

JOHANN FRICK

What We Owe to Hypocrites: Contractualism and the Speaker-Relativity of Justification

I. THE SPEAKER-RELATIVITY OF JUSTIFICATION

It is a familiar feature of our ordinary ethical discourse that the strength of a moral argument can depend on more than its content. When someone seeks to convince us to perform some action or to accept some policy, the appropriate response may be a function not just of the propositional content of the reasoning, but also of *who* is addressing the argument to us. In such a context of interpersonal justification, the justificatory force of a moral argument sometimes depends on the *identity* of the speaker.

Despite its importance to everyday practice, this *speaker-relativity of interpersonal moral justification* has received scant attention in contemporary moral philosophy. To the best of my knowledge, the only philosopher to have explored the phenomenon in any detail is G. A. Cohen in his Tanner Lectures, “Incentives, Inequality, and Community.”¹

For their feedback on earlier drafts of this article, I thank Eric Beerbohm, Joseph Carlsmith, Ben Ewing, Liz Harman, Louis-Philippe Hodgson, Mark Johnston, Frances Kamm, Adam Kern, Rahul Kumar, Trevor Latimer, Doug Lavin, Sarah-Jane Leslie, Tienmu Ma, Sarah McGrath, Richard Moran, Yascha Mounk, Ekédi Mpondo-Dika, Derek Parfit, Theron Pummer, Michael Rosen, Tim Scanlon, Pablo Stafforini, Lucas Stanczyk, Jonny Thakkar, Jay Wallace, Alec Walen, and Monique Wonderly, as well as the editors of *Philosophy & Public Affairs*. I would also like to thank audiences at Harvard University, Princeton University, Swarthmore College, the University of Oxford, and Yale University.

1. G. A. Cohen, “Incentives, Inequality, and Community,” Tanner Lectures in Human Values, delivered at Stanford University, May 21–23 (1991), reprinted as G. A. Cohen, *Rescuing Justice and Equality* (Cambridge, Mass.: Harvard University Press, 2008), chap. 1.

Cohen observes that a moral argument that passes muster when presented impersonally or in a suitable dialogical setting may entirely lose its force when transposed to a different interpersonal context. The reasons for this phenomenon are various. Consider the following four pieces of moral dialogue:

- (I) *B to A*: “We should vote C ‘coworker of the month’ for being so modest and self-effacing.”

Imagine, instead, *C himself* saying to A and B: “You should vote me ‘coworker of the month’ for being so modest and self-effacing.”

- (II) *B, to the host A*: “Please don’t be offended by the behavior of my friend C. He doesn’t know that in your culture it is considered an insult to leave one’s shoes on when entering a home.”

By contrast, at that very moment, C himself could not excuse his faux pas by saying: “Please don’t be offended by my behavior. I don’t know that in your culture it is considered an insult to leave one’s shoes on when entering a home.”

- (III) *Professor B*: “Although it is his turn, it would be downright irresponsible of us to elect C the new chair of our department.”

Professor A: “How so?”

Professor B: “Everyone knows that C stopped caring about the department a long time ago. He is sure to do a terrible job.”

The argument would wear an entirely different aspect if it came from Professor C’s own lips: “Although it is my turn, it would be irresponsible of you to elect me chair of our department. Everyone knows that I stopped caring about the department a long time ago. I’m sure to do a terrible job.”

- (IV) *Cowboy A*: “I’m afraid C won’t listen to reason. The only way we’ll get back the cow he stole from us last week is to *take* it back—if need be, by force.”

Cowboy B: “But you know C. He’d rather die than return our property! Are you really prepared to kill a man over a cow?”

Cowboy A: “You’re right, that *would* be excessive. . . . I guess we’ve got no choice but to let the matter go.”

Contrast this with a case in which C himself says to A and B: “The only way you’ll get back the cow I stole from you last week is by killing me. But, as you know, that would be excessive. So you really have no choice but to let the matter go.”

In all four examples, the argument, when presented by C, consists of the same factual and normative premises as that of speaker B.² Yet in each case, C, in particular, seems unable to successfully avail himself of the argument to provide the desired justification.

In the first example, the explanation is *conceptual*. C’s *claiming* the reward for himself by uttering the above argument would stand in performative contradiction with the supposed traits for which he demands to be honored, namely, his modesty and self-effacingness. Were C, in particular, to put forward the argument to press his claim, this would make the argument unsound.

In example (II), it is *epistemic* factors that prevent C from presenting B’s argument in his own voice. It is quite true, of course, that C’s ignorance of local customs should excuse his seemingly rude behavior. But precisely because he is presently unaware that his behavior might give offense, C cannot, at that moment, offer his ignorance as an excuse—entirely valid though it would have been.

The most interesting form of speaker-relativity, however, and the kind that will concern me in this article, is the one we encounter in examples (III) and (IV). Here, the objections against C appropriating B’s argument are *moral* in nature. It is not that the conclusion of C’s argument is *untrue* in either case: just in virtue of its bad consequences, I take it that it *would* be irresponsible of Professors A and B to elect C chair of their department, and that the two cowboys *would* be wrong to seek to recover their cow, knowing that this is likely to result in a deadly confrontation with C. Rather, as Cohen notes, the reason C cannot use B’s argument to provide his interlocutors with an adequate

2. Cohen calls this “an unimpeachable criterion of identity for arguments.” See Cohen, *Rescuing Justice and Equality*, p. 39.

moral justification is that it is his *own* readiness to act in a morally indefensible manner (by making no effort as department chair or by refusing to return the stolen cow without a bitter fight) that *makes* the conclusion of B's argument true. There is surely something awry with an agent's trying to offer up his own avoidable wrongdoing as a *justification* for why others should do as he says.

The phenomenon of speaker-relativity of justification, as it will concern me in this article, can thus be characterized by the following two propositions:

- (1) For moral reasons (the precise nature of which remains to be investigated), person C is unable to offer agent A an adequate moral justification for why A ought to perform some action ϕ (or omit some action ψ).
- (2) Nevertheless, it is *true* that A ought to perform ϕ (or omit ψ)—for instance, because *not* doing ϕ (or doing ψ) would lead to an outcome that is much worse. Moreover, there is another speaker B who *could* justify to A why he must so act.³

3. The speaker-relativity of justification should be distinguished from a related phenomenon that has begun to attract attention in recent years, namely, the fact that an agent may sometimes lack the standing to engage in speech-acts of *condemnation*, *blame*, or *complaint* vis-à-vis another person—despite the content of her criticism being entirely valid. Thus, in “Casting the First Stone: Who Can, and Who Can't, Condemn the Terrorists?” in *Finding Oneself in the Other* (Princeton, N.J.: Princeton University Press, 2013), G. A. Cohen notes that although certain actions are morally wrong, not all persons are in a position to condemn them. Saul Smilansky, in “The Paradox of Moral Complaint,” in *Ten Moral Paradoxes* (Oxford: Blackwell, 2007), pp. 90–99, notes that an agent who has, in the past, engaged in a certain type of wrongdoing may not have the right to complain when others treat him similarly in the present. Finally, T. M. Scanlon, in *Moral Dimensions: Permissibility, Meaning, Blame* (Cambridge, Mass.: Belknap Press of Harvard University Press, 2008), chap. 4, points out that while certain actions are unambiguously *blameworthy*, not everyone has the standing to *blame* the agent for performing them. Though these phenomena are clearly related to the speaker-relativity of justification, they are not two sides of the same coin. In particular, it seems to me (though I will not attempt to defend this claim here) that if C is unable to offer A an acceptable interpersonal justification for why she ought not to do ϕ , C cannot condemn A for doing ϕ . (Thus, in example (III) above, if Professor C is not in a position to justify to his colleagues that they ought not to elect him department chair, he is not in a position to condemn them for doing so either.) The reverse, however, is not necessarily true: C may lack the standing to condemn A for committing some wrong—say, stealing from other people—because C himself regularly engages in similar behavior. But this does not imply that C could not offer A a perfectly valid moral justification for why she ought not to steal.

II. OUTLINE OF THE ARGUMENT

If moral philosophy has, to date, largely ignored the phenomenon of speaker-relativity simpliciter, this is true, a fortiori, of the issues that it raises for the field of normative ethics. How well do established ethical theories accommodate, and account for, the phenomenon of speaker-relativity? And might the question “who can justify what to whom?” have implications for the rightness or wrongness of actions? The aim of this article is to begin to address these and other related questions.

One moral theory that gives a characteristically simple answer is act-consequentialism. Act-consequentialists, who hold that an action φ is right if and only if it has the best (expected) outcome of all actions available to the agent, are likely to regard the question “who can, or cannot, provide A with an adequate justification that he ought to do φ ?” as quite irrelevant to the *rightness* of doing φ . Who can *justify* to A that he ought to do φ does not, after all, affect the outcome of *doing* φ . Thus, if it is foreseeable that the consequences of doing φ will be best, then this is what A ought to do—regardless of who may or may not be able to offer A an adequate justification for this conclusion. For the act-consequentialist, the speaker-relativity of justification is at best normatively inert.

Indeed, act-consequentialists seem hard-pressed to make room for the phenomenon of speaker-relativity in the first place. The act-consequentialist might ask: If it is *true* that A ought to do φ (because this will produce the best outcome), then how could anyone *not* be able to justify to A that he ought to do φ ? For act-consequentialists, after all, the rightness or wrongness of actions is *prior* to the question of justification. An action or omission is justifiable to others, and its performance can justifiably be demanded of an agent, if and *because* it is morally right. Thus, if the consequentialist calculus determines that, because of C’s willful apathy, it would be morally right of A and B not to elect him chair of the department, then this can justifiably be demanded of them—including by C himself.⁴ Act-consequentialists are thus tempted

4. Of course, the fact that C could justify to A why he ought to perform φ does not imply that C ought to try to do so. It is conceivable that if C, in particular, presented A with an argument for why he ought to do φ , A would be less likely to do φ than if C said nothing. In such cases, act-consequentialism would oblige C to remain silent.

to dismiss the phenomenon of speaker-relativity as a “fallacy of relevance,” akin to that found in ad hominem arguments.

That act-consequentialism appears unable to account for the speaker-relativity of justification is unlikely to lose its defenders much sleep. This, after all, is just another in the long list of ways in which their theory parts ways with commonsense ethical thought, and consequentialists have long learned to bite bigger bullets than that. If act-consequentialism were the only prominent moral theory whose deliverances are at odds with speaker-relativity, there would appear to be little need for normative ethics to give this phenomenon much attention.

The main thrust of this article, however, is that act-consequentialism is not alone in being challenged by the speaker-relativity of justification. If I am right, this phenomenon represents an even more serious problem for one of consequentialism’s main systematic rivals: the contractualist theory defended by T. M. Scanlon.

Moral contractualism, as we will see in detail in Section 3, *reverses* the priority relation between justification and rightness that we observed in the case of consequentialism: for the contractualist, an action is morally right or permissible if and *because* it is justifiable to each person (or, in the terms of Scanlon’s formula, because it is “licensed by a principle that no one could reasonably reject”). Like attempts at justification, however, grounds for rejection must be presented in the form of arguments, addressed to the agent by those individuals who are adversely affected by his action. Hence, as I argue in Section 4, the phenomenon of speaker-relativity, which impacts *which* (if any) of those individuals can reasonably reject the principle that licenses the action’s performance, is therefore of great relevance to contractualism’s theory of rightness.

Building on this insight, Sections 5 to 7 raise a troubling possibility for contractualism: taking my cue from an example of G. A. Cohen’s, I argue that there are instances in which an action is intuitively wrong in virtue of its bad consequences—yet *none* of those negatively affected by it can reasonably reject the principle that licenses its performance. This finding threatens to drive a wedge through contractualism’s account of moral rightness. It suggests that the supposed *analysans* (justifiability to each person) and the *analysandum* (moral rightness) can sometimes come apart. In Section 8, I defend this result against

various objections, before concluding the argument in Section 9 with some suggestions on how Scanlon's theory could be amended to better deal with the problem of speaker-relativity of justification.

III. AN OVERVIEW OF SCANLON'S CONTRACTUALISM

Over the past three decades, the contractualist moral theory championed by T. M. Scanlon has emerged as a powerful and systematic challenge to consequentialism. As Scanlon writes in "Contractualism and Utilitarianism," the seminal article that grew into his book *What We Owe to Each Other*,⁵ one of the express aims of contractualism is precisely to provide "a clear account of the foundations of non-utilitarian moral reasoning."⁶

At the heart of Scanlon's contractualism lies his account of moral right and wrong. According to

Scanlon's Formula: An act is right just when it is licensed by a principle for the general regulation of behavior that no one could reasonably reject.⁷

What animates the contractualist formula is the thought that moral rightness—at least in the core domain of interpersonal morality—is essentially a matter of acting in a way that is *justifiable* to each person. As Scanlon writes: "the idea of justifiability shapes our thinking about right and wrong, and determines the content of this part of morality. Thinking about this part of morality is, I argue, thinking about what would be allowed by principles that others could not reasonably reject."⁸ An act is morally wrong, for Scanlon, just in case it cannot be justified to one or more persons, that is, if there exists some individual standpoint from which any principle licensing this act could be reasonably rejected.

5. T. M. Scanlon, *What We Owe to Each Other* (Cambridge, Mass.: Belknap Press of Harvard University Press, 2000).

6. See T. M. Scanlon, "Contractualism and Utilitarianism," in *Utilitarianism and Beyond*, ed. Amartya Sen and Bernard Williams (Cambridge: Cambridge University Press, 1982), pp. 103–28, esp. p. 103.

7. This formulation of the contractualist formula is a composite of the many subtly different versions of his formula that Scanlon presents throughout his book.

8. T. M. Scanlon, "Replies," *Social Theory and Practice* 28 (2002): 345.

Of course—as we saw in Section 2—the idea that an act is right if and only if it can be justified to others is one that even noncontractualists might embrace. An act-utilitarian, for instance, who believes that an act is right if and only if it maximizes pleasure, could maintain that in any given situation, the only act justifiable to each person is that which maximizes pleasure. He would then accept that an act is right if and only if it is justifiable to each person. The criterion of justifiability-to-each-person, however, is not fundamental to the utilitarian account of moral wrongness. Rather, it is only when and *because* an act conforms to utilitarianism's fundamental criterion of rightness, namely, pleasure maximization, that the utilitarian considers it justifiable to each person. For the act-utilitarian, moral rightness is the logically prior concept, on which the notion of justifiability-to-each-person rides piggyback. For the contractualist, by contrast, the idea of justifiability to each person plays a *fundamental* role in his account of moral right and wrong. Indeed, for Scanlon, the most basic account of what it *is* for an act to be wrong is simply that it is not justifiable to each person.⁹

But what does it mean for an act to be justifiable to every person? Scanlon's answer is contained in the second half of his formula, and it is here that his theory acquires its distinctly anticonsequentialist bent. According to Scanlon, an act is justifiable to each person if and only if

9. So much at least is true for the part of morality that falls within the scope of the contractualist formula. Following Frances Kamm, "Owing, Justifying, and Rejecting," *Mind* 111, no. 442 (2002): 323–54, I shall refer to this part of morality as "M1." Scanlon does not claim that his contractualist account of right and wrong *exhausts* morality "in the broad sense in which most people understand it." Scanlon, *What We Owe to Each Other*, p. 6. For instance, an action may be morally criticizable, in a *broader* sense, because it fails to respond appropriately to the impersonal value of some wonder of nature (pp. 218–23), or to the badness of animal suffering (pp. 177–87). (Call the part of morality not captured by Scanlon's formula "M2.") Scanlon does maintain, however, that his contractualist theory provides an attractive and reasonably complete account of the morality of right and wrong in our dealings *with other persons*, that is, with that part of morality "having to do with our duties to other people, including such things as requirements to aid them, and prohibitions against harming, killing, coercion, and deception" (p. 6). What distinguishes persons from most other animals, according to Scanlon, is that we have what he calls "judgment-sensitive attitudes," that is, attitudes that are adopted or withdrawn in accord with appreciation of *reasons*. According to Scanlon, appreciating the value of the person must involve recognizing and respecting her capacities to assess and act on reasons. And the way to value the rational and self-governing capacity in a person is to treat her in accord with principles that she could not reasonably reject.

it is permitted by at least one principle that no one could reasonably reject.

“Reasonable rejection,” for Scanlon, is a partly moral notion. Thus, in order for our rejection of a moral principle to count as *reasonable*, we must give appropriate weight to the moral claims of other people. In particular, Scanlon argues that the reasonable rejection of moral principles should be subject to the following crucial restriction:

The Individualist Restriction: Reasonable rejection of a moral principle must occur by appealing to the principle’s implications only for oneself, or for some other single person.

In Scanlon’s words: “the justifiability of a moral principle depends only on *individuals’* reasons for objecting to that principle and alternatives to it.”¹⁰

The bulk of the literature on contractualism has focused on the *anti-aggregationist* implications of the Individualist Restriction: the Individualist Restriction excludes the combination of reasons or claims across individuals. What matters in determining whether a moral principle is reasonably rejectable is not the sum of different persons’ reasons for rejecting the principle, but merely how the strongest individual reason for rejecting this principle compares to the strongest individual reason for rejecting any alternative principle.

For our purposes, however, it is more important to note what I shall call the *personal* nature of the Individualist Restriction: for a principle to be reasonably rejectable, there must be someone who can reject it from a *personal* standpoint, that is, by appealing to the implications that the principle has for someone in *her* position. An act is wrong, Scanlon thinks, only if there is someone who can argue that we failed to treat her in a way that was justifiable to *her*.¹¹

Scanlon does not mean to rule out that people may object to some action or omission *on behalf of* another person (as her spokesperson or proxy, so to speak). However, Scanlon is explicit that what the

10. Scanlon, *What We Owe to Each Other*, p. 229.

11. See *ibid.*, p. 218: “In order for a principle to be reasonably rejectable there must be some relevant standpoint from which people typically have good reason either to refuse to accept that principle as part of their own practical thinking or to refuse to recognize it as a ground that others may use to justify their conduct.”

spokesperson can say on behalf of the person they he or she speaks for is still a function of what that person herself could say. As he puts it, grounds for reasonable rejection “must be reasons that such a person would have ‘on his or her own behalf.’ This is to say, there must be reasons arising from the way such a person would be affected by following the principle, or by having others do so.”¹² In other words, if a person could not, in principle, make the case herself, by appealing only to personal reasons, that she can reasonably reject a principle licensing you to perform some action or omission, then her spokesperson could not do so on her behalf either.

First among the factors that may make a principle unjustifiable to an individual are the burdens that it imposes on her (either by making her worse off or by causing her to forgo potential benefits). An individual might also reasonably reject a principle because it treats her *unfairly*, by “singl[ing] others out, without justification, for a privileged moral status.” Even if fairness-based reasons for rejecting a principle do not necessarily appeal to a person’s well-being, Scanlon maintains, “they are still . . . *personal* reasons, since they have to do with the claims and status of individuals in certain positions.”¹³

By contrast, there are two things an individual cannot appeal to in reasonably rejecting a principle. First, the Individualist Restriction bars us from rejecting a principle on the grounds that it would lead to an outcome that is impersonally bad, that is, bad simpliciter, independently of its effects on particular people. As Scanlon puts it: “Impersonal values do not provide, in themselves, reason for rejecting principles of right and wrong.”¹⁴ Instead, we must be able to point to some person who could reasonably reject a principle based on its implications for *her*.

Second, an individual could not claim that a principle is reasonably rejectable because the acts it licenses are *wrong*. Such a use of “reasonably rejectable” would treat it as downstream from the concept of moral wrongness. The aim of contractualism, however, is to give an account of morally wrong action *in terms of* reasonable rejectability. On pain of circularity, the contractualist therefore cannot invoke his

12. Scanlon, “Replies,” p. 345.

13. Scanlon, *What We Owe to Each Other*, p. 219.

14. *Ibid.*, p. 222.

belief that an act would be morally wrong as grounds for reasonable rejection.

Of course, that contractualism bars us from appealing to our moral intuitions about the wrongness of some action as grounds for reasonably rejecting a principle that licenses this action does not prevent us from consulting our intuitions when evaluating the success of contractualism itself. Indeed, Scanlon regards it as one of the key attractions of his theory that it provides a theoretical underpinning for our intuitive opposition to crude forms of consequentialist aggregation.¹⁵ In the same measure, however, contractualism's plausibility as a moral theory will depend on whether or not Scanlon's formula is capable of tracking certain other moral intuitions of ours.

IV. SPEAKER-RELATIVITY AS A CHALLENGE TO CONTRACTUALISM

One very widely held intuition is that actions or omissions are sometimes morally wrong if they will foreseeably have consequences that are very bad. We need not be consequentialists, for whom the rightness of an action is solely a function of its consequences, in order to hold that an action's consequences often matter to its rightness. However, as we just saw, Scanlon's theory restricts the way in which consequences can matter to the rightness of an action. According to the Individualist Restriction, the fact that an action or omission will have consequences that are very bad simpliciter does not constitute, in itself, grounds for reasonably rejecting a principle licensing us to perform that action or omission. Within the domain of interpersonal morality, the action or omission will only count as wrong if somebody has a justified *personal* complaint against our action or omission based on how this affects *him*. Lest contractualism clash with our moral intuitions, Scanlon must therefore hope that in any instance in which intuition tells us that omitting φ or doing ψ is wrong in virtue of its foreseeably bad consequences, there is someone who can reasonably reject a principle that licenses us to omit φ or do ψ .

It is in this context that the phenomenon of speaker-relativity threatens to raise a serious problem for contractualism. Recall that in

15. See, for instance, Scanlon's famous "transmitter room" case in *ibid.*, pp. 235ff.

Section 1, I characterized the instances of speaker-relativity that concern me in this article as having the following features:

- (1) For moral reasons, person C is unable to offer agent A an adequate moral justification for why A ought to perform some action φ .
- (2) Nevertheless, it is *true* that A ought to do φ , because *not* doing φ would lead to an outcome that is much worse.

Now, if C is unable to *justify* to A why A ought to perform φ , then, I submit, C cannot *reasonably reject* a principle that licenses A not to do φ . Without being able to make an adequate case to A why he ought to do φ , C can have no reasonable complaint against a principle that licenses A not to do φ . Indeed, since Scanlon maintains that the reasonable rejection of moral principles must proceed on the basis of *personal* reasons, he is committed to a stronger claim. For ease of future reference, I will label it the “Cannot Justify/Cannot Reject” principle:

Cannot Justify/Cannot Reject. Person C (or a spokesperson for C) can reasonably reject a principle that licenses A not to do φ if and only if C is able to offer A an adequate moral justification, appealing only to C’s *personal reasons*, for why A ought to do φ .

This connection between the speaker-relativity of justification and the reasonable rejection of principles would spell trouble for Scanlon’s theory if there were cases in which both of the following propositions are true:

- (3) There is *no one* who can reasonably reject a principle that licenses A not to do φ , because no one is able to offer A an adequate moral justification, appealing only to their *personal* reasons, for why A ought to do φ .
- (4) Nevertheless, it is *true* that A ought to do φ , because not doing φ would lead to an outcome that is much worse. Indeed, A’s failure to perform φ would, intuitively, be morally *wrong*.

Since the reasonable rejection test *is* Scanlon's criterion of rightness and wrongness in interpersonal morality, the truth of (3) would imply that A's failure to perform φ could not be morally wrong. But if the case were one in which (4) was *also* true, then this would be the intuitively wrong answer. A case in which propositions (3) and (4) are both true would therefore represent a troubling counterexample to contractualism's account of moral wrongness. It is such a case that I shall present in Section 7. Before we come to my counterexamples, however, we must pause and take a second, closer look at the phenomenon of speaker-relativity of justification.

V. COHEN'S KIDNAPPER

How exactly can it be that two moral arguments, consisting of the same factual and normative premises, differ in their justificatory force, just because of who is making the argument? How could a person be unable to provide another with an adequate moral justification for why she ought to perform some action, if that action is, in fact, morally required of her? Indeed, what more *is* there to successful interpersonal moral justification than showing another person, by means of a sound argument, that she has all-things-considered moral reason to do as you say?

I propose to explore these questions, and the contractualist's answers to them, by considering an illuminating example from Cohen's Tanner Lectures. Consider the following case:

Kidnapper: A kidnapper has taken Smith's daughter hostage. He is demanding that Smith pay him \$10,000, or else he credibly threatens to kill the child.

There is no question that the kidnapper is making an immoral threat against Smith. In many cases, Cohen remarks, there will be pro tanto moral reasons for resisting such threats: paying the ransom might encourage more kidnappings; and, quite independently of any negative consequences, ceding to a vile threat might be considered inherently bad. Still, considering the awfulness of what the kidnapper is threatening, and assuming that there is no other way of freeing his daughter, most would agree that Smith is morally required to pay the ransom. A

friend, wishing to justify this conclusion to Smith, might then present him with the following argument:

- (1) It would be terrible if the kidnapper killed your daughter.
- (2) The kidnapper will kill your daughter unless you pay him \$10,000. Therefore, you morally ought to pay the kidnapper \$10,000.

In putting forward premise (2), the friend is treating it *as given* that the kidnapper will kill Smith's daughter, unless he pays him the ransom. Although the friend knows that the kidnapper *could* release Smith's child unharmed without being paid a ransom, he foresees that, as a matter of fact, the kidnapper will not do so. For the purpose of reaching the right decision, he may therefore enjoin Smith to reason *as if* the kidnapper had no choice but to kill his daughter if he refused to pay him the ransom.¹⁶

Coming from the friend, the suggestion to treat the kidnapper's future behavior as conditionally given is surely legitimate, as is the argument that builds on it. When I can foresee that some intervening agent will react to my doing φ by doing ζ , this knowledge often bears on whether my doing φ would be morally right—even if it were morally

16. To take someone's future wrongdoing as conditionally "given," to regard him as someone whom we cannot realistically expect to respond even to clear moral reasons, is in one sense akin to viewing him with what P. F. Strawson calls the "objective attitude." See P. F. Strawson, "Freedom and Resentment," in *Freedom and Resentment, and Other Essays* (London: Routledge, 2008), pp. 1–28. According to Strawson, "To adopt the objective attitude to another human being is to see him . . . as an object of social policy; as a subject for what, in a wide range of sense, might be called treatment; as something certainly to be taken account, perhaps precautionary account, of; to be managed or handled or cured or trained; perhaps simply to be avoided" (p. 9). At the same time, Cohen's kidnapper is, of course, very different from the typical objects of Strawsonian objective attitudes. Unlike young children or the mentally ill, Cohen's kidnapper is (by assumption) perfectly *capable* of recognizing and responding to moral reasons. The mere fact that we cannot rely on him to do so (indeed, that we can rely on him not to do so), should not bar us—as in most cases where the objective attitude is appropriate—from holding toward the kidnapper "reactive attitudes" such as anger, resentment, or reproach. Strawson himself is prepared to allow that "even in the same situation, [the objective attitude and reactive attitudes] are not altogether exclusive of each other" (p. 9). Though not the paradigmatic objects of the objective attitude, Strawson admits that "we can sometimes look with something like the same eye on the behaviour of the normal and the mature" (p. 10). This, perhaps, is one kind of case that Strawson may have had in mind.

impermissible for the intervening agent to perform ζ and she were entirely free to act otherwise.¹⁷

Cohen's crucial insight is this: In the context of interpersonal moral argument, the kidnapper *himself* cannot legitimately treat his future behavior as given. That is, the kidnapper himself cannot cite his willingness to kill Smith's daughter as a "justification" for why Smith is morally required to pay the ransom. Imagine the kidnapper trying to present Smith with the following first-personal version of the friend's argument:

- (1*) It would be terrible if I killed your daughter.
- (2*) I will kill your daughter unless you pay me \$10,000.
Therefore, you morally ought to pay me \$10,000.

Though both premises in the kidnapper's argument express the same factual and normative claims as those of the friend's argument above, coming from the kidnapper the argument lacks the justificatory force that it had when presented by the friend. Although the kidnapper obviously creates a moral *reason* for Smith to pay him the \$10,000 by informing him what will happen to his daughter if the ransom is not paid, this, Cohen stresses, is not to be confused with giving Smith an acceptable *moral justification* for why he ought to pay him this sum of money.¹⁸ Indeed, as Cohen remarks, presenting his claims under the guise of a moral argument would serve to further discredit the kidnapper, beyond his already reprehensible actions.

One of the reasons, Cohen argues, why the kidnapper's argument cannot serve as an adequate justification for why Smith ought to pay him is that the kidnapper himself is *responsible* for the fact that Smith must pay for his daughter's release—and this is something he can be called on to justify. Unlike Smith's friend, who may treat the kidnapper's intransigence *as given* and thus needs only an *epistemic* justification for advancing premise (2) of his argument, the kidnapper owes us a *moral* justification for *making* (2*) true—a justification that, needless to say, he cannot supply.

Cohen's kidnapper is engaged in a form of bad faith. At least on one way of reading the vignette, the kidnapper presents his unwillingness to

17. I discuss these issues in my Oxford BPhil thesis, *Morality and the Problem of Foreseeable Non-Compliance* (2008), available upon request.

18. See Cohen, *Rescuing Justice and Equality*, chap. 1.

return Smith's daughter without a ransom as if it were a mere psychological *datum* about himself, for which he—like Smith's friend—bears no responsibility. Though his argument is phrased in first-personal terms, the kidnapper takes an essentially *impersonal* attitude toward his own future actions, casting them as something outside his control and thus immune to demands for justification. But given that there is nothing actually preventing the kidnapper from releasing Smith's daughter without receiving a ransom, this retreat to the impersonal point of view seems morally illegitimate—an evasion of responsibility for his own actions.¹⁹

Put in general terms, the kidnapper's argument falls afoul of what we may call

Cohen's Principle: If I (will) deliberately and avoidably make it the case that *p*, then I cannot justify to you why you ought to do *φ* by *appealing* to the fact that *p*, unless I can also morally *justify* making it the case that *p*.

The kidnapper himself deliberately and avoidably makes premise (2*) of his argument true, while lacking any justification for doing so. Hence, he, of all people, cannot *invoke* the fact that (2*) is true to give Smith an acceptable moral justification for why he ought to pay him.²⁰

19. In *Authority and Estrangement: An Essay on Self-Knowledge* (Princeton, N.J.: Princeton University Press, 2001), Richard Moran makes a very similar point. He writes:

If I consider myself impersonally, as someone with a certain character and history, with certain beliefs and desires which then enter into my moral deliberation along with the character and desires of others, I leave out of consideration the fact that I am *responsible* for that character and those desires. But these cannot be for me a set of data for which I must simply make room in my deliberations, as I may have to accommodate the empirical fact of other people's beliefs and desires. For in my own case I remain responsible for their endorsement, justification, and continued role in my life, and this determines whether or not they are to have any place at all in my deliberations. . . . Hence, they cannot enter into my thinking as the fixed beliefs and desires of some person or other, who happens to be me. (p. 164)

20. The following, by contrast, is an argument that satisfies Cohen's Principle. Sheriff to bandit: "If you commit this robbery, I will have to arrest you. But if I arrest you, there won't be anyone to provide for your wife and two small children, which would be terrible. So, just for this reason, you ought not to commit the robbery." Here, the sheriff *can* legitimately appeal to something that he will deliberately and avoidably make true, namely, the fact that he will arrest the bandit if he commits the robbery and the bad consequences for his family that this would entail, because arresting the bandit would be justifiable.

Cohen's Principle also provides a plausible account of what went wrong in the third and fourth pieces of moral dialogue in Section 1. Take the former example: Professor C attempted to justify to A and B that they ought not to elect him chair of the department by appealing to the fact that he would not make any effort once in office. But whether C makes an effort or not is, we assume, largely up to him; and as a member of the department, called on to do his fair share, he surely *ought* to do so. Since C lacks any moral justification for not making an effort as chair, Cohen's Principle implies that he cannot appeal to the fact that he will not make an effort to *justify* to A and B why they ought not to elect him. This seems exactly right.²¹

And yet, as it stands, Cohen's analysis is not fully satisfactory. The Kidnapper case demonstrates that our pretheoretical notion of interpersonal justification and the idea of showing someone, by means of a sound argument, that they ought to do as we say can come apart. Cohen has also given us a partial explanation of what generates this failure of coextensiveness: intuitively, giving someone an adequate justification for why they ought to do φ is subject to a more stringent necessary condition, namely, Cohen's Principle.

What Cohen has not provided us is a deeper explanation of *why* interpersonal justification is subject to these more stringent conditions. Indeed, for all the intuitive appeal of Cohen's discussion, a skeptic might wonder: What more *could there be* to interpersonal moral justification than showing another person, by means of a sound moral argument,

21. It is important to note that by affirming Cohen's Principle, I do not take a stand on a related but importantly distinct debate in normative ethics, namely, that between so-called *actualists* and *possibilists* about moral "oughts." For a seminal discussion, see Frank Jackson and Robert Pargetter, "Oughts, Options, and Actualism," *Philosophical Review* 95 (1986): 233–55. The central question in that debate is this: Do facts about what I *will* do make a difference to what I *ought* to do? Actualists hold that, in evaluating my options, I should consider what would *actually* be the case were that option to be chosen, "where what would be the case includes what the agent would simultaneously or subsequently in fact do" (*ibid.*, p. 233). By contrast, possibilists claim that predictions about how I will actually behave in the future are not relevant to which option I ought to choose; rather, I should evaluate each option according to what it is *possible* for me to do. The question that Cohen's Principle addresses is different. It is not about the "first-personal" question whether, and how, predictions about my future behavior bear on how *I* myself ought to act; rather, it is about the "inter-personal" question whether, and how, I can appeal to claims about my future behavior to *justify* to others that *they* ought to act in a certain way.

that they ought to do as you say? To answer this skeptical challenge, we need an account of the *nature* of interpersonal justification that supports and explains our account of the conditions under which interpersonal justification comes apart from showing someone that they ought to do as you say.²² It is to this task that I turn in the next section.

VI. INTERPERSONAL JUSTIFICATION AS JOINT DELIBERATION AND THE SOURCES OF SPEAKER-RELATIVITY

There is, of course, a sense in which the argument of Cohen's kidnapper *does* succeed in justifying that Smith ought to pay. This is the sense in which *all* sound deductive or ampliative arguments justify their conclusions. Given the truth of the argument's premises, we are warranted in accepting its conclusion as well. We can call this sense of justification "propositional justification," since the aim of justification in this sense is simply to give us reason to believe some proposition—in this case, the proposition that Smith ought to pay the kidnapper.

By contrast, there is a second sense of justification—justifying *to* another person that he ought to perform some act—that is different, and more demanding. This is the sense of justification that I have been referring to as "interpersonal justification," and that, I submit, is speaker-relative in certain cases.²³ But what exactly is the nature of interpersonal justification, and how does it differ from showing one's interlocutor, by means of a sound moral argument, that they ought to do as one says?

22. It is a further sign of moral philosophy's relative neglect of questions surrounding the dialogical presentation of moral arguments that such an account is almost wholly absent from the literature. Indeed, I am not aware of any detailed analysis of the concept of interpersonal moral justification.

23. To be precise, we can distinguish (at least) two senses of interpersonal justification, what I call "self-directed" and "other-directed" justification. In instances of self-directed justification, the object of justification is the speaker's *own* behavior. The aim of self-directed justification is to establish, vis-à-vis his interlocutor, that the speaker acts in a morally defensible manner. In instances of other-directed justification, the object of justification is a *prescription* for another agent. The aim is to justify to one's interlocutor that *they* ought to act in a certain way. Speaker-relativity of justification is a phenomenon that affects interpersonal justification in the other-directed sense. (*Audience*-relativity of justification—if there is such a thing—might also affect justification in the self-directed sense.) My account in the following will be concerned primarily with interpersonal justification in the other-directed sense; but I will indicate below how my account applies, in a truncated sense, to cases of self-directed justification as well.

Here is my attempt at an answer. I follow Scanlon and others in maintaining that for it to be the case that someone morally ought to do φ just is for it to be the case that she has conclusive moral reason to do φ .²⁴ Further, I support the relational account of reasons defended by Scanlon in *Being Realistic about Reasons*. Specifically, I follow Scanlon in asserting that “is a conclusive reason for” is a four-place relation, CR(p , A , c , φ) holding between a fact p , an agent (or agents) A , a set of circumstances c , and an action φ (or set of actions φ_1 , φ_2 , ...). This is the relation that holds just in case p is a conclusive reason for a person A in situation c to do φ .²⁵

The set of circumstances c may include information about how *other* persons will behave, in some cases *conditional* on the agent acting in certain ways. This can include behavior that runs counter to conclusive moral or prudential reasons. Moreover, it is uncontroversial that such facts about the relevant background conditions may determine whether a given fact p is a conclusive reason for the agent to do φ or not. Thus, the fact that the kidnapper will let Smith’s daughter go unharmed if he pays the ransom is a conclusive reason for Smith to pay the kidnapper in conditions where, unless the ransom is paid, the kidnapper will kill Smith’s daughter. It would not be a conclusive reason in circumstances where the kidnapper would release the child even without receiving a ransom.

To show someone that they morally ought to do φ therefore just is to establish that the relevant CR-relation holds in their case. This is something that *anyone* in possession of the relevant information can do, for instance, by means of a sound argument. Even the *kidnapper*, in Cohen’s case, can be the source of the father’s moral knowledge that he ought to pay him. Hence, as far as showing someone that they have decisive reason to do φ , or morally ought to do φ , there is no phenomenon of speaker-relativity.

Things are quite different in the case of interpersonal justification. For speaker B to claim that agent A morally ought to do φ in circumstances c is to make a normative claim about A alone: A has a conclusive moral reason to do φ in circumstances c . By contrast, I will argue that for B to claim that he has given A a successful interpersonal justification that A

24. See T. M. Scanlon, *Being Realistic about Reasons* (Oxford: Oxford University Press, 2014), chap. 1.

25. *Ibid.*, p. 4.

ought to do φ in c is implicitly to make a claim not just about what A has reason to do, but about B's *own* reasons for action as well. Interpersonal justification, I maintain, is a form of *joint deliberation*: For it to be true that B successfully justifies to A that A ought to do φ in circumstances c , this must be (part of) the correct answer to the question "what do *we*, A and B, have reason to do in circumstances c^* ?"

What A and B have reason to do in circumstances c^* need *not* be a "joint action," in the sense of the philosophical literature on shared agency.²⁶ A and B's jointly deliberating about how they each ought to act in c^* does not entail that the correct answer must involve an instance of joint action. Rather, what distinguishes B's deliberating *jointly* with A about how to act from A's deliberating alone is what counts as part of the *circumstances* of action in either case (hence the notational variation between " c^* " and " c "). In the latter case, the circumstances c of A's action include information about B's foreseeable future behavior; it is part of the background conditions against which A must decide how to act. By contrast, in the former case, B's future actions are not part of the *circumstances* c^* of action for A and B, but rather *part of* A and B's future actions. They are part of what A and B are trying to *settle* through joint deliberation.

In support of my account of interpersonal justification as joint deliberation, let us review some of our pretheoretical ideas about interpersonal justification. Unlike giving someone a conclusive moral reason to perform φ —which, as the case of Cohen's kidnapper illustrates, may involve coercion or other kinds of misbehavior on the part of the speaker—successful interpersonal justification is a form of moral *suasion* that does not rely on such strong-arm tactics. Interpersonal justification, it might be said, is how we get one another to do things *qua moral agents*.

This contrast between interpersonal justification and "mere reason-giving" is reflected in the manner in which it is appropriate for an agent to respond to instances of either. If B succeeds in justifying to A that he ought to do φ , this, morally speaking, "closes the case": A has nothing to retort to B, and, insofar as A is morally motivated and reasoning correctly, he will not hesitate to form the intention to do φ . Moreover, in terms of A's reactive attitudes toward B, A has no basis for resenting B for asking him to do φ , or feeling hard done by their interaction, however onerous or unpleasant doing φ may be.

26. I thank Alec Walen for pressing me to clarify this point.

Things are quite different in the case of “mere reason-giving.” As a practical matter, it may be that Smith has no choice but to meekly give in to the kidnapper’s demands. Any further discussion might only serve to aggravate the kidnapper and put his daughter’s life in greater danger. But this hardly closes the case *morally speaking*. While it is true that, *as things stand*, Smith has a conclusive moral reason to pay the ransom, it would be entirely appropriate for Smith to angrily remonstrate with the kidnapper, and to resent him for making his demand. More importantly, suppose that Smith responded to the kidnapper’s argument with the following “counterproposal”: “Why don’t you just let my daughter go free *without* me paying you a ransom? It’s perfectly within your power to do so, you know?” The kidnapper has no acceptable basis on which to reject this counterproposal in favor of the course of action that his argument supports. The fact that he *will not* let the child go without receiving a ransom is neither here nor there in this regard.

We can thus say: it is the mark of successful interpersonal justification that it is immune to counterproposals of this kind. For B to successfully justify to A that he ought to do φ in circumstances c , it must be the case that A cannot propose a different course of action, jointly accessible to A and B, such that (i) this course of action does not involve A doing φ and (ii) this course of action is a more plausible answer to the question “what are *we*, A *and* B, to do in circumstances c *?”

To find a course of action that is immune to such counterproposals requires B to engage in precisely the exercise of joint deliberation with A that I described above. B has to show A that the action that he is calling on A to perform is (part of) the correct answer to the question “what do *we*, A *and* B, have reason to do in circumstances c *?”²⁷

27. I promised above to show how my account of other-directed justification can be applied to the case of self-directed justification as well. For other-directed justification to succeed, I have argued, it must be the case that the action that the speaker B proposes to the agent A is (part of) the correct answer to the question “how do agent *and* speaker have reason to act in the relevant circumstances?” But in the case of self-directed justification, that is, justification that aims at showing that the speaker *himself* acts in a morally defensible manner, speaker and agent are *one and the same*. So the question of how speaker and agent have reason to act in the relevant circumstances reduces to the question “how do I, the speaker, have reason to act in these circumstances?” The speaker succeeds in justifying his action just in case he shows that his proposed action is a plausible answer to this question.

Cohen's Principle falls naturally out of this account of interpersonal justification as joint deliberation. If B attempts to justify to A that he ought to do φ , but does so—in violation of Cohen's Principle—with an argument that relies on some fact that he makes true by his own deliberate, avoidable, and *morally indefensible* behavior, he has disengaged from joint deliberation with A. In effect, B demands that A accept some indefensible aspect of his behavior as a brute datum, for A to work around. By constraining the courses of action jointly available to A and B in this manner, B can no longer purport to be proposing the most plausible answer to the question “what are *we*, A and B, to do in circumstances c^* ?” This is why satisfying Cohen's Principle is a necessary condition for successful interpersonal justification.

This concludes my general account of the concept of interpersonal justification. To arrive at a fully fleshed-out *conception* of interpersonal justification, we must fill in certain elements. In particular, what counts as a more plausible answer to the question “what are we, A and B, to do in circumstances c^* ?” will depend on one's substantive moral theory. For the purposes of this article, we are interested in the contractualist's answer. For the contractualist, how A and B morally ought to act in circumstances c^* is a function of what principles governing their actions could or could not be reasonably rejected. A and B morally ought to choose some course of action φ_A ; φ_B just in case there is someone who could reasonably reject a principle licensing A and B *not* to do φ_A ; φ_B .

Armed with this account of interpersonal justification, we are now in a position to give a deeper explanation of the sources of speaker-relativity in Cohen's Kidnapper case. Cohen's kidnapper attempts to justify to Smith that he ought to pay him the ransom. But, coming from the kidnapper, this attempted justification fails, and, I will now argue, *must* fail. For one thing, as we have already seen, the kidnapper's argument falls afoul of Cohen's Principle: the kidnapper cannot morally defend the fact that he makes his argument's second premise true. Smith's daughter could reasonably reject a principle that licenses the kidnapper to kill her if he does not receive a ransom.

But, I believe, there is a further reason why the kidnapper is unable to justify to Smith that he ought to pay him, which goes beyond Cohen's analysis. For lack of a better term, let us say that attempts at interpersonal justification that violate Cohen's Principle are, for that reason, *procedurally defective*. Whether or not B *could* justify to A that

he ought to do φ , *this* is not the way to do it. However, the fact that B's actual argument for the conclusion that A ought to do φ fails Cohen's Principle (and is therefore procedurally defective) does not entail that this speaker could not, in principle, provide A with a successful justification for why he ought to do φ .²⁸

What a contractualist analysis of the Kidnapper case reveals is the fact that there is more to the failure of interpersonal justification in this case than the procedural matter of violating Cohen's Principle. Rather, as I will now show, the kidnapper's case for why the father ought to pay him is *substantively defective*: it is impossible for the kidnapper to successfully justify to Smith that he ought to pay him; for the most plausible answer to the question "what are *we*, you and I, to do in these circumstances?" is not one in which the father pays the kidnapper.

In the Kidnapper case, there are three possible courses of action for Smith and the kidnapper:

- (I) Smith pays the ransom and the kidnapper releases Smith's daughter.
- (II) Smith does not pay the ransom and the kidnapper kills Smith's daughter.
- (III) Smith does not pay the ransom and the kidnapper releases Smith's daughter.

28. Consider an overzealous campaigner against tax fraud who seeks to get known tax evaders to pay up by threatening harm against their family members if they do not pay. Like Cohen's kidnapper, he might cloak his threat in the guise of a moral argument:

- (1) It would be a very bad thing if I harmed your family members.
- (2) I will harm your family members unless you pay your taxes.

Therefore, just for that reason, you ought to pay your taxes.

Though, as in the case of Cohen's kidnapper, the vigilante's threat gives the tax evader a moral reason to pay up, his argument does not constitute a successful justification. It fails Cohen's Principle, because it relies on a premise that it is morally indefensible for the speaker to make true: even if the tax evader fails to pay his taxes, it would be morally impermissible for the vigilante to harm his family members. But although this argument fails as a justification for why the tax evader ought to pay, this does not entail that the vigilante could not, by means of a different argument, justify to the tax evader that he ought to pay. Surely, the most plausible answer to the question "what are *we*, you and I, to do in these circumstances?" would, in this situation, include the fact that the man pays his taxes.

It is clear that Smith's daughter can reasonably reject a principle that licenses option (II). This course of action imposes on her a far greater burden, namely, death, than either of the other two courses of action would impose on anyone.

But, for similar reasons, Smith can reasonably reject a principle calling for option (I). Smith can reasonably reject option (I) since it imposes a greater burden on him—the cost of paying the ransom—than another available course of action, (III), imposes on *anyone*. By contrast, the kidnapper's argument in support of option (I) is marred by its extreme *hypocrisy*. The kidnapper seeks to show that Smith is morally required to pay him the ransom by appealing to a value—the well-being of Smith's daughter—which he himself is willing to disregard entirely. Double standards of this kind, Scanlon would agree, are a paradigm of unreasonableness in moral argument. The kidnapper fails to treat the father as his moral equal, by unfairly and avoidably imposing on him the burden of protecting a value that he himself is prepared to wantonly disregard.²⁹ This is something that the father could reasonably reject, for “singl[ing] others out, without justification, for a privileged moral status.”³⁰ “Quod licet Iovi non licet bovi” is not an acceptable attitude among persons of equal moral standing.

I conclude that the only principle governing how Smith and the kidnapper should act in c^* that no one can reasonably reject is one that calls for course of action (III).³¹ It follows from this that the kidnapper cannot justify to Smith that he ought to pay the ransom.

What now needs to be explained is why there is *speaker-relativity* of justification in this case: Why can the *friend* justify to Smith that he ought to pay, when the kidnapper was not able to do so? According to the joint deliberation account defended above, the friend can justify to

29. For a recent analysis that locates the central moral wrong of hypocritical moral address in a failure to treat the addressee as one's moral equal, see R. J. Wallace, “Hypocrisy, Moral Address, and the Equal Standing of Persons,” *Philosophy & Public Affairs* 38 (2010): 307–41.

30. Scanlon, *What We Owe to Each Other*, p. 219.

31. This includes Smith's daughter. What she has personal reasons to care about—her own survival—would be safeguarded if Smith and the kidnapper follow course of action (III) just as much as under course of action (I). The fact that the kidnapper predictably will not abide by course of action (I) is not germane in this context. In evaluating whether a moral principle can reasonably be rejected, we must look at what will happen if the principle is *followed*, not if it is *disobeyed*.

Smith that he ought to pay just in case the friend can show Smith that his paying the ransom is part of the most plausible answer to the question “what are *we* to do in the relevant circumstances *c*?” To keep matters simple, let us assume that Smith’s friend cannot *independently* influence the kidnapper’s behavior and could not himself pay the ransom (so that possible questions about sharing in the cost of the ransom do not arise). Joint deliberation about the best course of action for *these two* then reduces to thinking about what Smith himself ought to do in these circumstances.

Moreover, and crucially, since joint deliberation takes place between Smith and the *friend*, and not Smith and the *kidnapper*, this alters what must be taken as given and counted as part of the circumstances of action. Specifically, since neither Smith nor the friend can control the kidnapper’s behavior, the fact that the kidnapper will kill Smith’s daughter unless paid a ransom forms part of the background circumstances *c*, against which the friend and Smith must jointly deliberate.

There are thus two courses of action available to Smith and his friend:

- (IV) Smith pays the ransom.
- (V) Smith does not pay the ransom.

Given these options, the friend can successfully justify to Smith that he ought to pay the kidnapper. A principle licensing Smith not to pay could be reasonably rejected by Smith’s daughter: Smith’s refusing to pay will foreseeably get her killed. Hence, Smith cannot make a successful counterproposal to the friend’s exhortation to pay, and therefore the friend’s justification succeeds.

I have discussed Cohen’s Kidnapper case at some length, but not because I believe that this instance of speaker-relativity poses, *in itself*, a problem for Scanlonian contractualism. There is a straightforward contractualist explanation for why it would be morally wrong of the father not to meet the kidnapper’s demands: for much the same reason that the *friend’s* justification succeeds, Smith’s *daughter* could herself justify to Smith why he must pay the kidnapper. Moreover, in making this argument, the daughter would only be appealing to *personal* reasons. Hence, the Cannot Justify/Cannot Reject Principle from Section 4 is met in this case, and Smith’s daughter could reasonably reject a

principle licensing her father not to pay. If Smith refused to accede to the kidnapper's threat, his daughter could claim to have been wronged by his intransigence.

Rather, in the context of my present argument, the value of Cohen's case is that of a stepping-stone. The insights we have gleaned in discussing Kidnapper will enable me to show why, in certain other instances of speaker-relativity, which I term cases of *deliberate self-harm*, there is *no one* who could reasonably reject a principle that licenses an intuitively wrong action.

VII. THE PROBLEM OF DELIBERATE SELF-HARM

In Cohen's Kidnapper case, contractualism had no difficulty in accounting for our intuition that Smith acts wrongly if he refuses to meet the kidnapper's demands. For there existed an individual standpoint—that of Smith's daughter—from which any principle permitting Smith not to pay the kidnapper could be reasonably rejected.

Cases of deliberate self-harm, by contrast, are cases in which the agent can foresee that unless she does some action φ , another person will deliberately harm *himself* (but no one else). In such cases, the person threatening harm and the person who will suffer if we do not accommodate his threat are, thus, *one and the same*. Consider the following variation on Cohen's Kidnapper case:

The Man Who Took Himself Hostage: Jones, a single man with no dependents or friends, comes to your doorstep. Brandishing a pistol, he credibly threatens to kill himself unless you pay him \$20.

Let us stipulate that Jones's threat is credible—but not because he is insane or has in some way disabled his will, so that he will “automatically” shoot himself if you do not pay him. Instead, assume that (for whatever reason) Jones simply is stubbornly determined to harm himself if he does not get your \$20.³²

32. To make this, admittedly fanciful, case psychologically more credible, you might imagine that, in making his demand, Jones is motivated not by personal greed, but by some cause in which he strongly believes (but which you have no independent duty to support); for instance, suppose that Jones wants to donate the money to his church.

Like the kidnapper's actions in Cohen's example, Jones's behavior is nothing if not a form of extortion. And as in Kidnapper, this fact might be thought to give you a *pro tanto* reason against ceding to Jones's threat. Nonetheless, given the great evil that Jones's death would represent, I believe a strong case can be made that it would be all things considered wrong of you to refuse Jones's demands. The \$20 is only a small sacrifice for you, and by giving this money to Jones you have it in your power to avert a terrible human tragedy. As I noted in Section 4, we need not be consequentialists to hold that if not doing φ would foreseeably lead to an outcome that is very bad, this sometimes gives us conclusive moral reasons to do φ . Nonconsequentialists, too, should concur that when we can avert an awful outcome at a trivial cost to ourselves or others, and there are no other ends of comparable moral importance in play, it would be morally wrong not to do so. The problem for Scanlonian contractualism is that it seems committed to rejecting this strong intuition in the present case.

Recall that for an act or omission to be wrong according to Scanlon's formula, there must be some individual who has *personal* reasons for rejecting a principle that licenses us to perform the act or omission. Someone, that is, must be able to *justify* to you, appealing only to personal reasons, why you ought not to perform the act or omission. But as I shall now show, unlike in the Kidnapper case, there is no one who could claim this in The Man Who Took Himself Hostage.

By assumption, Jones has no dependents or friends who might suffer as a result of his death. Therefore, if *anyone* has a personal complaint against your refusal to pay, it would have to be Jones himself. My challenge, then, is to explain why Jones could not reasonably reject a principle that permitted you to ignore his threat.

It is here that the parallel to Cohen's Kidnapper case becomes relevant. Suppose Jones were to attempt to justify to you that you ought to pay him, since refusing to pay would impose a great burden on him. His argument might take the following form:

- (1) It would be very bad for me if I killed myself.
 - (2) I will kill myself unless you pay me \$20.
- Therefore, you ought to pay me \$20.

I maintain that, coming from Jones's lips, this argument does not constitute a successful justification for why you ought to pay him. The reason is not, I submit, that Jones violates Cohen's Principle in advancing his argument. While foolish in the extreme, it would not be *morally unjustifiable* for Jones to kill himself if you refuse to pay him (at least not according to Scanlon's contractualism).³³ Nonetheless, Jones's attempted justification is *substantively* defective, because it does not pass the joint deliberation test of the previous section: as a mooted answer to the question "how are we, you *and* I, to act in these circumstances?" Jones's proposal is hostage to a counterproposal that he cannot reasonably reject.

There are three courses of action available to you and Jones:

- (I) You pay Jones, and he does not kill himself.
- (II) You do not pay Jones, and he kills himself.
- (III) You do not pay Jones, and he does not kill himself.

Option (II) imposes on Jones the severe burden of premature death. Now, if Jones were somehow "programmed" to shoot himself unless you paid him \$20, he would have a strong case for reasonably rejecting a principle that licensed you not to pay. Under the actual circumstances, however, Jones's attempted justification appears preposterous. Given the ready availability of option (III), Jones cannot deny that if he dies, the proximate cause of his death will have been his deliberate decision to shoot himself. Moreover, Jones's ability to choose appropriately and avoid death is in no way impaired by your refusal to pay him. In killing himself, Jones may be acting foolishly. But his action is entirely *deliberate*. When he puts the gun to his head and pulls the trigger, Jones acts with a full awareness of the harm that this will cause him; indeed, his action has no other purpose.

In this connection, it is important to stress the disanalogies between cases of deliberate self-harm, such as The Man Who Took Himself Hostage, and what we may call cases of *unintentional self-harm*, such as the following:

Complicated Policy: A policy maker is contemplating the implementation of a policy that has the potential to significantly improve the

33. A detailed explanation of this claim is given in the following section.

lives of his constituents. Every citizen who responds correctly to the policy once it is promulgated will become better-off. The policy, however, is complicated, and easy to misunderstand. As a result, it is foreseeable that a sizable percentage of the politician's constituents will fail to respond correctly. These citizens (call them the "noncompliers") will foreseeably become very much worse-off as a result of their failure to respond appropriately—so much worse-off, in fact, that implementing the complicated policy would lead to a far worse overall outcome than not implementing the policy.

Like Jones, the noncompliers in Complicated Policy could have avoided the harm that befalls them by acting appropriately. But unlike Jones, their self-harm is not deliberate. Rather, it springs from features of their psychological constitutions—their disposition to *misunderstand* the complicated policy—that are beyond their control. It is precisely for this reason that the noncompliers in Complicated Policy can, in good faith, appeal to their likely future self-harm, as a reason for rejecting the policy. They (or an advocate for their cause) might complain: "Given that we are disposed to misunderstand this complicated policy and consequently harm ourselves, a principle that allows you to promulgate this policy would impose a significant burden on us." In the face of this argument, the retort "you don't *have* to harm yourselves" seems weak. It is tantamount to saying: "you don't *have* to misunderstand." But misunderstanding (like forgetting, or acting compulsively) is not a deliberate action that we can refrain from at will.

Compare this to Jones's complaint: "Given that I will deliberately shoot myself if you do not pay me, and this would be very bad for me, a principle that allows you not to pay me would impose a significant burden on me." Here, it *does* seem a perfectly adequate retort that Jones does not, after all, *have* to harm himself.

Jones's attempt at justification—like that of Cohen's kidnapper—is rendered unreasonable by its blatant hypocrisy. As we saw above, it would be hypocritical of Cohen's kidnapper to appeal to the child's well-being in attempting to justify to Smith that he ought to pay him; for the very threat to the child against which the kidnapper demands that Smith protect her by paying him only arises because *he himself* refuses to govern his future actions by a concern for the child's well-being. An analogous objection can be made to Jones's complaint: Jones

could not reasonably demand that you pay him \$20 out of a concern for his own well-being. Why? Because the very threat against which Jones demands that we protect him out of a concern for his well-being only arises because *he himself* refuses to govern his future actions by a concern for his well-being. Jones's proposed course of action is unreasonable, because he enjoins you to make a material sacrifice in order to protect a value that he himself could safeguard at no cost at all, but that he threatens to deliberately undermine.

I conclude that Jones lacks the appropriate personal grounds to reasonably reject a principle that licenses you not to pay him. Jones is not "burdened" by such a principle, since even if you refuse to pay, Jones remains entirely free to avoid any ill consequences for himself, simply by deciding not to shoot himself.³⁴ I believe that a Scanlonian contractualist is thus forced to embrace the counterintuitive proposition that it would not be morally wrong to refuse Jones the small sum of money he demands, even though doing so could avert a foreseeable human tragedy. For most people, I have stressed, this is a troubling conclusion.

VIII. OBJECTIONS AND REPLIES

It might be objected that the case of The Man Who Took Himself Hostage is psychologically too fanciful to serve as the basis of a serious critique of contractualism. No one, surely, is *so* irrational that he would kill himself if he did not receive \$20!

This criticism, I believe, can be met. I have designed the details of my case so as to make the relevant parallels with Cohen's Kidnapper case maximally perspicuous. If some readers feel that this renders my case unbelievable, it is easy to think of other examples that share the same structure: from unhappy lovers threatening to kill themselves if their partner leaves them to political hunger strikes aimed at forcing a politician's hand, the real world is replete with cases in which agents seek to apply moral pressure to others by threatening harm to themselves. Consider the following case:

34. Introducing a spokesperson or proxy for Jones would not solve this problem. Recall that for Scanlon, grounds for reasonable rejection "must be reasons that ... a person would have 'on his or her own behalf.'" Since Jones could not reasonably reject a principle licensing you not to pay on his own behalf, it follows that neither could a spokesperson do so in Jones's name.

Political Suicide: A politician is contemplating implementing a policy that would improve the lives of most of his constituents. A small group of extremists, however, is virulently opposed to the policy for ideological or religious reasons. Fifty members of this group have credibly threatened that if the policy is implemented, they will publicly take their lives in protest. Given this threat, we can assume that the consequences of implementing the policy will, on the whole, be very much worse than the consequences of not implementing the policy.

Like Jones, the extremists are engaged in a form of moral blackmail. By threatening to take their own lives, they are seeking to pressure the politician into abandoning his policy. But, like Jones, the extremists could not deny that if any harm befalls them, this will be because they decided to deliberately harm themselves—*not* because this was, in any way, an ineluctable consequence of enacting the policy. It would seem, therefore, that the extremists in Political Suicide could not reasonably reject a principle permitting the politician to enact his policy, based on how it affects them. Once again, it appears that the contractualist is committed to the conclusion that refusing to cede to the moral blackmail could not be wrong.

Again, however, this conclusion seems too strong. Our moral reasons for accommodating the extremists' threat are certainly *weakened* by the fact that their deaths would be the result of their own *deliberate* actions.³⁵ Nonetheless, if ignoring their threat would lead to an awful outcome, we intuitively have at least a *prima facie* obligation to accede to their demands. If the costs of accommodation are very small, this may even be what we morally ought to do all things considered.³⁶

The Political Suicide case can also be used to counter a different rejoinder that a contractualist might make to my critique of his

35. I shall say more about how speaker-relativity affects the strength of our moral obligations in the following section.

36. Of course, as in most instances of symbolic political violence, the extremists in Political Suicide believe that the politician has independent moral reasons, unrelated to a concern for their welfare, to cede to their demands. Probably one aim of their protest is, precisely, to draw the public's attention to these supposed reasons. Sometimes such independent reasons for accommodation may well exist. The point is, however, that even if (as in my example) such reasons do *not* in fact exist, we may still have a *prima facie* obligation to accommodate the extremists' threat. It is this conclusion that contractualism cannot share.

position. In response to the original case of The Man Who Took Himself Hostage, a contractualist may suggest, in an ecumenical spirit, that while it would not be morally wrong to refuse Jones the money (despite the foreseeably terrible consequences of doing so), it would certainly be good and generous to cede to his demand.³⁷ On this view, acceding to Jones's demand is *supererogatory*. Though an agent could reasonably reject a principle that *required* him to give Jones the money, it may well be morally praiseworthy to do so.

Though promising at first blush, this reply falters in a case like Political Suicide, where the politician must sacrifice not just his own well-being, but that of *other* citizens as well, in order to prevent the extremists from harming themselves. Not implementing his policy because of the extremists' threat means imposing on the rest of the citizenry a burden that the extremists could not justify to them. These burdens give the citizens grounds for rejecting a principle that licenses the politician to retract his policy. The extremists, by contrast, lack reasonable personal grounds for rejecting a principle that licenses the politician to implement his policy. Under these circumstances, where the citizens' justified personal complaints against retracting the policy are not counterbalanced by weightier personal reasons for accommodating the extremists' demands, the citizens could reasonably reject a policy of accommodating the extremists' demands. Since the reasonable rejection test is Scanlon's criterion of moral wrongness, a contractualist who stays true to his view is committed to the conclusion that accommodating the extremists in this case would not be supererogatory, but *must* be morally wrong. This, I have argued, is an implausible conclusion to reach.

A third objection to The Man Who Took Himself Hostage case contends that by accepting that it would be morally wrong to turn down Jones's demand for \$20, you would, in effect, be opening yourself up to limitless opportunities for future moral blackmail. Surely no plausible moral theory should require us to be *suckers*?

Here is a first quick response to this objection: while a government may regard it as a weighty consideration against paying hostage takers that this may lead to more hostage takings occurring in the future, a parallel argument has considerably less traction in the case of The Man

37. T. M. Scanlon proposed this response to me in conversation.

Who Took Himself Hostage and other cases of deliberate self-harm. The psychological constitution required to make a credible threat against *oneself* is quite different from, and far harder to acquire than, that required to credibly threaten harm against others. Even if you revealed a willingness to accommodate Jones's threat of deliberate self-harm, it is therefore unlikely that this would cause such behavior to become dramatically more widespread in the future. Unlike Jones, most people simply lack the psychological constitution that would allow them to capitalize on your willingness to pay in such cases.

Derek Parfit and Doug Lavin have suggested to me the following rejoinder on behalf of proponents of the sucker objection: Even if the psychological constitution required to pull off a credible threat of self-harm is not very widespread, we do know that *Jones* has it. If we give in to his threat this once, what is there to stop him from coming back tomorrow and every day in the future, threatening each time to shoot himself unless we pay him \$20?

I believe that rather than raising a *sui generis* problem for our moral intuition in The Man Who Took Himself Hostage case, the Parfit-Lavin rejoinder touches on a far more general problem in the theory of beneficence—that of *iteration*. We often accept that doing φ would be morally obligatory when doing φ would save someone else from a great harm, comes at little cost to ourselves, and could be performed only by us in the circumstances. Nonetheless, we balk at the thought that we might be obliged to perform φ an *indefinite* number of times whenever the situation is relevantly similar to the one in which we thought ourselves first obliged to do φ . Consider the following illustration from Garret Cullity:³⁸ We all agree that in Peter Singer's famous "drowning child" example, we have a moral obligation to wade into the pond and rescue the child. But now imagine the nightmarish vision of a huge sea, full of an indefinite number of drowning children, each of whom we could save at a cost that is individually negligible. It is much less clear that we must spend practically every waking moment of our lives saving these children. But, surely, this should not weaken our original conviction that it would be wrong not to save the single child in Singer's case.

38. Garret Cullity, *The Moral Demands of Affluence* (Oxford: Oxford University Press, 2004).

I believe that we would have an even stronger moral justification for resisting iteration in *The Man Who Took Himself Hostage*, without having to give up our original judgment. For in contrast with Garret Cullity's example, we would here be called on not to save not a large number of *different* lives, but the same life *over and over* again.

The problem in an iterated version of *The Man Who Took Himself Hostage* is that if we know that if we save the man now by giving him \$20, he is going to return the next day and again make the same threat, we are achieving much less by saving him today than we might have thought. Instead of ensuring that the man will live "happily ever after," we are merely buying time—one more day, to be precise. But, of course, it is much less clear that we are morally required to pay \$20 to postpone a threat to someone's life by just one day than to avert it indefinitely. Hence, if and when it becomes clear that we are facing an iterated version of *The Man Who Took Himself Hostage*, it may well be permissible to stop giving money. But as I argued above by analogy with Singer's drowning child example, this intuition must not be thought to simply carry over into the *noniterated* original case.

I have so far been writing under the assumption that if Jones carries out his threat to kill himself, his action—while foolish and misguided in the extreme—is not morally wrong in the contractualist sense, since (ex hypothesi) Jones's death affects no one negatively but himself. This assumption, however, could be challenged; and, if successful, this challenge might then form the basis of a final attempt to defend contractualism against the force of my argument in the preceding section. For suppose it could be shown that by shooting himself, Jones acts morally wrongly, namely, by violating a duty *to himself*. Jones's *threat* to kill himself would then confront you with a problem of *partial compliance*: how to respond to immoral behavior, or the threat of such behavior, on the part of another agent. Further, it might be suggested that if Jones's killing himself would be not just foolish but morally wrong, this would give you an especially weighty moral reason for acceding to Jones's threat, lest you yourself become *complicit* in his immoral action. Indeed, perhaps Jones himself could reasonably reject a principle that allowed you to go along with his immoral scheme. So perhaps the contractualist could, after all,

account for our intuition that you act wrongly by refusing to pay Jones the \$20 he demands.³⁹

The first thing to say in response to this attempt to defend contractualism is that Scanlon himself would be unlikely to endorse its initial assumption. Scanlon expressly doubts that there are duties or wrongs to oneself in the contractualist sense.⁴⁰ According to Scanlon, “the part of morality that [his] contractualism describes is essentially interpersonal.”⁴¹

Scanlon is willing to allow that some instances of suicide might be called “immoral” in a broader sense. He writes:

Just as murder shows a lack of respect for human life, there can be cases in which suicide does so as well. We might say, of a person who commits suicide out of a cynical conviction that nothing is worth doing, or out of disappointment at being rejected by a lover, that he or she showed a lack of understanding of the value of life and allowed it to be wasted. [What makes such suicides criticizable] is a failure to see the reasons they have to go on living, reasons provided, for example, by their possible accomplishments, by the good they might do for others, and by the pleasures they could attain.⁴²

But such failure to respond appropriately to one’s reasons to go on living, Scanlon thinks, is not wrong in the narrower sense of “morally wrong” that contractualism is describing. As he stresses, an agent’s having decisive reasons against performing some action is compatible with that action’s being allowed by a valid moral principle. The moral permissibility of the action “just reflect[s] the absence of a specific kind of further reason not to do such things: we don’t owe it to anyone not to do this.”⁴³

39. I am grateful to Jay Wallace for pressing me to consider this objection.

40. See T. M. Scanlon, “Replies,” *Social Theory and Practice* 28 (2002): 337–58, in particular the section “Duties to Oneself,” pp. 349–52. In that article, Scanlon’s main example of a putative duty to self is a duty to not let one’s talents go to waste. But Scanlon’s reasoning is quite general, and can be straightforwardly applied to the case at hand (as I have done in the following). Indeed, in a brief passage from *What We Owe to Each Other* that I quote below, Scanlon touches on the issue of suicide, and suggests that killing oneself, absent its potential effects on other persons, does not constitute moral wrongdoing in the contractualist sense.

41. *Ibid.*, p. 349.

42. Scanlon, *What We Owe to Each Other*, pp. 104–5.

43. Scanlon, “Replies,” p. 350.

But why not say that Jones *does* owe it to someone not to kill himself—namely, to himself? If properly understanding the value of his life would lead Jones to realize that he has strong reason not to kill himself, then is not killing himself under these circumstances something that Jones could not justify to himself? If so, would his action not be wrong in the contractualist sense as well?

Scanlon considers this line of argument, but rejects it. The fact that Jones's killing himself is not something that he could successfully justify to himself "is a mere consequence of the prior claim about the reasons [he has to go on living]. It provides no support for this prior claim and adds nothing to it."⁴⁴ This is in contrast with the way that Jones's reasons to go on living would give him grounds to reasonably reject a principle that allowed *others* to *kill him*. In such a case, the reasons that Jones has to go on living give rise to a *further* reason for others not to kill him. It is not just that Jones's life has value. In addition, we *owe it* to him not to kill him.

Suppose, however, that Scanlon is wrong to deny that Jones would act wrongly, in the contractualist sense, by killing himself. How would this bear on what *you* (the person blackmailed by Jones) ought to do in this case? Does the claim that Jones would be acting immorally as opposed to merely foolishly alter the strength of *your* reasons to give in to Jones's threat? That is, over and above the fact that it would be *bad for* Jones to lose his life, does the fact that Jones would *wrong himself* in taking his life add weight to your reasons to prevent this from happening, as the above objection assumes?

I believe we should answer this question in the negative. To see why, consider, first, a more general question: Does the fact that a person would suffer a harm as the result of *wrongdoing* (where it is not *you* who is inflicting the wrongful harm) give you a stronger moral reason to prevent this harm than to prevent a person from suffering a harm of the same size that is not the result of wrongdoing?

I believe it is a mistake to think that which of two harms one should prevent is a function of which would involve greater wrongdoing. Preventing others from doing wrong matters primarily because we thereby prevent the *consequences* of their wrongdoing. Cases where we have reason to be concerned with preventing wrongdoing, independently of

44. *Ibid.*, pp. 350–51.

reasons to prevent its consequences, are primarily cases where we *ourselves* would be acting wrongly. We have special *agent-relative* reasons to avoid *doing* wrong. But we have little or no reason to prevent wrongdoing by others *apart* from the reasons we have to prevent the bad consequences of wrongdoing.⁴⁵

Scanlon seems to concur, when he writes that

even those who believe that there is a special prohibition against intentional killing do not think that one must, for this reason, prevent [a] murder rather than [a fatal] accident. The idea that the murder is a much worse thing to have happen would seem to imply that one must try to prevent it, even if one has a slightly greater chance of being able to prevent the accidental death. But this seems wrong.⁴⁶

If Scanlon is correct, then the assumption that Jones would be acting wrongly, as opposed to merely foolishly, does not strengthen your reasons to prevent him from killing himself.

Indeed, if it were the case that Jones would act wrongly by killing himself, I believe this would *weaken*, not strengthen, Jones's case for rejecting a principle that allowed you not to pay him. In Section 7, I argued that Jones could not reasonably reject a principle licensing you not to give in to his demand, since in doing so he would be requiring you to bear a cost in order to protect a value—his well-being—that he is able to safeguard at no cost to himself, but that he is willing to deliberately endanger. Now, if, in addition, Jones would be violating a duty to self by killing himself, this would put him in an even *worse* position to justify to you why you ought to pay him. For, in that case, his attempt at justification would also fall afoul of Cohen's Principle. That principle, recall, states that, in trying to justify to you why you ought to act in a certain way, an agent can appeal to some fact *p* that he himself will deliberately and avoidably cause to obtain, only if he can also justify *making it the case* that *p*. The relevant fact, in our case, is that Jones will shoot himself if he is not paid. But if it is morally wrong for Jones to shoot himself if he is not paid, this is something that he could not

45. I am indebted to Jeff McMahan for a discussion of these points.

46. Scanlon, *What We Owe to Each Other*, p. 83.

justify. Hence, under Cohen's Principle, he could not appeal to it as a premise in an argument meant to justify that you ought to pay him.

In sum, I maintain, against these objections, that cases of intentional self-harm pose a serious challenge to the contractualist account of moral rightness. On the one hand, considerations of supererogation and iteration do not explain away or erase the intuition that it would be *prima facie* morally wrong not to give in to the blackmailer's demand. On the other hand, even a modified contractualism allowing for duties to self would have to conclude against that moral intuition.

IX. TOWARD A PLURALIST CONCEPTION OF MORAL RIGHTNESS

The Man Who Took Himself Hostage and Political Suicide therefore present the contractualist with precisely the problem that I outlined at the end of Section 4. Omitting to perform an action φ is morally permissible, according to the contractualist formula, because doing so is licensed by a principle that no one could reasonably reject from a personal standpoint; nonetheless, failing to do φ is intuitively morally wrong, at least on some ways of fleshing out these cases, in virtue of the bad consequences that refusing to do φ would have.

This discrepancy between the deliverances of the contractualist formula and our moral intuitions confronts the contractualist with the following dilemma: *Either* he bites the bullet and concedes that in some cases, contractualism has openly counterintuitive implications. By making this concession, however, the contractualist risks forfeiting one of his theory's major selling points, namely, the right to claim that it is *consistently* better at tracking our moral intuitions than its consequentialist rival. *Or* he admits that the two halves of the contractualist formula—the *analysans* (reasonable rejectability) and the *analysandum* (moral wrongness)—can sometimes come apart. That is, he accepts that there are cases where it would be morally wrong to refuse to do φ , despite the fact that a principle permitting us not to do φ could not be reasonably rejected from an individual standpoint.

My advice to the contractualist is to take the second horn of this dilemma. What my discussion of the problem of deliberate self-harm suggests is that our moral intuitions embody a stronger concern for the consequences of our actions than can always be captured by Scanlon's notion of justifiability-to-each-person. Besides our desire to be able to

justify our actions to every person, we are right to be *independently* concerned with the value of their outcomes. For this reason, Scanlon's theory would, I believe, be greatly strengthened if, instead of maintaining that his contractualist formula provides an exhaustive account of *what it is* for an action to be right or wrong in the domain of interpersonal morality, he were to claim: (i) that being disallowed by some unrejectable principle is one, but not the *only*, property that *makes* an action wrong; and (ii) that being permitted by an unrejectable principle is one, but not the *only*, property that *makes* an action right.⁴⁷

This revision would allow Scanlon to maintain that even in our dealings with those beings whom we owe a justification for our actions, there are other properties besides justifiability to each person—such as the goodness or badness of our action's consequences—that independently contribute to making our actions right or wrong. According to this *pluralist* proposal, while the contractualist formula captures an important class of pro tanto moral reasons that *contribute* to making actions right or wrong, reasonable rejectability in the contractualist sense is not the only relevant consideration in determining whether an action is right or wrong *all things considered*. When these different right- or wrong-making properties conflict, they must be weighed against each other. And sometimes the fact that an act would have very bad consequences might morally outweigh the fact that it is licensed by a principle that no one could reasonably reject from an individual standpoint.⁴⁸

47. The suggestion that Scanlon should claim that his reasonable rejection criterion gives us not an account of what it *is* for an act to be wrong but a higher-order description of some of the wrong-making properties of actions is advanced, for broader reasons, by Derek Parfit in *On What Matters*, vol. 1 (Oxford: Oxford University Press, 2011), pp. 368–70. More precisely, according to Parfit's proposal, being disallowed by a principle that no one could reasonably reject should be understood as a *higher-order* wrong-making property of actions, under which other first-order wrong-making properties, such as, for instance, being a lying promise, can be subsumed. In more recent writings, Scanlon appears willing to consider this proposal; see T. M. Scanlon, "Replies," *Ratio* 16 (2003): 424–39.

48. In her review of *What We Owe to Each Other*, "Owing, Justifying, and Rejecting," Frances Kamm argues that "what Scanlon thinks is an account of 'wrongness' is better understood as an account of 'wronging' (p. 335). I am sympathetic to this proposal, though I do not have the space to pursue it further here. Suffice it to say that Kamm's proposal is compatible with my pluralist account. The resulting picture would be one according to which the contractualist formula does double duty: as an account of *wronging* another person and as an account of a central class of pro tanto moral reasons that contribute to making an action wrong *all things considered*."

Such a pluralist account of moral wrongness, I believe, proves superior at tracking our moral intuitions about cases of deliberate self-harm than both the pure consequentialist and the pure contractualist accounts. Consider again the case of The Man Who Took Himself Hostage. An act-consequentialist, who holds that the rightness of an action is determined exclusively by the value of its outcome, and who therefore rejects the phenomenon of speaker-relativity, is committed to maintaining that our moral reasons for giving Jones the money he demands in The Man Who Took Himself Hostage are *as strong* as in a case where Jones needed \$20 to save himself from some external threat. (Imagine Jones has contracted a deadly virus, and you are the only person who can give him the \$20 he needs to buy the antidote. Call this case *Antidote*.) This, I think, is very implausible. Intuitively, you have a much more stringent obligation to give Jones the money in Antidote. Note in particular that while you may be acting wrongly by refusing to give Jones \$20 in either scenario, only in Antidote would it seem that you also *wrong* Jones by turning down his request. In The Man Who Took Himself Hostage, Jones is in no position to complain if we fail to meet his demand, since if anything bad happens to him as a result, this will be only because he deliberately decided to make it happen. Consequentialism's account of The Man Who Took Himself Hostage fails to give this important fact any normative weight.

Contractualism, as we have seen, faces the reverse criticism. It takes the fact that Jones could not reasonably reject a principle allowing you not to pay him *so* seriously that this is taken to imply that you could not be acting wrongly if you refused to accede to Jones's demand—even if doing so would have awful consequences. This, too, is an implausible conclusion. While consequentialism generates counterintuitive results because it takes only impersonal reasons into account, contractualism's exclusive focus on personal reasons for rejection equally blinds it to morally relevant features of this case.

The pluralist view that I am proposing avoids the shortcomings of both pure consequentialism and pure contractualism in dealing with the problem of self-harm. It can allow that the foreseeably bad consequences of not giving Jones \$20 give us *some* moral reason to give Jones the money in both Antidote and The Man Who Took Himself Hostage. At the same time, the fact that Jones's demands in the latter case are

unreasonable and place an unfair burden on you is seen as reducing the strength of your moral obligation. The pluralist view thus implies that the maximum amount of money that you could be required to give to Jones in a case like *The Man Who Took Himself Hostage* is considerably lower than what you could be called on to give him under the conditions described by Antidote. And this, it seems to me, is precisely the right conclusion to reach.

Finally, let me point out that the pluralist view I am advocating is far from being an ad hoc solution to an isolated problem for contractualism. On the contrary, combining Scanlon's contractualist account with an independent concern for the goodness of the consequences of our actions seems the key to solving a number of other problems that have been raised for contractualism in the recent literature. Thus, in a recent article for this journal, I have argued that contractualists will be able to accommodate certain plausible judgments concerning risk imposition and moral aggregation, only if they adopt precisely the kind of pluralist view that I have here defended.⁴⁹ Likewise, Derek Parfit, in volume 2 of *On What Matters*, argues cogently that the "Non-Identity Problem," which he famously raised in *Reasons and Persons*, cannot be solved with the resources of Scanlonian contractualism.⁵⁰ The *explanandum* in nonidentity cases is our intuition that it is morally wrong to deliberately create a future person with a life that is worth living, but whose life will go significantly less well overall than the life of a numerically different person whom we could have created instead at no greater cost to anyone. We have this intuition, despite the fact that creating this person is not bad or worse for anyone who ever lives and, arguably, cannot be said to *wrong* anyone by harming them or violating their rights.⁵¹ Given these features of nonidentity cases, it is hard to see how anyone could have personal reasons for rejecting a principle licensing the agent to perform the intuitively wrong action. Pure contractualism thus seems ill equipped to handle the Non-Identity Problem. By contrast, amended along the

49. See Johann Frick, "Contractualism and Social Risk," *Philosophy & Public Affairs* 43 (2015): 175–223, in particular Sections 7–9.

50. Derek Parfit, *On What Matters*, vol. 2 (Oxford: Oxford University Press, 2011), in particular Sections 78 and 79.

51. I defend these claims in detail in my "Future Persons and Victimless Wrongdoing" (forthcoming).

lines that I have proposed, a pluralist contractualist can provide a relatively straightforward answer: although a principle permitting the action in question could not be reasonably rejected for personal reasons, and thus, perhaps, could not be said to *wrong* anyone, the action could still be morally wrong *all things considered*, because it produces consequences that are much worse than those of an available alternative.

The fact that all these seemingly disparate problems with contractualism—the problem of deliberate self-harm, problems concerning risk imposition and aggregation, and the Non-Identity Problem—seem amenable to a common solution seems to me strong evidence that the pluralist proposal is on the right track.

X. CONCLUSION

In this article, I set out to explore some of the ethical implications of a neglected phenomenon in interpersonal moral discourse: the speaker-relativity of interpersonal justification. To demonstrate the interest of this phenomenon for normative ethics, I focused on one paradigmatic instance of speaker-relativity, the problem of deliberate self-harm. The intuitions we have about cases like *The Man Who Took Himself Hostage* and *Political Suicide*, I argued, raise new and serious problems for two prominent systematic moral theories, act-consequentialism and, in particular, Scanlonian contractualism.

Besides revealing the value of speaker-relativity as a test case for established ethical theories, I hope that my discussion has also hinted at the *constructive* potential for ethical theorizing that a detailed examination of this phenomenon may hold. Normative ethics often makes progress by turning its attention to new problems. In this article, I have argued that the difficulties that both consequentialism and contractualism have with the problem of deliberate self-harm may be remedied by shifting to a *pluralist* conception of moral rightness, which combines what appears sound in each approach.

I do not pretend to have given anything approaching a full account of the problem of speaker-relativity in this article. A fuller discussion would no doubt seek to give a more complete taxonomy of the sources of speaker-relativity of justification, and attempt to link it to the

emerging literature on a person's standing for blame, condemnation, or complaint that was briefly mentioned at the outset of the article. To the extent, however, that the best way of responding to a neglected problem is, first of all, to *raise* it, I hope that this article has at least taken a step in the right direction.