5

Moral Indulgences

When Offsetting is Wrong

Rebecca Chan and Dustin Crummett

0. Introduction

Suppose that it would be wrong for Scott to eat meat, absent a successful defeater. (Roughly, a successful defeater is a factor that nullifies or overrides the considerations that normally make the act wrong.[1]) Scott realizes this. But Scott also realizes that there are many very effective, underfunded, animal welfare charities. For instance, Animal Charity Evaluators (2017) estimates that one of their top-rated charities, the Humane League, may do the equivalent of saving as many as 180,000 animals from being factory farmed for every thousand dollars it receives through advocacy work. Scott realizes that even if this estimate is off by a few orders of magnitude,[2] he could easily do more good by making fairly small donations to such charities than he does harm by eating meat. (This is especially true if he limits himself to eating beef: since cows are very large and produce a lot of meat, eating beef contributes to far fewer animal deaths than, say, eating chicken or shrimp.) And he prefers making such donations to giving up meat altogether. So he does that.[3]

---

[1] This is McMahan's (2001, 236) gloss on moral defeaters. See also footnote 7.

[2] A critical discussion of Animal Charity Evaluators' methodology can be found at https://medium.com/@harrisonnathan/the-problems-with-animal-charity-evaluators-in-brief-cd56b8cb5908.

[3] Scott is named after Scott Alexander, author of the blog Slate Star Codex, who (inspired by another blogger, Katja Grace) once advocated a course of action similar to the one Scott takes. His post is what inspired us to consider the permissibility of the course of action described (2015a). After writing this paper, we learned that, in a separate post, Alexander similarly puzzled over the moral distinction between pollution and murder offsetting, which we discuss in the next paragraph (2015b). There, Alexander considers that universalizability might be an *objection* to offsetting meat-eating, but does not defend our claim that it might be a solution to explaining the moral difference between these different types of offsetting. He also addresses the question of the morality of offsetting in Alexander (2017).

Presumably, there is a point at which such donations aren't otherwise obligatory. (If Scott was obligated to donate part of his money anyway, suppose he already donated that part, and we're talking about additional donations; we assume that Scott isn't obligated to give *all* of his disposable income to charity.) Of course, Scott could donate the money *without* eating meat, and that would be better than what he does. But Scott could also permissibly refrain from either donating or eating meat, and that would be worse. Since Scott does better than he was obligated to do, he figures that he does nothing wrong—he figures, that is, that his donation defeats what would otherwise be the wrongness of his eating meat. But does it?

Notice that while there may be uncertainty with respect to the permissibility of Scott's actions, there is little uncertainty with respect to some other, structurally similar actions. Consider Billy Kincaid, who Sherlock Holmes discusses in episode two of season three of the BBC's *Sherlock*. When Watson asks Holmes to be his best man, Holmes misunderstands and thinks Watson is asking who the best man he knows is. Holmes responds:

> Billy Kincaid, the Camden garrotter. Best man I ever knew. Vast contributions to charity, never disclosed. Personally managed to save three hospitals from closure and ran the best and safest children's homes in North England. Yes, now and then there would be some garrottings, but stacking up the lives saved against the garrottings, on the balance . . .

For obvious reasons, Sherlock's response is played for laughs. Suppose that Billy performed his good works with the intention of offsetting his garrottings. Billy's reasoning might then be pretty similar to Scott's, but Billy's actions seem clearly impermissible. One shouldn't murder another person, even if one then saves two lives by donating a few thousand dollars to an anti-malaria charity. Even acts of heroism don't make up for a "garrotting now and then."

While Billy Kincaid acts impermissibly, there are structurally similar actions that appear permissible. Consider Pauline, who as part of an otherwise modest lifestyle, likes to take a lot of nice vacations to foreign countries. This travel generates a lot of pollution. In fact, it generates far more than her fair share—an amount of pollution that it's wrong to generate, absent a defeater. However, Pauline is conscientious about her pollution, and donates money to pollution fighting efforts with the intention of offsetting the pollution she creates. Her donation is otherwise supererogatory, has a positive effect that outweighs the badness of her pollution, and accomplishes a good outcome that could have also been accomplished without her

traveling. Pauline's actions have the same structure as Billy's, but unlike Billy's, her actions seem permissible. Indeed, purchasing pollution offsets is a common practice amongst factories, and we generally regard this practice as a morally good thing. We might have reservations about the particular details of some pollution offset system—about, say, whether the offset mechanisms involved are really effective, whether it distributes the burdens of pollution fairly, etc.—but few people have a problem with such a system in principle.[4]

Finally, consider *God*. Suppose that, for any world God could make, there's another, arbitrarily better world which God could make. (Perhaps God could always make a world much better by adding many additional good things.) And suppose that there's at least one world—call it *W1*— which God can permissibly make (cf. Kierland and Swenson 2013). Suppose the world God does make is like *W1*, except for two features. First, it contains some gratuitous evil, i.e., evil that exists either for no particular reason or for some morally insufficient reason (maybe God allows some terrible atrocity, knowing it will produce some good literature). Second, God adds a bunch of additional good things in order to ensure that the world God does make is much better than *W1*. Does God do anything wrong?

This chapter is about the above actions and others that are structurally similar to them. Specifically, it explores the permissibility of courses of action like these, ultimately with the aim of shedding light on the well-known question of what God should do if worlds just get better and better. First, it introduces the concept of *moral indulgence* to characterize sets of actions that bear the structure of those above (Section 1). Roughly speaking, an agent is morally indulgent when they do something which, absent a defeater, is wrong, and, in order to offset this, do something which is supererogatory and which is more good than the bad action was bad.[5] Second, this chapter proposes an explanation of when and why being morally indulgent is permissible (Section 2). As our above cases suggest, some morally indulgent sets of actions are permissible while others are impermissible. We seek to find an explanation for why this is. Our proposed explanation appeals to universalizability of the sort found in certain forms of Kantianism, contractualism, and rule consequentialism. Finally, we explore

---

[4] There are exceptions who take a significantly more negative view than we do (see, e.g., Sandel 2012), but they're in the minority.

[5] A similar concept is independently developed in a currently unpublished paper, "Moral Offsetting," by Tyler John and Amanda Askell.

the implications of our account for the question of what God should do, and why God should do it, if there are no unsurpassable possible worlds (Section 3). This discussion of divine morally indulgent actions contains an interesting upshot: while the Kantian, contractualist, and rule consequentialist principles underlying universalizability converge in non-divine cases of moral indulgence, they come apart in the divine case (and perhaps other structurally similar cases).[6] Thus, our discussion of divine indulgence, in addition to its interest for philosophers of religion, may provide a reason to prefer some of these theories over others.

1. Moral Indulgence

Scott, Billy, and Pauline are *morally indulgent*. An agent is morally indulgent if:

- (a) an act A is wrong, absent a defeater for its wrongness;
- (b) there is some other act B, such that:
  - (i) B is supererogatory (unless it is obligatory only because one performs A),
  - (ii) the goodness of B-ing outweighs the badness of A-ing, and
  - (iii) the goodness of B-ing can be realized independently of A-ing;
- (c) a person performs B with the intention of offsetting the badness of A.

For instance, eating meat is, we supposed, wrong, absent a defeater; donating to animal charities is supererogatory, outweighs the badness of eating meat, and can be done independently of eating meat; and Scott donates in order to offset the badness of his eating meat. So another way to put the

---

[6] We can imagine extremely unrealistic cases where a non-divine agent is placed in a situation similar to the one in which we imagine God being. For instance, an extremely powerful genie might be able to choose between options which are relevantly similar to creating $W1$, $W2$, or $W3$, or might give a human being the opportunity to do so. But there are at least two reasons for focusing specifically on the divine case. The first is that the question of what worlds God ought to create if there is no best world is of independent interest to philosophers of religion, in light of the fact that God may actually be in this situation. (Similarly, theological fatalism could be reframed as oracle fatalism, but theological fatalism is still of special interest since it's the subject of a live debate in philosophy of religion.) The second is that many people think that in moral theorizing, we ought to be more confident about, and give more evidential weight to, judgments about realistic cases as opposed to extremely unrealistic ones. Assuming that there's at least some real possibility that God may exist and be in the situation we describe, this may be a reason for ethicists qua ethicists to pay this case special attention, even independently of its interest for philosophers of religion.

question we're interested in is whether, when, and why actions which meet condition (b) can defeat what would otherwise be the wrongness of actions which meet condition (a) when performed with an intention that satisfies (c).

A few features of this definition deserve comment. First, saying that "*A* is wrong, absent a defeater" means what it sounds like. As mentioned, defeaters are factors that nullify or override the considerations that normally make the action wrong; without them, the normally wrong act remains wrong.[7] Regarding (b)(i), the clause "unless it is obligatory only because one performs *A*" is needed because we might think there are cases where, *given* that we perform *A*, we are obligated to perform *B* in order to defeat the wrongness of *A*. The point is just that, if we hadn't performed *A* or *B*, we wouldn't have done anything wrong. Regarding (b)(ii), the "goodness of *B*-ing" and the "badness of *A*-ing" are intentionally left broad, and can mean whatever plausible thing we want them to mean. The most obvious way to interpret the terms is as referring to the value of the states of affairs which result from the actions, but if we wanted, we could instead interpret them in terms of, say, the virtuousness or viciousness of the actions. Additionally, we talk about how *B*-ing *outweighs* *A*-ing because we have in mind cases of harm outweighing rather than harm blocking.[8] Finally, note that (b)(iii) is essential to differentiating moral indulgences from other cases which have received more attention from ethicists. In general, there's nothing odd about the idea that performing a supererogatory action might defeat what would otherwise be the wrongness of a different action: for

---

[7] Nullifying defeaters are *undercutting*: what would normally count as a reason for action does not count as a reason at all, given the presence of a nullifying defeater. For instance, that an action would promote pleasure is generally a reason to perform that action. But if an agent will derive pleasure from torturing someone, perhaps promoting this sadistic pleasure does not even count as a reason in favor of carrying out the torture—the reason is nullified. Overriding defeaters are *rebutting*: reasons for action still carry normative weight, but are outweighed by other reasons. For example, it is normally wrong to cause a child to experience pain by poking them with a needle, but if one is doing so to provide the child with a vaccination that will prevent even more severe suffering, causing the pain is permissible because it is overridden by the associated good. Here, there is still a factor that counts against vaccinating the child, but other, countervailing considerations are more important. This distinction will not be too important in what follows: we are concerned only with whether the wrongness is defeated (so that performing an action meeting condition (a) is permissible), not with the nature of this defeat. (It does seem clear that Pauline has *a* reason, even given her making the donation, not to do the travel—namely, that she will cause a lot of pollution. It may be less clear whether she has *deontic moral reason*—the sort of reason that creates moral *obligations*—in light of having made the donation. For an explanation of the difference between deontic moral reasons and other reasons, see Wallace 2013.)

[8] In other words, if *B*-ing defeats the wrongness of *A*-ing by preventing *A*-ing from having any bad effect at all, we don't count that as a case of moral indulgence.

instance, suppose I fail to make my promised lunch date with you because I'm busy supererogatorily saving someone from a burning building. But here, breaking the promise is necessary for saving someone, so that this is just an ordinary case of moral aims conflicting. However, in moral indulgence cases, there is no *conflict*; the agent *could* realize both moral aims. They just don't. It's this feature that makes them deserving of special attention.

Notice that the above conditions for moral indulgence were given as *sufficient*, rather than necessary, conditions. That's because we want to allow an agent to be morally indulgent by performing only one act which possesses, in the right way, both right- and wrong-making features.[9] Specifically, an agent can also be morally indulgent if:

(a*)   in virtue of possessing property p, an act A is wrong, absent a defeater for its wrongness;

(b*) A also possesses a property p^* such that:

(i*)   performing an act with p^* is supererogatory (unless it is obligatory only because one performs an act with property p),

(ii*) p^*'s goodness outweighs p's badness, and

(iii*)   the goodness of performing an act with property p^* can be realized without performing an act with property p;

(c*)   The agent performs A with the intention of having their performance of an act with p^* offset their performance of an act with p.

Most ordinary cases of moral indulgence involve performing two separate actions. However, the single act version will be important in Section 3.2. Perhaps God's actualizing a world consists in a single act which fulfills condition (a*) (in virtue of actualizing gratuitous evil) as well as condition (b*) (in virtue of containing lots of additional goodness). If so, we want to treat God's act as structurally analogous to what Scott does, even though it involves only one act, rather than two.

There is one more thing to note before moving on. Scott harms animals by eating meat, and then helps animals by donating money. It seems natural to say that there's some sense in which Scott's bad action and good action affect the same cause in a way that they wouldn't if, say, he ate meat and later donated to an anti-malaria charity. Call this "cause-matching."

---

[9] Also, we want to remain neutral with respect to how to individuate actions. So having one- and two-action conditions for being morally indulgent allows us to characterize action(s) as morally indulgent without having to dive into the metaphysics of act individuation.

Cause-matching often seems natural and appropriate, but we face at least two questions about its moral significance. First, why match causes, instead of just donating to whatever cause will produce the most good? Second, is there some metaphysically privileged way of individuating causes? (If we think of the problem to which Scott contributes as the deaths of the particular animals he eats, then his good act *doesn't* help the same cause. If we think of it as the fact that bad things happen, then the anti-malaria donation would cause-match, too.) If there is no metaphysically privileged way of individuating causes, then, since whether someone cause-matches is relative to which of infinitely many arbitrary ways of carving up causes we consider, any moral significance which attaches to cause-matching may also seem unacceptably arbitrary or relativistic.

We don't treat cause-matching as *essential* to moral indulgence, or to its permissibility. However, we do share the sense that it sometimes seems natural and appropriate. Fortunately, our account will be able to explain both why this is and how, for relevant purposes, to individuate causes. We'll return to this point in Section 2.4.

2. The Moral Status of Moral Indulgence

2.1 Two desiderata and what can't meet them

As we noted in the Introduction, it seems clear that morally indulgent agents sometimes act wrongly, and sometimes don't: Billy's murders are wrong, while Pauline's polluting isn't. Meanwhile, the moral status of Scott's actions at least doesn't seem *very* clear to us; we expect that individual intuitions about it will differ, and that they will usually be much less strong than in those concerning the cases of Billy and Pauline.

In providing our own account of the moral status of moral indulgence, we'll attempt to meet at least two desiderata. First, the explanation should do justice to our intuitions about the paradigmatic cases. It should yield the result that Billy acts wrongly while Pauline acts permissibly. It also should yield the result that the moral status of Scott's action is less clear than Billy or Pauline's. Second, the explanation should not be ad hoc. Rather, whatever it appeals to as constituting the morally important differences between the cases should be things which we have independent motivation for thinking are morally important.

We don't claim that the explanation we provide is the *only* account which meets these two conditions. Probably, many different accounts can do this,

so that selecting between them will ultimately need to be done on other grounds (for instance, various reasons for accepting one moral theory or another which have nothing to do with moral indulgence). But notice that meeting these conditions isn't *trivial*. Some moral theories, at least in their most common forms, won't be able to do so.

Famously, maximizing act consequentialism is unable to accommodate supererogation. This will make moral indulgence impossible, since no actions will be able to meet condition (b). Of course, this also seems intuitively wrong: for instance, of note here is that it implies that Pauline acts wrongly. One way for an act consequentialist to accommodate supererogation is to adopt a satisficing view, on which I'm morally required only to produce *enough* (not necessarily the *most possible*) good. Actions which produce more than enough good, when it was possible for me to instead produce just enough good, would then be supererogatory. However, since normal forms of satisficing act consequentialism *only* care about staying above the line, they'll allow too many cases of moral indulgence: Billy's actions (and others relevantly like them)[10] are permissible, just like Pauline's are. (In fact, that satisficing consequentialism might license gratuitous murder or assault, provided one stays above the satisficing line, is a classic objection against it; see Kagan 1989.)

But in other cases, we will be unable to accommodate the intuitive judgments about these cases for reasons which have nothing essentially to do with supererogation. Consider a view which holds that, when an action has both good and bad effects, it is a necessary condition for its being

---

[10]   We can imagine two types of satisficing view. One says that the sum total of all our actions must fall above a line. Another says that every one of our individual actions must fall above a line. Philip Swenson pointed out to us that the latter view would not allow actions like Billy's, since murdering would fall below the line. However, this view is still too permissive in morally equivalent single-action indulgence cases. If Billy is permitted to donate only a certain amount, $X, could make a donation of $X+Y which would save several more lives than a donation of $X, but instead makes a donation which saves these additional lives but also somehow kills his rival, this is clearly impermissible, but will still fall above the satisficing line (which was met by a donation of just $X). Further, this view will probably be too strict in other cases. Assuming that, when I can avoid it, it forbids my performing actions which are not merely suboptimal but are actually fairly bad, it will rule out Paula's going on her trip, which is fairly bad taken on its own, despite her offsetting donation. (If it *doesn't* rule this out, then it would allow me to do nothing but needlessly somewhat bad actions throughout my entire life, which doesn't seem correct.) Finally, it's a little hard to see why an act consequentialist would be drawn to the individual-action form of satisficing: Why care about the effects of individual actions, if the total effects of two series of actions are the same? So we think the individual-action form of satisficing act consequentialism is less inherently plausible while also facing counterexamples in moral indulgence cases similar to those faced by other forms of act consequentialism.

permissible that it meet the criteria laid out by the doctrine of double effect. Generally, these criteria are taken to be something like the following from McConnell (2003, 880):

1. The act itself must be morally good or at least indifferent.

2. The agent may not positively will the bad effect but may permit it. If the agent could attain the good effect without the bad effect, they should do so. The bad effect is sometimes said to be indirectly voluntary.

3. The good effect must flow from the action at least as immediately (in the order of causality, though not necessarily in the order of time) as the bad effect. In other words the good effect must be produced directly by the action, not by the bad effect. Otherwise the agent would be using a bad means to a good end, which is never allowed.

4. The good effect must be sufficiently desirable to compensate for the allowing of the bad effect.

Of course, there are some slight variations to the doctrine of double effect, but they won't make a difference here. Suppose that the good (enjoyment, neat Facebook pictures, etc.) achieved by Pauline's travel is not itself sufficient to justify causing the pollution (which is clearly possible, provided travel causes enough pollution). Her action then fails condition 4, and so is wrong regardless of her later donation.[11] Since this is a counterintuitive result, we seem to have a counterexample to the principle of double effect which has no clear counterpart outside of moral indulgence cases. Perhaps

---

[11] We could imagine a more complicated case in which Pauline ensures that the action of traveling and the action of making the donation are *the same act*. In that case, the act would meet condition 4, since, we supposed, the donation was much more desirable than the pollution was undesirable. However, for the principle of double effect to be very useful, it must be interpreted as requiring that the agent, if possible, achieve the good without the evil, *even if* that means sacrificing other goods which are not proportional to the evil. (For instance, suppose I can bomb the dictator now, ending the war but killing hundreds of civilians who have gathered to hear his speech, or I can the bomb dictator in an hour, killing only him but causing me to miss the season premiere of *The Good Place*. I must wait an hour, achieving the proportional good—ending the war—even though this means missing out on a less important good—seeing the season premiere.) McConnell builds this into condition 2 ("If he could attain the good effect [i.e., the one which, as per condition four, is proportional to the evil] without the bad effect he should do so"), while Walzer (1977, 151–6) builds it into condition 3. Either way, Pauline's action will fail this condition, since she could achieve the good—donating—without the pollution, even though it would cost her the travel which, we supposed, was not proportional to the pollution. We could say the same if Pauline claims that, in the ordinary version of the case, her donation is actually a good effect of her travel, since she wouldn't make it without traveling. (Of course, that claim might also be problematic for other reasons.)

there are other moral views which similarly cannot meet the desiderata. But rather than dwelling on those, we'll turn to our own explanation.

2.2 Universalizability and the second desideratum

We propose that whether a case of moral indulgence is permissible depends, roughly speaking, on whether it would be acceptable if people generally believed and were motivated to comply with principles that allowed them to be morally indulgent in the relevant way.[12] This is an appeal to a certain understanding of *universalizability*. This account fulfills our second desideratum of not being ad hoc, since a number of leading moral theories accept universalizability, i.e., that an act is permissible if and only if it would be acceptable for people to generally believe and be motivated to comply with principles that allowed them to perform the act in those circumstances. We'll focus on three moral theories that accept some form of universalizability here: rule consequentialism, contractualism, and a certain form of Kantianism. Roughly speaking, rule consequentialists will understand acceptability in terms of producing good consequences, and so think an act is permissible if and only if in accordance with principles which are such that they would have the best (or at least sufficiently good) consequences if accepted by everyone (e.g., Hooker 2000). Contractualists will understand acceptability in terms of being justifiable to all, where that is understood in a certain way, so that an act is permissible if and only if in accordance with principles for the regulation of behavior which everyone in some kind of fair decision situation could not reasonably reject (e.g., Rawls 1971; Scanlon 1998). And Kantians of a certain stripe will understand acceptability of terms of what the agent, or all agents, could rationally will, and so will think an act is permissible if and only if it is in accord with principles which the agent, or all agents, could rationally will to be universally accepted.

---

[12] Notice that this is somewhat different from a view which focused on which principles would be acceptable if people in general *successfully complied* with the relevant principles. Focusing on acceptance means that we must take into account the possibility that people might believe and be motivated to comply with principles without *actually* complying with them—due to ignorance, self-deception, weakness of will, or whatever—and so entails that acceptable principles will be determined in a way that minimizes the possibility of failures like this. We think, and it has been argued elsewhere, that it is better for people who accept moral theories of this general sort to go with our route, rather than the perfect compliance route (cf. Hooker 2000, chs. 3.2 and 4.2).

We should say a bit about the kind of Kantianism we'll focus on, because it differs importantly from orthodox Kantianism. For Kant himself, there are two ways a maxim might be such that one could not will that it be a universal law. First, it might be such that it "cannot even be thought without contradiction as a universal law of nature" (Kant 1997, 33). So, famously, I supposedly cannot conceive without contradiction of a situation where a maxim allowing false promising is universalized, since "in accordance with such a law there would properly be no promises at all, since it would be futile to avow my will with regard to my future actions to others who would not believe this avowal or, if they rashly did so, would pay me back in like coin," so that "my maxim, as soon as it were made a universal law, would have to destroy itself" (Kant 1997, 15). Of course, there is a question here about what, exactly, not being able to conceive of such a situation without contradiction comes down to (cf. Korsgaard 1985). Second, it might be that, though there is no "inner contradiction" in the maxim's being universalized, I nonetheless cannot rationally will that it be universalized because this would conflict with something else which I necessarily will, so that such a will would necessarily "contradict itself" (Kant 1997, 33). So, for instance, when someone considers universalizing a maxim allowing them to spend their life in leisure without developing their talents, they can conceive this situation without contradiction ("a nature could indeed always subsist with such a universal law"), but nonetheless "cannot possibly will that this become a universal law," since a rational being "necessarily wills that all the capacities in him be developed" (Kant 1997, 33). Of course, there is a question here about what, if anything, agents necessarily will, and why.

Derek Parfit (2011) suggests some important modifications to Kantianism. He suggests (chs. 1–2, 12–14) that we should simply start with the claim that we have reasons to want certain things (that we avoid pain, for instance), and understand what we can rationally will in terms of what's in accordance with those reasons. This allows us to avoid claims about what we can or can't conceive without contradiction, or what we do or don't necessarily will. And he suggests that we should consider what *everyone* could rationally will to be a universal law, not just what the agent could will, in order to avoid worries about whether it could be rational for an agent to will a clearly immoral maxim which would harm others but benefit themselves (2011, 334–8). Accordingly, he suggests that Kantians should adopt the principle that an act is wrong "if and only if, or just when, such acts are disallowed by some principle that is…one of the only principles whose being universal laws everyone

could rationally will" (2011, 412–13), where what everyone can rationally will is understood in the way just described. We think Parfit's Kantianism is more plausible than Kant's Kantianism, so, from now on, when we discuss Kantianism, we'll really have in mind Parfit's modification of it. Of course, if it turns out that Kant's Kantianism can also yield the right results about the paradigm cases, this only helps us, since it only strengthens our claim that the feature we're appealing to is found in many different moral theories.

With this kind of Kantianism in mind, we can now turn to the way in which the three theories we're focusing upon converge—or, as the case may sometimes be, diverge. There is some dispute over whether the acceptability conditions posited by each of these theories wind up yielding co-extensional results. For instance, are the principles that could be justified to everyone those that would have the best consequences? Or are the principles that have the best consequences those that every agent rationally could will? Parfit (2011) argues at great length that the best forms of rule consequentialism, contractualism, and Kantianism will turn out to be co-extensive with one another with respect to their judgments about the permissibility of actions. Unsurprisingly, others disagree.[13] We don't accept the claim that the three theories converge; in fact, we'll argue in Section 3.3 that this claim is false. However, we do think that the best forms of these theories exhibit a great deal of *overlap*. Fortuitously for us—or so we argue—they overlap in a way that yields agreement when it comes to the permissibility of moral indulgence in the paradigm cases which we wish to explain. (Thus, we won't need to worry too much about what the right way to flesh out what "acceptability" is, even in Section 2.3.) There are two reasons one might doubt this, which have to do with the fact that our examples involve, respectively, pollution and non-human animals. We'll address those in turn.

Some people think that nature has some sort of intrinsic value which is harmed by pollution. If this is right, rule consequentialism will recognize a prima facie duty not to pollute which is grounded in this value. It's not obvious that the same can be said for contractualism and Kantianism. For instance, Scanlon (1998, 218–23) argues that contractualists should not recognize "impersonal" values, which are not directly tied to the interests of any person, as directly giving rise to moral obligations. The thought is that in the contracting situation, no one will have standing to reject principles

---

[13] Parfit's argument comprises Volume I of *On What Matters*; for critical responses, see the essays by Susan Wolf, Allen Wood, Barbara Herman, and T. M. Scanlon at the beginning of Volume II.

on the grounds that they permit acting against these impersonal values, since no one directly has a stake in their being respected. Whatever intrinsic value nature has is supposed to be an example of such an impersonal value. If this is right (we make no claim about whether it is), then contractualists will not recognize the prima facie duty which rule consequentialists recognize here. However, both contractualists and Kantians will still recognize *a* duty against pollution which is grounded, not in the intrinsic value of nature, but in the interests of others who rely on nature for survival, enjoyment, etc., and who will be harmed by disease (e.g., in the case of toxic pollutants) or natural disaster (e.g., in the case of CO_2, which contributes to global warming) if pollution is excessive. Accordingly, rule consequentialists, contractualists, and Kantians can all agree that *some* amount of polluting will be wrong absent a defeater, even if they disagree somewhat about what that amount is or how strong the relevant obligation is. Accordingly, they will agree that it's possible to set up a case like Pauline's, where someone pollutes so much that they act in a way which is wrong, absent a defeater. And the explanation we give in Section 2.3 for why acceptable principles allow Pauline's indulgence won't be affected by the differences between the theories.

The issue with animals is a little more complicated. Provided that the flourishing and suffering of animals has value and disvalue, consequentialists think we have obligations to them. But contractualists and Kantians don't always agree: for instance, Scanlon (1998, 177–87) is "inclined" (184) to think that animals should be excluded from the contracting situation, and Kant (2001, 240) thinks we have no obligations to them since they lack a rational nature. This would be an important difference, when it comes to Scott's case. Such people might agree that we have reasons, derived from our obligations to humans, not to do certain things to animals: maybe we shouldn't eat them for environmental reasons, or shouldn't torture them so as not to cultivate cruel dispositions. But they might also think that Scott's case isn't morally much different than Pauline's case, whereas we thought it was much less obvious that what Scott does is permissible. For our part, we think it's obvious that we have obligations to animals. Fortunately, there are various ways contractualists and Kantians might explain this. For instance, contractualists might give animals trustees in the contracting situation to whom we must justify the principles we select (cf. Rowlands 1997; Scanlon 1998, 183–5). Kantians might say that non-human animals themselves possess the kind of agency necessary to be worthy of treatment as ends in

themselves, and so to have moral status within the Kantian framework (Korsgaard 2004). Since we think it's obvious that we have obligations to animals, we think the *best* forms of contractualism and Kantianism will be compatible with this fact. (Anyone who agrees with our judgment about Scott's case will probably agree.) We'll therefore assume that we're discussing animal-friendly versions of contractualism and Kantianism in what follows.

2.3 Application of universalizability and the first desideratum

Our explanation involving universalizability, which appeals to a component of many popular moral theories, is not ad hoc, and so meets our second desideratum. It also meets the first. Consider Billy, our indulgent, philanthropic murderer. It's hard to imagine that a society could last very long if people thought it was acceptable to murder others, provided that they donated a few thousand dollars to the Against Malaria Foundation. The results would be catastrophic, probably leading to the collapse of society altogether. People would live in fear, distrust, and resentment of each other. They might feel a need to preemptively strike against their enemies, and already competitive arenas like the academic job market would become fraught with danger.[14] Notably, this line of reasoning is similar to a classic rule consequentialist explanation of why one shouldn't butcher a patient in order to save five others with the patient's organs (cf. Parfit 2011, 363–4). If doctors generally did this, society would be filled with fear, distrust, and resentment, and people would fail to seek needed medical treatment. The resulting situation would be unacceptable by any plausible standard. For instance, it would have bad consequences, and some people would be able to reasonably reject it and unable to rationally will it. Thus, if

---

[14] One might wonder why "'don't freak out and dissolve society when people morally indulge in murder" couldn't be added to the rule so that offsetting murder with saving many would be universalizable. It's important to remember that in evaluating rules, one of the things we should consider is the effort needed to get people to accept and be motivated to comply with these rules (e.g., Hooker 2000, ch. 3). If a rule makes acceptance and motivation to comply much harder or impossible, that's a reason to think it isn't universalizable. Given contingent facts about human psychology, it's hard to see people accepting and being motivated to comply with such a rule. See also footnote 9.

universalizability is necessary for permissibility, Billy's donations don't defeat the wrongness of his killing. We get the intuitively correct result.

Meanwhile, consider Pauline. A certain amount of pollution inevitably will be created by any modern society, and there are great benefits to producing some. The question for acceptable, universalizable principles is not whether to pollute, but how to regulate the pollution that does occur. It's important that we not produce so much that the marginal costs exceed the marginal benefits, and, if someone produces more than their fair share, this will be wrong, absent a defeater. (Principles which allowed people to just produce as much pollution as they wanted, without needing a defeater for doing so, would result in too much pollution being created, with the attendant catastrophes.) However, people do sometimes have reasons for wanting to produce more than their fair share—economic reasons, say, or, in Pauline's case, personal enjoyment. If someone is willing to use resources which they otherwise could permissibly have kept to do things which more than offset their additional pollution, it seems acceptable to allow this to serve as a defeater for the wrongness of their polluting: they get what they want, and the pollution situation is better than if they'd done nothing. Further, unlike with murder, it is generally not possible to target particular individuals by polluting. (Polluting may impose a very small risk—say, of getting a disease—on a very large number of people, but that would be an extremely ineffective method of targeting one's rivals.) Therefore, the same reasons for fear, preemptive strikes, and avoiding the academic job market are not applicable to pollution offsets. So, under the right circumstances, offsetting will promote value, and no one has reason to object. As mentioned, we may have issues with particular pollution offset systems (e.g., they may be ineffective or unfairly disadvantage poor people), but having a fair and effective system in place to manage pollution seems, in principle, universalizable. Again, we get the intuitively correct result.

Finally, Scott's situation is more complicated. On the one hand, unlike Billy's case, allowing people to behave as Scott behaves wouldn't spread distrust and resentment, provoke preemptive strikes, or, generally speaking, do anything else that might lead to the collapse of civilization. Further, it would produce some good, insofar as it might promote donations and other good acts which, in individual cases, more than offset the harm of eating meat. On the other hand, we might worry, for instance, that it would be impossible to cultivate and sustain the attitudes needed to eliminate the widespread abuses of animals which exist in our society as long as people feel free to eat them, even if they feel obligated to make up for doing so. And, of course, an

outright prohibition on eating meat would have the benefit of reducing the killing of animals for meat. So the question winds up being complicated. We won't try to settle the question here, though we hope we've provided a framework in which one might do so, and think it's worth doing, given the question's practical relevance. What matters for our purposes now is that, again, our explanation gives the result we were looking for—Scott's case is less clear than either Billy's (which is rendered impermissible) or Pauline's (which was permissible).

2.4 Cause-matching

As seen above, universalizability yields the correct result in all three cases and meets the two desiderata. Our account also provides a plausible account of the significance of what we earlier called "cause-matching," or directing your good action towards helping the same issue that your bad action exacerbated. Recall that, on the one hand, cause-matching often seems appropriate, while, on the other, it seems puzzling for two reasons. First, it isn't immediately clear why one would cause-match instead of just doing whatever was best. Second, attempts at cause-matching might seem arbitrary, absent some metaphysically privileged way of individuating causes (and it might seem implausible that there is such a way).

Consider the first puzzling feature. There are various reasons why acceptable principles might recommend or sometimes even require cause-matching, and these help explain why it sometimes seems especially appropriate. For instance, cause-matching might help ameliorate coordination problems. There are reasons to do things that cause pollution, but it's important that we not cause too much, and whether we cause too much depends, in part, on individual actions performed by lots of different people. One reasonable way to help make sure that we don't cause too much, and to help make sure that people causing excess pollution aren't doing it for totally trivial reasons, is to make individuals responsible for offsetting their pollution if they cause more than their fair share. (Of course, in practice, this will also need to be supplemented with other measures.) Or consider the importance of indicating concern for various issues (cf. Crummett forthcoming). If we learn that Scott eats meat, we might reasonably suppose that he either rejects strong claims about the value of animal lives, or accepts but doesn't act in accordance with them. Since people take their cues about what's acceptable partly from other people, it will be very hard to cultivate and sustain good

attitudes towards animals if many people send such signals.[15] Perhaps, though, one way for Scott and others to offset this impression is to ensure that they do other actions such that they benefit animals on the whole. Acceptable principles might therefore recommend signaling in this way, if they allow what Scott does at all. With all that being said, while cause-matching is often appropriate, we don't claim that it's *essential* to the permissibility of moral indulgence. Perhaps there are cases where considerations like these don't apply, or where they're outweighed by the potential for doing *much* more good by not cause-matching. And, again, this seems like an intuitively reasonable position.

Now consider the second puzzling feature: individuation. Given our understanding of why cause-matching is appropriate, there doesn't need to be any metaphysically privileged way of individuating causes. Instead, whether we should classify actions as affecting "the same cause" for purposes of cause-matching will depend on contingent considerations about the results that principles mandating that I cause-match in a certain way will have in practice. The coordination problem in which Pauline is involved has to do with limiting the amount of pollution. If there is no deep metaphysical reason to classify "pollution" as a single cause without classifying "pollution and criminal justice reform" as one, it's nonetheless the case that Pauline's donating to criminal justice reform won't help us with the relevant coordination problem. Or in Scott's case, that he eats meat makes it natural to suppose that he must not care too much about farmed animals; it's much harder to draw inferences from this to his views about, say, nuclear disarmament. So cancelling the inference by signaling concern *for animals* is what's morally salient. This view provides a way of individuating causes based on their morally salient features, regardless of whether individuating them based on morally salient features is more metaphysically joint-carving than alternatives.

As we said earlier, we don't take cause-matching to be an essential feature of morally indulgent actions. However, it does often seem appropriate, and our view explains why this is. Furthermore, our view offers natural ways of responding to the two puzzles associated with cause-matching. In light of this, we do take the ability of our view to account for cause-matching as a major advantage of our view.

---

[15] Would the appeal to universalizability render this consideration irrelevant, since, in evaluating principles, we assume that people generally accept and attempt to follow them? No, see footnotes 12 and 14.

3. Divine Indulgence

3.1 Background assumptions

In Section 2, we argued that universalizability gives plausible answers about which cases of moral indulgence are permissible and which are not. Furthermore, this kind of universalizability is an independently credible candidate for being morally important, since it is found in certain forms of Kantianism, contractualism, and rule consequentialism. In this section, we turn to the case of divine indulgence. Interestingly, while Kantianism, contractualism, and rule consequentialism converge in their verdicts on the morally indulgent actions we've discussed thus far, they diverge when it comes to the divine case (and any structurally similar imaginary cases). Thus, the discussion of whether it's permissible for God to perform morally indulgent actions might, in addition to being interesting on its own, provide a reason to prefer some of those theories over others.

Before getting underway, there are two background assumptions about God's obligations which we would like to make explicit. First, we assume that God has moral obligations. Second, we assume that the same fundamental principles that govern God's obligations also govern our obligations (so that the account we developed to handle human moral indulgence can also be applied to God). These assumptions are not wholly uncontroversial (e.g., Adams 2017; Murphy 2017). But they comprise the dominant view, and, anyway, we don't have space to defend them here.

We'll also make two more assumptions about the nature of modal space and its impact on God's obligations. First, we assume that for each world that God could make, there is an arbitrarily better world that God could make instead. Here are two ways that might be true. First, perhaps God could improve on any world by adding an additional happy person, while holding everything else fixed. (This requires that two things be true: First, the modal claim that, for any world, there's another world which is just like it except that it has one more happy person, and, second, the axiological claim that this additional person makes the world better.) God could then make any world arbitrarily better by adding an arbitrarily large number of additional happy people. Second, perhaps there is no limit on how well-off a person can be, and God could improve on any world by making every person in that world slightly better off—for instance, maybe there is no metaphysical limit on how much pleasure you can experience, or to how strong a desire can be when it's satisfied or to how many desires you can have which are satisfied, or maybe an objective list theory of well-being is true, and there's no limit on

how many objective goods you can be positively related to. (This requires that two things be true: First, the modal claim that, for any world, there's another world which is just like it except that everyone is better off, and, second, the axiological claim that making everyone better off makes the world better.) God could then make any world arbitrarily better through an arbitrarily large number of improvements to individual well-being. Again, these claims are not uncontroversial (e.g., Kraay 2010; Climenhaga 2017), but they are widely accepted and seem fairly plausible to us.

We also assume that even if the prior claim is true, there is at least one world which God can permissibly make. (Given that at least this world exists, the theist better hope that it was created permissibly!) Some philosophers (e.g., Rowe 2004, ch. 6) have argued that if there are no unsurpassable worlds, God cannot exist, since a perfect being could not make a world which was less than the best. Brian Kierland and Philip Swenson (2013) argue that this claim is false, since it contradicts a true version of the ought-implies-can principle. There must be at least one world which God can actualize, consistent with God's perfection. Assuming that God's perfection entails God's not violating any moral obligations, it follows that there's at least one world God can permissibly make. We think Kierland and Swenson are correct, and so accept this implication.

3.2 Divine indulgence

With all this in the background, we now can construct two cases of divine indulgence. Here's the first case. Imagine three possible worlds. The first, *W1*, is a totally paradisiacal world. *W1* is good enough that God can permissibly make it, despite the fact that God could have made a better world by adding more happy people. The second, *W2*, is like *W1*, with two exceptions. First, it contains Bob, an innocent person who suffers horribly and pointlessly throughout his life before being annihilated. Second, the world also contains 500 trillion additional happy people. There are so many additional eternally happy people that, overall, *W2* is better than *W1*.[16] Finally, the

---

[16] One might claim that *W2* can't really be better than *W1*: Bob's terrible suffering lexically outweighs any possible set of additional good things which God might add. You might think this is so because, in actualizing *W2*, God would wrong Bob, and divine wrongdoing would be so horrible as to lexically outweigh any set of created goods. If so, set that aside. Claims about the supposed disvalue of the wrong involved in making *W2* can't be involved in the answer to the question we're concerned with—whether and why it's wrong to make *W2*—on pain of circularity. Maybe you instead think *W2* can't be better than *W1*, not because of any wrongdoing,

third world, *W3*, is like *W2*, except that Bob also gets an eternal, happy life. *W3* is better than *W2*. Given that *W1* is a world that God permissibly can make, *W3*, which is a Pareto improvement upon *W1*, should also be a world that God permissibly can make.

Notice that creating *W2* seems to fit the single action model of moral indulgence. Doing so has a property—namely, giving Bob a pointlessly horrible existence—that makes performing it wrong, absent a defeater. Furthermore, since Bob's horrible existence is totally pointless, nothing serves as a defeater, unless creating more goodness than was otherwise necessary does. Creating *W2* also has another property—namely, causing the 500 trillion happy people to exist—which is such that (i) it's supererogatory, (ii) it outweighs the badness of giving Bob a horrible existence, and (iii) in light of the possibility of *W3*, the relevant good could have been achieved without the relevant bad, or anything similar to it. Finally, creating *W2* might also meet the intention criterion; assuming God intends to act in accordance with God's obligations, God's creating more goodness than was otherwise necessary will presumably be partly intended to offset God's creating gratuitous evil. If this is right, then the creation of *W2* fits the structure of morally indulgent actions.

So, could God permissibly create *W2*, condemning Bob to a life of nothing but pointless, unimaginable torment? If so, this would seem to undermine any version of the problem of evil which appeals to the claim that God would act wrongly in allowing evil.[17] After all, for any gratuitous evil we observe, it might, for all we know, be offset by something wonderful somewhere else. On the other hand, the permissibility of *W2* might have unnerving implications—what if God treats *us* like Bob? For our part, we think the creation of *W2* seems clearly impermissible: If God wants the

---

but because Bob's suffering just directly outweighs any possible additional goods. We think this claim is implausible. After all, there are people on earth who live lives full of horrible suffering. If philosophical naturalism is true, their earthly lives represent their entire existences. But very few naturalists think this *alone* makes the existence of life on earth regrettable (even if they think it is, in fact, regrettable due to the *proportion* of people who lead such lives). And taking this view would have implausible implications for certain ethical issues. For instance, it would imply that people who don't believe in an afterlife should think procreation generally makes the world vastly worse, since, unless your line quickly dies out, it's virtually certain that at least one of your descendants will have a life of misery. But thinking this on this ground alone would be very odd.

[17] Other versions might, for instance, appeal to the claim that God *loves* us, and therefore wouldn't allow certain evils, rather than the claim that God would act *wrongly* in allowing those evils. For a criticism of versions which appeal to something other than God's moral obligations, see Murphy (2017, ch. 2).

500 trillion additional happy people, God should just make *W3* and give Bob a nice life, too. (Note, too, that if God were to turn the creative decision over to us, it would seem clearly impermissible to create *W2* rather than *W3*. Combined with our earlier assumption that God's obligations are similar at the fundamental level to our own, this suggests that it would also be impermissible for God to create *W2* rather than *W3*, unless there is some morally important disanology between ourselves and God which is relevant to this case.) If this judgment was in tension with our best theory of the permissibility of moral indulgence, then perhaps we would need to rethink it. On the other hand, if a theory can account for this judgment, that might be a reason to accept the theory. We'll return to this point shortly.

Here's a second moral indulgence case. The case of moral indulgence involving *W1–W3* employs the first procedure for improving worlds by adding additional happy people. A morally indulgent case can also be constructed out of the second procedure involving increasing the happiness of existing people. (This second example is important primarily because a certain form of rule consequentialism will have something different to say about this case than about the prior case.) Again consider three worlds: *W4*, *W5*, and *W6*. *W4* is like *W1*. It's a totally paradisiacal world that God can permissibly make, even though God could have made everyone much better off. *W5* is like *W4*, except for two features. First, it contains Jack, an innocent person who suffers horribly and pointlessly throughout his life before being annihilated. Second, it contains large improvements in everyone else's (already great) lives—improvements that are sufficient to outweigh the disvalue of Jack's suffering. Finally, *W6* is like *W5*, except that Jack has a life as eternally wonderful as everyone else's. Like *W4*, *W6* is a world that God can permissibly make.

Creating *W5* intuitively appears impermissible for essentially the same reasons that creating *W2* did. After all, it contains Jack, an individual who has a pointlessly horrible life. Given *W6*, it's clear that Jack's suffering is not necessary for achieving some greater good. Nevertheless, *W5* is significantly better than *W4*, which is permissible to make. In light of this, creating *W5* looks morally indulgent, again for basically the same reasons that creating *W2* does.

3.3 Universalizability and divergence on the divine

So what does our account imply about the permissibility of these cases of moral indulgence? Unlike in the cases of Scott, Billy, and Pauline, we think

there will be a deep, essential difference depending on which *particular* moral theory's account of universalizability we employ. Start by considering rule consequentialism. To know what rule consequentialism says about the cases in question, we must distinguish between particular forms of it more precisely than we have so far. Specifically, we must make two distinctions. The first is between *maximizing* and *satisficing* rule consequentialism. *Maximizing* rule consequentialism tells us to follow the rules that would have the *best* consequences if people generally believed and were motivated to comply with them. *Satisficing* rule consequentialism tells us we need only follow rules which would have *sufficiently good* consequences if people generally believed and were motivated to comply with them. Next, distinguish between *person-affecting* and *total* rule consequentialism (cf., e.g., Parfit 1984, ch. 18; Singer 1993, 103–4). *Person-affecting* rule consequentialism tells us that what matters is the sum total of consequences *for people who actually exist*. It denies that there is any obligation to create *additional* happy people. *Total* rule consequentialism claims that what matters is just the overall value of the world. It recognizes an obligation to create additional happy people, if one can do so without causing offsetting harm.[18] These are cross-cutting distinctions, so that one can be a maximizing person-affecting, maximizing total, satisficing person-affecting, or satisficing total rule consequentialist.

The maximizing total view implies that, if there are no unsurpassable worlds, God cannot permissibly create any world. For any world God creates, a principle which forbade creating that world and instead required creating a better world would produce a better result. This entails that *W1* and *W4* are impossible (since they were defined as being permissible), and that the other worlds mentioned are also impossible (since they were defined relative to *W1* or *W4*). Since we thought that, given ought-implies-can, there should always be at least one world which God can permissibly actualize regardless of what modal space looks like, this seems to us like a bad result.

Meanwhile, *if* there is a limit to how well-off an individual can be, then the maximizing person-affecting view implies that *W1* and *W3* are possible. That God could create additional happy people doesn't count against their

---

[18] Singer (1993, 103–4) defines the total view as implying that "that it is good to increase the amount of pleasure in the world by increasing the number of pleasant lives, and bad to reduce the amount of pleasure in the world by reducing the number of pleasant lives," and the person-affecting (or, as he calls, the "prior existence") view as denying "that there is value in increasing pleasure by creating additional beings." Our usage tracks his, except that we put things in terms of what's obligatory rather than what's valuable.

permissibility, since there's no obligation to create happy people. This view also implies that *W2* is impermissible, since it fails to maximize the well-being of the people who actually exist. By making *W3*, God could increase Bob's well-being without hurting anyone else. We agree with all these judgments. However, if there is *not* a limit to how well-off an individual can be, then this view implies that God cannot permissibly create any world. For any world God creates, a principle which forbade creating that world and instead required creating an otherwise similar world where the inhabitants were better off would produce a better result for the people who actually exist. This implies that *W4–W6* are impossible. Since we thought that there should always be at least one world which God can permissibly actualize regardless of what modal space looks like, this again seems to us like a bad result.

Meanwhile, any form of satisficing view probably implies that all of *W1–W6* are permissible. Consider the satisficing total view. If all we need is for our principles to result in a world which is good *enough*, they can allow God to create worlds like *W1*, *W3*, *W4*, and *W6*. However, since *W2* and *W5* are better than *W1* and *W4*, respectively, presumably they will also allow the creation of those worlds. After all, these worlds are better than *W1* and *W4*, respectively, which were good enough, and what mattered was that things were good enough. Of course, because, on the rule consequentialist picture, we need to consider the consequences of people generally believing that a certain set of principles is the correct one, we will also need to factor in the consequences of people believing that God is permitted to allow horrible, pointless evils to befall them. This might cause some anxiety, and so on. However, God can always make up for this by adding more good to the world. Principles which tell God to make a world which is good enough and which forbid God from allowing horrible, pointless suffering will not produce better consequences than principles which tell God to make a world which is good enough, factoring in any pointless evils God allows *and* any bad consequences which result from people knowing God can allow pointless evils. Both will satisfice. Since we thought *W2* and *W5* were intuitively impermissible, we think this is a bad result.

Similar things can be said about the satisficing person-affecting view. Since this view doesn't recognize an obligation to produce additional happy people, this possibility is no threat to the permissibility of *W1* or *W3*. Since it only requires that God produce *enough* good for actually existing people, there will presumably be permissible worlds like *W4* and *W6*. However,

since it only requires that the sum total of good for existing people be good enough, it will allow the misfortune of Bob and Jack to be offset by additional goods to others, and so will allow *W2* and *W5*. Satisficing rule consequentialism here seems to license too much, just as satisficing act consequentialism licensed too much when it allowed Billy's moral indulgence.

We saw that every form of rule consequentialism gave what we regarded as implausible judgments of one form or another. If this was a general feature of attempts to explain the permissibility of moral indulgence by appealing to universalizability, it might threaten our account. If this was a feature of every plausible attempt to explain the permissibility of moral indulgence, it might cause us to rethink our initial judgments. However, both contractualism and Kantianism seem to yield what we regarded as the intuitively correct judgments about the two cases of divine moral indulgence we presented.

First, consider contractualism. Plausibly, contractualism renders both *W2* and *W5* impermissible, since, plausibly, Bob and Jack could reasonably reject principles allowing them to be given pointlessly horrible lives. If their suffering was *necessary* for the additional goods found in their worlds, perhaps one could argue that reasonableness required them to take one for the team. However, given that isn't so, intuitively, it seems perfectly reasonable for them to insist that, if God wants the additional goods involved, God should make a world like *W3* or *W6* instead.

What about *W1* and *W3*? Generally, contractualists have thought that we have an obligation to act in ways which are justifiable to all and only *actual* people, without having any obligation to bring merely possible people into existence (cf. Scanlon 1998, 177–87). If this is the right route, then contractualism would be more like person-affecting than total consequentialism, and the possibility of additional happy people would be no threat to the permissibility of these worlds. What about *W4* and *W6*? One might suppose that, if there is no limit to individual well-being, individuals could reasonably reject whatever God did, since they always could have been made happier. But the contracting procedure is supposed to model a decision-making process among reasonable agents who are attempting to reach agreement on principles which are acceptable to all parties. In such a situation, agents who will object to any principle put before them don't seem to be acting in good faith. It therefore seems plausible to think that some lives will be good enough for God to give us, even though it may be difficult to say exactly

where the line is. (And like we said, it seems intuitive to think the line will rule out a life consisting of nothing but horrible torment.) Contractualism therefore gives what we thought were the correct answers about the permissibility of *W1–W6*.

Now consider Kantianism (of the Parfit-flavored sort we have in mind). Could Bob and Jack rationally accept their treatment? Answering "no" makes a slightly stronger claim than saying that they could rationally reject it (it could be that they could rationally accept or reject it). However, given the availability of *W3* and *W6*, it still seems fairly clear to us that they shouldn't accept the treatment they get in *W2* and *W5*. As in contractualism, what matters is plausibly what *actual* people can rationally accept, so there is no threat to the permissibility of *W1* from the possibility of additional happy people. Can the people in *W4* accept their position, despite the fact that they could have been happier? Consider the famous case of EverBetter Wine, which gets better and better over time without limit (Pollock 1983). Fortunately, I am immortal, so I can wait as long as I like to drink the wine. Unfortunately, whenever I drink the wine, I would have been better off had I waited. It seems to us that there will be a point at which I'm rationally permitted to drink the wine—presumably, after it's gotten really, really, really good—even though it's hard to say exactly where that point is, and even though I forsake arbitrarily great gains by drinking it at that point rather than later. If I'm rationally able to give myself a benefit which is good enough in the wine case, it seems plausible that I'll also be rationally able to accept God's giving me a life which is good enough. Therefore, *W4* and *W6* will be permissible, which, again, seems to us like the right result.

To sum up: Our initial judgments were that God should be permitted to actualize some world or other, even if there are no unsurpassable worlds, but that God shouldn't actualize worlds where someone has a pointlessly horrible existence. The forms of rule consequentialism we considered each violated one of these claims. However, Kantianism and contractualism both implied them. That we can justify these judgments on a theoretical level may be a reason to hold on to them. That Kantianism and contractualism can do so while rule consequentialism cannot may be a reason to favor these theories over rule consequentialism. And that our universalizability-focused account of the moral status of moral indulgence can, when interpreted in the Kantian or contractualist way, give us plausible judgments in these cases may be a reason to accept it.

4. Closing Thoughts

We close by noting an important question for future research. A feature of Bob's and Jack's cases was that they not only suffered horribly, but had horrible *existences*: their suffering was uncompensated and their lives were bad for them on the whole. However, not every case of moral indulgence needs to be like this. Suppose there is no limit to how well-off a person can be, so that, for any life they actually have, they could have had instead an arbitrarily better life. And suppose that, in such a situation, there is nonetheless at least one life God is permitted to give us. We can now construct a case of moral indulgence where we start with *L1*, a fantastically good life which God is permitted to give us, then consider *L2*, a life which is like *L1* except that it contains (i) some horrible pointless suffering and (ii) offsetting goods so that we're better off than in *L1*, and finally consider *L3*, which is like *L2* except without the pointless suffering. Can God give us *L2*? If so, this would be hugely significant. Just like the claim that *W2* and *W5* are permissible, it would seem to undermine any version of the problem of evil which appeals to the claim that God would act wrongly in allowing evil. After all, for any gratuitous evil we observe, it might, for all we know, be offset *for that person* by some heavenly bonus. And the potential implications may not be as disturbing as those which follow from saying that *W2* and *W5* are permissible. We might suffer pointless evils, but we might at least be assured of not pointlessly being denied an existence which is extremely good on balance.

On the Kantian and contractualist views, whether God is permitted to give us *L2* will depend on whether we could, respectively, rationally accept or not reasonably reject God's giving us *L2* when we were fine with God's giving us *L1*. Whether this is so will depend on whether we have reason to care about harms apart from their impact on our overall well-being (since, with *L2*, we are harmed but better off overall). One might suppose that we do. Return to the EverBetter Wine. Suppose someone else needs to decide when to open the wine and give it to me. One might think that, intuitively, it would be rational for me to accept their opening the wine now, and rational to accept their opening it a month from today, but not rational to accept their opening it a month from today while slamming my hand in a door, even if I'm better off in the final scenario than in the first. On the other hand, it might also seem somewhat odd in itself to think that I should care about harms apart from their impact on my overall well-being, and perhaps there is a way to argue that our intuitions about door slamming and similar

things aren't appropriately responsive to the bizarre features of the case. We don't think we have a decisive way to settle this issue here. However, given its potentially great significance, it deserves further attention.[19]

Bibliography

Adams, Marilyn McCord (2017), "A Modest Proposal? Caveat Emptor! Moral Theory and Problems of Evil," in *Ethics and the Problem of Evil*, ed. James Sterba, Indiana University Press, pp. 9–26.

Alexander, Scott (2015a), "Vegetarianism for Meat-eaters", *Slate Star Codex*. http://slatestarcodex.com/2015/09/23/vegetarianism-for-meat-eaters/. (accessed August 19, 2018).

Alexander, Scott (2015b), "Ethics Offsets," *Slate Star Codex*. http://slatestarcodex.com/2015/01/04/ethics-offsets/ (accessed August 19, 2018).

Alexander, Scott (2017), "Contra Askell on Moral Offsetting," *Slate Star Codex*. http://slatestarcodex.com/2017/08/28/contra-askell-on-moral-offsets/ (accessed August 19, 2018).

Animal Charity Evaluators (2017), "The Humane League." https://animalcharityevaluators.org/charity-review/the-humane-league/#rf1-12708 (accessed March 15, 2018).

Climenhaga, Nevin (2017), "Infinite Value and the Best of all Possible Worlds," *Philosophy and Phenomenological Research*, 97(2), pp. 367–92.

Crummett, Dustin (forthcoming), "Expression and Indication in Ethics and Political Philosophy," *Res Publica*.

Herman, Barbara (2011), "A Mismatch of Methods," in *On What Matters*, Vol. 2, ed. Samuel Scheffler, Oxford University Press, pp. 83–115.

Hooker, Brad (2000), *Ideal Code, Real World*, Oxford University Press.

Kagan, Shelly (1989), *The Limits of Morality*, Oxford University Press.

Kant, Immanuel (1997), *Groundwork of the Metaphysics of Morals*, trans. and ed. Mary Gregor, Cambridge University Press.

Kant, Immanuel (2001), *Lectures on Ethics*, trans. Peter Heath, ed. Peter Heath and J. B. Schneedwind, Cambridge University Press.

Kierland, Brian and Philip Swenson (2013), "Ability-Based Objections to No-Best-Worlds Arguments," *Philosophical Studies*, 164: 669–83.

Korsgaard, Christine (1985), "Kant's Formula of Universal Law," *Pacific Philosophical Quarterly* 66(1–2): 24–47.

Korsgaard, Christine (2004), *Fellow Creatures: Kantian Ethics and Our Duties to Animals*, Tanner Lectures on Human Values.

Kraay, Klaas (2010), "Theism, Possible Worlds, and the Multiverse," *Philosophical Studies* 147(3): 355–68.

McConnell, F. J. (2003), "Principle of Double Effect," in *New Catholic Encyclopedia*, Vol. 4, The Catholic University of America Press.

McMahan, Jeff (2001), *The Ethics of Killing*, Oxford University Press.

Murphy, Mark (2017), *God's Own Ethics: Norms of Divine Agency and the Argument from Evil*, Oxford University Press.

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

Parfit, Derek (2011), *On What Matters*, Vol. 1, Oxford University Press.

Pollock, John (1983), "How do you Maximize Expectation Value?" *Noûs* 17(3): 409–21.

Rawls, John (1971), *A Theory of Justice*, Harvard University Press.

Rosen, Gideon, (2010), "Might Kantian Contractualism be the Supreme Principle of Morality?", in *Essays on Derek Parfit's On What Matters*, ed. Jussi Suikkanen and John Cottingham, Wiley-Blackwell, pp. 97–115.

Rowe, William (2004), *Can God be Free?*, Clarendon Press.

Rowlands, Mark (1997), "Contractarianism and Animal Rights", *Journal of Applied Philosophy*, 14: 235–47.

Sandel, Michael (2012), *What Money Can't Buy: The Moral Limits of Markets*, Farrar, Straus, and Giroux.

Scanlon, T. M. (1998), *What We Owe to Each Other*, Harvard University Press.

Scanlon, T. M. (2011), "How I am not a Kantian," in *On What Matters*, Vol. 2, ed. Samuel Scheffler, Oxford University Press, pp. 116–40.

Singer, Peter (1993), *Practical Ethics*, Cambridge University Press.

Wallace, R. Jay (2013), "The Deontic Structure of Morality," in *Thinking About Reasons: Themes from the Philosophy of Jonathan Dancy*, ed. David Bakhurst, Brad Hooker, and Margaret Olivia Little, Oxford University Press, pp. 137–67.

Walzer, Michael (1977), *Just and Unjust Wars*, Basic Books.

Wolf, Susan (2011), "Hiking the Range," in *On What Matters*, Vol. 2, ed. Samuel Scheffler, Oxford University Press, pp. 33–57.

Wood, Allen (2011), "Humanity as an End in Itself," in *On What Matters*, Vol. 2, ed. Samuel Scheffler, Oxford University Press, pp. 156–68.