontology as such.[4] Among the most commonly discussed is Scheffler's *paradox of deontology*: the idea that there is an "air of irrationality surrounding the claim that some acts are so objectionable that one ought not to perform them even if this means that more equally weighty acts of the very same kind or other comparably objectionable events will ensue."[5] For convenience, I'll refer to cases of this form—where an agent's refraining from a morally objectionable act will result in a greater number of comparably objectionable acts being performed—as "Scheffler cases." Many dismiss Scheffler's paradox on the grounds that it implicitly assumes that killing is first and foremost a *bad* to be minimized, whereas deontologists instead hold it to be a *wrong* to be avoided in each particular instance (even if one's avoidance of killing has the causal upshot that more killings ultimately occur). The argument thus strikes them as question-begging.[6]

There's room for reasonable disagreement about the extent to which Scheffler's challenge undermines deontology.[7] Still, it's always an advantage for an argument to grip a broader range of interlocutors. So I hope to extend the paradox of deontology to have broader dialectical "reach" than the original version. In particular, I'll show that we can grant the deontologist's response to Scheffler cases, and *still* raise a further challenge that needs to be addressed. Whatever your view of Scheffler's paradox, you should conclude that this new one is strictly stronger in force.

---

[4] Besides the paradox of deontology that I go on to discuss, three other families of objections that I'm aware of are: (1) arguments based on *ex ante Pareto* principles such as Caspar Hare, "Should We Wish Well to All?" *Philosophical Review* 125, no. 4 (2016): 451–72, https://doi.org/10.1215/00318108-3624764 and (arguably) Michael Huemer, "A Paradox for Weak Deontology," *Utilitas* 21, no. 4 (2009): 464–77, https://doi.org/10.1017/s0953820809990227; (2) the "paralysis" objections of Howard Nye, "Chaos and Constraints," in *Dimensions of Moral Agency*, ed. David Boersema (Newcastle upon Tyne, England: Cambridge Scholars, 2014), 14–29 and Andreas Mogensen and William MacAskill, "The Paralysis Argument," *Philosophers' Imprint* 21, no. 15 (2021): 1–17, http://hdl.handle.net/2027/spo.3521354.0021.015; and (3) objections relating to how to treat *uncertain* constraints violations (Frank Jackson and Michael Smith, "Absolutist Moral Theories and Uncertainty," *Journal of Philosophy* 103, no. 6 (2006): 267–83, https://doi.org/10.5840/jphil2006103614.)

[5] Samuel Scheffler, *The Rejection of Consequentialism*, rev. ed. (Oxford: Clarendon, 1994), 82. See also Robert Nozick, *Anarchy, State, and Utopia* (New York: Basic Books, 1974), 30: "If nonviolation of C is so important, shouldn't that be the goal? How can a concern for the nonviolation of C lead to the refusal to violate C even when this would prevent other more extensive violations of C?"

[6] See, e.g., Timothy Chappell, "Intuition, System, and the 'Paradox' of Deontology," in *Perfecting Virtue: New Essays on Kantian Ethics and Virtue Ethics*, eds. Lawrence Jost and Julian Wuerth (Cambridge: Cambridge University Press, 2011), 271–88.

[7] Scheffler's paradox is so widely discussed precisely because it highlights a structural feature that many reasonably regard as a *cost* of deontology. The mere fact that dedicated defenders of the view have already "priced in" this cost does not mean that there is no philosophical value to highlighting it for undecided parties to consider. Similar things may be said of the standard objections to consequentialism, which consequentialists find similarly unconvincing.

This paper thus aims to revive the heart of the paradox in a new and improved form. The key moves are (i) to focus on what *preferences* we ought to have over various possible worlds (or what we should *hope* to see happen), and (ii) to compare worlds in which infringing agents successfully minimize violations to worlds in which their attempts fail and *all* of the potential victims are killed.

There's a neglected general puzzle about what preferences and attitudes deontologist bystanders ought to have towards acts of optimific wrongdoing.[8] To introduce some convenient terminology, let's say that deontic constraints are "robust" if they give bystanders sufficient reason to prefer that the constraint be respected by the agent, and "quiet" if they speak *exclusively* to the agent (giving bystanders no reason to disprefer their optimific violation). As we'll see, there's a lot to be said for the robust conception of constraints on which bystanders may oppose such wrongdoing. But either verdict raises interesting further challenges.

This paper undertakes two tasks. Section II offers a critical overview of the choice between quiet and robust deontology, highlighting significant costs of quiet deontology that the robust view successfully avoids. By this point in the dialectic, the robust view looks to be an extremely attractive option for deontologists. Section III then undercuts this good news by presenting a new and seemingly decisive objection to the robust view of deontic constraints. The surprising result: either deontic normativity is "quiet", or deontology is false. *Preferring that others respect constraints* is no longer on the table.

II. ROBUST VERSUS QUIET CONSTRAINTS

II.A. Agent-Neutral Deontology and Robust Constraints

Deontic constraints differ from a "consequentialism of rights" which seeks to *minimize* rights-violations.[9] Deontologists hold that you should not violate another's rights, even if this violation would have the effect of overall minimizing rights violations. It is sometimes assumed that deontic constraints must thus be purely *agent-relative* (and time-relative too): that the agent must have a special concern to avoid *themselves* (now) violating rights, rather than sharing the moral concerns of an impartial spectator. But this assumption may be questioned: an alternative—and in many ways more attractive—option is available.

---

[8] Prior authors to raise this question include Andreas Mogensen, "Should We Prevent Optimific Wrongs?" *Utilitas* 28, no. 2 (2016): 215–26, https://doi.org/10.1017/s0953820815000345 and Reém Segev, "Should We Prevent Deontological Wrongdoing?" *Philosophical Studies* 173, no. 8 (2016): 2049–68, https://doi.org/10.1007/s11098-015-0596-9.

[9] Nozick, *Anarchy, State, and Utopia.*

As Tom Dougherty[10] and Kieran Setiya[11] make clear, a concern to minimize violations is *not* the only form that general (agent-neutral) concern could take. An agent-neutral deontologist could instead oppose rights-violations *in each instance* (no matter the agent),[12] and for *each* rights-violating act, prefer that the agent had instead acted permissibly—no matter the downstream consequences (at least within moderate limits).[13]

Some philosophers may be interested in the project of formulating an *exclusively agent-neutral* view of constraints. That is not my concern here. The question I'm interested in is whether deontic constraints have *robust* normative authority, granting bystanders sufficient reason to prefer that agents act permissibly (even when wrongdoing would have better consequences). Positing sufficient agent-neutral reasons is compatible with positing additional agent-relative reasons. Compare: impartial observers have sufficient reason to prefer that a child not suffer terribly (all else equal), even if the child's parents have even more reason to hope the same. Deontologist bystanders may similarly hope that a child not be murdered by her surgeon for the "greater good," even if they think that the surgeon has *even stronger* reasons to avoid such villainy.

II.B. The Costs of Quiet Deontology

On its face, the robust view seems to offer deontologists a very attractive conception of constraints. This section will survey reasons to prefer it over the "quiet," *exclusively* agent-relative alternative, on which deontic constraints make no difference to what bystanders should hope to see happen.

First, the robust view preempts any possible charges of egoism, self-indulgence, or "clean hands" fetishism, fitting well with an attractively "patient-centered"

---

[10] Tom Dougherty, "Agent-Neutral Deontology," *Philosophical Studies* 163, no. 2 (2013): 527–37, https://doi.org/10.1007/s11098-011-9829-8.

[11] Kieran Setiya, "Must Consequentialists Kill?," *Journal of Philosophy* 115, no. 2 (2018): 92–105, https://doi.org/10.5840/jphil201811525.

[12] I follow Dougherty's terminology in calling this "agent-neutral deontology." Setiya rejects the label of "deontology"—but his theory incorporates genuine deontic constraints, making it a form of "deontology" as I'm using the term here. Of course, the terms don't matter so long as we're clear on what we mean by them.

[13] This is a loose characterization. As Matthew Hammerton, "Is Agent-Neutral Deontology Possible?," *Journal of Ethics and Social Philosophy* 12, no. 3 (2017): 323, https://doi.org/10.26556/jesp.v12i3.267 notes, there may be weird symmetric cases where one killing will occur if and only if another does not, and neither is causally upstream of the other. In such a case, there are no agent-neutral grounds for a preference either way. Bystanders should presumably be indifferent between the symmetric possibilities. Deontologists may need to appeal to agent-relative reasons to establish that each agent is still bound by a constraint against killing. I don't take this possibility to undermine the intuitive grounds for taking constraints to be robust in other (causally asymmetric) cases, where it is logically coherent to do so. Thanks to an anonymous referee for prompting me to discuss this issue.

conception of moral concern. If, as many deontologists claim, the inviolable nature of human dignity calls for *respect* rather than *promotion*, for example, it would seem rather unprincipled to suddenly deny that this extends to third party attitudes. (Why shouldn't bystanders *also* respect one's inviolable dignity, by opposing one's being treated as a mere means to the greater good? Side-constraints should constrain attitudes as well as actions.) Consider how absurd it would seem for a committed deontologist bystander to mentally cheer on constraint-violating acts of utilitarian sacrifice for the greater good.[14] Such attitudes seem unfitting by the lights of deontological principles.

Second, robust deontology offers the strongest possible response to Samuel Scheffler's underlying concerns about constraints.[15] Scheffler argued that there's something deeply obscure about how merely agent-relative moral reasons could really have the normative authority to simultaneously override both the agent's own personal preferences and the verdicts of a more impartial moral perspective.[16] If deontic constraints themselves have decisive impartial significance, then this concern is perfectly pre-empted.

Third, it seems intuitively undeniable, as Setiya[17] writes, that "in general, when you should not cause harm to one in a way that will benefit others, you should not want others to do so either." Quiet deontology violates this appealing form of moral universalizability.

We should generally want others to act rightly—especially when the moral stakes are high.[18] At a minimum, if constraints truly matter, we surely *may* prefer that they be respected. If you take deontic constraints seriously, you should not want the agent in a Scheffler case to violate their constraint, even though the circumstances make it "optimal" by consequentialist lights. Non-consequentialists explicitly hold that there's

---

[14] If the deontologist is prohibited from such "cheering" as an expressive *act*, focus instead on the underlying mental state or attitude that *would* be expressed by such a cheer. That still seems like a very odd attitude for a deontologist to hold in this situation, even if it remains unexpressed.

[15] Samuel Scheffler, "Agent-Centred Restrictions, Rationality, and the Virtues," *Mind* 94, no. 375 (1985): 415–16, https://doi.org/10.1093/mind/xciv.375.409.

[16] To make the challenge vivid, imagine that you can save the lives of five loved ones only by killing a stranger. From what perspective is it preferable to let your loved ones die? Not the impartial perspective: the quiet view concedes that constraints have *merely* agent-relative significance. But also not the agent's own perspective: let's suppose that you care more about your family. In order for morality to have the *authority* to override your personal preferences, it must be able to claim the mantle of a larger, more legitimate and impartial perspective. Yet such ambitions are precisely what the quiet view of constraints disavows.

[17] Setiya, "Must Consequentialists Kill?," 97.

[18] Some possible exceptions, involving comparatively unimportant or "low stakes" obligations, will be discussed in II.C.

more that matters morally than just promoting value. Their overall moral preferences may surely reflect these broader concerns.

Fourth, only the robust view respects the datum that *moral perfection is not lamentable*: fully-informed agents are not morally required to do things that an ideal spectator (or God) would prefer that they not do. Consider: what could be the point of such a lamentable morality as quiet deontology posits? We'd be better off casting it into the flames and speaking no more of the accursed thing. At least, we should want all others to do so (and they should want the same of us): quiet deontology is deeply self-effacing in this way.

These are hefty costs, and every one of them is straightforwardly avoided by endorsing the robust view of constraints instead.

To fully appreciate these costs, it's worth reflecting on other "quiet" normative views that deny agent-neutral reasons entirely. These are views suggestive of deep normative conflict, like rational egoism, on which different agents could easily regard another's normative competence as extremely unfortunate. If one egoist is about to exploit another, the victim may well say of their oppressor: "Damn! I grant that he's doing exactly as he has most reason to do, but how I *wish* he would do otherwise!" Egoism renders one's normative competence—*successfully* doing as one truly ought to do—lamentable to others.

It's a striking feature of agent-relative views (like egoism) that *we generally shouldn't want others to believe or follow them*. That makes sense for egoism (which makes no claim to "morality"), but fits ill with commonsense moral thought. Quiet deontologists would seemingly have most reason to want other agents to comply with consequentialism (since the impermissibility of their acts would be of no interest to anyone but the agent, whereas the impartial value thereby secured *is* of broader interest).

That is, the shift to a quiet view of constraints renders deontology surprisingly self-effacing. And this isn't just the superficial point (sometimes erroneously presented as an "objection" to consequentialism)[19] that some people may better achieve moral goals by believing and aiming at something other than the moral truth. Quiet deontology is lamentable in the deeper sense that we shouldn't even want others to *successfully* follow it. We all have decisive moral reason to prefer that others instead comply with a moral code that is better supported by agent-neutral reasons.

[19] See Katarzyna de Lazari-Radek and Peter Singer, "Secrecy in Consequentialism: A Defence of Esoteric Morality," *Ratio* 23, no. 1 (2010): 34–58, https://doi.org/10.1111/j.1467–9329.2009.00449.x for critical discussion.

Given how publicly hostile to consequentialism many deontologists are, this result is big news that should change their attitudes and behavior. Even if they are personally constrained against lying for the greater good, they should at least be happy to see sincere consequentialists winning out in the marketplace of ideas. Depending on the details of their view, it may even be wrong for them to interfere by discouraging consequentialist thought (and action) in others. "Government house" utilitarianism was criticized for wanting only an elite few to know the truth. Quiet deontologists shouldn't even want elites to hear it, no matter how competent they may be. Robust deontology avoids this cost, and allows us to always hope that agents do the objectively right thing. Quiet deontology (unlike both consequentialism and robust deontology) bizarrely makes *rightness itself* lamentable.

Now, insofar as most deontologists don't seem to have any such pro-consequentialist desire—if they saw the agent in Trolley Footbridge decide *not* to kill the one as a means to saving five, they would surely react with more relief than disappointment—that's evidence that they're actually committed to the robust view.[20]

Plausibly: if killing one to save five is so morally objectionable that it ought not to be done, then it is sufficiently morally objectionable that bystanders should—or at least may—*prefer* that it not be done. It's hard to make sense of a serious morality denying this principle.

One could imagine a somewhat amoral (explicitly egoistic) form of deontology that denied this. If we ought generally to care about others' well-being, but then ought to have an overriding concern *not to get our own hands dirty*, for example, then this could coherently yield a *purely* agent-relative account of constraints. Bystanders would now plainly have most moral reason to hope that agents violate constraints—and indeed stop obsessing about their "clean hands" altogether—in order to do more good. This seems a deeply unappealing view, on which the reasons to respect constraints are no longer recognizably *moral* in nature at all. If deontologists are forced to abandon

[20] Suppose that the five victims on the tracks were deliberately tied there by five distinct villains, turning this into a Scheffler case. As an anonymous referee flagged, it would be possible to hold a robust view of constraints only in cases of saving from accidental death, while retreating to quiet deontology when the beneficiaries of a wrongful killing would themselves otherwise be wrongfully killed. I grant that such a bifurcated view is possible, but I wouldn't expect it to have widespread appeal. After all, it's hard to see what could motivate treating the different cases of wrongful optimific rescue so differently from each other, so long as both rescues are indeed still wrong despite maximizing impartial value. One would have to think that wrongful killing creates a special kind of *deontic disvalue* that (in contrast to ordinary disvalue) makes saving five from this fate—even via wrongful means—impartially *preferable*, without making it *permissible* to so act. There seems little motivation to adopt such a view of deontic (but not ordinary) disvalue.

the robust view of constraints, then they look to be back on the hook for clean-hands fetishism: the distinctive commitments of quiet deontology constitute a lamentable normative obsession that we should all prefer that others avoid.

What unifies these objections is that they reveal how the quiet view of constraints *forsakes deep normative authority*.[21] From clean-hands fetishism to Scheffler's paradox, the underlying philosophical insight is that there's a special kind of normative authority that's tied to impartiality—to acting in a way that is justifiable to, and preferable from, a neutral point of view that can be shared by all moral agents. The great promise of the robust view is that it alone allows deontology to meet this aspiration: to posit constraints that we all should want to see respected—constraints that truly *matter*.

II.C. Assessing Robust Deontology

Against all this, there seems comparatively little reason to favor the quiet view of constraints. In this section, I'll survey three prima facie objections to the robust view and argue that none of them are decisive.

First, *some* aspects of commonsense deontology may have purely agent-relative significance.[22] Special obligations may be like this, at least when they involve low stakes. It could be that a parent is obligated to buy their child a birthday present rather than donating the money to save a stranger's life, whereas an ideal observer should prefer that the agent wrongly prioritize saving the stranger. Perhaps (or perhaps not; see below). Promises are another candidate case: you might think that an ideal observer would want you to break a promise when doing so would overall be for the best.[23]

Rather than denying that authoritative obligations must be "robust," the above observations may instead form the basis of an intriguing argument against (suboptimal) special obligations and promissory duties. It's deeply puzzling to think

[21] I've found, in conversation, that readers vary markedly in the extent to which this claim resonates with them. I include it because I think it's an important truth that many readers are now in a position to successfully grasp. But I acknowledge that some others will deny it, and even express bafflement at the claim that there's any missing "authority" to purely agent-relative reasons. I don't currently have anything more to say, beyond the contents of this section, to those who find this claim baffling. Hopefully others will pick up the torch from here, and future research will further clarify the dispute.

[22] Thanks to two anonymous referees for prompting me to discuss this. Note that Setiya himself thinks "it is vital to distinguish general restrictions, which lend themselves to agent-neutral treatment, from special restrictions, which do not." (Setiya, "Must Consequentialists Kill?," 98)

[23] For this reason, I'm inclined to categorize promises as a form of special obligation. They are not the sort of "important deontic constraint" that inspires the robust view. A broken promise is not relevantly like killing someone.

that it could really be *wrong*, in any sense that matters, to do what an ideal God would most *want* us to do.[24]

Alternatively, such concerns may lead one to doubt that even special obligations are best understood as *purely* "quiet." Consider the contempt that many feel for Dickens' Mrs. Jellyby, with her neglected family and "telescopic philanthropy"—able to "see nothing nearer than Africa."[25] Many people do seem to prefer that others meet their special obligations, even if it results in less overall good. Presumably they expect that an ideal God would share their disdain for the Mrs. Jellybys of the world. Perhaps this is the only way to vindicate the normative authority of special obligations: insist that ideal bystanders should also want to see these obligations met. More would need to be said for a critic of robust normativity to decisively rule out this view.

But whatever one thinks of that, for present purposes it suffices to note the intuitive disanalogy between these special cases and the serious deontic violations that this paper is concerned with. There's nothing *so* strange-seeming about a deontologist bystander breathing a sigh of relief upon learning that another's promise was optimally broken, or that a special obligation was optimifically violated. Those wrongs lack the intuitive impartial significance of *killing an innocent person*. So the intuitive plausibility of robust constraints—restricted in scope to moral *atrocities* of this sort—is not threatened by the possibility that some more mundane wrongs may be "quiet," or have merely agent-relative significance.

Second, in cases involving ignorance, we clearly may prefer that an agent acts *subjectively* wrongly (unsuccessfully attempting murder, say), if this would have the effect of saving lives without actually harming anyone.[26] This suggests an important clarification of robust deontology. What may reasonably be taken to be impartially dispreferable is not mere *subjective* (or evidence-relative) wrongdoing, but specifically *objective* (fact-relative) wrongdoing. The greater significance of the latter is also observable first-personally. If an oracle tells you that you will soon do something either extremely subjectively wrong or extremely objectively wrong, a virtuous person will obviously hope for the former! Accordingly, this paper is concerned with objective rather than subjective wrongdoing.

Third, recall that my case against quiet deontology drew (in part) on intuitions about prospective or contemporaneous preferences—for example that it would seem

[24] Indeed, on an "ideal observer" conception, such desirability constitutes the very *meaning* of the objective ought.

[25] Charles Dickens, *Bleak House* (1853; The Literature Network), chap. 4, http://www.online-literature.com/dickens/bleakhouse/5/.

[26] Thanks to an anonymous referee for suggesting this objection.

contrary to the spirit of deontology to hope or prefer that an innocent person be pushed in front of a trolley for the greater good. It would seem cheap and superficial to endorse deontic constraints in a way that made no difference to what actions you should hope that others choose. But even deontologists may find themselves preferring the consequentially-best outcome *in retrospect*, which seems inconsistent with robust deontology. For example, we plausibly ought to prefer to read in the newspaper that one person was killed as a means to saving five others, rather than learning that the five were killed.[27] Outcomes that are worse by consequentialist lights seem like *bad news*. That someone was killed as a means seems far less *important*, objectively speaking, than that more lives were saved.

As a consequentialist, I happen to endorse that intuition. But others are free to reject it; we've seen strong theoretical reasons why deontologists ought, in principle, to reject it; and it's worth emphasizing that accepting it raises serious theoretical challenges for deontologists. (It's old news that deontologists are committed to granting deontic constraints priority over maximizing value. This at least makes sense if certain non-consequentialist properties are just *more important* than maximizing value. But it's another matter entirely to say that you endorse prioritizing *less important* properties over *more important* ones. Robust deontologists successfully avoid such self-confessed fetishism.) I don't here claim that the challenge is strictly unanswerable, just that it suggests significant costs that are unique to quiet deontology.

Let's pause to survey the options for how to respond to the intuitive clash between prospective/contemporaneous and retrospective preferability judgments. Each of quiet and robust deontology abandon one set of intuitions or the other;[28] but then, as we've seen, quiet deontology has additional theoretical costs on top of this. So I don't see much reason to favor quiet over robust deontology on this basis.[29]

---

[27] Thanks to an anonymous referee for prompting me to discuss this example.

[28] Specifically, robust deontology rejects the retrospective intuition; quiet deontology rejects the prospective/contemporaneous intuitions previously discussed.

[29] There is a third option worth mentioning. We could endorse the retrospective judgments in a different way than quiet deontology. Parity of reasoning may lead us to expect that whatever explains the *erroneous* appeal of deontic constraints to bystanders also explains why agents themselves erroneously find deontic constraints to be intuitive. (For example, perhaps the victims of agency seem more salient than "background" victims when considered prospectively. Or perhaps we don't fully trust the stipulation that a highly risky action really will work out for the best, until we have the aid of hindsight.) Whatever the most plausible story turns out to be, there's little reason to expect the debunking explanation to debunk *only bystanders'* reasons to want deontic constraints to be respected. So even if we trust retrospective judgments of the desirability of optimifically violating constraints, that may be taken to support consequentialism rather than quiet deontology.

Overall, then, there look to be strong theoretical reasons for deontologists to favor the robust view of constraints. It's the most principled form of deontology on offer. As such, reasons to reject this view are, to some extent, reasons to accept consequentialism (this being the most principled competitor view). Quiet deontology is left in an awkward middle-ground position, with comparatively little going for it. That's my read of the normative landscape, at least. Others may, of course, view things differently. But it is at least a defensible take on the dialectic which (if accepted) entails that the loss of robust deontology as an option would be devastating to the plausibility of deontology as such.

Whatever your views on the merits or appeal of quiet deontology, you may still be surprised to learn that the robust view can (as Section III argues) be so decisively refuted that it is no longer even *available* as an option. It's surprising, and philosophically valuable, to learn that (as bystanders with no personal interests at stake) we *must* prefer that others comply with consequentialism.

III. THE NEW PARADOX

III.A. Clarifying Preferability

The robust view of constraints invokes the notion of *preferability* from the perspective of an "ideal observer"—a fully informed, morally ideal, neutral bystander. I will use the symbol "$\succ$" to indicate such ideal preferability, prefaced by "$\Diamond$" to indicate a *permissible* (rather than required) preference.

Some readers may be unfamiliar with the concept of preferability used throughout this paper—it's not a concept often employed by deontologists—so allow me to offer three quick clarifications before launching into my central argument.[30]

First, I use "preferability" in the *fitting attitudes* sense: $W_2 \succ W_1$ if and only if it is uniquely fitting (for an ideal observer) to prefer W_2 over W_1, all things considered.[31] This claim is *not* conceptually reducible either to claims about value or to claims about permissibility and impermissibility. It's a conceptually wide-open question *what it is fitting to prefer*, and how these fittingness facts relate to questions of permissibility and value. Deontologists typically hold that we should care about things other than promoting value, whereas consequentialists may doubt whether we should care about deontic status at all. If one were to employ only subscripted, reducible concepts of $\text{preferability}_{\text{value}}$ and $\text{preferability}_{\text{deontic-status}}$, it would be difficult to even articulate this

[30] Thanks to several anonymous referees for prompting this.

[31] By "uniquely fitting", I mean that it is fitting to have this preference, and *unfitting* to have any incompatible preference, such as preferring W_1 over W_2, or indifference between the two. A prefatory "$\Diamond$" cancels the claim to uniqueness, instead indicating a merely permissible preference.

disagreement.[32] But we can further ask about preferability in the irreducible *fitting attitudes* sense, on which it is a conceptually open (and highly substantive) normative question what we should all-things-considered prefer, desire, or hope to see obtain. This is the notion of preferability employed throughout this paper. (Section IV.C. addresses the objection that deontologists need not be committed to making any claims about preferability, so understood.)

This paper thus assumes psychological realism. Some philosophers treat preferences as *reducible* to choice dispositions. I reject such behaviorism. Instead, I start from the assumption that (all-things-considered) *desires* and *preferences* are real and familiar psychological states that reflect an agent's overall *concerns*,[33] that can take as their objects even unchoosable states of affairs,[34] and that can be (or fail to be) warranted by their objects. This does not beg any questions, since it places no restrictions on what agents may be concerned *about*. Preferences may reflect deontological concerns just as well as they may reflect utilitarian ones.[35]

Second, I take preferability (like desirability) to come in degrees. It seems intuitive that some preferences are stronger than others, and justifiably so. For example, you may prefer to receive a lollipop than a papercut. You may also prefer being left alone over being tortured to death. I think the latter preference could justifiably be *much stronger*. If you cannot make sense of this claim, then my argument will have no traction on you.[36]

Third, this paper discusses preferability over *act-inclusive* states of affairs or possible worlds. Some readers may suspect that morally evaluating states of affairs or possible

---

[32] Arguably, these subscripted notions are not truly *preferability* rankings at all, but merely *rankings*—of value and deontic status, respectively. They don't appear to have anything essentially to do with preferences.

[33] For a skeptical take, see Ralph Wedgwood, "Must Rational Intentions Maximize Utility?," *Philosophical Explorations* 20, no. S2 (2017): 73–92, https://doi.org/10.1080/13869795.2017.1356352. Note that such skepticism is beyond the scope of this paper.

[34] For example, we should prefer the state of affairs in which *wild animals live happy lives, without any agents choosing this* over that in which *wild animals suffer horribly, without any agents choosing this*. But, by definition, neither of these states of affairs could be chosen.

[35] In contrast to some who seek to "consequentialize" deontology, I make no claims here about *explanatory priority*. I'm happy to allow that deontological preferences may reflect antecedent judgments about moral rightness or reasons for action. So, I take my assumptions here to be extremely weak, and congenial to any reasonable deontologist.

[36] A simple way to make sense of it: the strength of one's preference for one state of affairs over another may be given by the net difference in one's total desire satisfaction between those two states of affairs. Since being tortured to death frustrates much stronger desires than receiving a papercut does, whereas receiving a lollipop is only slightly more desirable than being left alone, we should expect that being left alone rather than tortured is the comparison that involves by far the greater difference in net desire satisfaction. James Fanciullo, "Preference as Desire," *Journal of Philosophy,* forthcoming defends a similar reduction of preference to desire.

worlds is an *inherently* consequentialist endeavor. But this suspicion is misguided, as I'll now explain.

There is a possible ambiguity in talk of "outcomes." Sometimes we may speak of *general outcomes* (in abstraction from how they are brought about). For example, *preventing five killings* is a general outcome that is clearly pro tanto desirable. But preferability over general outcomes has no necessary connection to permissibility in specific instances, since the central idea of deontology is precisely that generally good outcomes may not be pursued via impermissible means. For deontologists, evaluating general outcomes tells us little about what ought to be done.

In this paper, I am instead concerned with the more specific notion of an *act-inclusive* outcome, state of affairs, or possible world. This includes not just the general outcome, but also the specific action that brought it about. For example, preventing five killings *by means of killing a distinct innocent person* is an act-inclusive outcome that a deontologist may overall *disprefer* to the alternative of permissibly allowing the five killings. It's an important feature of deontology that act-inclusive worlds or outcomes may be evaluated *radically differently* from mere "general outcomes." The specific details of how the general outcome was brought about matter immensely.[37]

So, while deontologists will naturally reject any suggestion that the preferability of a general outcome entails the permissibility of bringing it about, I'm aware of no good reason for anyone (deontologist or otherwise) to deny that we should prefer the *specific, act-inclusive world* that results from acting as we ought, over any alternative specific world that diverges only as a result of our *choosing wrongly* from that same choice-point. (We'll revisit this question in Section IV.B.)

As such, this is a *theory-neutral* conception of preferability. It begs no questions. It allows for features of actions (including their deontic status) to take priority over consequent value in determining which specific outcome (from among those that the agent *can* realize) one should overall prefer to see realized. Indeed, as we saw in Section II.B., by upholding respect for deontic constraints as impartially preferable, robust deontology is able to secure many theoretical benefits and defang what would otherwise be serious threats to the normative authority of deontic constraints.

---

[37] This also explains why *preferring that agents not act wrongly* does not immediately entail that a third party ought to *interfere to prevent their wrong action*. It remains open to the deontologist to prefer that the agent freely choose to act rightly, without preferring that third parties interfere to ensure this. (I take no stand on this further question.) The rational preferability of a general outcome doesn't entail that it would be rational to choose that outcome, because the introduced action *modifies* the (act-inclusive) state of affairs in ways that deontologists may consider morally relevant. So the general link between rational preferability and rational choice is only guaranteed when considering specifically *act-inclusive* states of affairs.

Alas, as we'll now see, robust deontology cannot be sustained in conjunction with substantively undeniable judgments about what preferences moral decency requires of us.

III.B. Setup

Let us begin with some cases. In all of these cases, the background setup involves five other agents who are each about to murder a different innocent victim. Protagonist may be in a position to prevent these five murders, by means of herself killing a sixth individual. Against this background, compare the following four (act-inclusive) possible outcomes:

> **Five Killings**: Protagonist does nothing, so the five other murders proceed as expected.
> **One Killing to Prevent Five**: Protagonist kills one as a means, thereby preventing the five other murders.
> **Failed Prevention**: As above, Protagonist kills one as an intended means, but in this case fails to achieve her end of preventing the five other murders. So all six victims are killed.
> **Six Killings**: Instead of attempting to save the five, Protagonist simply murders her victim for the sheer hell of it, just like the other five murderers. So all six victims are killed.

Agents should find wrongdoing to be morally dispreferable to some permissible alternative. If they can bring about either W_1 or W_2, and it would be wrong for them to bring about W_1 (but not W_2), then they should prefer world W_2 over W_1. (This claim is further defended in Section IV.B.) Further, when the wrong act involves serious rights violations, then (by the robust view of constraints) neutral bystanders should—or at least *may*—likewise prefer W_2.

Note that, by explicitly allowing that a violation-minimizing violation makes sense to disprefer, these starting assumptions are noticeably weaker than the assumptions underlying the traditional paradox of deontology. They're positively *friendly* to robust deontology.

III.C. The Argument

Recall that I use "$>$" to denote preferability to an ideal observer, prefaced by "$\Diamond$" to indicate a *permissible* (rather than required) preference.[38] In addition, I use the "$>>$"

[38] The argument may be easier to follow if you initially ignore the "$\Diamond$" weakenings and instead imagine that every preference mentioned is *required*. But it's striking that the argument goes through even when robust constraints are weakened as indicated, since it makes them even harder to reject.

symbol to indicate *vast* preferability, understood as follows. Consider how strongly bystanders should prefer *one* less generic murder to occur in the world (such as the difference between Five Killings and Six Killings). I will use "vast" to indicate preferences that are *strictly stronger* than that.[39] Now, to formulate my argument:

(1) Protagonist acts wrongly in One Killing to Prevent Five, due to violating an important deontic constraint, and ought instead to bring about the world of Five Killings. (For reductio)

(2) If an agent can bring about just W_1 or W_2, and it would be wrong for them to bring about W_1 (but not W_2) due to violating an important deontic constraint, then $W_2 \lozenge\!\succ W_1$. (Weak robust constraints)

(3) Five Killings $\lozenge\!\succ$ One Killing to Prevent Five. (From 1, 2)

(4) One Killing to Prevent Five $\succ\!\succ$ Failed Prevention. (Premise)

(5) Failed Prevention $\succcurlyeq$ Six Killings. (Premise)

(6) Five Killings $\lozenge\!\succ\!\succ$ Six Killings. (3–5, transitivity)

(7) It is not the case that Five Killings $\lozenge\!\succ\!\succ$ Six Killings. (Definition of '$\succ\!\succ$')

Contradiction (6, 7).

Explanation: We assume (1) for reductio. (2) presents a weak (permitting rather than requiring) version of the robust view of constraints, which we've seen has a lot of prima facie appeal (and is arguably required for deontic constraints to have full-blown normative authority). Putting the two together yields (3), a claim closely related to the target of the traditional paradox of deontology. A flat-footed consequentialist may stop here, and ask how you could reasonably prefer a larger number of killings to a smaller

[39] This simple formulation assumes that there is a reasonably precise answer to the question of how strongly bystanders should prefer one less generic murder to occur in the world. If you reject this assumption, then we instead need to give "vast" a context-relative interpretation. To guarantee the truth of premise (7), we must consider how strongly our ideal observer *happens* to prefer Five Killings over Six Killings, and define "vast" (in this context) as preferences that are strictly stronger than *that*. (And interpret "$\lozenge$" as indicating preferences that can reasonably be held *concurrently* with this one. (7) then claims, trivially, that *holding fixed* the strength of their preference for Five Killings over Six Killings, they cannot have this very preference be stronger than it is.) So understood, premise (4) then affirms a comparative requirement, that bystanders ought to care *more strongly* about the *five* additional gratuitous killings in Failed Prevention (relative to One Killing to Prevent Five) than they do about the *one* additional gratuitous killing in Six Killings (relative to Five Killings). Since they have no reason to comparatively neglect any of the six victims (let alone five of them), and comparatively neglecting some victims for no reason is less than ideal, this seems just as much a datum as the context-independent interpretation of (4) discussed in the main text. So, for ease of exposition, I will bracket this complication. Thanks to Eden Lin for suggesting this challenge.

number, whatever other minor details may differ between the two outcomes (so long as those details make no instrumental difference to the total welfare). But let's grant the deontologist that constraints make sense, and that the traditional Schefflerian "paradox" provides no reason to doubt this. Even granting this, a new (and deeper) problem emerges once we bring our two remaining cases into the picture. The rest of the argument reveals the surprising truth that (3) is strictly unacceptable: no bystander can permissibly have this paradigmatically "deontological" preference.

I take (4) to be a moral datum: on any sane view, One Killing to Prevent Five must be *vastly* preferable to Failed Prevention. In the former scenario, Protagonist kills one person to prevent five other killings. The latter scenario contains this killing, and everything that's morally objectionable about it, *plus* five additional, completely gratuitous murders, as Protagonist now fails in her attempt to prevent them. However severely the deontologist may judge Protagonist's moral violation, they must surely agree that *once* the choice is made (and her victim killed) we have immensely strong moral reasons to hope that this act—however objectionable it might have been—at least succeeds in preventing the five other murders. For if the five other murders happen *in addition*, that is *so* very much worse. Specifically, this is a substantially greater moral difference (i.e., yields a world that is morally dispreferable to a greater extent) than would generally be obtained by just adding one gratuitous murder to a scenario.

It's worth emphasizing that adding five gratuitous murders has got to be a big deal on *any* moral view. I'm not smuggling in anything contentious in making this evaluation. I'm not, for example, supposing that we must prefer the smaller number of killings even when there are differences in the causal structure, or different individuals spared, which could provide *some* reason to prefer the other option. We're talking about a case of Pareto inferiority, where there is literally *no reason at all* to prefer the extra killings. So we should regard (4) as an unassailable moral datum, the rejection of which would entail severe moral disrespect to the five extra murder victims.

However, I'll now explain why, once deontologists accept our previous premises, they cannot also accommodate (4). We'll see that once they permit a bystander to *prefer that Protagonist comply* with the deontic constraint, they thereby permit that bystander to have an *indecently low* strength of preference for One Killing to Prevent Five over Failed Prevention. On the face of it, this seems an extraordinarily surprising result. But consider:

(5) Failed Prevention $\succcurlyeq$ Six Killings.

When the same six killings occur either way, it's an interesting question whether the killers' motivations matter. Perhaps they don't, in which case we may be indifferent

between the two outcomes in which these same six killings occur. Alternatively, if motivations do matter, it's surely better for Protagonist to be beneficently motivated (despite not actually achieving any good) than for her to gratuitously murder someone just for the sadistic glee of it. There's no way that being beneficently motivated could be intrinsically morally *worse* than being evilly motivated, so we may safely conclude that Failed Prevention is either equivalent to or preferable to Six Killings.[40]

If we try to include (4) alongside (3) and (5), we get the (permitted) preference chain:

(*) Five Killings ◊> One Killing to Prevent Five >> Failed Prevention ⩾ Six Killings.

And thus (by transitivity):

(6) Five Killings ◊>> Six Killings.

This is inconsistent with the stipulated meaning of '>>' (or "vast") to denote moral chasms that strictly *exceed* the magnitude of one additional typical murder. By transitivity, the magnitude of preferability between any two adjacent links of the chain must be strictly weaker than the preferability of the first item over the last. But the first and last items of the chain are Five Killings and Six, which differ by *exactly* one additional typical murder. Hence, as (7) makes explicit, they cannot differ in (permitted) preferability by *more* than this.

The challenge for deontologists is that there just isn't enough moral room between Five Killings and Six to accommodate the moral gulf that ought to lie between One Killing to Prevent Five and Failed Prevention. As a result, when they allow Five Killings to be morally preferred to One Killing to Prevent Five, they are unable to accommodate our moral datum (4), that Failed Prevention is *vastly* dispreferable to One Killing to Prevent Five.[41]

---

[40] Since there would not seem any respect in which Six Killings is preferable, it would not seem plausible to insist that the two are instead *on a par* in a way that would threaten my subsequent use of transitivity. Parity requires some grounds for ambivalence. But if one retains doubts on that point, my argument may be recast by skipping the Failed Prevention case altogether and instead directly appealing to the alternative moral datum (4*) One Killing to Prevent Five >> Six Killings. For again, if Protagonist is going to kill her victim either way, it would clearly be *vastly* preferable that she thereby save five other innocents than that she just kills the sixth victim for the hell of it, without even bothering to save the other five while she's at it.

[41] Setiya, "Must Consequentialists Kill?," 102 notes a related puzzle about the ethics of killing. Given constraints against killing (which he endorses), we must prefer two random killings over One Killing to Prevent Five. While this initially sounds odd, Setiya defends this verdict as follows: "The situation in which someone is going to be killed unless they are saved [by wrongly killing another innocent] is as bad as the situation in which they are going to be killed. Ethically speaking, the damage has been done… It makes

In effect, the deontologist is committed to holding that, once an agent kills one in an impermissible attempt to prevent five other killings, it matters scandalously little whether their attempt succeeds or fails.[42] That is, it doesn't matter sufficiently whether five additional—and entirely unmitigated—killings occur or are prevented. This seems incompatible with treating killing as extremely morally serious: the very commitment that motivates deontic constraints against killing in the first place. Deontic constraints are thus "paradoxical" in the strong sense of being self-undermining.

Ideally, deontologists should want to resist calls for utilitarian sacrifice by means of "upgrading" the importance of *not violating the one's rights* rather than by "downgrading" the importance of *saving the five*. My argument casts doubt on whether they can avoid doing the latter.

IV. DEFENDING THE ARGUMENT

In this section, we'll explore three ways that deontologists might try to resist my argument: by rejecting transitivity, preferring to act wrongly, or refusing to countenance preferability facts altogether.

IV.A. Rejecting Transitivity

Since my argument relies on the transitivity of preferability (across the specified cases), and some philosophers deny that preferability is universally transitive, one might hope to reject the argument on this basis.

But this is clutching at straws. Compare: some philosophers accept, as their preferred solution to the Liar paradox, some true contradictions. But it would be silly to think that you could get out of *any* objection just by accepting true contradictions

---

things worse, not better, that the button is pushed, so that the innocent stranger dies. That is why One Killing to Prevent Five is worse than Five Killings: it starts out the same and then declines. If we think through the temporal unfolding of events in One Killing to Prevent Five, we can explain why Five Killings, and thus Two Killings, should be preferred." (104–105) This defense of constraints contains the seeds of its own refutation. For the claim that "the damage has been done" when victims are *threatened* (in this way)—rather than when the threat is *realized*—implies, plainly enough, that little is gained by averting the threat and saving their lives (in this way). But this is unacceptable. Compare Christopher Howard, "Consequentialists Must Kill," *Ethics* 131, no. 4 (2021): 742, https://doi.org/10.1086/713952: "Call me old-fashioned, but I think that further damage is done when these people die."

[42] The same holds of putatively "consequentialist" views that mirror the verdicts yielded by deontic constraints, such as Matthew Oliver's view that radically discounts benefits obtained by using another person as a means. Oliver, "The Means and the Good," *Analysis* 81, no. 4 (2022): 665–74, https://doi.org/10.1093/analys/anab027. Clearly, to radically discount the value of the innocent victims' lives in this way is incompatible with our datum (4), that One Killing to Prevent Five >> Failed Prevention.

willy-nilly. True contradictions are theoretically costly, and need to be well-motivated on any occasion of acceptance.

So it goes, I suggest, for violations of the transitivity of preferability. Nobody holds that preferability is *never* transitive. On the contrary, transitivity is clearly the default assumption, and to reject it in any given case requires motivation. And as far as I can tell, there is no reason at all to expect a violation of transitivity in the particular cases this paper is concerned with.

IV.B. Preferring to Act Wrongly

One may be initially dubious of our assumed link between deontic status and fitting preference.[43] Why couldn't non-consequentialist agents share the consequentialist's preferences for higher-value outcomes, while insisting that they're nonetheless obligated to act against those preferences? Certainly, deontologists may have a *general* preference that fewer people die rather than more, all else equal; they aren't monsters. But problems arise when we consider their preferences between more specific, act-inclusive states of affairs, such as Five Killings vs. One Killing to Prevent Five.

IV.B.1. Wrongness and What's Worth Caring About

Could a deontological Protagonist prefer One Killing to Prevent Five over Five Killings, while still maintaining that it would be wrong for her to kill? Such a combination of attitudes seems of questionable coherence. For consider the other emotional states and attitudes that go along with all-things-considered preferences. In regarding One Killing to Prevent Five as preferable, it seems that Protagonist would also need to *hope* that she chooses to realize this state of affairs, and subsequently feel regret and disappointment if she does not.[44] This seems incompatible with regarding that choice as truly wrong (at least in any sense that matters, implying authoritative and decisive normative reasons to avoid so acting).

Our concept of *mattering*—what's *worth caring about*—seems intimately connected with fitting attitudes. So even if deontic constraints could be coherently combined with

[43] Recall that our first background assumption was that *agents* should prefer not to (themselves) act wrongly. We then extended this preference to bystanders via the robust view of constraints. In the current section, we're revisiting the initial assumption about agents.

[44] Recall, from III.A., that we are comparing specific act-inclusive states of affairs here. I'm not objecting to combining the deontic judgment in question with a *general* preference for fewer killings: obviously one could generally hope for fewer killings without hoping to secure this via immoral means. Rather, here we are talking about an agent who prefers the world of One Killing to Prevent Five, *in which they kill an innocent person as a means*, over that of Five Killings (in which they don't kill). The tension is in holding this *specific* act-inclusive preference in combination with the deontic judgment that the distinguishing act is wrong. Thanks to an anonymous referee for prompting me to emphasize this point.

utilitarian preferences, the upshot would seem to be that deontic constraints don't really matter. Sure, the deontologist may maintain that there is an "obligation" not to kill. But this would seem a merely verbal victory if it turned out that we shouldn't really *want* to fulfill such obligations, and that what's truly preferable is to kill one to save five. Put another way: if we're all agreed that maximizing happiness is what we should most want and care about, then any residual disagreements about "obligation" would seem no more threatening to the utilitarian than residual disagreements about what's "honorable" (when we all agree that we've no reason to care about "honor" as such).[45]

Section II.B. concluded that the quiet view of constraints was missing deep normative authority. The connection we now observe between *mattering* and fitting attitudes reinforces this conclusion. If deontic constraints are truly important, we cannot generally prefer that they be violated. Such wanton disregard indicates a matter of moral indifference.

IV.B.2. Act-Directed vs. State-Directed Preferences

I have so far assumed that the preferability of an act entails the preferability of the corresponding (act-involving) state of affairs or possible world. So, for example, if Protagonist ought to prefer not to kill in our scenario, then they ought to prefer the world of Five Killings over that of One Killing to Prevent Five. But this crucial assumption could be questioned. Howard Nye, David Plunkett, and John Ku[46] distinguish *act-directed* and *state-directed* motivations, and suggest that the fittingness conditions of the two may diverge:

> It might seem strange at first to think that we should hope that we will act as we should not act. But it is actually a familiar phenomenon that we should hope that we will have motives that it is unfitting to have (e.g. unwarranted anger towards one if that is the only way to prevent an evil demon from killing five).

But the analogy to reasonably wanting unfitting (but beneficial) motives breaks down because of the Motivations-Actions Principle introduced by these authors on p.12: a fittingness reason to *want* to ϕ is ipso facto a reason to *actually* ϕ. If you've reason to want and hope that you push the one, these are equally reasons to *bring it about* that you push the one—most straightforwardly, by *pushing* them.[47] By contrast, there is no

---

[45] If one adopts an agent-relative view, despite the costs noted in II.B., then one could hold that these obligations should "matter" *just* to the agent. They then have *some* non-verbal significance. But as noted in that earlier section, we ordinarily think that constraints against killing have wider import than this; a *purely* agent-relative account of their significance seems not to take them sufficiently seriously.

[46] Howard Nye, David Plunkett, and John Ku, "Non-Consequentialism Demystified," *Philosophers' Imprint* 15, no. 4 (2015): 17, n40, http://hdl.handle.net/2027/spo.3521354.0015.004.

[47] They could also be reasons to perform other actions that raise the probability that you'll perform the desired action: taking a "pushy" pill, perhaps, or reading compelling arguments for consequentialism.

such general link between fitting reasons to desire *some other desire* and the fittingness of that other desire. This is why there's nothing remotely "strange"-seeming about wishing for beneficial motives that it is unfitting to have.

At this point, it's worth reiterating our earlier observations about *what matters*. In general, when faced with a conflict between having fitting attitudes or having (unwarranted) attitudes that we ought to want and hope to possess, we should act to bring about the preferable attitudes (especially if lives are on the line). Preferability trumps warrant in practical importance.[48] So if it *were* ever the case that "we should hope that we will act as we should not act," it would follow that we should act so as to bring about the prohibited action (most straightforwardly, by *doing* it) after all—a contradiction that goes well beyond merely "seem[ing] strange at first."

It thus seems safe to conclude that fitting act-directed and state-directed motivations cannot diverge. The two kinds of *motivation* can diverge, but only when one of the attitudes is unwarranted and ought to be overridden. Consider that the most intuitively compelling examples of such divergence in Nye, Plunkett, and Ku[49] seem to involve a kind of weakness of will: "wanting to yell at someone in a fit of anger vs. wanting it to be true that one has yelled at him so he doesn't walk all over you," and "wanting to exercise now vs. wanting the world to be such that one exercises now." In both of these cases, whether one *should* perform the act in question seems entirely determined by the fittingness of the *state-directed* preference. The absence of the act-directed motivation may make acting more *difficult*, but doesn't change what's really warranted.

So, while it's perfectly *understandable* to want not to kill one as a means—in just the same way that it's perfectly understandable to want not to exercise—these narrowly act-directed motivations are apt to be overridden precisely because they tend not to be responsive to a sufficiently broad range of important considerations. An agent may

[48] This is generally observable as a matter of form, no matter one's particular ethical views. For example, suppose fear is warranted when faced with a wild bear. But suppose the bear is more likely to attack you if you are scared, so it would be preferable not to feel fear. You don't have to be a consequentialist to think that in such a case you have every reason to take a fear-blocking pill. Whether one's resulting emotional state is warranted or not is clearly far less decision-relevant than whether the resulting state is one that you have good reason to want to be in. To deny the verdict that you should take the fear-blocking pill, one would have to argue that you actually have most reason to want *not* to be in an unwarranted emotional state: that this should be your primary concern in the circumstances. This is a possible view, but it shows that the question of warrant is not *directly* decision-relevant; as a matter of form, it only becomes decision-relevant if it can be shown to inform one's verdicts about preferability. (This does not assume consequentialism: it's open to deontologists to hold that verdicts about preferability are explanatorily *downstream* of verdicts about right and wrong action.)

[49] Nye, Plunkett, and Ku, "Non-Consequentialism Demystified," 8.

need to steel their will in order to bring themselves to do what they have most overall reason to do.[50]

In sum: If Protagonist ought to prefer One Killing to Prevent Five over Five Killings, then she ought to kill the one when faced with just this choice. To maintain constraints against killing, deontology must furnish agents with decisive reason to prefer to abide by those constraints.

Crucially, to insist on congruence (between what we ought to *prefer* and what we ought to *bring about*) is not to presuppose consequentialism. For one way to understand the consequentialism-deontology distinction is precisely in terms of which of these two normative verdicts serves to *explain* the other.[51]

Consequentialists *begin* with a conception of what's antecedently desirable (i.e., the good), and from this derive their conclusions about what we ought to do. Deontologists can coherently reverse this order of explanation (while retaining harmony between right action and fitting preference) by taking at least some facts about what we should prefer to be explanatorily downstream of facts about our obligations. It would seem most natural for them to hold that Protagonist should prefer Five Killings over One Killing to Prevent Five precisely *because* the distinguishing act (killing one) would be *wrong*.[52] That's a recognizably non-consequentialist position, and I think it is the best option available to deontologists. The only problem is that the resulting view is vulnerable to the central argument of this paper.

IV.C. Rejecting Preferability

The most sweeping way to object to my argument would be to reject the very notion of *preferability*. A deontologist may claim that they are *only* concerned to elucidate our

---

[50] This suggests an interesting debunking explanation of the prima facie appeal of deontology. Consequentialism is unappealing in the same way that eating your vegetables is unappealing. It requires you to take into account less salient interests and reasons, at the cost of more salient ones. As biased, limited agents, we naturally find this to be unpleasant. But as we teach our kids, they really should eat their vegetables even so.

[51] Richard Yetter Chappell, "Fittingness: The Sole Normative Primitive," *Philosophical Quarterly* 62, no. 249 (2012): sec. VI, https://doi.org/10.1111/j.1467–9213.2012.00075.x. This account is contestable, and my argument here does not require that it turn out to be the correct account of the consequentialism-deontology distinction. It suffices that *some* deontological views maintain congruence between what we ought to do and what we ultimately ought to prefer, by taking the latter to be downstream of the former.

[52] I take this possibility to undermine Howard's argument ("Consequentialists Must Kill," 749) for incongruence on the basis that "these reasons to act have their source in the value of particular people, and so aren't given by value-making features of outcomes, [hence] they're reasons to act which don't derive from, or correspond to, reasons for preferring some outcomes to others." They may correspond to such reasons without deriving from them, as our reasons for desire may instead derive from these reasons for action.

duties, and so outright refuse to entertain any further normative claims about fitting attitudes. They may thus seek to reject the premises of my argument, not on the grounds that some alternative preferences are better warranted, but on the grounds that they are not committed to *any* particular claims about what we should prefer.

Such a move strikes me as unsatisfactory for several reasons. First, we may note that even a view that makes no *explicit* claims about preferences or preferability may nonetheless find itself *implicitly* committed to further claims involving these (or similar) concepts.[53] For example, in claiming that something *matters*, we seem to be suggesting that one *ought to care* about it, and it is difficult to make sense of how one could genuinely care about something without having any associated wants, hopes, desires, preferences, or the like. So, as previously noted, if deontic constraints really matter, then it seems we should want them to be respected.

But even if it is possible to remain strictly neutral on all questions of preferability or the ranking of states of affairs, we may nonetheless insist that such an incomplete moral perspective must, in principle, be coherently *completable*. That is, if certain verdicts about the deontic statuses of actions cannot be coherently combined with *any* plausible further claims about the relative preferability of various possible worlds, then those initial verdicts cannot all be true. But this incompatibility is precisely what my argument seeks to establish. So the deontologist would seem to be in trouble even if they themselves make no positive claims about preferability, for the very fact that they *cannot* combine their view with other true claims suffices for rejecting their view.

After all, it would not be plausible to deny that there *can* be truths about preferability. We clearly should prefer that innocent people not suffer, all else equal. Compare a possible future in which a child gets struck by lightning with an alternative in which they don't. Nobody has acted wrongly in either case. But we clearly ought to hope and prefer that the child *not* be struck by lightning (all else equal). This is a datum of common sense, and any theorist who denies it is in the grip of a theory. There's *undeniably* more to morality than just the deontic assessment of actions. As moral agents, we should care about more than just right- and wrong-doing. A complete moral theory must have something to say about what broader preferences are fitting, virtuous, or morally called-for, and some of these preferences will concern states of affairs, not just actions, that affect people's well-being. No sane and decent view can deny this.

---

[53] Richard Yetter Chappell, "Fittingness Objections to Consequentialism," in *Consequentialism: New Directions, New Problems*?, ed. Christian Seidel (Oxford University Press, 2019), https://doi.org/10.1093/oso/9780190270117.003.0005 similarly argues against complacent consequentialists that they need to take fittingness objections more seriously.

But perhaps a more limited rejection of preferability could be defended. Inspired by Philippa Foot's [54] claim that a phrase like "the *best* state of affairs" becomes meaningless in a context where beneficence is in conflict with justice or other virtues, one might claim more broadly that *there is no fact of the matter* about all-things-considered preferability in cases of deep moral conflict (e.g., between beneficence and deontic constraints). A proponent of this view will allow that they have reasons of beneficence to prefer One Killing to Prevent Five, and reasons of respect for the inviolable value of the one's humanity to prefer Five Killings, and then simply insist that there is no way for them to balance these reasons and form an overall preference or verdict on the case. They are not just torn (as even utilitarians should be to some extent),[55] but *irreparably* torn.

I grant that this is a coherent view, and one that my argument cannot rule out. But it comes at great cost. For, given the connection between all-things-considered preferability and rational choice, it would follow that there is similarly no fact of the matter regarding what we ought, all things considered, to *do* in these cases. Rather than affirming deontic constraints, this view transforms them into (indeterminate) moral dilemmas. One might say that we "deontologically ought" to respect the constraint, but it would be equally true to say we "consequentially ought" to violate it, and the view under consideration rules out the claim that we definitively ought to care more about the constraint than about the consequent benefits of violating it. All we can say is that we ought to feel (irreparably) torn, which leaves us entirely lacking in practical normative guidance.

V. CONCLUSION

Deontologists hold that killing is so morally serious that it should not be done, even to prevent more killings. One would expect bystanders to be able to endorse such an important constraint, and so prefer Five Killings over One Killing to Prevent Five. Surprisingly, this preference turns out to be incompatible with caring sufficiently about whether or not the prevention attempt turns out to be successful. That is, robust constraints against killing turn out to permit callous disregard for whether some potential victims are actually killed. Given that we ought to view such killings as extremely morally serious, we must reject deontic constraints (or at least prefer that others optimifically violate them).

---

[54] Philippa Foot, "Utilitarianism and the Virtues," *Mind* 94, no. 374 (1985): 196–209, https://doi.org/10.1093/mind/XCIV.374.196.

[55] Cf. Richard Yetter Chappell, "Value Receptacles," *Noûs* 49, no. 2 (2015): 322–32, https://doi.org/10.1111/nous.12023.

We saw that one might escape this argument by proposing that deontic constraints and impartial value give rise to incomparable reasons. But then there is no fact of the matter regarding how we should adjudicate the relevant trade-offs. Rather than embracing constraints, the resulting moral view would be guilty of normative abandonment: leaving us without practical guidance.

The upshot of this paper is a deep paradox for ethical theory, as the following four features turn out to be mutually inconsistent:

1. Deontic constraints,
2. Robust normative authority,
3. Normative guidance, and
4. Adequate respect and concern for those who can be rescued at no *further* cost.[56]

This is an extremely surprising result. Setiya,[57] for example, presents an attractive-looking view with the first three features, unaware that this commits him to violating the fourth. Even deontologists who are less drawn to an agent-neutral conception of constraints may be surprised to learn that (at cost of permitting moral disrespect) they cannot even *permit* bystanders to prefer that Protagonist rightly refrain from committing a violation-minimizing violation. The robust authority of constraints is thus lost. Whether we end up endorsing consequentialism or quiet deontology for ourselves, we must all prefer that *others* consign deontology to the flames.

[56] That is, *after* one has already been, perhaps illicitly, killed as a means to initiate the rescue attempt.
[57] Setiya, "Must Consequentialists Kill?"

Acknowledgments

Thanks to Matthew Adelstein, Sarah Buss, Krister Bykvist, Tim Campbell, Clinton Castro, David Chalmers, Julia Driver, James Goodrich, Peter Graham, Johan Gustafsson, Matthew Hammerton, Eden Lin, Will Lugar, Fiona Macpherson, Angra Mainyu, Douglas Portmore, Theron Pummer, Peter Railton, Connie Rosati, Kieran Setiya, Peter Singer, Keshav Singh, Dean Spears, Elliott Thornley, Helen Yetter-Chappell, several anonymous referees, students at the University of Miami, and audiences at Stockholm University, the University of Toronto, UT-Austin, NTU Singapore, and Oxford University. I also thank blog commenters at PEA Soup, on Facebook, at philosophyetc.net and at goodthoughts.blog, for helpful discussion and comments.

Competing Interests

The author has no competing interests to declare.