Forthcoming in *Philosophical Studies*

Daniel Immerman

How should we accommodate our future misbehavior? The answer turns on how bad it will be

Abstract. *Professor Procrastinate receives an invitation to review a book. Best would be to accept it and then write the review. But if he accepts it, he will never get around to writing. And this would be worse than declining. Should he accept? Possibilists say yes, Actualists say no, and I say we need more information. In particular, we lack some information about the level of goodness of the various options. For example, we lack information regarding how much better it would be to accept and write than it would be to decline. In the course of defending my view, I discuss its implications for ethical theory and our everyday actions.*

Keywords. Actualism; Possibilism; Moral Obligations; Non-Ideal Theory

Here is a widely discussed case:[1]

> PROFESSOR PROCRASTINATE. Professor Procrastinate receives an invitation to review a book. The best thing that can happen is that he says yes, and then writes the review when the book arrives. However ... were Procrastinate to say yes, he would not in fact get around to writing the review. Not because of incapacity or outside interference or anything like that, but because he would keep on putting the task off. Thus, although the best that can happen is for Procrastinate to say yes and then write ... what *would* in fact happen were he to say yes is that he would not write the review. Moreover ... this ... is the worst that can happen. [Jackson and Pargetter, 1986, 235] [Emphasis theirs].

[1] This case, or ones similar to it, are discussed in e.g. [Baker, 2012, 641], [Cariani, Forthcoming], [Goldman, 1978, 185-6], [Jackson and Pargetter, 1986, 235], [Littlejohn, 2009], [Portmore, 2011, 151], [Portmore, Forthcoming], [Timmerman, 2015, 1512], [van Someren Greve, 2013, 482-3], [Vessel, 28, 166], [Woodard, 2008, 18].

Discussion of this case focuses on the question: ought Procrastinate accept the invitation, in the objective all-things-considered moral sense of "ought"?[2]

This question is of practical interest; such cases are widespread. We regularly make plans or commitments and then don't live up to them. We buy gym memberships only to fail to use them and make plans to socialize during busy periods only to lack the mental energy to fully engage.

In addition, solving the puzzle promises to yield theoretical insights. Arguments regarding these cases turn on important principles in ethical theory – principles related to obligations and advice, intentions and ideals. Thus identifying which of these arguments are successful, which are unsuccessful, and why promises to yield insight into ethical theory, at both the normative and meta-ethical level.

This paper defends a new response to the question: ought Professor Procrastinate accept the invitation? I argue that we cannot answer it because we are missing key information. This information concerns the level of goodness of the various options. For example, while we know that it would be better for Procrastinate to accept the invitation and then write the review, we do not know *how much better* it would be.

This paper has three sections. The first clarifies my position and situates it within the literature. The second offers a case-based argument for it. The third develops an additional argument for it. This argument turns on similarities between our future selves and other people.

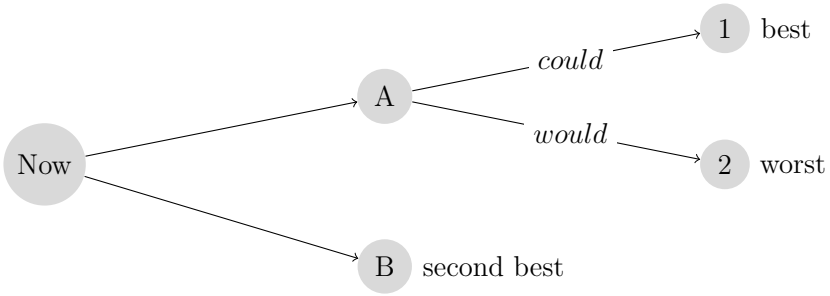
1 Clarifying my position

The Professor Procrastinate case shares structure with a number of related cases. I refer to all such cases as "could-but-won't" cases.

In could-but-won't cases, an agent – "the focal agent" – decides between two alternatives, A and B. If she chooses A, she will decide between two further alternatives: 1 and 2. Best is to choose A and then 1, second best is to choose B, and worst is to choose A and then 2. But if she chooses A, she will choose 2.

Here is the structure in graphical form:

[2]By the "objective all-things-considered moral sense of 'ought' ", I mean a moral sense of "ought" that takes all relevant information into account, not merely the information that the subject possesses, and is not overridden by some other moral ought.



Let me show how Professor Procrastinate fits this structure. He is currently deciding between accepting the invitation (A) and declining (B). If he accepts, he will either write (1) or not (2). Best is to accept and then write (A and then 1), second best is to decline (B), and worst is to accept and then not write (A and then 2). If he were to accept, he would not write.

Given this structure, we face a question: ought the focal agent choose option A or B? There are three possibilities: i) option A,[3] ii) option B,[4] or iii) it depends on further information.[5]

The two historically dominant positions in the literature are i) and ii). These two positions are known as Actualism and Possibilism. By contrast, I defend iii).

Roughly speaking, *Actualists* say one ought to perform some action just in case what would happen if one do so is better than what would happen if one didn't. Because of this, Actualists say the focal agent ought to choose option B. So, for example, they say Professor Procrastinate ought to decline the invitation. They reason as follows: what would happen if he declines is better than what would happen if he accepts. After all, if he accepts, he will fail to write the review, which is worse than declining.

Meanwhile, roughly speaking, *Possibilists* say that one ought to perform some action just in case it is part of the sequence of actions that forms the

[3]For those who endorse this position, see e.g. [Feldman, 1986], [Goldman, 1978], [Greenspan, 1978], [Humberstone, 1983], [Thomason, 1981], [Vessel, 2009], [Vorobej, 2000], [Zimmerman, 1996].

[4]For those who endorse this position, see e.g. [Goble, 1993], [Goldman, 1976], [Jackson and Pargetter, 1986], [Sobel, 1976], [Sobel, 1982].

[5]See e.g. [Carlson, 1999], [Cohen and Timmerman, Forthcoming], [Gustafsson, 2014], [McKinsey, 1979], [Portmore, 2011, 208], [Timmerman, 2015, 1525].

best sequence one can perform. Because of this, Possibilists say that in could-but-won't cases, the focal agent should choose option A. So, for example, they say that Professor Procrastinate ought to accept the invitation. They reason as follows: agreeing to write the review is part of the sequence of actions that forms the best sequence Procrastinate can perform. After all, the best sequence he can perform is to accept the invitation and then write the review.

Despite this disagreement, both Actualists and Possibilists agree that we have enough information to answer the question. But some people disagree with them – these people say we do not have enough information to determine what the characters ought to do.[6] In other words, they say: merely knowing that someone is in a could-but-won't case isn't enough for us to know whether they ought to do A or B. Instead, what they ought to do depends on further details.

A natural question such people are asked is: "What sort of information would be enough to allow us to determine what Procrastinate should do?" Two types of answer to this question are defended in the literature.

First, some respond that it would be enough to have further *modal* information. This is information about what will happen if the focal agent does certain things and what would have happened if she had done certain things in the past.

So, for example, one might ask: could Procrastinate have done anything in the past so as to become less of a procrastinator. One might think that if there's nothing he could have done – it procrastination is an immutable fact about his personality – then he has less of an obligation in the present to take on the review.[7]

Second, some say we need what I will call *action-theoretic* information – information about what is under the focal agent's control in the present, the past, and certain counterfactual situations, about what they can and could do intentionally in these situations, and so on.

So, for instance, one might wonder if there's some intention or commitment Procrastinate can presently form that will lead to his following through.

[6]See e.g. [Carlson, 1999], [Cohen and Timmerman, Forthcoming], [Gustafsson, 2014], [McKinsey, 1979], [Portmore, 2011, 208], [Timmerman, 2015, 1525].

[7]Compare: suppose I promise to pick up my friend from the airport at 6 and then stay home because I don't feel like going. Suppose it's now 5:55 and I can't get to the airport in time. It seems like I nonetheless have the obligation to pick her up at 6 and this is because there are things I could have done – driving to the airport – that would have made it possible to pick her up.

If it is indeed possible for him to presently form an intention or commitment that will lead to his writing, it seems plausible to say that he ought to accept the invitation.

In sum: some people agree with me that we need further information before we can decide what the focal agent should do. They then go on to say that the information we need is modal or action-theoretic information.[8]

While it may be true that modal or action-theoretic information is *necessary* – I do not take a stand on this particular issue – I think that modal or action-theoretic information is not *sufficient* for determining what someone like Professor Procrastinate ought to do. In particular, according to my position, even if we had all of the relevant modal and action-theoretic information, there is still more information we'd need in order to determine if the focal agent ought to choose A or B.[9]

In particular, as I've already indicated, I will argue that we need information concerning the *goodness* of the various options. Recall that the only information we have is the following: the best option is better than the intermediate, which is in turn better than the worst.

One missing piece of information concerns the *relative* goodness of the options. In particular, we don't know how much better it would be to choose A and then 1 than it would be to choose B. And we don't know how much better it would be to choose B than it would be to choose A and then 2.

Another piece concerns the absolute goodness of the options. For instance, we're missing information about the absolute goodness of choosing option B. Is this very good? Slightly good? Neutral? Slightly bad? Very

[8]Note that I'm not assuming the distinction between these is perfectly sharp. Thanks to an anonymous referee for pressing me on this.

[9]This negative position does not seem to have been defended in the literature. One possible exception is Christopher Woodard, who has developed a view according to which an agent is obliged to do X just in case he has overall reason to do so, and in which we have both actualist and possibilist reasons; see [Woodard, 2008], [Woodard, 2009]. While this view is somewhat non-committal on how these reasons combine to form overall reasons, his writings suggest that he thinks adding modal and action-theoretic information will not be enough to settle on a verdict regarding what Professor Procrastinate ought to do.

It is perhaps worth noting that in endorsing my view, one need not endorse Woodard's view; I am not committed to the view according to which one has an obligation to do X just in case one has overall reason to do something, nor am I committed to the existence of both actualist and possibilist reasons. As I note later on in this paper, there are many different types of normative ethical position one can endorse while accepting my view. And many of these do not (or need not) have Woodard's commitments. For example, there are various forms of deontology and virtue theory that are consistent with my view.

bad?

In sum, then, my position lies within a broader class of views. This broad class says the following: if all we know is that someone is in a could-but-won't case, we don't have enough information to determine whether they ought to choose A or B. But while those in this broad class tend to claim that this uncertainty could be resolved if we simply had more modal or action theoretic information, I argue that we need information concerning the goodness of various options.

2 A case-based argument

This section develops a case-based argument for my view. The goal is to find cases that hold constant those features that others think are relevant and shift the features that I think are relevant and then show that our judgments shift.[10]

[10]First, and most obviously, they will all be could-but-won't cases. In addition, as I noted earlier, some theorists think that certain modal and action-theoretic information is relevant to determining what characters in such cases ought to do. And, as I also noted, information like this isn't typically given when describing could-but-won't cases. Nonetheless, I would like to try to hold this sort of information consistent between cases.

In order to ensure parallel structure between the cases I will present, I hereby stipulate that they all share the following features: (i) There are things that the characters could have done in the past so as to avoid their current situation. (ii) Likewise, there are ways they could have formed their character in the past so as to bring it about that if they chose option A they would choose option 1. (iii) But they have not done these things. And thus, there is nothing they can do in the present, at the very moment, that would ensure that if they were to choose option A, they would then choose option 1. (Or, at the very least, there is nothing they can do that is not so extreme as to be obviously morally unacceptable.)

So, for example, regarding Procrastinate, (i) There are things he could have done in the past – e.g. not becoming a professor – that would have ensured that he didn't find himself in the situation he presently does. (ii) And there are also ways he could have formed his character – developing better habits – that would have made it such that if he accepted the invitation, he would have followed through. (iii) But alas he chose to be a professor and did not form his character in this way. And thus there is nothing he can do at this very moment to ensure that if he accepts the invitation, he will follow through.

One might ask: why have I chosen to specify the modal and action-theoretic information as I have. The reason is because it seems the most natural way to understand this sort of case. So, for example, it would be weird to think there was nothing Procrastinate could have done to keep himself from finding himself in this situation or to change his habits. And, of course, if there was something he could reasonably do right now to ensure that he completed the review, he should accept and do it. For example, suppose that if he were to

My strategy in selecting cases has been to focus on ones already present in the literature. One reason for doing so is that it shows that others share my judgments (I have picked cases in which the original authors are explicit regarding their judgments).

The first two papers written on the Actualism/Possibilism debate defended Actualism. One, by Holly Goldman, centers on a case which runs as follows:

> [S]uppose a faculty member, Jones, is currently trying to decide whether to go to the office or to stay at home. If she goes to the office, she can either attend a faculty meeting [regarding a language requirement for undergraduates that will pass if she attends the meeting and not otherwise] or talk with one of her students. If she stayed home she would do research for her lectures and then write up her notes ... If she talks with her student, he will ask her advice about seeking psychiatric counseling ... if she doesn't talk with him ... he will seek and obtain effective help ... if she went to the office, she would then [unwisely] elect to talk with the student and discourage him from seeking psychiatric aid. Thus going to the office would result in the student's having an emotional breakdown. ... we may suppose in addition that Jones' voting for the language requirement would produce greater utility than any of the other acts available to her. [Goldman, 1976, 454-8].

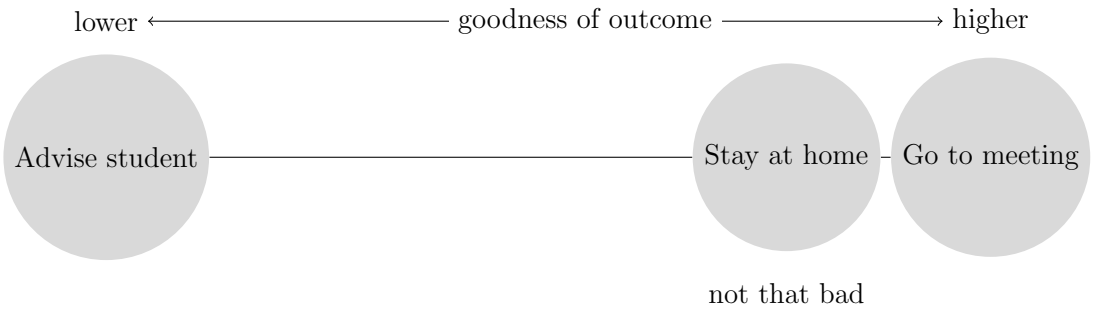
Goldman thinks that in this case the professor ought to stay home, and I agree. I think it is worth emphasizing some features of the case that can help explain this judgment. First, the worst option is far worse than the intermediate option. That is, it would be far worse for the professor to meet with the student than it would be for her to stay home and get her lecture notes typed up. After all, if she goes to meet the student, she will convince the student not to seek care, which will trigger an emotional breakdown. Second, the intermediate option is not that much worse than the best option; it is not much worse to type up lecture notes than to vote on the language requirement.[11]

---

attach a sticky note to his computer which said "write the review" then he would follow through. In such a case, he should obviously accept the invitation and write the note.

[11] As I read the case, Goldman is taking it as implicit that whether the language requirement passes doesn't matter too much. She says that the benefit it provides is "pleasure the

In addition, it's worth noting that the middle option is not that bad; it is not that bad to stay home and type up the lecture notes. In graphical form, the case looks like this:



It's not a coincidence that this case has these features; the cases that other Actualists tend to focus on also have them. For instance, Fred Feldman offers a case involving three potential treatments by a doctor, which he titles "Best", "Good", and "Horrible" [Feldman, 1986, 52] and a case from Sobel involving either drinking milk, eating pickles, or first eating pickles and then drinking milk. Regarding the last option, Sobel assures us that mixtures of pickles and milk is vomit-inducing and thus far worse than the other options [Sobel, 1976, 199-200].

Meanwhile, Lou Goble offers a case that is even more extreme. In it, a character named Fred must decide whether to stay inside on Sunday or go outside. If he goes outside, he will face a further choice: either work in his garden or to daydream. Best would be to go outside and work his garden, but in fact, if he were to go outside, he would choose to be idle and daydream. Unfortunately, there is a murderer on the loose who deeply hates daydreamers, so if Fred goes outside and then daydreams, he will be killed

students will ultimately take in being able to use a foreign language" [Goldman, 1976, 460] and seems to treat this as being a less serious matter than the graduate student having "a serious emotional breakdown" as a result of not seeking counseling [Goldman, 1976, 454].

That said, and at the risk of belaboring the obvious: nothing about my view commits me to saying that language requirements are unimportant. Rather, my view says the following: the more inclined we are to say that passing the language requirement is incredibly important, the more inclined we should be to say the faculty member ought to go to campus.

by the murderer.[12]

Here, then is a particularly stark case. It is far worse for Fred to go outside, daydream, and be murdered than it is for him to stay inside; this difference is much bigger than the difference between staying inside and going outdoors and working. Meanwhile, staying inside is not that bad. Because of this, my account predicts that Goble will think that Fred ought to stay inside. And he does so, writing: "Knowing what will happen, would we, should we, advise Fred to pick up his hoe? If he does do that and meets the fate described, will we judge that he did the right thing? The answer to these questions to me seems clearly No" [Goble, 1993, 139].

Meanwhile, Possibilists tend to focus on cases which follow a different pattern. The first published Possibilist paper, written by Patricia Greenspan, focuses on a case that runs as follows:

> Suppose I have been appointed placement director [and] that all the work of a placement director is done at a single frantic set of APA meetings. Now, in fact, if I went to these meetings I would

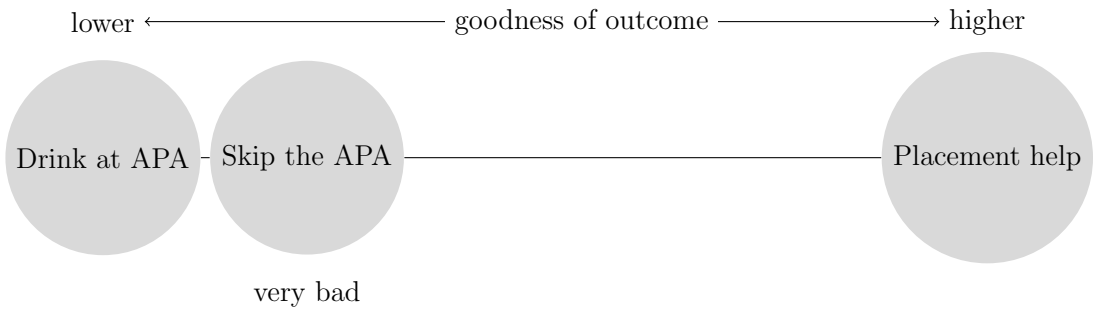
[12]Here is the case in Goble's words:

> Consider: Fred is lazy and given to daydreaming. He knows he ought to work in his garden on Sunday afternoon and that to work in his garden he must have his hoe. He means to work in his garden and so he picks up his hoe and goes out to his garden plot. There, however, instead of actually working the soil around his radishes, he pauses to contemplate the wonder of growing things. Fred can become quite rhapsodic. In his reverie he reflects how the garden unites himself, a man, to the earth, and how indeed all people are bound to the earth ... Fred is given to such lapses. Of course, they are avoidable. ... Fred's daydreams, in most cases, are harmless; he comes out of them in a few minutes and works in his garden with greater vigor than if he did not have those moments. This day, however, it matters a great deal. This day there is a lunatic loose in the neighborhood. This person too worships gardens. Carrying a great scythe over his shoulder, he stalks the suburbs, seeking people at work in their flower beds or vegetable patches. When he comes upon someone diligently laboring in his or her garden, he will pause to chat a while about cabbages or calendula and then go on his way. But he despises all idleness; when he finds someone standing around daydreaming instead of working, then he becomes violent. He wields his scythe, and cuts the idler down to size. So it is with Fred. The Reaper strikes. It's all over very quickly (though not painlessly). ... I grant that Fred ought to work in his garden. But is it also true that he ought to pick up his hoe, when doing so he will meet this terrible fate? [Goble, 1993, 138-9].

> simply get drunk and socialize, ignoring my duties as placement director ... But if I stayed home from the meetings, I would write a paper – something I am not obliged to do, let us grant, as I am obliged to work on placement; but something that will at least be more profitable than goofing off at the meetings. [Greenspan, 1978, 78].

Greenspan thinks that she ought to attend the APA and perform her duties as placement director and I agree. But it's worth emphasizing some features that help me make the judgment. First: the worst option is not much worse than the intermediate one. That is, attending the meetings, getting drunk and socializing is not that much worse than staying home and writing a paper. After all, she is under no obligation to write the paper. Second, the intermediate option is far worse than the best option; it would be much better to attend the meetings and help the students get placed than to stay home and write a paper that she doesn't even have to write.

Furthermore, the intermediate option is pretty bad; as placement director, it's pretty bad to skip the APA and stay home. (Or at least it was at the time the paper was written, at which point attending the APA meeting as placement director was very important for helping students get jobs.) In graphical form, the case looks like this:



As the literature has progressed, the examples have gotten even more extreme, regularly featuring rape and murder in the two worst options. Those offering these cases seem to be aware that this feature helps drive our intu-

itions.[13]

The very best cases for making this point are – unfortunately – ones that are rather brutal and disturbing. For example, Ralph Wedgwood offers a case in which an evil person has abducted a child and is deciding whether or not to deeply harm her right now. If he does not harm her now, he will face the choice later on whether to release her or do things that are even worse to her. Unfortunately, he would decide to do things that are even worse.[14]

Here, then is a particularly stark case; the best option – letting the girl go – is far better than the two worse options, which are much closer than they are to the best option. And, as my account predicts, Wedgwood thinks that the pedophile ought to choose not to harm the girl right now. Not only does he say this, he thinks the judgment about this case is clear; he calls the case a reductio ad absurdum against the alternative view.

Let me stop and take a step back. As I have shown, if we hold constant the information that others think is relevant, but vary the relative goodness of the options, we discover the intuitions about cases that my view predicts.

2.1 An objection

In my case-based argument, I suggested (or at least hinted) that our intuitions can be explained by a principle along the following lines:

> Relative Goodness: Other things being equal, the greater the

[13] See e.g. [Baker, 2012, 645]. [Timmerman, 2015, 1511].

[14] Here is the fuller – and more graphic – description of the case that Wedgwood provides:

> For example, imagine a wicked paedophile, who has just abducted a 10-year-old girl and imprisoned her in his secret cellar. Suppose that it is still possible – though unfortunately quite unlikely – that the paedophile will repent of his evil plans, and return the girl unharmed to her parents. Surely, if anything is clear about this case, it is clear that it is not true that the paedophile ought to rape the girl.
>
> ...
>
> Suppose that it is also true in this case that if the paedophile did not rape the girl, he would torture her to death, whereas if he did rape her, he would not subject her to any additional torture, and would not kill her. So, presumably, the paedophiles conduct would be at least somewhat better if he raped her than if he didn't. Hence actualists must say that the paedophile ought to rape the girl. This seems to me a reductio ad absurdum of the actualist view. [Wedgwood, 2009].

> distance between the best option and the intermediate option, and the less the distance between the intermediate option and the worst option, the less acceptable it is to choose the action with the intermediate outcome.

One might have the following worry: this principle seems hard to situate in a broader theoretical framework. More concretely: even if this principle matches intuitions, it seems hard to see how one could derive it from more fundamental normative ethical principles. Meanwhile, stipulating it without deriving it from such principles seems ad hoc.[15]

My response: it seems to me that Relative Goodness can be grounded in a broader ethical framework. To illustrate this, I will briefly describe several normative frameworks that harmonize nicely with Relative Goodness. (Of course, given the wide array of normative views on offer, I cannot hope to discuss them all. And even with regards to the frameworks I discuss, there is much more to be said.)

My first example of an ethical framework from which we can derive Relative Goodness is a utilitarian view according to which one ought to do whatever maximizes expected utility.

I will take a relatively simple version of the view and focus on how it applies to the Professor Procrastinate case. To do so, I need to start by making a point about chances. In the Professor Procrastinate case, it's stipulated that if he accepts, he will not actually write the review. That said, it's not as if it's impossible for him to write the review. On the contrary, it seems like a natural interpretation (or, at the very least, a possible interpretation) that when he accepts, there is a chance – albeit a low one – that he will go on to write the review.

Let me suppose, to make things concrete, that if Professor Procrastinate accepts the invitation, there will be an eighty percent chance that he will fail to write it, which would be rather bad (-10 utility) and a twenty percent chance that he writes the review, which would be rather good (10 utility). Suppose furthermore that if he turns down the invitation then there will be a hundred percent chance that a new reviewer will have to be found, which will be moderately bad (-5 utility).

In such a case, the expected utility of his accepting the invitation is $[(.8 * -10) + (.2 * 10)]$ which is -6 whereas the expected utility of his turning

[15]Thanks to an anonymous referee for expressing this worry.

down the invitation is $(1 * -5)$ which is -5. So according to the utilitarian view under discussion, he should turn down the invitation, seeing as doing so maximizes expected utility.

By contrast, suppose the best outcome yielded a higher utility, e.g. a utility of 20 (suppose that he would write an absolutely terrific review). In such a case, the expected utility of his accepting is $[(.8 * -10) + (.2 * 20)]$ which is -4, and the expected utility of turning down the invitation is still -5 so he should accept the invitation.

This case provides an illustration of how, according to this particular utilitarian view, altering the level of goodness between the options, while holding other factors fixed, affects what one ought to do.[16]

While the expected utility view is one view that can help ground Relative Goodness, it is by no means the only sort of ethical theory that can do so. To illustrate this, let me look at a second example of an ethical framework that can do so – in particular, an Aristotelian framework. Here I will draw on some themes from a paper by Rebecca Stangl titled "Taking Moral Risks and Becoming Virtuous". The paper discusses a character named Nancy:

> Nancy is thinking of going into nursing, because she sincerely wishes to relieve the pain and suffering of others. And were she to go into nursing, she almost certainly would relieve much pain and

[16]As an aside, I should note that the view also predicts another related phenomenon, which I will call *Relative Likelihood*: If we hold fixed the level of goodness of outcomes (that are, or will be, under your control), but adjust the chances that they will occur, we can alter what you morally ought to do.

For example, suppose that if I go to Thanksgiving at my in-laws, even though it's possible for me to control myself, I'm extremely likely to fly into a rage. In this case, I probably shouldn't go. By contrast, suppose I have an extremely low chance of flying into a rage when I visit them at Thanksgiving. In this case, it is fine for me to go.

The expected utility theory explains this result because it says that what you ought to do turns on both the goodness of the outcomes and the likelihood that they will occur. By contrast, Actualists and Possibilists cannot accept Relative Likelihood or the cases it helps explain. For Actualists, all that matters is knowing which outcome will actually obtain, and for Possibilists, all that matters is knowing the best possible outcome that could obtain. So neither can endorse the claim that if we hold fixed the level of goodness of outcomes (that are, or will be, under your control), but adjust the chances that they will occur, we can alter what you ought to do. Likewise, neither can accept the claim that altering the chances that I will fly into a rage alters whether that I should go to Thanksgiving. All that matters – if one is an Actualist – is whether I will fly into a rage, and all that matters – if one is a Possibilist – is whether it is possible for me to avoid the rage. Neither cares what the chance are that I will fly into the rage.

> suffering. But Nancy also rightly fears that the exposure to pain and suffering might make her a less compassionate, and thus less virtuous, person. She might even get to the point where she does the job only for the money. If that were to happen, her ability to live an active life of virtue in the future would be seriously compromised. Should she pursue a career in nursing? [Stangl, 2015, 216].

This seems to have a similar structure to a could-but-won't case; the main character is deciding between two options (going into nursing or not), and there are three possible outcomes (Nancy doesn't go into nursing [intermediate]; Nancy goes into nursing and maintains compassion [best]; and Nancy goes into nursing and loses compassion [worst]).

Stangl describes this case as a case of "moral risk" and raises the question: when ought one take a moral risk? Her verdict, which bears a close similarity to Relative Goodness, is the following:

> ... it seems that the best thing to do is to aim for some sort of balance: when one is considering taking a moral risk, one needs to know how likely it is that one will improve, how likely it is that one will deteriorate, and what magnitude the improvement or deterioration might be. A very large risk seems advisable only if the prospects of improvement are very great. And if the improvement is negligible, then the risk ought not be great. [Stangl, 2015, 224-5].

It is worth emphasizing some aspects of Aristotelian virtue ethics that Stangl draws on, because they also apply to Relative Goodness.

First, she emphasizes that her case involves a conflict between virtues – beneficence pushes Nancy towards becoming a nurse, so that she may help others, whereas perfectionism (becoming a more virtuous agent) pushes her away from becoming a nurse, seeing as becoming a nurse risks her becoming a less virtuous person. [Stangl, 2015, 220-9]. When it comes to weighing this against her potential loss of compassion, it matters how beneficence is involved in being a nurse. As Stangl puts the point, "If the need for qualified medical personnel is especially dire, then the ... goods she will promote by becoming a nurse are very fundamental indeed" [Stangl, 2015, 229]. In short, then, it seems that in weighing the different virtues against each other – the

added beneficence in being a nurse against the decreased perfectionism that might arise – one thing that matters is how much beneficence is involved in being a nurse.

This sort of point applies to could-but-won't cases and Relative Goodness. In the Professor Procrastinate case, for example, an Aristotelian might see us as having to weigh different virtues. Perhaps, for example, beneficence and service speak in favor of accepting the invitation, given that writing the review would be helpful and a good way to serve the academic community. Meanwhile, integrity and compassion speak against agreeing to it, given that Procrastinate will fail to write it. In weighing these against each other, an Aristotelian could say, it matters how much beneficence and service are indeed provided in agreeing and then writing.[17]

In addition, Stangl emphasizes that her view fits nicely with several other core aspects of Aristotelean ethical theory, viz. (i) the doctrine of the mean and (ii) particularism. As she writes:

> On my account, whether the virtue of perfectionism tells one to take a moral risk depends upon very particular facts about the situation and the person in question ... It is part of the canonical understanding of that approach to morality that the "right" or "mean" in action and feeling is relative to the individual, and can only be determined by the person of practical wisdom. No one ought to take inappropriate risks. But what constitutes an inappropriate level of risk for one person may not constitute an inappropriate level of risk for another person. ... [Stangl, 2015, 225].

Likewise, an Aristotelian could support Relative Goodness by saying that one ought to aim for the mean when it comes could-but-won't cases. One shouldn't always choose the option that is part of the sequence that forms the best possible sequence of actions – doing so is problematic when it puts one at a serious risk of catastrophic failure. But neither should one always choose the option that would be the best, given what one will actually do in the future. Sometimes one should take moral risks, especially if there is much to gain, and not much cost if one fails to act for the best. An Aristotelian

[17]I should note that it's not merely Aristotelians that could say this; a similar story could be told on a Rossian deontological view, where we weigh various duties against each other [Ross, 1930].)

could emphasize that there are no ironclad rules; only with practical wisdom comes the ability to know how to act in particular could-but-won't cases.

In sum, then, it seems that Relative Goodness is defensible from within various meta-ethical views, and thus doesn't require one to adopt an ad-hoc principle that is ungroundable from within a more broader ethical context. Of course, I should hedge that I have only discussed a couple of views and there are many more details to be given. So, at best, I think I've offered some prima facie support for the claim that there are various packages of views in ethical theory that meld nicely with Relative Goodness. In the next section, I'll return to offense, looking at at some additional theoretical considerations favoring Relative Goodness.

3 Multi-person cases

This section offers another argument for my position. It harmonizes nicely with my first argument while bringing out some additional theoretical points favoring my view.

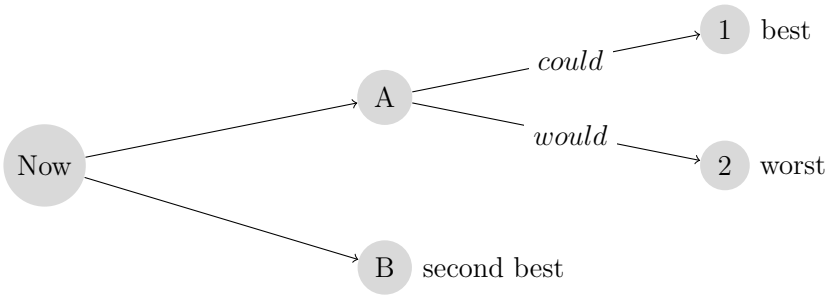
The argument involves cases that are similar to could-but-won't cases but with an important twist: they involve two different decision-makers. I call them "multi-person cases".

Here's an example: Suppose I'm trying to decide whether to invite my brother to my wedding. Best would be to invite him and have him behave. Second best would be to not invite him. And worst would be to invite him and for him to cause a major scene. If I invite him, he could choose not to cause a scene. But as a matter of fact, if I invite him, he will cause a scene.

In general, then, multi-person cases have the following structure: one person –"the focal agent" – makes an initial decision between A and B. If the focal agent chooses A, then another agent makes a decision between 1 and 2. Best is for the focal agent to choose A and then for the second agent to choose 1, second best is for the focal agent to choose B, and worst is for the focal agent to choose A and the second agent 2. But if the focal agent were to choose A, then the second agent would choose 2.

So, for example, in the wedding case, I am the focal agent, option A is inviting my brother to the wedding, option B is my not inviting him, option 1 is him behaving at the wedding and option 2 is him causing a scene. Best would be to invite him and have him behave (A and then 1), worst would be to invite him and have him cause a scene (A and then 2) and

intermediate would be to not invite him (B). This case, then, has the same familiar structure from could-but-won't cases:



Of course, I do not wish to put too much weight on this similar structure. After all, structural similarities can mask underlying differences. Instead, I will be defending my conclusion – when it comes to could-but-won't cases, whether the focal agent should choose A or B turns on the level of goodness of the options – via the following argument:

> 1. When it comes to multi-person cases, whether the focal agent should choose A or B turns on the level of goodness of the options.
>
> 2. If so, then when it comes to could-but-won't cases, whether the focal agent should choose A or B turns on the level of goodness of the options.
>
> –
>
> 3. When it comes to could-but-won't cases, whether the focal agent should choose A or B turns on the level of goodness of the options.

Let me defend the premises in turn, starting with Premise 1.

3.1 Defense of the first premise

Premise 1 says that when it comes to multi-person cases, whether the focal agent should choose A or B turns on the level of goodness of the options.

Premise 1 is very similar to the conclusion of the argument. The only difference is that Premise 1 talks about multi-person cases while the conclusion talks about could-but-won't cases.

Views similar to Premise 1 have been discussed in the context of discussions of non-ideal theory. In *A Theory of Justice*, John Rawls distinguishes between "ideal theory" which concerns cases in which all relevant agents comply with the demands of justice applying to them, and "non-ideal theory" which concerns cases in which not all agents do so [Rawls, 1971, 245-6]. Multi-person cases fall under non-ideal theory – they are cases in which not everyone acts as they ought. In particular, the second agent in these cases behaves in a non-ideal manner.

Rawls and others have discussed at length what one ought to do in non-ideal cases. And the principles they articulate typically sound similar to Premise 1, thus suggesting that this premise is intuitive.

For example, in discussing what to do when others do not behave appropriately, Christine Korsgaard writes: "the common-sense approach to this problem uses an intuitive quantitative measure: we depart from our ordinary rules and standards of conduct when the consequences of following them would be 'very bad' " [Korsgaard, 1986, 345].[18] In addition, Richard Brandt writes: "One is to follow the rule optimal when there is perfect (or near perfect) compliance, except when departing from it will avoid serious harm." [Brandt, 1988, 359]. Both suggest that in cases in which others are non-compliant, what one ought to do turns on the level of goodness of the options.

In addition to these sorts of general statements, we can also case-based support for Premise 1 by examining cases.

First, there are multi-person cases in which it is intuitive that the focal agent ought to choose option A. As with last section, I think the best sort of example have following structure: the focal agent is choosing whether or not to do something awful, and if they don't choose to do the awful thing, someone else will do something even more awful.

One such example is that of a sheriff deciding whether to frame an innocent man, where, if he doesn't do so, a mob will have a race riot and kill a

[18] I should note that Korsgaard, in her article, does not simply endorse this common-sense view and move on. Rather, she thinks that the expression "very bad" is somewhat unhelpful – how are we to determine if something counts as very bad or not? For this reason, she goes on to offer a theory, similar in structure to that of the common-sense approach, that more precisely fills in what it takes to be "very bad". [Korsgaard, 1986].

number of other innocents. [McCloskey, 1957, 468-9].[19]

Next, there are multi-person cases in which it is intuitive that the focal agent ought to choose option B. Here, as with the previous section, I think the best sort of example have the following structure: the focal agent is choosing whether or not to do something that is not that bad, and if they don't choose to do the thing, someone else will do something extremely bad.

Probably the most famous example of this is the case of the murderer at the door, which was given by a critic of Kant, Benjamin Constant. In it, a dangerous murderer is pursuing your friend, who has taken refuge in your house. The murderer arrives at your door and asks whether your friend is hiding there and you have to decide whether to lie or tell the truth [Kant, 1996]. Best, of course, would be to tell the truth and for the murderer not to kill your friend. But, unfortunately, if you tell the truth, the murderer will choose to kill your friend. By contrast, if you lie the murderer will go away and your friend will live. It seemed intuitive to Constant – and to most people who consider the case, even those following in the Kantian tradition – that you should lie to the murderer [Cholbi, 2009, 18], [Korsgaard, 1986, 327].

In short, the first premise enjoys intuitive support with some labeling it common sense; in addition, there is a strong case-based argument for it.

3.2 Defense of the second premise

Let us turn to the second premise. It says that what goes for multi-person cases also goes for could-but-won't cases.

Let me be more precise. As a reminder, the first premise said: When it comes to multi-person cases, whether the focal agent should choose A or B turns on the level of goodness of the options. And the second premise says: If the first premise is true, then when it comes to could-but-won't cases, whether the focal agent should choose A or B turns on the level of goodness of the options.

I should acknowledge that the second premise is somewhat controversial and I do not think that I can offer a decisive argument in its favor. Nonethe-

[19]Cases with this structure – those in which I can act badly so as to prevent someone from acting even worse – can have a lot of intuitive pull. For instance, Doug Portmore, in his book *Commonsense Consequentialism*, indicates that cases with this sort of structure – in particular, a case in which murdering someone else would lead to an ever-so-slightly better outcome than not doing so – were what convinced him to no longer be a utilitarian [Portmore, 2011, 3-4].

less, I think there are a number of considerations that speak in favor of it. The consideration I shall focus on comes from an examination of cases. It involves cases in which our intuitions do not seem to be affected by whether it is a could-but-won't case or a multi-person case.

I think the fact that there are cases in which our intuitions aren't affected in this way should be fairly unsurprising. It rests on an insight emphasized by various philosophers – most notably Derek Parfit – that in certain cases, we relate to our future selves in almost the same way as we do to other people [Parfit, 1984]. (I should note that one can endorse this insight without taking on some of Parfit's stronger claims about personal identity across time.) Using this insight, I will offer two cases. The first is a multi-person case. The second is a could-but-won't case with a similar structure to the first. And in the second case, the focal character relates to his future self in almost the same way as he does to another person.

First, the multi-person case:

> MAKING A SCENE – MULTI-PERSON: Avery had been dating Sawyer for a while, but they recently split up. This is hard for me because I live with Sawyer but am also friends with Avery. Today, Avery has stopped by to invite me to a holiday party she is throwing tonight for members of her office. It sounds a little boring, but I have nothing else going on, and so I'll probably go. Unfortunately, Sawyer saw her invite me. In everyday life, he is an extremely calm and reasonable and right now he is taking their break-up well. But he likes to drink and when he has drunk a lot, it feels as though he turns into a different person from his normal self; he becomes loud, angry, and aggressive. I am worried that this evening Sawyer is going to get drunk and then, in his aggressive state, he is going to try to get the details of the party from me and then go to the party and cause a major scene. This would lead to a lot of embarrassment and quite possibly Avery losing her job. So I am thinking about getting rid of the details of the party. That way, if Sawyer does get drunk and aggressive, he won't be able to get the details from me. (If I destroy them, I won't remember them and thus won't be able to give them to him). This worry is well-grounded – as it turns out, if I don't get rid of the details, Sawyer will get drunk, get the details from me, and then make a scene, causing Avery to lose her job.

In this case, best would be for me to keep the details and for Sawyer not to get drunk, get them from me and cause a scene. Worst would be for me to keep the details and for Sawyer to get drunk, get the details, and cause a scene. And intermediate would be to destroy the details. In this case, it seems as though I ought to destroy the details. The worst option – Sawyer getting drunk and causing a scene – is very bad. Meanwhile, the intermediate – my destroying the details – is not so bad; the party sounds a little boring. And it is not that much worse than the best – my keeping the address and Sawyer behaving himself.

Compare this case with a single-person version:

> MAKING A SCENE – SINGLE-PERSON: Everything in the above case is the same, but with a twist: I am Sawyer. That is, Avery has decided to split up with me and – because we are still friends – invited me to the holiday party. I am normally an extremely calm and reasonable person and right now I am taking the news well. But I like to drink and when I have drunk a lot, it feels as though I turn into a different person from my normal self; I becomes loud, angry, and aggressive. I am worried that this evening I am going to get drunk and then, in my aggressive state, I am going to go to the party and cause a scene. And, in fact, my worry is well-grounded; if I don't destroy the details, I will get drunk, go to the party, and cause a scene, leading Avery to lose her job.

Again, it seems intuitive that I should destroy the details. And, more to the point – our intuitions do not shift much between cases. That is, the intuition about how I should act – in particular, whether I should destroy the details or not, does not seem to shift much as we shift from the multi-person case to a single-person case. And this is not surprising, given that I relate to my drunken future self in a similar way to the way I relate to others.

3.3 The theoretical upshot of this argument

At this point, I have defended the two premises of my second argument. I did so in part by appeal to intuitions. And so I would like to step back a little bit and describe some theoretical features that I think this argument brings out.[20]

[20]Note that discussion in this section bears similarities to discussion in [Woodard, 2008].

As my first premise brings out, in choosing how to act in a world in which others are not ideal, we have to take an appropriate balance between two extremes. One extreme is acting in the way that would be best if all others acted in an ideal manner. This sort of thinking is illustrated by Kant and his counter-intuitive suggestion that we should not lie to the murderer at the door. This extreme is too idealistic; others sometimes act non-ideally. Because they do so, we sometimes should act in ways that differ from how we could behave if all acted ideally. In particular, sometimes we ought to prevent really bad things from happening as a result of the non-ideal actions of others. For example, we should lie and tell the murderer that our friend is not at home, thus saving the friend's life.

Another extreme is acting in a way that would be best given how others will actually act. This extreme is not idealistic enough. If someone else will be doing something really bad and we can lessen this badness by doing something ever-so-slightly less bad ourselves, we need not do so. In other words, sometimes it is ok to act for the best, even though others will not follow suit. So, for example, as Williams emphasized, it is not acceptable to murder a protester, even if others will kill other protesters if one doesn't.

My suggestion in this section is that a similar thing holds regarding our future selves. When it comes to thinking about how to act now, we should acknowledge that our future selves will not behave ideally. And in choosing how to act, we should take an appropriate balance between two extremes. One extreme is acting in the way that would be best if our future self acted in the best way possible. This extreme is too idealistic; our future self will not always act in an ideal manner. And so we should sometimes avoid doing things that will lead to our future selves behaving poorly.

For instance, sometimes we ought to avoid putting ourselves in situations that we know might tempt ourselves to act in bad ways. If I know that I often choose to make poor decisions when in the company of a certain acquaintance, it is prudent to choose to avoid spending time with the acquaintance, even if it's the case that, if I behaved ideally, it would be fine to spend time with him.

Another extreme is acting in a way that will lead to the best outcome given how our future self will act. While this is sometimes fine to do, it is not always acceptable. That is, it is sometimes unacceptable to behave poorly now, even if behaving poorly now will mean one doesn't behave even worse later on. For example, if I am currently buying coffee and in a foul mood, it is not acceptable for me to be extremely rude to the barista, even if doing

so will allow me to let off some steam and thus avoid being even ruder to a colleague later on in the day.

In sum, this second argument has important implications regarding non-ideal theory. In particular, it brings out parallels between the cases in which others will be non-compliant and the cases in which one's future self will be non-compliant. This subject – the subject of what one should do if one's future self will be non-compliant – is an important one. But unfortunately, it is a subject that has been neglected in the philosophical literature on non-ideal theory [Rivera-López, 2013]. And thus it is significant that the second argument I have offered has important implications regarding it, implications that are worth pursuing further.

4 Conclusion

In this paper, I have defended the claim that what one ought to do in a could-but-won't case depends on further information – information about the level of goodness of the various options. I defended this claim via two arguments. If my thesis is right and my arguments successful, where should we go from here? I have three suggestions.

First, in this paper, I mostly played offense. That is, I focused on offering arguments in favor of my view. So some future work to be done is to play defense – to consider arguments for opposing views, discovering whether they go awry and what this reveals.[21]

Second, as I noted, my paper defends the conclusion that a certain sort of information is *necessary* for determining what focal characters ought to do in could-but-won't cases. But it leaves open whether the information I have focused on – information about the level of goodness of the various options – is sufficient, or whether there are yet further pieces of information that need to be provided.[22]

Third, and finally, I have offered some initial suggestions regarding how

[21] For example, some of these arguments turn on the sorts of advice it would be appropriate to give to characters like Professor Procrastinate [Feldman, 1986, 55-6], [Jackson and Pargetter, 1986, 237], [Zimmerman, 1996, 198]. And, as the reader can perhaps anticipate, I'm inclined to say that those endorsing (and criticizing) this argument haven't considered the view that what advice one should give someone turns on the degrees of goodness and badness of the various options.

[22] For example, I suggested in section 2.1 that perhaps information about chances of outcomes is also relevant.

my view connects up with non-ideal theory. But there is work to be done in fleshing out these suggestions, in examining their implications, and in defending them more fully. So, for example, there is work to be done in precisely drawing out what ideal and non-ideal theory can tell us about our relationships to our future selves and the exact ways in which this is similar and different to what it tells us about how we should relate to others.

5 Acknowledgments

Thanks for helpful comments to Asia Ferrin, Ting Lau, Doug Portmore, Fritz Warfield, Chris Woodard, attendees at a philosophy club meeting (Thomas Ferguson, Nate Fulton, Greg Glatz, Tino Petrocelli, Dan Rabinoff, Elena Rabinoff, Casey Woolwine), an audience at the Great Plains Philosophy Conference, and an anonymous referee.

References

[Baker, 2012] Baker, D. (2012). Knowing yourself – and giving up on your own agency in the process. *Australasian Journal of Philosophy*, 90(4):641–56.

[Brandt, 1988] Brandt, R. B. (1988). Fairness to indirect optimific theories in ethics. *Ethics*, 98(2):341–360.

[Cariani, Forthcoming] Cariani, F. (Forthcoming). Consequence and contrast in deontic semantics. *The Journal of Philosophy*.

[Carlson, 1999] Carlson, E. (1999). Consequentialism, alternatives, and actualism. *Philosophical Studies*, 96(3):253–68.

[Cholbi, 2009] Cholbi, M. (2009). The murderer at the door: What kant should have said. *Philosophy and Phenomenological Research*, 79(1):17–46.

[Cohen and Timmerman, Forthcoming] Cohen, Y. and Timmerman, T. (Forthcoming). Moral obligations: Actualist, possibilist, or hybridist? *Australasian Journal of Philosophy*.

[Feldman, 1986] Feldman, F. (1986). *Doing the Best We Can*. D. Reidel.

[Goble, 1993] Goble, L. (1993). The logic of obligation, 'better' and 'worse'. *Philosophical Studies*, 70(2):133–63.

[Goldman, 1976] Goldman, H. S. (1976). Dated rightness and moral imperfection. *The Philosophical Review*, 85(4):449–87.

[Goldman, 1978] Goldman, H. S. (1978). Doing the best one can. In Goldman, A. I. and Kim, J., editors, *Values and Morals*. D. Reidel.

[Greenspan, 1978] Greenspan, P. S. (1978). Oughts and determinism: A response to goldman. *The Philosophical Review*, LXXXVII(1):77–83.

[Gustafsson, 2014] Gustafsson, J. E. (2014). Combinative consequentialism and the problem of act versions. *Philosophical Studies*, 167:585–96.

[Humberstone, 1983] Humberstone, I. L. (1983). The background of circumstances. *Pacific Philosophical Quarterly*, 64:19–34.

[Jackson and Pargetter, 1986] Jackson, F. and Pargetter, R. (1986). Oughts, options, and actualism. *The Philosophical Review*, 95(2):233–55.

[Kant, 1996] Kant, I. (1996). On a supposed right to lie from philanthropy. In *Practical Philosophy*, pages 605–15. Cambridge University Press.

[Korsgaard, 1986] Korsgaard, C. M. (1986). The right to lie: Kant on dealing with evil. *Philosophy and Public Affairs*, 15(4):325–349.

[Littlejohn, 2009] Littlejohn, C. (2009). Critical notice of michael zimmerman's *Living With Uncertainty. Philosophical Books*, 50(4):235–47.

[McCloskey, 1957] McCloskey, H. J. (1957). An examination of restricted utilitarianism. *The Philosophical Review*, 66(4):466–485.

[McKinsey, 1979] McKinsey, M. (1979). Levels of obligation. *Philosophical Studies*, 35:385–95.

[Parfit, 1984] Parfit, D. (1984). *Reasons and Persons*. Oxford University Press.

[Portmore, 2011] Portmore, D. W. (2011). *Commonsense Consequentialism: Wherein Morality Meets Rationality*. Oxford University Press.

[Portmore, Forthcoming] Portmore, D. W. (Forthcoming). Perform your best option. *The Journal of Philosophy*.

[Rawls, 1971] Rawls, J. (1971). *A Theory of Justice*. The Belknap Press of Harvard University Press.

[Rivera-López, 2013] Rivera-López, E. (2013). Nonideal ethics. In LaFollette, H., editor, *The International Encyclopedia of Ethics*. Wiley-Blackwell.

[Ross, 1930] Ross, W. D. (1930). *The Right and the Good*. Clarendon Press.

[Sobel, 1976] Sobel, J. H. (1976). Utilitarianism and past and future mistakes. *Noûs*, 10(2):195–219.

[Sobel, 1982] Sobel, J. H. (1982). Utilitarian principles for imperfect agents. *Theoria*, 48(113-26).

[Stangl, 2015] Stangl, R. (2015). Taking moral risks and becoming virtuous. In Miller, C., Furr, R. M., Knobel, A., and Fleeson, W., editors, *Character; New Directions from Philosophy, Psychology, and Theology*, pages 215–232. Oxford University Press.

[Thomason, 1981] Thomason, R. H. (1981). Deontic logic and the role of freedom in moral deliberation. In Hilpinen, R., editor, *New Studies in Deontic Logic*, pages 177–86. D. Reidel.

[Timmerman, 2015] Timmerman, T. (2015). Does scrupulous securitism stand-up to security? two problems for moral securitism and how we might fix them. *Philosophical Studies*, 172:1509–28.

[van Someren Greve, 2013] van Someren Greve, R. (2013). Objective consequentialism and avoidable imperfections. *Ethical Theory and Moral Practice*, 16(3):481–92.

[Vessel, 2009] Vessel, J.-P. (2009). Defending a possibilist insight in consequentialist thought. *Philosophical Studies*, 142:183–95.

[Vessel, 28] Vessel, J.-P. (28). Against securitism, the new breed of actualism in consequentialist thought. *Utilitas*, 2(164-78).

[Vorobej, 2000] Vorobej, M. (2000). Prosaic possibilism. *Philosophical Studies*, 97:131–6.

[Wedgwood, 2009] Wedgwood, R. (2009). Against actualism. *http://peasoup.us/2009/09/against-actualism/*.

[Woodard, 2008] Woodard, C. (2008). *Reasons, Patterns, and Cooperation*. Routledge.

[Woodard, 2009] Woodard, C. (2009). What's wrong with possibilism. *Analysis*, 69(2):219–26.

[Zimmerman, 1996] Zimmerman, M. (1996). *The Concept of Moral Obligation*. Cambridge University Press.