

- Mortensen, C. and G. Priest. 1981. The truth-teller paradox. *Logique et Analyse* 95–96: 381–88.
- Read, S. 1979. Self-reference and validity. *Synthese* 42: 265–74.
- Sorensen, R. 2001. *Vagueness and Contradiction*. Oxford: Oxford University Press.
- Sorensen, R. 2005a. Précis of vagueness and contradiction. *Philosophy and Phenomenological Research* 71: 679–85.
- Sorensen, R. 2005b. A reply to critics. *Philosophy and Phenomenological Research* 71: 712–28.

## *The non-consequentialist moral force of promises: a response to Sinnott-Armstrong*

MICHAEL B. GILL

That we have a duty not to break a promise is often presented as a reason to reject consequentialism. Sinnott-Armstrong has argued, however, that in fact consequentialism provides the best account of the duty not to break a promise (2009; see also Sinnott-Armstrong 1992). I will argue that Sinnott-Armstrong's argument fails. The phenomena of promising remain a reason to reject the view that all of morality consists in our duty to produce the best consequences. That is not to say that we do not have a duty to produce the best consequences. Rather, the duty to produce the best consequences is, like the duty not to break a promise, one of a plurality of moral duties.

According to Sinnott-Armstrong, the best explanation of the strength of one's duty to keep a promise is that the duty's strength is based on the consequences of breaking the promise. This is because if the consequences of keeping a promise vary, then the strength of one's duty to keep the promise will vary as well; while if the consequences of keeping a promise do not vary, the strength of the duty will not vary even if all the other circumstances surrounding the promise do vary. And, applying Mill's method of concomitant variation, Sinnott-Armstrong maintains that 'since the strength of a moral obligation goes up and down as the harms in violating it go up and down, this correlation supports the hypothesis that the harms of violating it are what makes the moral obligation as strong as it is' (2009: 440). Moreover, Sinnott-Armstrong argues, if the strength of the duty to keep a promise is based on the consequences of breaking it, then the best explanation of the existence of the duty must be consequentialist as well. The only way to rebut this consequentialist argument is 'to explain why some moral obligations are stronger than others without invoking the harmful

consequences of violating those moral obligations' (2009: 442). Sinnott-Armstrong doubts that this can be done. I will try to show that it can be, by explicating cases in which the strengths of moral obligations vary even while the consequences remain the same.

Sinnott-Armstrong bases his argument on cases in which you have promised a friend to do something that will benefit him in the future – such as take him to the airport, or meet him for lunch. The duty to keep the airport-promise is stronger than the duty to keep the lunch-promise. And, according to Sinnott-Armstrong, the best explanation of this difference is that the consequences of breaking the airport-promise are worse than the consequences of breaking the lunch-promise. But a problem with the way Sinnott-Armstrong supports his position is that the cases he uses to pump our intuitions are ones in which you have promised to perform some future action that it is very natural to think will improve the well-being of the promisee – cases in which a very salient feature of what you have promised to do is its improvement of the promisee's future mental states.<sup>1</sup> But people also care about things independent of their future mental states, and promising is a way we can become entrusted with those other things.<sup>2</sup> If we focus only on promises that very conspicuously affect the promisee's future mental states, it may be difficult to discern the non-consequentialist force a promise can exert. But when we attend to promises to do those other sorts of things, the non-consequentialist force becomes easier to spot. We can then see that the strength of the duty not to break a promise varies as the importance or value of the thing promised varies, even while the consequences may stay the same. Consider the following examples.

A has suffered severe, irreparable brain damage. He is not aware of his surroundings, has no memories, and is unable to perform even the most basic tasks. But he is not in physical pain and does not seem to be suffering emotional distress. When A was competent, you promised him that if he was ever to become severely brain damaged, you would do what you could to hasten his death so as to prevent him from lingering in that state. As it now happens,

1 Sinnott-Armstrong is arguing for consequentialism, not utilitarianism or any other view that commits to a particular kind of consequence having intrinsic value. So, his position is not dependent on welfare being the only thing of ultimate moral significance. My point here is that it just so happens that in each of the examples he uses to make his case the future welfare of an individual seems to be the most conspicuous moral concern, and this feature of his examples makes them illegitimate test cases for the question of whether future consequences are the basis of the moral obligation to keep promises. Below I will address two consequentialist responses that take the consequences to be maximized to be something other than individuals' future welfare.

2 For a powerful account of moral concerns that cannot be accommodated by consequentialist thinking, see Hill 1984.

however, to hasten A's death you will have to undertake a course of action that poses significant risks to your own well-being.<sup>3</sup>

In situation 1, A is a person who cared deeply and conspicuously about never being in an irreversible mentally-impaired state. The issue was one he went out of his way to discuss with you, and your promise to prevent him from lingering in such a state was made in a solemn moment.<sup>4</sup>

In situation 2, A is a person who did not give any more thought to the possibility of being irreparably mentally damaged than most people. The topic came up when the two of you happened to hear about a similar case on TV. A did make clear his desire not to linger in such a state, and you really did promise him that you would do what you could to hasten his death if such a situation ever came to pass. But the moment was more casual than in situation 1.

The consequences of keeping or breaking your promise are the same in 1 and 2. But your duty to keep Promise 1 is stronger than your duty to keep Promise 2. This is because A1 cared more about not being in a brain-damaged state than A2. That's not to say that A2 didn't care about or place any value on not lingering in a brain-damaged state. Nor is it to say that your promise to A2 is *faux* or lacking any moral force. If it were possible to end A2's life without any risk, the promise you made to him would obligate you to do so. But it is to say that the moral weight of the trust A1 placed in you is greater than the moral weight of the trust A2 placed in you, and that difference affects the relative strengths of the duties to keep your promises in A1 and A2 even while the consequences of keeping the two promises remain the same.

If these non-consequentialist considerations explain the differing strengths of your duties not to break the promises, moreover, then they may also constitute part of the best explanation of the existence of those duties. You are obligated to do something for A in these cases, at least in part, because A valued that the thing be done and because he entrusted you to do it, and not only because of the consequences that would flow from your keeping or breaking the promise.

3 This scenario may be thought to be akin to a deathbed promise. Sinnott-Armstrong warns against overly fanciful deathbed promise scenarios, telling us that 'we should not trust our moral intuitions in such odd cases, because they did not evolve to fit such weird circumstances' (2009: 441). I take it, though, that the cases I propose are not bizarre or even unrealistic. Moreover, unless we consider cases in which keeping or breaking a promise has no effects on the future mental states of the promisee, it will be difficult to test the non-consequentialist force of a promise, as the salience of the future mental states can interfere with our assessment. This makes consideration of deathbed-promise-like scenarios (which, as I've said, do not all seem to me to be weird) inevitable.

4 Sinnott-Armstrong writes, 'The source of strength [of a promise] is not the solemn tone in which I made the promise' (2009: 439). But the solemnity of the making of a promise can signal the importance of the thing promised, and the strength of a duty to keep a promise can be sensitive to the importance of the thing promised.

It is also true, of course, that the consequences of keeping or breaking a promise can affect whether or not you ought to keep or break it. If it had been easy to keep your promise to A2, then it would have been the case that you would have been obligated to do so. And if it had been incredibly costly for you to keep your promise to A1 (if, for instance, it would have meant your going to prison and leaving uncared for your three young children), then it may have been the case that you would not have been obligated to do so. But the best explanation of these phenomena is that there are two independent duties at play in the cases: the duty to keep your promise, and the duty to produce the best consequences. When the consequences of doing something are extremely significant, the consequentialist duty can become very strong. And the stronger that duty becomes, the more likely it will be to override the duty not to break a promise. But that does not imply that the duty not to break your promise is merely a subset of, or collapses into, the duty to produce the best consequences. Whether you ought to keep your promise in cases in which the consequences would be better if you did not keep it could instead be determined by the relative strengths of the two different duties.

A possible response consequentialists could make is that among the good things we should seek to maximize is promise-keeping itself. On this consequentialist view, the more promises that are kept, the better, because the keeping of a promise is intrinsically good. This promise-consequentialist view would imply that it is right to break a single promise if that is the only way to ensure that multiple other promises are kept but wrong to break a promise if it reduces the number of promises kept overall. There are three problems with this response. First, this promise-consequentialist view doesn't actually provide us with an explanation of the duty not to break a promise. It takes that duty to be as basic as – and thus sheds no more explanatory light on it than – Rossian pluralism. Kantianism, moreover, gives a much richer explanation of the duty than both. But Sinnott-Armstrong claims to be showing that consequentialism gives a better explanation of the duty.<sup>5</sup> Secondly, we will think the duty to keep the promise to A1 is stronger than the duty to keep the promise to A2 even when we suppose that no other promise-keepings are differentially affected by our keeping or not keeping the promises to A1 and A2. Thirdly, there is empirical evidence that commonsense morality does *not* think a person ought to break a single promise in order to ensure the keeping of multiple other promises, which contradicts what is implied by an

5 I do not think Sinnott-Armstrong would want to claim that the duty to keep a promise is explained by the goodness of that one promise's being kept (i.e. by the fact that when that promise is kept, the world is better simply because it contains one more kept promise). For as he explains elsewhere, the distinction between consequentialist and non-consequentialist moral reasons does not make sense unless we 'specify that the consequences of an act must be distinct from the act itself' (1992: 409; see also 417).

(agent-neutral) consequentialism about promises (i.e. a consequentialism that holds that promise-keepings are what we should seek to maximize) (Lopez et al. 2009).

Another possible response consequentialists might make is that you harm someone more by not fulfilling a more intense desire of hers than by not fulfilling a less intense desire. Now A1 had a more intense desire not to linger in death's antechamber than A2. So by not hastening the death of A1 you harm him more than you do by not hastening the death of A2, and this fits perfectly well with the consequentialist view that we ought to minimize harm. But in order to provide the best account of promise-keeping, the consequentialist must show not merely that A1 is more seriously harmed than A2 because his unfulfilled desire was more intense than A2's. The consequentialist must show, rather, that A1 is more seriously harmed because a promise made to him was broken. For if the consequentialist view relies merely on the intensity of the desire that was unfulfilled (and the harm that follows from that), then it will not be able to explain why your duty to hasten the death of A1 is stronger than your duty to hasten the death of A3, who is just as desirous as A1 of not lingering in death's antechamber but to whom you have made no promise whatsoever. Consequentialists may consider responding to this objection by arguing that the stronger duty to hasten the death of A1 rather than A3 is based on the duty to fulfill an additional desire of A1's, namely the duty to fulfill his *de dicto* desire that promises to him be kept. But pointing to the harm done to a promisee when that *de dicto* desire is not fulfilled fails as an account of the differing strengths of the duties to keep different promises. For that *de dicto* desire would be the same in every case of promising, and thus would provide no explanation of the differing strengths of our duty to A1 and our duty to A2. I suppose the consequentialist could then give a more complicated account such as: the difference of the strengths of your duties to A1 and A2 is explained by the differing intensity of their desires not to linger in death's antechamber, while the difference between your duty to A1 and A3 is explained by the existence in the one case but not in the other of the potential non-fulfillment of the *de dicto* desire that promises made to him be kept. But this view would fail because that *de dicto* desire does not constitute an adequate explanation of the difference between the strength of your duty to hasten the death of A1 and the strength of your duty to hasten the death of A3. For if the *de dicto* desire were the basis of that difference, then the same difference in strength would exist between [1] your duty to pick up a bottle of milk for me on the way home when you've promised me that you would, and [2] your duty to pick up a bottle of milk for me on the way home when you've made no such promise to me but when I have the same intensity of desire for a bottle of milk as in [1]. But I think it's clear that the difference between your duties in milk cases [1] and [2] is not as great as the difference between your duties in the cases of A1 and A3.

If the consequences of breaking a promise are significantly enough better than keeping it, then you may be obligated not to keep it. But if the consequences are only a teeny bit better, you may be obligated to keep it. As Ross put it,

Suppose, to simplify the case by abstraction, that the fulfillment of a promise to A would produce 1,000 units of good for him, but that by doing some other act I could produce 1,001 units of good for B, to whom I have made no promise, the other consequences of the two acts being of equal value; should we really think it self-evident that it was our duty to do the second act and not the first? I think not. (Ross 2003: 34–35)

Ross's abstract point can be made more concrete by supposing that the promise involved is to pay A back \$10 I owe him on the date that I had promised to repay him. I realize that if I gave the \$10 to B instead (to whom I have made no promise), B could do something with the money to turn it into \$10.01 tomorrow. Clearly, it would be wrong for me to give the money to B and break my promise to A. But what if by giving the \$10 to B, B could turn it into \$100,000 tomorrow? Then it might very well be morally acceptable to give the money to B and break my promise to A. I have no doubt that Sinnott-Armstrong can mount a consequentialist explanation of the phenomena Ross points to. But the two-duties-of-varying-strength view looks, at least at first sight, to be simpler and more coherent than an explanation that takes there to be only the single duty to produce the best consequences.

Sinnott-Armstrong says that his argument applies not only to promising but to other apparently non-consequentialist duties as well, such as the duty not to lie. But another pair of cases shows that the strength of duties other than promising can also vary even while the consequences stay the same.

B is a well-known person who is also your friend. You learn something personal and highly sensitive about her (maybe it's something about a medical condition of hers, or her sexual history, or her family background). You never promise B you won't reveal the secret; perhaps B doesn't even know that you know. But you do know that B prefers that no one else find out. At a later time, you are talking with a young person who is going through a very rough time as a result of personal circumstances that are similar to B's. You realize that if you tell the young person what you know about B, it will give the young person considerable solace. The young person then asks you, 'I've heard rumors that B went through something similar. Do you know if those rumors are true?' You realize that the young person is the very soul of discretion and will never tell anyone else B's secret if you reveal it.

You can lie to the young person, telling her that you have no idea whether B went through the same thing. Or you can tell the truth, which is that you do in fact know. If you say that you do in fact know, you will not strictly

speaking have revealed B's secret. But the young person will then ask, 'So are the rumors true?' and unless you lie in response to *that* question, the secret will come out. (If you say, 'Yes, I know, but I'm not going to tell you', the young person will rightly conclude that the rumors are true. You can try saying, 'Even I knew I wouldn't tell you; let's talk about something else'.<sup>6</sup> But you may very well lack the ability to deliver that line in a manner that doesn't let the cat out of the bag.)

In situation 1, B has taken great steps to conceal this particular secret. It is clearly something she cares deeply that others not discover. Her privacy in this particular matter matters tremendously to her.

In situation 2, B is someone who generally prefers that people mind their own business and that her friends not talk about her to others. She values her privacy, although she is not consumed by protecting it and has never placed any more emphasis on this particular secret than on many other things about her life.

Even if we suppose that the consequences of lying to the young person are the same in both cases, the relative weights of the duties not to lie and to protect your friend's privacy are different in B1 and B2. (Indeed, it might be that your all-things-considered duty in B1 is to lie and in B2 is not to lie.) That's because B1 cares more about others' not knowing her secret than does B2; B1's caring for this may very well be independent of any consequences that might flow from others' knowing; and your duties to your friends are sensitive to what your friends care about.

If Sinnott-Armstrong's argument appears to have traction, it's because his examples draw our attention to promises to do things that affect the future well-being of the promisee (whether she gets a ride to the airport, or is met for a lunch date). But sometimes the morally relevant thing that matters to a person toward whom you have a duty is independent of her future well-being. It is the latter sort of case that reveals the non-consequentialist force certain moral duties exert.<sup>7</sup>

*University of Arizona*  
*Tucson, AZ 85721-0027, USA*  
*gillm@email.arizona.edu*

## References

- Hill, T. 1984. Autonomy and benevolent lies. *Journal of Value Inquiry* 18: 251–67.
- Lopez, T., J. Zamzow, M. Gill, and S. Nichols. 2009. Side constraints and the structure of commonsense ethics. *Philosophical Perspectives* 32: 305–19.
- Ross, W. 2003. *The Right and the Good*. Oxford: Oxford University Press.

6 Sinnott-Armstrong suggested this response.

7 Thanks to the *Analysis* referee, Mark Timmons and Walter Sinnott-Armstrong for their helpful comments on this article.

Sinnott-Armstrong, W. 1992. An argument for consequentialism. *Philosophical Perspectives* 6: 399–421.

Sinnott-Armstrong, W. 2009. How strong is this obligation?: an argument for consequentialism from concomitant variation. *Analysis* 69: 438–42.

## *Counterfactuals and downward causation: a reply to Zhong*

JONAS CHRISTENSEN AND JESPER KALLESTRUP

### 1. *The downward causation argument*

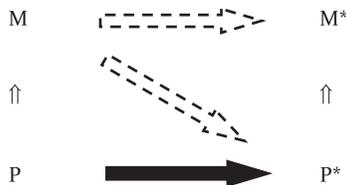
Lei Zhong (2012) propounds a new causal exclusion argument against non-reductive physicalism. Suppose mental property M causes mental property M\*. Given strong psychophysical supervenience:

(S): If any system *s* instantiates mental property M, then there necessarily exists a physical property P such that *s* instantiates P, and necessarily anything instantiating P instantiates M.

M and M\* have the distinct physical properties P and P\* as their respective base properties. Then, according to the principle of downward causation:

(DC): If property A causes property B, then A would cause any base property of B.

M must also be a cause of P\*. And the physical world is causally closed such that the physical effect P\* also has P as a sufficient physical cause. But P\* cannot have more than one sufficient cause, assuming the absence of psychophysical overdetermination. Hence, M is causally inert.<sup>1</sup> See Diagram 1 below:



Zhong assumes that (S), non-overdetermination and closure are undisputed, which means the non-reductive physicalist has no choice but to

1 The foregoing simplifies in three respects. First, talk about property A causing property B is short for an instance of A causing B to be instantiated. Second, we ignore time-indices in (S) and elsewhere. Third, as all physicalists are committed to the metaphysical necessitation of the non-physical by the physical, we shall henceforth take occurrences of ‘necessity’ to express metaphysical necessity.