

Wrongfulness Rewarded?

a normative paradox

[penultimate draft; please cite published version: <https://doi.org/10.1007/s11229-021-03098-4>]

David O'Brien
dobrien10@tulane.edu
Department of Philosophy
Tulane University
New Orleans, LA, USA

Ben Schwan (corresponding author)
b.e.schwan@gmail.com
Department of Bioethics
Case Western Reserve University
Cleveland, OH, USA

Contribution Statement: Each author contributed equally to this paper.

Abstract

In this paper, we raise and discuss a puzzle about the relationships among goods, reasons, and deontic status. Suppose you have it within your power to give someone something they would enjoy. The following claims seem platitudinous: (1) you can use this power to reward whatever kind of option you want, thereby making that option better and generating a reason for that person to perform it; (2) this reason is then weighed alongside and against the other reasons at play; and (3) altogether, the reasons determine the deontic statuses of that person's options. We show, however, that in a certain class of cases at least one of these apparent platitudes must be false. In particular, we show that in a certain kind of case *wrongfulness* cannot be rewarded. In some cases, if one tries to reward wrongfulness, something surprising must go awry: either what you attempt to give as a reward would not, in fact, be good, or it would not generate a reason, or it would have a surprising effect (or non-effect) on the deontic status of the relevant options. The upshot is that the relationships among goods, reasons, and deontic status are complicated in ways that have not previously been remarked.

Keywords: paradox, normativity, reasons, goods, deontic status

Acknowledgements

We are grateful to Jeff Behrens for feedback on an earlier draft of this paper and to Shyam Nair and Dan Hausman for helpful conversation.

1. Introduction

There's a *reward* for your ϕ ing only if you'll get something good for ϕ ing and this fact is a reason for you to ϕ .¹ Rewarding, then, alters the normative landscape. Once you're offered a reward for doing something, you gain a reason to do that thing, and, sometimes, this additional reason is enough to change what you ought to do.

Now suppose you have some good to give and you'd like to use it as a reward. It's natural to think that it's possible for you to use your good to reward whatever kind of option you want—i.e., that your power to reward is, in this sense, unrestricted. But, we will argue, given some very plausible principles, this natural thought leads to a contradiction in a range of cases in which you attempt to use a good to reward what's *wrongful*. And so there is a puzzle about rewarding:

THE REWARD PUZZLE. Either the power to reward is not unrestricted, or at least one other very plausible principle about normativity is false.

We should note here at the outset that we don't take our puzzle to be primarily of interest as a puzzle in applied ethics—as a puzzle about everyday attempts to reward. Rather, we think it should be of interest, first and foremost, to ethical theorists and metaethicists because it shows that *some* aspect of our understanding of the normative domain needs revision—either our understanding of the good, or of reasons, or of deontic status, or of the relations that hold among them. And, while there are suggestive parallels between our puzzle and other well-known paradoxes (e.g., the liar and set paradoxes), the fact that our puzzle arises within the normative domain complicates the search for a solution. So, at any rate, we shall suggest.

Our paper proceeds as follows. In §2, we present and argue for THE REWARD PUZZLE. In §3, we canvass the various unpalatable options with which we are faced if THE REWARD PUZZLE is true. In §4, we conclude with some remarks on the interest and upshot of our arguments.

2. The argument for THE REWARD PUZZLE

Consider:

The Pickle Popsicle. Nadia is deciding between flossing (which would moderately improve her dental health) and not flossing (which would slightly improve her mood), and she *loves* pickle popsicles. Worried that Nadia is becoming a bit of a goodie-two-shoes, Nadia's mother Frances says to her: "You'll get a reward—this pickle popsicle—but just on the condition that you do something *wrongful* (that is, an action that is forbidden)."

Let f_1 be the fact that flossing would moderately improve Nadia's dental health; let f_2 be the fact that not flossing would slightly improve Nadia's mood; and let R_ϕ be the propositional function that Nadia will be rewarded with the pickle popsicle if and only if she ϕ s (thus R_{floss} would be the fact, if it is one, that Nadia

¹ As we mention just below, the purpose of the ensuing project is to raise and examine a puzzle about normativity, not to critique or defend a particular account of rewards. So if you disagree that being good and being reason-generating are necessary conditions for being a reward, then treat our use of 'reward' in this way as stipulative.

will be rewarded with the pickle popsicle if she flosses). The assumption that Frances can reward Nadia in this way, together with some unexceptional principles about normativity, yields a contradiction. To see this, consider:

- (1) REWARD FOR WRONGFULNESS. For any option φ , R_φ just in case it's forbidden for Nadia to φ .
- (2) REASONS ARE UNBALANCED. (i) f_1 is a reason to floss and f_2 is a less weighty reason to not floss; (ii) for any option φ , if R_φ , then the reasons to φ are weightier than the reasons to not φ .
- (3) FORBIDDEN IFF WORST. For any option φ , it's forbidden for Nadia to φ just in case the reasons to not φ are weightier than the reasons to φ .

REWARD FOR WRONGFULNESS expresses the natural thought that Frances's offer constitutes a genuine reward—i.e., that it would be good for Nadia to get the pickle popsicle and that the fact that she would get it is a reason for her to do whatever option is forbidden. REASONS ARE UNBALANCED encodes some apparently unremarkable principles about the weights of individual reasons and combinations of reasons. FORBIDDEN IFF WORST is an instance of a plausible principle about normativity that links the reasons for an action and that action's deontic status. But, given 2 and 3, 1 must be false.

Here is a proof: Suppose for *reductio* that (4) it is forbidden to floss. Then, (5) the reasons to floss are weightier than reasons to not floss (from 4, 1, and 2). And so, (6) it's not forbidden to floss (from 5 and 3). But 4 and 6 are together a contradiction. So, (7) it is not forbidden to floss (from 4-6). By a similar *reductio* sub-proof, beginning from the assumption that it is forbidden to not floss, (8) it is not forbidden to not floss. So (9) it is neither forbidden to floss nor forbidden to not floss (from 7 and 8). However, (10) it is neither forbidden to floss nor forbidden not to floss only if it's true *both* that the reasons to floss are not weightier than the reasons to not floss *and* that the reasons to not floss are not weightier than the reasons to floss (from 3). But (11) either the reasons to floss are weightier than the reasons to not floss or the reasons to not floss are weightier than the reasons to floss (from 2). So, (12) it's not the case that it's neither forbidden to floss nor forbidden not to floss (from 10 and 11). Since 9 and 12 are together a contradiction, 1-3 is an inconsistent triad. In other words: either Frances cannot reward Nadia in this way, or at least one of the apparently unexceptional normative principles encoded by 2 or 3 is false.

Before turning to consider possible responses to the inconsistency, it will help to ward off some preliminary worries.

First: As stated, FORBIDDEN IFF WORST encodes the assumption that options that are worst by even a very slight margin are forbidden. This may seem implausible. According to satisficing views, an option is permissible so long as it is good enough. And it might plausibly be thought that, in our case, even the worst option is good enough. According to other views, some reasons are not weighty enough, or not of the right kind, to generate (either moral or rational) requirements.² Our case, in which a child is deciding between self-regarding options with little at stake either way, might be thought to be just such a case. But these are incidental features of our case. Our puzzle does not arise only when the stakes are low or the worst option is good enough. It arises whenever the weight of a putative reward is greater than the difference in weight

² See, for example, Bedke (2011, pp. 130–131).

between the antecedent reasons for the forbidden option and the next best alternative. Thus our puzzle could re-emerge in a high-stakes case in which Nadia is choosing between two options favored by reasons quite different in weight and in which some third party offers a reward weightier than the difference between these reasons.³

Second: Our puzzle contrasts with puzzles or arguments that threaten or undermine ordinary normative judgments or practices. Singer's (1972) drowning-child argument threatens many people's judgments about the extent of our moral duties. Parfit's (1984) harmless-torturers cases seem to undermine our practices of holding people responsible only for the marginal difference made by their individual actions. To our eyes, our puzzle has no such clear first-order normative upshot. Very plausibly, most ordinary reward schemes are quite unlike the one in *The Pickle Popsicle*.⁴ But, as we shall next argue, the puzzle is nevertheless of theoretical interest. Its interest lies in what it tells us about the structure of the normative domain—the character of goods, reasons, and deontic status, and the relations among them.

3. What can give way?

At least one of 1-3 must go. But, considered individually, each seems plausible. In this section, we canvass the various options for avoiding or accounting for the inconsistency among them.⁵

The various ways of avoiding or accounting for the inconsistency can usefully be divided into two camps. According to one kind of response, what Frances says in some way *misfires*, in the sense that it leaves the prior deontic landscape unchanged. According to another kind of response, what Frances says does *not* misfire, in the sense that it induces some change in the prior deontic landscape. Here, we consider instances of these two kinds of responses, motivating each, and explaining which revisions in 1-3 each requires.

³ In §3.2.2 below, we consider the possibility of a disjunctive response that rejects FORBIDDEN IFF WORST, on different grounds, in low- and high-stakes cases, respectively.

⁴ Even an idiosyncratic parent like Frances, who wants to reward wrongful behavior, could avoid generating our puzzle simply by selecting an appropriately *small* reward. For any given forbidden action, there is, in principle, *some* good thing that could be used to reward it without causing the trouble that arises in *The Pickle Popsicle*. (Thanks to an anonymous reviewer for pressing us on this point.)

⁵ Helpful comments from reviewers, *inter alia*, brought to our attention Bradley's (2007) discussion of a related puzzle for a certain kind of desire-satisfaction theory of welfare. This theory (together with some plausible background assumptions) yields a contradiction when, in addition to some unsatisfied first-order desires, an agent has a sufficiently intense desire that her life go badly. Some of the proposals we consider in this section are analogous to proposals discussed by Bradley (see also Feldman (2004) and Skow (2009)). But, despite this, we take our puzzle to be of independent interest for at least three reasons. First, as we discuss below, our puzzle raises distinctive metaethical issues that merit discussion in their own right. Second, our puzzle suggests that, rather than a local problem for a particular version of desire-satisfactionism, Bradley's puzzle is an instance of a general class of puzzles about normativity. Third, and relatedly, if there is such a class of puzzles, then it seems reasonable to hope that considering our related puzzle might help to point toward a general diagnosis and solution.

(3.1) Misfire responses

Prior to Frances's offer, not flossing is forbidden and flossing is permissible. If what Frances says misfires, this is still the deontic profile of the case. And, since Frances is offering Nadia a pickle popsicle just on the condition that she does what's forbidden, this means that Nadia will receive a pickle popsicle if and only if she does not floss. But, given this offer, how is it possible that the deontic profile of the case is unchanged? There are three possibilities:

- (i) although getting the pickle popsicle is a reward—i.e., it would be good for Nadia to get it and that she'd get it is a reason for her to not floss—it is not a reason that changes the deontic status of either of her options; or
- (ii) although getting the pickle popsicle would be good for Nadia, getting the pickle popsicle is not a reward, because it is not a good that gives her a *reason* to not floss; or
- (iii) getting the pickle popsicle is not a reward, because it would not be *good* for Nadia to get it.

We motivate and discuss these three possibilities in turn.

(3.1.1) First misfire proposal: the reasons-deontic status connection

If Frances succeeds in giving Nadia a reason to not floss, how could it be that not flossing remains forbidden? First, it might be that, although the balance of *all* the reasons favors not flossing, there is some blockage between at least some of the reasons to not floss and this option's deontic status, such that not flossing remains forbidden. To accept this view would be to reject FORBIDDEN IFF WORST. Second, it might be that, when combined with the other reasons in the case, the balance of Nadia's reasons still favors flossing. To accept this view would be to reject REASONS ARE UNBALANCED—it would be to claim that the misfire is explained by some fact about how reasons interact, such that the previously forbidden option remains forbidden. Clearly, each of these explanations of the misfire is surprising and, we argue next, hard to motivate as a solution to our puzzle.

To motivate rejecting FORBIDDEN IFF WORST, one might follow Wodak's (2019) suggestion that some reasons are 'redundant', in the sense that, although they favor a required action, they do not—even partly—explain why the action is required. In particular, Wodak argues that reasons can be in this sense 'redundant' when they are grounded in the very fact that the action is required. Suppose, for example, that the gods offer you a reward if you do what you ought to do. That fact, Wodak claims, counts in favor of your doing what you ought to do, but it does not—even partly—explain why you ought to do whatever it is you ought to do. Generalizing Wodak's argument, one might propose that *whenever* a reason is partly grounded in the fact that an action has a given deontic status, that reason does not—even partly—explain why that action has the deontic status that it does.

On this reading of Wodak's view, an action's deontic status depends only on those reasons that are not grounded in the very fact that the action has the deontic status it does. So whether it's forbidden for Nadia to floss cannot be grounded—even in part—in the fact that not flossing would be rewarded, since that fact, in turn, depends on the fact that flossing is forbidden. FORBIDDEN IFF WORST, to the contrary, implies an

action is forbidden just in case the set of *all* the reasons that favor it is outweighed by the set of all the reasons that favor the action's complement. So, to accept this version of the misfire response is to reject FORBIDDEN IFF WORST.

In other work (ms), we raise doubts about the possibility of redundant reasons. But, whatever the cogency of Wodak's arguments, it is dubious whether they gain purchase in the present context. Wodak argues that a reward for doing what one ought to do is a redundant reason, on the basis that such a reward could not *make a difference* to which action the balance of reasons favors. Roughly speaking, Wodak's view is that such a reward necessarily leaves the deontic status of each of the agent's possible actions untouched because the reward is only attached to what there is most reason to do, *regardless of the reward* (i.e., to what the balance of reasons favors, even excluding the reward). But it is not at all clear that the fact that doing something wrongful would get Nadia a pickle popsicle is, in this sense, redundant. Even if it is correct that a reward for doing what one *ought* to do could at most 'reinforce' the case for doing this action, a reward for doing what one ought *not* to do—again, on the assumption that the pickle popsicle *is* a reason—does not appear to be similarly incapable of making a difference to the balance of reasons. Thus, it appears that the considerations that motivate Wodak's argument can't motivate a corresponding argument in the case at issue here.

Let us turn next to a different version of the misfire response according to which—although FORBIDDEN IFF WORST is correct and although Frances succeeds in giving Nadia a reason to not floss—the deontic landscape is unchanged because the balance of reasons still favors flossing. It is hard to see how one might motivate rejecting REASONS ARE UNBALANCED in this way. Of course, it is a familiar thought in the literature that there may be various kinds of 'interaction effects' among reasons.⁶ Thus, for example, some writers have argued for the following possibility: that one reason *p* is weightier than another reason *q* absent other morally relevant considerations, even though *q* is weightier than *p* given an additional reason *r*. To take a standard example, suppose that some explosion harms Boris moderately and Glinda mildly. Absent any other morally relevant facts, the reason to relieve Boris's suffering is weightier than the reason to relieve Glinda's (because Boris's is moderate and Glinda's is mild). But supposing that it was Boris who maliciously caused the explosion, it's plausible to think that the relative weights are reversed—that the reason to relieve Glinda's suffering is weightier than the reason to relieve Boris's.⁷ But, as this example suggests, if there are any such cases, there is some *explanation* why *r* interacts with *p* and *q* in a way that changes their relative weights—e.g., that the suffering of a more deserving person matters less. In *The Pickle Popsicle*, however, it is hard to see what explanation could be given for the claim that the pickle popsicle reward, when combined with an improvement to Nadia's mood, cannot together outweigh an improvement to Nadia's dental health.⁸ The only possibility that springs to mind is that rewards *as such* never have normative weight sufficient to change an act's deontic status. But clearly that's not right. You shouldn't break your promise to be home by 7:00. But if someone offers world peace for you to stay out

⁶ See Dancy (2004) for a canonical account.

⁷ Kagan (1988, p. 20).

⁸ It's especially difficult to see given that, on this reading, the reward fact would have to pull double duty: both counting in favor of not flossing *and* impacting the relative weights of f_1 and f_2 to ensure that (contra REASONS ARE UNBALANCED) f_1 is weightier than $\{R_{\text{not floss}}$ and $f_2\}$.

until 7:30, then surely that's what you ought to do. So rewards are capable of changing an act's deontic status, which makes it especially puzzling why a pickle popsicle reward couldn't in Nadia's case.

(3.1.2) *Second misfire proposal: the good-reasons connection*

The second possibility locates the source of the misfire, not in the connection between the reasons for an option and that option's deontic status, but rather in the connection between what's good and a person's reasons. There are two ways to motivate the view that Frances's offer of something good if Nadia does what's forbidden does not give Nadia a reason to do the forbidden option. Accepting either view corresponds to rejecting REWARD FOR WRONGFULNESS.

First, many philosophers endorse a deliberative constraint on reasons, maintaining that a given fact is a reason for an agent to perform some action only if that fact is eligible to serve as a premise in the agent's deliberation about what to do.⁹ But one might doubt that the fact that Nadia would get a pickle popsicle if she does something forbidden is eligible to play this role. As the proof in §2 demonstrates, the fact that she would get a pickle popsicle counts in favor of her flossing only if it doesn't, and it counts in favor of her not flossing only if it doesn't. So, with respect to each of her possible options, it might seem that there is no pattern of good reasoning that starts from the belief that she will get a pickle popsicle if she does what's forbidden and terminates in her performing that option (or intending to perform it or whatever is the conclusion of practical reasoning).¹⁰ If this is right, then there's a well-motivated story to tell about why Frances's offer of a pickle popsicle doesn't generate a reason and hence fails to constitute a genuine reward. And this is a plausible basis for rejecting REWARD FOR WRONGFULNESS.

But recall that this is a *misfire* response—a response to our puzzle which supposes that, after Frances makes her offer, the deontic statuses of Nadia's options remain unchanged. Given this, since not flossing remains forbidden, Nadia will receive a pickle popsicle just in case she doesn't floss. And, in this light, it's hard to see why the fact of Frances's offer couldn't play a role in Nadia's deliberation about what to do. Remember: Nadia *loves* pickle popsicles. So it seems perfectly intelligible that Nadia might factor the fact of Frances's reward into her decision about whether to (not) floss.¹¹

In addition, this response makes whether Frances's offer constitutes a reward depend in (what strikes us as) a bizarre way on just how good Frances's offered popsicle is. To illustrate, suppose that, rather than a pickle popsicle, Frances was offering an aggressively mediocre grape popsicle, such that any reason generated by the fact that Nadia would get a grape popsicle would be too weak to change the overall valence

⁹ See, among others, Snedegar (2019), Way (2017), Silverstein (2016), Schroeder (2007), and Setiya (2007).

¹⁰ There might be other ways of developing this kind of response to the puzzle. One might, for example, maintain that facts about the deontic status of an option over which one is deliberating can never serve as premises in (good) deliberation about what to do because they represent the *conclusion* of practical deliberation, and hence such deliberation would be (in some sense) circular. This strikes us as an implausibly strong version of the deliberative constraint. But, in any case, we take the criticisms that follow—perhaps with some modest variation—to apply to it.

¹¹ Notably, it's of no help to try to conceive of this as a *non-misfire* response. If we suppose the deontic statuses of Nadia's options *do* change, then this change calls out for an explanation. And, presumably, this explanation should involve some appeal to a change in Nadia's reasons. But the response to our puzzle currently under consideration is a response that attempts to *deny* that Frances's offer is reason-generating. So, it's hard to see how the deliberative constraint on reasons could be employed to motivate any kind of *non-misfire* response, either.

of her reasons, no matter which option it attached to. On this version of the case, the reasoning that previously appeared to show that the original offer couldn't generate a reason—specifically, that the fact of the offer was ineligible to play a role in deliberation because it would favor flossing only if it didn't and favor not flossing only if it didn't—wouldn't apply. So the advocate of this response to our puzzle appears committed to the claim that while the offer of the pickle popsicle doesn't generate a reason, the offer of the grape popsicle does. Whether the offer of some good generates a reason, then, would sometimes depend on *how* good it is relative to the other reasons in play. And that would be surprising indeed.

Finally, even if these concerns can be overcome, there are variations of our puzzle that simply don't depend on the generation of reasons. Suppose for instance, that Frances were to offer to make the world better only if Nadia does what's worst. By reasoning parallel to our proof in §2, it's worst for Nadia to (not) floss only if it's not and it's not worst for Nadia to (not) floss only if it is. So even if the deliberative constraint on reasons can be successfully utilized to show that certain types of offers aren't reason-generating (again, we're skeptical), it wouldn't dissolve a purely axiological version of the puzzle. So it seems that appealing to a deliberative constraint on reasons can be, at most, a partial solution.

A different way in which one might explain why the good-reasons connection misfires is by appealing to the *deontic* status of the action to which the good is attached. In support of this claim, consider yet another variant of the case:

The Wrongful Promise. Nadia is deciding between flossing (which would moderately improve her dental health) and not flossing (which would slightly improve her mood). Worried that she is becoming a bit of a goodie-two-shoes, she promises to her mother Frances that she will do something *wrongful* (that is, an action that's forbidden).

In this case, Nadia keeps her promise just in case she does an action that is forbidden for her to do. But one might claim that, although she has successfully promised to engage in wrongful behavior, and although it would be in one respect good were Nadia to keep her promise, Nadia cannot in this way give herself a *reason* to engage in wrongful behavior. And this thought does not seem wholly implausible. A promise to do what is wrongful might reasonably be claimed either to have no force at all or, at least, to generate a reason of very little weight. Absent the existence of some such restriction, we would (implausibly) have the normative power to alter the reasons we have by mere promissory fiat.¹² Some similar restriction, so one might add, must limit our ability to reward. If Nadia cannot give *herself* a reason to do what's forbidden merely by promising to Frances that that's what she will do, then Frances cannot give Nadia a reason to do what's forbidden merely by offering her a reward for doing so. Just as our normative powers to promise are in this way limited, the thought goes, so too are our normative powers to reward.

Once again, however, this response comes with significant costs. First, it is a substantive and controversial commitment to suggest that our power to promise (in a reason-implying way) is limited. Second, and more importantly, to whatever extent it seems plausible that a power to *promise* might be in this way limited, it seems heroic to try to extend this thought to the case of rewarding. While it might seem plausible that we cannot always alter the deontic profile of our situation merely by performing a certain kind of speech act, offering something good for the performance of some action is not merely performing

¹² See Smith's (1997) discussion and motivation of a principle which would have the effect of "prohibit[ing] individuals from imposing obligations on themselves to do what is independently morally objectionable" (158).

a speech act. Frances's offer means that Nadia will get something genuinely good for acting wrongly. And this fact seems to satisfy any plausible requirement for being a reason. It would, as just mentioned, be good for Nadia to get a pickle popsicle, she knows it would be good, and she's motivated by the fact that it would be good. So, beyond the controversial assumptions made here, it's not clear what else could disqualify Frances's offer from being a reward.

(3.1.3) *Third misfire proposal: the wrongfulness-good (mis)connection*

We just suggested that, if it would be good for Nadia to get the pickle popsicle (and she knows it and is motivated by it), it is hard to deny that the offer of a pickle popsicle is reason-generating. This suggests a third way out of the inconsistency: to deny that it actually *is* good for Nadia to get the pickle popsicle. The suggestion here is that what Frances says misfires because it is not possible to attach something (that might otherwise be) good to an action with this deontic status. To some, of course, this will seem an exotic suggestion. To motivate it, recall the thought that, in *The Wrongful Promise*, there is a misfire because of the *content* of the promise: that it is a promise to do something *wrongful*. If one has this thought, a natural way of vindicating it is to propose that Nadia's keeping *this* promise would not be in any way *good*. And then, transposed to *The Pickle Popsicle*, the corresponding thought is that it would not be in any way *good* for Nadia to get the pickle popsicle, *because* it is attached to performing a forbidden action. To accept this view would again be to deny that the pickle popsicle qualifies as a reward, but this time on the grounds that the putative reward isn't good (rather than that it doesn't generate a reason).

Again, however, even if it is true that it's not good to keep promises to do what's forbidden (or even if such 'promises' are not promises), it is harder to see how to make sense of this thought in the context of rewarding. Since Nadia loves pickle popsicles, it's difficult to deny that her life is improved by getting one. One might attempt to motivate this response by accepting a hybrid or moralized view about individual welfare, according to which, very roughly, one's welfare is improved only if one takes pleasure in what is (or is appropriately related to) something that is in itself good.¹³ Nadia's enjoying this pickle popsicle—one that she gets only if she does something *wrongful*—might be thought to fail to satisfy this condition. And so, one might propose, Frances cannot successfully offer it to Nadia as a *reward*, since Frances cannot, of course, change the facts about what is and is not prudentially valuable. But this escape route, if intriguing, is far from costless. Quite apart from the substantive commitment to this view of individual welfare, it is questionable whether such a view can plausibly be made to deliver the desired verdict in our case. (Nadia would not be taking pleasure *in* doing wrong, and the pickle popsicle is not in *itself* bad.) Moreover, it is questionable whether this proposal is appropriately general. There is a mirror-image version of our puzzle in which Frances proposes to *punish* Nadia for doing what's *required*. Extending the present proposal to that case would commit us to accepting that whatever hard treatment Frances proposes to mete out would not make Nadia's life go worse. And that is very hard to accept.

¹³ See Adams (2002), Feldman (2004), and Kagan (2009). Would such a view entail that getting the pickle popsicle is *bad* for Nadia? Not necessarily. For rich discussion of the various ways in which such views of welfare might be extended to the case of illfare, see Kagan (n.d.).

(3.1.4) *Taking stock*

Insisting that what Frances says *misfires*—that the deontic profile of Nadia’s options before and after the offer of the pickle popsicle are the same—might at first appear to be an attractive way out of the puzzle. But this one way turns out to be at least three, and each comes at substantial cost because of its surprising implications. To insist that Frances’s offer constitutes a genuine reward is either to insist that some reasons (like the one generated by Frances’s offer) can’t make a difference to what one ought to do (but without a good explanation as to why), or to accept an apparently unmotivated view about how Frances’s reward interacts with Nadia’s other reasons. To insist that Frances’s offer doesn’t constitute a reward is either to deny a very plausible connection between what’s good and one’s reasons, or to maintain that otherwise good things sometimes aren’t when they’re entangled in an attempt to reward. So appealing to a misfire isn’t as attractive as it may have first appeared. Maybe, then, there’s a way out to be found by claiming that what Frances says *doesn’t* misfire.

(3.2) *Non-misfire responses*

Each of the misfire responses carries some cost or has surprising implications. And this motivates investigating alternative responses on which what Frances says does not wholly misfire but, instead, in some way alters the deontic profile of the case. Recall that, prior to Frances’s offer, not flossing is forbidden (and flossing not forbidden) for Nadia. We can categorize different kinds of non-misfire responses by what they take the resulting deontic profile of the case to be.

As we explain below, some versions of the non-misfire response have it that Frances’s offer induces quite radical alterations in the prior normative landscape. But let us begin with the less radical versions of the response.

(3.2.1) *First non-misfire proposal: Contextualism*

To start, one might attempt to avoid our puzzle by endorsing a contextualist reading of ‘forbidden’ on which to say that some option is ‘forbidden’ is to say that the reasons to not perform that option are weightier than the reasons to perform it [*in view of some subset of reasons*]. According to contextualism, generically construed, what it means for an option to be ‘forbidden’ is determined in part by at least one contextually set parameter—a *modal base* which determines a relevant set of propositions, and/or an *ordering source* which ranks the possible worlds consistent with the modal base—such that ‘an agent is forbidden to ϕ ’ is true if and only if the agent ϕ s in none of the top ranked worlds, according to the ordering source, consistent with the modal base.¹⁴ The contextualist, then, might propose to solve our puzzle by maintaining that, given the context, the modal base of ‘forbidden’ in Frances’s offer does not include the fact that Nadia will receive a pickle popsicle for doing what’s forbidden. And, indeed, there’s a plausible reading of *The Pickle Popsicle* on which what Frances means when she tells Nadia “You’ll get a pickle popsicle just on the condition that you do what’s forbidden” is that Nadia will get a pickle popsicle just on the condition that she does what’s

¹⁴ Contextualism, so characterized, is inspired by Kratzer’s (2012) work on the semantics of modals. Our gloss here follows Worsnip (2019), who helpfully distinguishes “generic contextualism” from the specific versions of the view that appear in the metaethical literature. Since the contextually determined parameters are rarely explicit, we’ll place them in brackets and precede them—following Kratzer—with “in view of”. (We’re grateful to an anonymous reviewer for pressing us to discuss how contextualists might respond to our puzzle.)

forbidden [in view of the facts independent of this very offer].¹⁵ The relevant facts independent of Frances's offer are just the facts that flossing would moderately improve Nadia's dental health (f_1) and that not flossing would slightly improve Nadia's mood (f_2). And since f_1 is weightier than f_2 (as stated in REASONS ARE UNBALANCED), and since it's forbidden for Nadia to ϕ just in case the reasons to not ϕ are weightier than the reasons to ϕ (as stated by FORBIDDEN IFF WORST), what it's forbidden for Nadia to do [in view of the facts independent of Frances's offer] is not floss. So, read this way, Frances's offer amounts to an offer to give Nadia a pickle popsicle for *not flossing*. And, the contextualist can maintain, this fact is a reason for Nadia to not floss that makes it the case that Nadia is forbidden to floss [in view of *all* the facts]. The contextualist, then, can attempt to dissolve our puzzle by maintaining that our argument equivocates between two distinct meanings of 'forbidden'—forbidden [in view of the facts independent of Frances's offer] in REWARD FOR WRONGFULNESS and as meant by Frances, and forbidden [in view of *all* the facts] in FORBIDDEN IFF WORST.

This is, we think, an initially attractive approach to solving our puzzle. The contextualist's interpretation of 'forbidden' is not implausible, and it provides the resources for identifying distinct meanings of 'forbidden' at play in our argument. But note that endorsing contextualism—generically construed—is not enough to solve our puzzle. While there are indeed *some* interpretations of Frances's offer on which our puzzle does not arise (like the one mentioned in the previous paragraph), in order to escape our puzzle, the contextualist must maintain that there could be *no version* of our case in which Frances offers Nadia a pickle popsicle for doing what's forbidden [in view of *all* the facts]. But this, we think, would come as a surprise to Frances. Again—being concerned about Nadia becoming a bit of a goodie-two-shoes—Frances plausibly might *intend* to reward Nadia for doing what's forbidden in exactly this all-things-considered sense. And, unsure of her meaning, we can imagine pressing Frances: "Do you mean 'forbidden *independently of your offer*'?" And she might well respond: "No; I want Nadia to do what's forbidden in light of *all* her reasons, and that's exactly what I'll reward her for doing." With these details filled in, it's difficult to see how a contextualist parameter could exclude the fact of Frances's offer from the modal base of 'forbidden' as meant by Frances.

Importantly, this isn't to reject contextualism, and it's not to deny that the contextualist has resources to explain how, in many quotidian versions of the case, Frances's offer *could* generate a reason in an unproblematic way.¹⁶ Rather, it's to cast doubt on the claim that the contextualist can flesh out the character of 'forbidden' in such a way that, when uttered by Frances, its modal base *necessarily* excludes the fact of her offer. And unless the contextualist can do this, our puzzle remains unsolved. So, for a contextualist solution to our puzzle to succeed, more would have to be said about exactly how the modal base of 'forbidden' is contextually determined in such a way that it *invariably* excludes Frances's offer. Perhaps

¹⁵ Or, alternatively, [in view of the facts that obtained *before* this very offer], or any other modal base that excludes the fact of Frances's offer.

¹⁶ In precisely this way, the contextualist gives us a straightforward way to understand everyday offers of rewards that *do not* generate our puzzle. But our project doesn't require our puzzle to arise in any *actual* instance of reward-offering—only in a *possible* one. And the point here is that we don't see how the contextualist can show that our preferred reading of Frances's offer is *impossible*.

such a story can be told.¹⁷ But, even if it can, it would constitute what we take to be a surprising restriction on our power to reward by placing a surprising restriction what it's possible for us to mean when we say that an act is forbidden in the context of a reward.

(3.2.2) *Second non-misfire proposal: everything forbidden or everything permissible*

Next, one might propose that Frances's offer renders both flossing and not flossing forbidden. To accept this view would be to reject either FORBIDDEN IFF WORST or REASONS ARE UNBALANCED. And this response may not seem wholly implausible. After all, there are arguably independent reasons to think that agents can face irresolvable moral dilemmas. However, the typical motivations for believing that a case is an irresolvable dilemma—that whatever one does sacrifices an important value—do not seem to be present in *The Pickle Popsicle*. Unlike the kinds of cases that standardly motivate positing dilemmas, Nadia's choice is not very morally weighty, the options need not be equally significant, they do not differ importantly in kind, and opting for one does not involve an important sacrifice of a competing weighty value (moral or otherwise). So, to avoid the inconsistency by appealing to dilemmas is to expand the range of cases in which dilemmas might arise. Thus there is some cost, beyond that typically associated with positing dilemmas, in avoiding the inconsistency in this way.

One might reply that this proposal could be combined with another thought canvassed in §2—namely, that in low-stakes variants of the puzzle there are no rational (or other normative) requirements at all—to yield an attractive disjunctive response to the puzzle. By adding an independently-motivated solution to 'low-stakes' versions of the puzzle, and positing dilemmas only in 'high-stakes' versions, this disjunctive response might be thought not to force us to greatly expand the range of cases in which dilemmas arise. But we remain, for two reasons, skeptical that this is an attractive way out of the puzzle. First, even in 'high-stakes' versions of the puzzle, we would lack a plausible explanation for *why* Frances's offer should be thought to induce a dilemma. Consider, for instance, a version of the case in which the only anterior considerations at stake are a very intense pleasure that someone would derive from Nadia's ϕ ing and a somewhat less intense pleasure that someone would derive from Nadia's not ϕ ing, and suppose that the reward on offer is weightier than the difference in the weights of the reasons given by these pleasures. Given this set-up and the assumptions in §2, we would again have an apparent contradiction. The present proposal would avoid the contradiction by positing a dilemma. But it cannot be the weighing of the two significant pleasures, alone, that explains the dilemma (or else we would, once again, have the objectionable cost of having greatly expanded the range of cases in which dilemmas arise). So it must be Frances's offer, in combination with the other normative considerations, that does so. But if Nadia would not otherwise face a dilemma, and if the reward on offer does not differ in kind from the anterior considerations, it is hard to

¹⁷ In fact, to tell such a story, one might draw on some of the other solutions to our puzzle canvassed above and below to explain why Frances can't mean what, on its face, it seems she does (or at least might) mean. Contextualism's ability to—in this way—*pair* with other solutions to our puzzle is, we think, a point in its favor. But it also emphasizes 1) the extent to which contextualism is, all by itself, an at best incomplete solution to our puzzle, and 2) the ways in which the theoretical upshots of a contextualist solution very much depend on *why* the meaning of 'forbidden', as uttered by Frances, is restricted in the way that the contextualist insists that it is.

understand how Frances's offer of the reward could have induced it.¹⁸ Explaining away the puzzle about *The Pickle Popsicle* in this way thus seems undermotivated.

One might instead propose that Frances's offer renders both options not forbidden. (On this view, Frances has simply offered Nadia something good for doing something that it is not possible for Nadia to do.) However, this view is on reflection hard to motivate. Suppose that the fact that Nadia will get a pickle popsicle for acting wrongly is neither a reason to floss nor a reason to not floss. To have rendered *both* options not forbidden, Frances's offer must somehow have either rendered f_1 and f_2 not reasons or altered their weights (such that REASONS ARE UNBALANCED is false), or it must somehow have severed the connection between the weights of reasons and an option's deontic status (such that FORBIDDEN IFF WORST is false). But none of these implications can easily be motivated. It's very difficult to see why, in the presence of a *non*-reason like the fact of Frances's offer, *either* the individual weights of f_1 and f_2 *or* their status as reasons *or* the connection between f_1 and f_2 and the deontic status of flossing should be thought to vary.¹⁹

We have just suggested that it is difficult to motivate the 'everything is permissible' solution, *given* the assumption that the (only) reasons that bear on Nadia's choice are f_1 , f_2 , and the reason generated by Frances's offer. But perhaps this assumption is questionable. To see how it might be questioned, consider an analogy between our puzzle and a paradox discussed by Fraser and Hawthorne (2015). Nostradamus believes at 11.59pm that nothing Arthur consciously believes at midnight will be true (call this proposition F), and tells Arthur F. On this basis, Arthur deduces and consciously believes at midnight the following disjunction: either F or Trump won the popular vote in 2020 (call this proposition T). These assumptions apparently yield a contradiction.²⁰ So it appears that, as Prior (1961), suggests:

[W]e must just accept the fact that thinking, fearing, etc., because they are attitudes in which we put ourselves in relation to the real world, must from time to time be oddly blocked by factors in that world, and we must just let Logic teach us where these blockages lie.

¹⁸ To illustrate our skepticism on this point, suppose that in this 'high-stakes' version of the case both options are forbidden. Then the prospect of getting a pickle popsicle favors each option. Either the set of reasons favoring flossing is weightier than the set of reasons favoring not flossing (and so this response amounts to rejecting FORBIDDEN IFF WORST) or it is not (and so this response amounts to rejecting REASONS ARE UNBALANCED). Suppose the former possibility. Then, given the nature of the considerations that bear on Nadia's choice and the fact that one option is more favored than the other, it is puzzling how Nadia could be thought to face a dilemma at all. Suppose now the latter possibility. Then, given the stipulated individual weights of the considerations that bear on Nadia's decision, it is puzzling how it could be the case that neither option is more favored than the other. In either case, it is hard to see how to motivate the dilemma response.

¹⁹ Each of these views recalls Temkin's (2012) thesis that reasons are sometimes 'essentially comparative'. But, again, whether or not this highly controversial thesis is correct, the considerations that Temkin takes to motivate the thesis do not appear to be present in the present case.

²⁰ Since T is false, the disjunction F or T is true just in case F is true. But if F is true, then since Arthur consciously believes at the midnight the disjunction F or T, and nothing he consciously believes at midnight is true, F is false. On the other hand, if F is not true, then something Arthur consciously believes at midnight is true; but at midnight he consciously believes (only) the disjunction F or T, so F is true.

But, as Fraser and Hawthorne (2015) note:

Logic can only teach us so much. Logic tells us that nothing conforms to our setup, and that nothing *could* conform to our setup. But it does not tell us what would happen *instead*, were we to try to bring about the impossible situation.

So what *would* happen? Fraser and Hawthorne entertain two possibilities. According to the *addition* proposal: despite appearances, at midnight Arthur consciously believes some proposition in addition to the disjunction F or T. Alternatively, according to the *subtraction* proposal: despite appearances, at midnight Arthur does not succeed in believing even this disjunction (but instead stands in some other relation to it, or believes some proposition other than it).²¹ Either proposal would block the reasoning, sketched in footnote 20, that leads to the contradiction.

By appealing to an analogy with Fraser and Hawthorne's case, one might propose analogous responses to our puzzle. In Fraser and Hawthorne's case, the addition and subtraction proposals avoid contradiction arising by blocking Arthur's coming to have (exactly) *one* conscious belief at midnight. The analogous responses to our puzzle would avoid the contradiction by blocking Nadia's being given (exactly) *one* new reason. In the spirit of the addition proposal: Perhaps, in making her offer, Frances also makes it the case that Nadia has some further reason M to not floss, such that M is exactly as weighty as the difference in weight between f_1 and f_2 , thereby rendering both flossing and not flossing permissible. Since, on this way of responding to the case, neither option would be forbidden, neither option would net Nadia the pickle popsicle. So contradiction would be avoided.²² Alternatively, in the spirit of the subtraction proposal: Perhaps Frances's offer also makes it the case that Nadia has *no* reason to floss or not to floss (i.e., the offer nullifies the weights of f_1 and f_2), thereby again rendering both flossing and not flossing permissible. Again, on this way of responding to the case, since neither option would be forbidden, the prospect of getting a pickle popsicle would fail to be a reason, and contradiction would be avoided.

One of these proposals may be worth taking seriously. But, quite plainly, they are very unintuitive. And, importantly, they are even *more* unintuitive in our normative context than in Fraser and Hawthorne's case. Although it would be very surprising, it's at least conceivable that Arthur stands in some subtly different relation to the disjunction F or T than belief, or that he has an 'illusion of thought' when he entertains it, or that he necessarily comes to believe some proposition in addition to it.²³ But in our case, in

²¹ As Fraser and Hawthorne also note, there are more abstruse possibilities that would be consistent with such a 'blockage'. One might suggest that, in any world in which Nostradamus tries to tell Arthur F, his tongue freezes or he slips on a banana peel. Or one might suggest that, even in worlds in which Arthur has the words that express F running through his mind, he simply does not succeed in consciously *believing* F. Transposing the former proposal to our case, one might suggest that, in any world in which Frances tries to make her offer to Nadia, her tongue freezes or she slips. Or, transposing the latter proposal to our case, one might suggest that the pickle popsicle would simply not be something Nadia would enjoy or that it would be something she would enjoy but would simply not be a reason. (These versions of the misfire proposal differ from those considered in §3.1, insofar as they each take the relevant misfire to be a brute fact, not admitting of further explanation.) We join Fraser and Hawthorne in finding these kinds of misfire proposals quite unsatisfactory, and hence we won't address them further.

²² This is an unusual sense of *addition*, since it has the effect of making what *would* otherwise have been a reason (the pickle popsicle) not a reason. See Fraser and Hawthorne (2015, p. 172) on analogous difficulties in spelling out the addition proposal, in their context, in a way that successfully avoids contradiction.

²³ For discussion of the prospects of making good on these claims, see Fraser and Hawthorne (2015, pp. 171–173).

virtue of what could Frances's offer generate the mystery reason *M*? And in virtue of what would *M*, whatever it is, *be* a reason for Nadia? (Would it, like the pickle popsicle, correspond to a fact about Nadia getting something she would enjoy? But in virtue of what would it be true that she *would* get that thing, whatever it is? What Frances says does not appear to make it true that she would get anything *other* than a pickle popsicle.) Likewise, in virtue of what could it be true that Frances's offer makes Nadia's improved dental health cease to be a reason for Nadia to floss? We find it hard to see how any of these questions could be given satisfactory answers. It might be that logic teaches us that there *must* be, in this kind of case, some such addition to or subtraction from Nadia's reasons. But, if so, this would force us to revise our understanding of the normative domain in ways at least as surprising as any of the other responses we discuss.

So, we conclude, neither of the misfire options discussed in this section is a costless way out of the puzzle.

(3.2.3) *Third non-misfire proposal: the analogy with other paradoxes*

By this point, a different thought will have occurred to many readers. Thus far, we have been searching for a particular *kind* of solution to the puzzle. We have been trying to find some consideration local to the normative domain that will yield a way out of the puzzle—e.g., by providing a basis for thinking that what Frances says does not generate a reason or that Nadia's getting a pickle popsicle isn't good.²⁴ But perhaps this is the wrong kind of consideration to have been looking for. One might suggest instead that our puzzle is an instance of a more general paradox—not peculiar to the normative domain—and that the solution to our puzzle will follow as a straightforward corollary of whatever is the correct solution to that more general paradox.

To motivate this idea, consider Russell's well-known paradox.²⁵ It is at first natural to think that, given any intelligible condition *C*, there is a set *A* whose members *x* are all and only those things that satisfy *C*—i.e.,:

$$(\dagger) \exists A \forall x (x \in A \leftrightarrow C)$$

But Russell showed that this yields a contradiction when *C* is chosen to be a set that is not a member of itself— $x \notin x$ —since (\dagger) then seems to entail:

$$\exists A (A \in A \leftrightarrow A \notin A)$$

To see the suggestive parallel with our puzzle, let us recharacterize REWARD FOR WRONGFULNESS. Let *x* range over Nadia's possible actions—i.e., flossing and not flossing—and let *A* be the set of actions in whose favor getting the pickle popsicle counts. Then, where *F_x* is 'Nadia is forbidden to *x*', REWARD FOR WRONGFULNESS can equivalently be written:

²⁴ Fraser and Hawthorne (2015, pp. 174–175 n2) suggest a similar distinction between solutions to a puzzle that are 'internal' to the domain in which it arises and solutions which are not.

²⁵ One might alternatively suggest analogies with other paradoxes (e.g., the liar paradox). We focus on Russell's paradox in part because of some particularly instructive parallels with our case. We are grateful to a reviewer for the suggestion.

(1') $\exists A \forall x (x \in A \leftrightarrow Fx)$

In words, 1' says that there is a set A, comprising the actions in whose favor getting the pickle popsicle counts, such that an action is a member of that set just in case it's an action that meets a certain condition—namely, being forbidden. As in Russell's case, 1' follows from the more general (†). And, as in Russell's case, 1' entails a contradiction.²⁶ Russell showed that a naïve assumption about unrestricted set formation yielded a contradiction; we have shown that a naïve assumption about an unrestricted power to reward yields a contradiction. Given the similarity between the puzzles, one might thus reasonably expect that the correct solution to Russell's paradox can simply be 'plugged in' to solve our puzzle.

So far as it goes, we find this suggestion promising. But we think there is no simple route from noting the analogy with Russell's paradox to a solution to our puzzle: what strike us as the most promising solutions in this vein all have non-trivial metaethical upshots. We offer three remarks in support of these claims.

First: Some analogues of Russell's paradox are mere pseudo-paradoxes. Analogous reasoning, for example, yields the conclusion that there cannot be a barber who shaves all and only those who do not shave themselves. But this is not surprising—there is no antecedent reason to think that there *is* such a barber. So the fact that there can't be such a barber does not call out for explanation. Russell's case, by contrast, is a genuine paradox because there *is* antecedent reason to think that there is a (possibly empty) set corresponding to those things that meet any given intelligible condition. Like Russell's case and unlike the barber case, there *is* antecedent reason to think that you can use a reward to sweeten an option that meets any given intelligible condition. So ours is, we believe, a genuine puzzle. If it turns out that there is not the ability to reward in this way, we will need an explanation why not.

Second: Some responses to Russell's paradox, when transposed to the normative context, would yield either surprising revisions to our understanding of the normative domain or our judgments about the effects of France's offer or both. Some non-classical responses to Russell's paradox, for example, propose allowing for propositions to be both true and false or neither true nor false.²⁷ Applying these solutions to the normative version of the paradox, we take it, would mean drastically revising our understanding of the normative domain—such that Nadia's flossing (or not flossing) is either both forbidden and not forbidden or neither forbidden nor not forbidden. So if our puzzle is a version of Russell's paradox, and this (or something like it) is the way out of Russell's paradox, then our puzzle has some surprising metaethical implications.²⁸

²⁶ The proof is a simple variant of the one given in §2. Given 1', along with REASONS ARE UNBALANCED and FORBIDDEN IFF WORST, Fx holds just in case $x \notin A$. So 1' entails $\exists A \forall x (x \in A \leftrightarrow x \notin A)$. And when instantiated with respect to either flossing or not flossing, this yields a contradiction.

²⁷ Analogous responses to the liar paradox have it that the truth predicate has gaps or gluts (i.e., is such that there are sentences that neither fall in its extension or anti-extension (gaps), or that fall in both its extension and its anti-extension (gluts)). Pursuing instead the analogy with the liar paradox, one might propose that the '... is forbidden' predicate likewise has gaps or gluts. Similar remarks apply to these responses.

²⁸ These implications are so surprising, in fact, that they count against the plausibility of this kind of solution to our puzzle. If neither flossing nor not flossing is in *either* the extension or anti-extension of the '... is forbidden' predicate,

Third: One might propose, however, that at least one solution to Russell's paradox—namely, Russell's own favored solution—would yield a satisfactory solution to our puzzle. Russell's diagnosis was that his paradox arose from a vicious circularity in the specification of the problem set. The solution, then, is to prohibit this kind of vicious circularity in specifying a set. That is what Russell's vicious circle principle does: it claims that there is no totality containing a member that can be specified only in terms of that very totality. Our puzzle arguably exhibits a similar kind of circularity. Consider the set S of facts that is the union of the set of reasons to floss (S_{ϕ}) and the set of reasons not to floss ($S_{\neg\phi}$). If REWARD FOR WRONGFULNESS were true, then one member of S (namely, the reason generated by Frances's offer of a pickle popsicle) could be specified only by way of a property of S itself (namely, whether one of S 's subsets, S_{ϕ} , is normatively weightier than another, $S_{\neg\phi}$). The spirit, if not the letter, of Russell's vicious circle principle rules out the existence of such a totality. So REWARD FOR WRONGFULNESS is not true.²⁹

We are sympathetic to this proposal. But we think that, quite apart from familiar problems of motivating the vicious circle principle,³⁰ its application in this context—much like the non-classical responses canvassed above—leaves much to be explained about how to interpret the normative upshot of this response. On one interpretation, the vicious circle principle rules out the existence of the set S once Frances makes Nadia her offer. Frances's offer would thereby make it the case that there *is* no set of reasons (and so there are no reasons) either to floss or not floss. Frances's offer would have the effect of inducing a kind of normative neutrality or paralysis. But it's difficult to take this possibility seriously—the reward hanging, Damocles-like, over Nadia, rendering her choice situation normatively paralyzed or inert.³¹

On the alternative interpretation of this response, what the vicious circle principle rules out is only the existence of a set S that contains the reason generated by Frances's offer. But no vicious circularity arises when S is taken not to include this reason. So, on this interpretation, what Frances says would entirely misfire; no one, including her, has the power to reward unrestrictedly in this way. But it is worth noting that the normative upshot of this proposal, too, is very surprising. For assume this solution is correct. Frances offers Nadia the pickle popsicle if and only if she does what's forbidden. But in so doing, she fails

then neither action is forbidden, and so the prospect of a pickle popsicle fails to favor either of Nadia's options because no action is forbidden. But if Nadia won't get a pickle popsicle no matter what she does, in virtue of what does Frances's offer alter the deontic landscape? In this sense, this solution makes Frances's offer a non-misfire—it induces a surprising change in the deontic landscape—and, moreover, one that cries out for further explanation.

²⁹ The vicious-circularity diagnosis may make it appear that our puzzle does not concern *wrongfulness* in particular. Suppose that, in our case, Frances had instead offered Nadia a pickle popsicle just in case Nadia does what is favored by an odd number of reasons. It may seem that a similar puzzle would then arise (see Fraser and Hawthorne (2015, p. 176 n15)). Perhaps this is so. (And, if it is so, it would be count against those responses, considered above in §3.1.2 and §3.1.3, that rest on the fact that Nadia is offered a reward to do what is *wrongful*.) But this puzzle, if it is one, would arise only given a particular view about how reasons are to be individuated. Our puzzle, by contrast, arises so long as actions can have the property of being forbidden.

³⁰ See Irvine and Deutsch (2020) and the references therein.

³¹ Similar remarks apply to the related thought that one might respond to our puzzle by appeal to a 'ramified' view of reasons or oughts, analogous to Russell's theory of types. (On this view, a 'type-0' ought would correspond to the balance of 'type-0' reasons; to 'type-1' reasons, namely those which depend on 'level-0' normative facts, would correspond a 'type-1' ought; and so on.) See Fraser and Hawthorne's (2015, pp. 174–175) remarks about how an analogous way out of their puzzle would imply implausibly radical revisions to our understanding of the epistemic domain.

to give Nadia a reward.³² Nevertheless, we can imagine Nadia thinking: ‘How lovely, to get a pickle popsicle’, and asking us which option is forbidden. Since Frances fails to give Nadia a new reason, the answer, we correctly inform her, is that it’s forbidden to not floss. ‘Great’, she replies, ‘then I will do that, so that I can get the pickle popsicle.’ We may insist that she is making a mistake here—that the pickle popsicle simply *cannot* be a reward. But when Nadia *gets* and *enjoys* the pickle popsicle because (as Frances truly believes) she *did* do what was forbidden, Frances and Nadia might justifiably look askance at our insistence that no *rewarding* could possibly have happened. What else, they might reasonably ask, would need to be added, to make this scenario an instance of rewarding?

4. Conclusion

In §2, we argued for our central claim: that there is a puzzle about rewarding.

THE REWARD PUZZLE. Either the power to reward is not unrestricted, or at least one other very plausible principle about normativity is false.

The puzzle forces us to make some unpalatable choices. In §3, we suggested that there are two kinds of response to the puzzle. According to *misfire* responses, there are normative, metaethical, or extra-normative grounds for thinking that an apparent good offered for acting wrongly sometimes either isn’t actually good, doesn’t generate a reason, or doesn’t generate a reason capable of altering the deontic landscape. To accept a misfire response to the case, then, would be to accept a surprising *limitation* on our normative powers. By contrast, according to *non-misfire responses*, there are normative, metaethical, or extra-normative grounds for thinking that certain rewards induce surprising changes in the normative landscape (e.g., normative indeterminacy, dilemmas, or changes in the weight or presence of other reasons). To accept a non-misfire response to the case, then, would be to accept a surprisingly *expansive* view of our normative powers. We have not attempted to offer a decisive argument in favor of one of these kinds of response. We have, however, suggested that some *versions* of each kind of response are more or less palatable than others. But even the responses to the puzzle that we take to be *least* unpalatable have their costs. There is no way out of the puzzle that does not involve some surprising revision of our understanding of the normative domain. That, we take it, is the interest of the puzzle.

³² One might of course try to *explain* this fact by appeal to one of the solutions canvassed in §3.1. But we won’t rehearse the costs of those responses here.

References

- Adams, R. M. (2002). *Finite and Infinite Goods*. Oxford University Press.
- Bedke, M. (2011). Passing the Deontic Buck*. In *Oxford Studies in Metaethics* (pp. 128–152). Oxford University Press.
- Bradley, B. (2007). A paradox for some theories of welfare. *Philosophical Studies*, 133(1), 45–53.
- Dancy, J. (2004). Ethics Without Principles. In *Ethics Without Principles*. Oxford University Press.
- Elizabeth Fraser, R., & Hawthorne, J. (2015). Cretan Deductions. *Philosophical Perspectives*, 29(1), 163–178.
- Feldman, F. (2004). *Pleasure and the Good Life*. Oxford University Press.
- Irvine, A. D., & Deutsch, H. (2020). Russell’s Paradox. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2020). <https://plato.stanford.edu/archives/win2020/entries/russell-paradox/>
- Kagan, S. (n.d.). *What is the Opposite of Well-Being?*
- Kagan, S. (1988). The Additive Fallacy. *Ethics*, 99(1), 5–31.
- Kagan, S. (2009). Well-being as Enjoying the Good. *Philosophical Perspectives*, 23(1), 253–272.
- Kratzer. (2012). What “must” and “can” must and can mean. In *Modals and Conditionals* (pp. 1–20). Oxford University Press.
- Parfit, D. (1984). *Reasons and Persons*. Oxford University Press.
- Prior, A. N. (1961). On a family of paradoxes. *Notre Dame Journal of Formal Logic*, 2(1), 16–32.
- Schroeder, M. (2007). *Slaves of the Passions*. Oxford University Press.
- Setiya, K. (2007). Reasons without Rationalism. In *Reasons Without Rationalism*. Princeton University Press.
- Silverstein, M. (2016). Reducing Reasons. *Journal of Ethics and Social Philosophy*, 10(1), 1–23.
- Singer, P. (1972). Famine , Affluence , and Morality. *Philosophy and Public Affairs*, 1(3), 229–243.
- Skow, B. (2009). Preferentism and the Paradox of Desire. *Journal of Ethics and Social Philosophy*, 3(3), 1–17.
- Smith, H. M. (1997). A Paradox of Promising. *The Philosophical Review*, 106(2), 153.
- Snedegar, J. (2019). Deliberation, Reasons, and Alternatives. *Pacific Philosophical Quarterly*, 100(3), 682–702.
- Temkin, L. S. (2012). *Rethinking the Good: Moral Ideals and the Nature of Practical Reasoning*. Oxford University Press.
- Way, J. (2017). Reasons as Premises of Good Reasoning. *Pacific Philosophical Quarterly*, 98(2), 251–270.
- Wodak, D. (2019). Redundant Reasons. *Australasian Journal of Philosophy*, 1–13.

Worsnip, A. (2019). 'Ought'-contextualism beyond the parochial. *Philosophical Studies*, 176(11), 3099–3119.