KNOWING YOURSELF—AND GIVING UP ON YOUR OWN AGENCY IN THE PROCESS [PRE-PRINT]

Derek Baker

NOTE: This is the penultimate draft of a paper appearing in *The Australasian Journal of Philosophy*. Please cite that version.

Abstract

Are there cases in which agents ought to give up on satisfying an obligation, so that they can avoid a temptation which will lead them to freely commit an even more significant wrong? *Actualists* say yes. *Possibilists* say no. Both positions have absurd consequences.

This paper argues that common-sense morality is committed to an inconsistent triad of principles. This inconsistency becomes acute when we consider the cases that motivate the possibilism–actualism debate. So the absurd consequences of both solutions are unsurprising: any proposed solution will have consequences incompatible with common moral practice.

Arguments for denying one of the principles are considered and rejected. The paper then suggests that the inconsistent moral commitments originate out of an inconsistent picture of human agency. Revisionary pictures of human agency are considered. It is argued that a quasi-Platonic picture of agency, similar to that advocated by Gary Watson [1977], is the most promising.

Keywords: Free will, responsibility, options, possibilism, actualism, Professor Procrastinate

1. The First Debate

The unfortunate case of Professor Procrastinate [Jackson and Pargetter 1986] goes like this: The Professor is invited to review a friend and colleague's new book. His review, if he wrote it, would bring out contributions of the book that other reviewers would miss. (His friend is a clear thinker but a murky writer.) So he seems obligated by both friendship and profession to accept the review and write it. But he's Professor Procrastinate. He knows he won't ever write the review if he accepts it, and no review at all is the worst outcome possible.

So should the professor turn down the offer to write the review?[1] (In order to avoid complications regarding the objective and the subjective ought, let's assume that the professor is not ignorant of any relevant facts, moral or non-moral, and that he can reason perfectly well about these facts. The objective and subjective ought will recommend that he do the same thing. Let's also assume that the other agents discussed in this paper are in a similar position.)

Actualists say an agent ought to choose the option that results in his performing the best course of action, given what, for each option, the agent would do if he chose it. If Procrastinate accepts, he won't write the review. This course of action is inferior to the one

[1] For more dilemmas see [Goldman 1977; Thomason 1981; Jackson and Pargetter 1986; Jackson and Altham 1988; Humberstone 1991; Carlson 1999; Louise 2009].

he would perform if he turned down the review. So he ought to turn down the review. When our vices will lead to problems, we ought to opt for less damage.

Possibilists believe an agent ought to choose the option consistent with the best course of action it is within the agent's power to pursue. Procrastinate is free to write the review: he suffers from a vice, not a compulsion. Accepting the offer is the only option consistent with writing the review, which is the best outcome within his power. So he ought to accept. Ought implies *can.* No one ever said ought implies *is likely to.*[2]

In this debate, both sides present reductios against the opposing view. Possibilists require us to blunder into foreseeable disasters. Actualists tell us we are justified in abandoning our obligations because we will abandon our obligations.[3] These reductios are successful. Both positions have absurd consequences. We are strongly committed to inconsistent moral principles, thanks to an inconsistent picture of free will, or so I will argue.

The exercise of will is a choice about what to do in a situation. That situation can be described as a set of options, along with the outcomes (probable or deterministic) of those options. The situation is determined, uncontroversially, by the objects, their dispositions, other people, their dispositions, and one's own skills and talents.

What's unclear is whether the agent's own agency—or will—is part of the situation. Procrastinate will not write the review if he accepts, but that's only because he will keep freely choosing to put off the work. This fact is, or will be, under Procrastinate's immediate control. Can he treat it as a given of his situation when he could, presumably, make it go away?

2. A Parallel Debate and an Inconsistent Triad

This problem—whether one's agency is a component of one's situation—comes up in another debate, about the relation between the choices of the perfect agent and the right thing to do. One family of views holds that the right thing to do in some situation is what a perfect agent would do in an otherwise identical situation [Korsgaard 1986; Hursthouse 1999]. There is an obvious objection, however, if an agent's vices are a component of the situation he faces. A perfect agent has no vices. So when we imagine what the perfect agent would do in an otherwise identical situation, we have potentially altered the situation.

Michael Smith [1995] objects to imitationist views with the following example: a tennis player—let's call him Bob—loses his match. After losing, the perfect agent would walk over to his opponent and shake his hand. But Bob is enraged by defeat. If he gets close to his opponent, he will attack him.[4]

So what should Bob do? Smith thinks it is obvious that he should not shake hands. So what the perfect agent would do, and what an imperfect agent should do, must come apart.

This is Smith's account. I'm torn. Without question, that Bob will attack his opponent if he shakes hands is a decisive reason against hand-shaking. But, at the same time, failing to shake hands is rude. And it seems wrong that Bob's rudeness is *justified* by the fact that he's a belligerent ape. But decisive reasons justify. So why this conflicted feeling?

[2] Professor Procrastinate's dilemma has a diachronic structure, as do the other examples that will be considered in this paper. The diachronic cases are the most troubling and interesting. But for examples of synchronic cases, see [Jackson and Pargetter 1986: 236; Louise 2009: 332].

[3] For examples see page 19.

[4] The case comes originally from Watson [1975].

This: if displays of bad sportsmanship are morally wrong, then they cannot suddenly become morally right simply because one is prone to violent tantrums. Vices do not downgrade one's obligations, permitting the bad, so long as one refrains from the terrible.

It seems that an agent's psychological dispositions must be part of his situation, because they obviously affect the consequences of current choices just like the dispositions of objects. But, at the same time, the agent is responsible for them. Treating one's psychological disposition as the disposition of an object, or the dispositions of another person, is inauthentic: one is treating one's agency as a foreign object, outside of one's control.[5] If Bob punches, it will be because he chose to.

Our feelings about what Bob or the Professor should do can be summed up in three conflicting premises:

1. A psychological state which will lead you to perform a morally wrong action in certain situations gives you decisive moral reason to avoid those situations, so long as nothing of greater importance is sacrificed, and even if you will perform the wrong act freely.

2. If you have decisive moral reason to X, you are morally justified in Xing; if you have decisive moral reason to X, X is a morally right action.

3. An otherwise morally unjustified (or wrong) action is not justified (or made right) by the fact that, if one did not perform it, one would freely perform a more serious wrong.

I will consider arguments for denying 1 and 3, which fail. 2 should be revised, but even with the revision the paradox stands. I will then consider Jackson and Pargetter's idea of relativized ought-claims as an alternative to 2 which could potentially solve the problem; unfortunately, ought-judgments do not function as their theory suggests.

The paper will conclude with two points. First, our moral commitments are inconsistent, so it isn't surprising that possibilists and actualists can run successful reductios against each other; but a real solution to the problem must be a principled rejection of one of these commitments. Second, we will most likely need a revisionary picture of agency to resolve the paradox. Our picture of the will and the self is incoherent.

3. Denying 1

How could we deny 1? Perhaps we cannot take a predictive stance towards ourselves, at least in those cases in which we are free.

But we predict other people's behaviour all the time, and still regard them as free agents, based on their desires, vices, virtues, and quirks. These are not normally understood to vitiate freedom. If they did, social rationality among free agents would be impossible. I know that if I suggest going out to eat tonight, my wife will say no, though I still believe her

[5] It's important that angry Bob suffer from a standard vice, rather than a pathology. Treating a pathology as a feature of the environment, rather than a part of oneself, is not obviously inauthentic [Humberstone 1991: 154].

free to say otherwise. So I can know my own future behaviour without believing myself compelled.[6]

Perhaps predicting one's future behaviour isn't an implicit denial of freedom, but a failure to regard the freedom as one's own. You are treating your future self as a distinct agent [Thomason 1981: 186]. Your future actions are not mere consequences of your current choices, but things which you have direct control over.

There's something to this. I am in control of what I will do in the future. But, at the same time, the control seems imperfect. Years of experience with my future self have taught me that, despite the intimacy of our relationship, he frequently lets me down. Deciding now to do right when temptation arrives doesn't allow me to reason as though temptation were already overcome. If Bob knows that he typically surrenders to his anger despite resolutions not to, approaching the net seems horribly irresponsible, no matter how decisive today's commitment might feel.

Moreover, self-management is a basic part of practical reasoning. Someone who frequently overeats should not keep junk food around the house. Even if there's a greater long-term benefit he could secure for himself (friends visiting more often, say) by keeping junk food around and exercising self-control, he still shouldn't keep the junk food around the house. But if we shouldn't reason like possibilists in the prudential case, why should we reason like them in the moral case [Woodward 2009]?

A denial of 1 is incompatible with the most elementary aspects of practical reasoning: predicting what will happen in the world when one makes a choice. Some bad behaviour is predictable. How could agents be required to choose in a way which ignores easily predictable events?

4. Denying 3

So what about denying 3? Well, Bob's options seem to be: *cause offense* or *risk a complete disaster*. The moral reasons seem pretty clearly to favour causing offense. And what would morality have Bob do, if not what he has most moral reason to do? So Bob must be justified in causing offense, and justified by a wrong he would freely choose.

This is essentially to argue from the truth of 1 and 2 to the falsity of 3. But one could just as easily move from the 2 and 3 to a rejection of 1. Bob's options are not *cause offense* or *risk a complete disaster*. We assumed at the beginning that Bob was free not to attack. So Bob's options are: *walk away*, *shake with violence*, or *shake without violence*. By choosing the last, Bob would meet all of his obligations. *Walk away* would violate an obligation. So the moral reasons pretty clearly couldn't favour *walking away*.

"What else would morality have Bob do?" Morality would have Bob suck it up, show some self-control and basic decency, and shake hands with a smile. Deliberating as though resisting temptation is not an option, but instead an unlikely consequence of a risky gamble, falsifies the situation. Risk entails that the ultimate outcome is not completely under

[6] It's worth remembering that *knowledge* in the ordinary sense doesn't require certainty. If I am highly confident that A will X, am justified in that confidence, and A will in fact X, then, absent additional strangeness (like Gettier cases or lotteries), I know that A will X.

Likewise, it's worth remembering that when we say that if Bob were to shake hands, he would swing, this is not to say that necessarily, if Bob shakes hands, he swings. He could shake hands without violence (it is possible), but he wouldn't.

But if this remains unconvincing (or if we assume indeterministic worlds), assume instead that Bob knows that violence on his part is extremely probable. His dilemma remains.

one's control. Here the outcome is up to Bob, and nothing else. So a list of his options which lacks *shaking hands without violence* treats Bob as incapacitated when he is not.

A denier of 3 must give some explanation, moreover, of why we're inclined to say characters like Bob and the Professor are unjustified, as soon as the stakes are raised. Consider Sarah, who is in the control room of the power plant during an emergency. If she pulls the lever and holds it down briefly, she will prevent the explosion that will kill everyone. Unfortunately, there's a spider on the lever, and she is arachnophobic (though not in a way that compromises her freedom), so if she tries to hold the lever down, she'll eventually choose to give up and flee the control room leaving everyone to die. If instead she pushes the alarm button and runs out of the room, half the people in the plant will survive [Louise 2009: 332].[7]

Sarah cannot justify her decision to let half her co-workers die on the grounds that if she didn't, she would have chosen to let them all die.

An opponent of 3 could say that our intuitions are misleading. Bob has a bad character (as does Sarah). He does the right thing, but we feel that he must be unjustified because his character traits are unjustified. So of course we have the intuition that Bob is unjustified in walking away. It's easy for us to misidentify the source of blameworthiness.

Bob might be blameworthy for his character, but walking away looks unjustified nonetheless. Let's say that Bob walks away. And let's also say he has a counterpart, Bob*, who shakes hands peacefully, but is otherwise as much like Bob psychologically as possible: he still becomes enraged, and must exercise self-control to avoid beating his opponent. It seems that we might blame both for having murderous thoughts. But Bob*'s action still seems superior to Bob's. Bob's opponent could rightly be annoyed by Bob's rudeness; Bob*'s opponent couldn't.

Now we have presupposed that Bob, with his current character, is free to shake hands peaceably—is free to do what Bob* did. And Bob and Bob* must either have identical characters, or different characters.

Let's say Bob and Bob* act with the same character. Even if Bob*'s character is as unappealing as Bob's, his action is superior. Both feel inappropriate anger at someone: that violates one obligation; that's one thing for which they are unjustified. But we judge Bob*'s action to be better, because Bob did an additional unjustified thing. He violated one more obligation, which Bob* did not.

But maybe there must be some slight difference between the characters of Bob* and Bob in order to explain the greater level of self-control exercised on this occasion by Bob*. Then, since we have assumed that Bob is free to act as Bob* acts, we must assume that Bob was free to act with a slightly better character than he did. So saying that we blame Bob for his character no longer implies any denial of 3.

To summarize, denying 3 commits us to an inauthentic style of deliberation. It denies us options we really are free to take, simply because we will decide not to take them. And it forces us to say that bad agents may permissibly perform evil acts, so long as they refrain from the monstrous.

5. Denying 2

How could 2 be denied? Perhaps by pointing out that obligations can conflict. And when obligations conflict all of one's options will be impermissible. So whatever choice one makes will be unjustified.

[7] The example has been modified to more closely resemble Procrastinate's case.

Consider a person who made conflicting promises. He is responsible for limiting his options to bad ones, and so he will have no justification for whichever of the wrongs he ends up choosing to commit.

Or someone could face a moral blind alley, a tragic and irresolvable moral dilemma [Williams 1965; Nagel 1976]. He is not responsible for the situation, but nonetheless he now must choose which obligation to abandon. And he will be responsible for this choice.

Moral reasons might still favour one alternative or another, even when obligations conflict. But they no longer serve to justify an action (otherwise one obligation would simply disappear). Instead, the reasons recommend moral damage-control: how to minimize the harm of one's inevitably wrongful behaviour.

So we should deny 2. But it remains unclear how this would apply to Bob's case. It's been stipulated, this entire time, that he is free to shake hands without swinging. Nothing we've said yet would justify the claim that all of Bob's options are impermissible [Jackson and Pargetter 1986: fn. 12].

We've assumed Bob is free to meet all his obligations. So attempting to reconcile Bob's plight with 'ought implies can' (by finding a prior action for which he's culpable), or attempting to weaken 'ought implies can' (by allowing blind alleys), simply misplaces the source of our confusion.

Which means that if we replace 2 with

2′. If you have most moral reason to X, you are morally justified in Xing (i.e., Xing is morally right), unless you cannot meet all of your obligations.

then we should add

4. Bob and company are able to meet all of their obligations.

The inconsistency in our moral commitments is still apparent.

6. An Alternative Denial of 2: Jackson and Pargetter's Relativized Oughts[8]

Frank Jackson and Robert Pargetter in their [1986] paper on the possibilism–actualism debate argue for an alternative replacement for 2.

They believe that the truth of ought-claims is always relativized. As they describe it, different ought-claims 'select' from different sets of options [ibid.: 244–5]. This doesn't simply mean that what one ought to do is relative to the situation one is in. That is uncontroversial. (Everyone would agree that if Officer Jones ought to chase the shoplifter, it's because he can't prevent any more serious crimes at the moment.) Instead, an ought-claim always recommends the action it prescribes, not as the absolutely best option, but as superior to what the agent would have done otherwise.

As they [ibid.: 247] put it:

> For each A, it ought to be done by an agent just if what he would do if he did A is better than what he would do if he did not do A, that is, if it is what ought to be done out of the set of options consisting of what would be done if A were done, and what would be done if A were not done.

[8] This section benefitted greatly from my discussions with Arthur Chen.

Bob ought to walk away, because if he doesn't walk away, he'll punch. *Walk away* is part of the set of alternatives [*walk away*, *shake hands with violence*], and is the preferable alternative. And Bob ought to shake hands without punching, because if he doesn't shake hands without punching he'll shake hands and punch. *Shake hands without violence* is part of the set [*shake hands without violence*, *shake hands with violence*] and is the preferable alternative as well.

Now these obligations may appear to be incompatible. But according to Jackson and Pargetter, they aren't. The first obligation holds relative to one set of alternatives, the second obligation relative to another set.

> ...[A]lthough incompatible prescriptions out of the same set of options are objectionable, there is nothing particularly puzzling in incompatible prescriptions out of different sets of options. …
>
> The impossibility of my both going to the doctor and staying home... shows that they cannot be both what I ought to do out of the same set of options, but they can be (and are) both what I ought to do out of different sets of options.
> [Ibid.: 245]

Characters like Bob can meet all of their obligations, moreover [ibid.: 242–3]. Bob is obligated to walk away because of the truth of the counterfactual *if Bob were to shake hands then he would punch.* A world in which he meets his obligation to shake hands without violence is, however, a world in which that counterfactual is false. So meeting this obligation causes the obligation to walk away to disappear. Bob can do all that he ought, because one obligation 'overrides' the other [ibid.: 243].

The Jackson-Pargetter replacement for 2, then, would be:

> 2″. If you have decisive moral reason to X instead of Y and Y is the relevant alternative to X (i.e., Y is what you would do if you did not X), you are morally justified in Xing relative to the set of options [X, Y].

Their theory, then, has the following advantages if it works. We can acknowledge Bob's freedom. We can advise him to be prudent. And we can still hold that he's obliged to behave like any other decent member of society. (Of course, acquiring these advantages will also require rewriting 1 and 3 as claims about relativized reasons and relativized justification.)

Unfortunately, ought-judgments do not work the way their account predicts.

Imagine that a guru suddenly appears at the tennis court and advises caution to Bob: 'You ought to walk away.' But then a second, sterner guru appears and tells Bob, 'You ought to shake hands without violence.' According to Jackson and Pargetter, both claims are true, and this means that the two gurus are not disagreeing with each other. But they certainly seem to be disagreeing with each other. I find it hard to hear them doing anything else; and it is hard to see how else Bob could interpret them.

The explanation of why they are not disagreeing is that they are prescribing for different sets of options. But remember, these ought-claims are meant to apply to one and the same situation. And if the situation is the same, then the agent's options must be the same. Her abilities are the same. The physical environment in which she acts remains unchanged. And one's options are the things one is free to do, things within one's power.[9] So what do

[9] Thanks to John Maier for this point.

Jackson and Pargetter have in mind when they say that these ought-claims prescribe for different sets of options, if they are meant to apply to one and the same situation? At first glance, it seems that at least one of the gurus must be advising Bob on what to do in a situation other than the one he really faces.

Jackson and Pargetter do not explicitly address this problem. They write that, 'Actualism comes into play not just in evaluating the options, but also in determining which options are to be evaluated' [ibid.: 247]. But this cannot mean that options like *shake hands without violence* are never to be evaluated. 'Bob ought to walk away without violence' is supposed to be true, after all. The option is not to be evaluated, given certain statements or questions [ibid.: 249–52].

Jackson and Pargetter might mean that certain options are conversationally irrelevant. Presumably, then, for a given prescription, the conversationally relevant options are the prescribed option, and what the agent would do otherwise. But if this is what they mean, it's unclear what evidence they have in support.

Let's return to the two gurus, advising Bob: the stern guru's advice does not seem conversationally inappropriate. He seems to be contradicting the advice of the cautious guru; he does not seem to be changing the subject, or answering a question nobody asked. Bringing up the possibility of shaking hands without violence doesn't seem pedantic. On the other hand, if the cautious guru were to respond to his advice with, 'We weren't talking about that!' *that* really would sound bizarre.

For contrast, imagine that the stern guru instead says, 'You ought to watch less TV.' That does sound conversationally inappropriate; it does sound like changing the subject; it is not in disagreement with the cautious guru's advice. And in that case the response, 'We weren't talking about that!' sounds completely correct.

So if the Jackson-Pargetter account is meant as a claim about conversational rules governing advice-giving, we need more evidence that these are in fact the rules. If not, we need more detail on what the theory is claiming; specifically, we need to know how a prescription can apply to an agent in a situation, when it prescribes from out of a set of options other than the actual set of things the agent is free to do. And in either case, they need to provide some explanation of why we hear the two gurus as disagreeing with each other, when their theory states that they are not. Without such an explanation, their theory seems to conflict badly with our linguistic intuitions, and not just in some strange, peripheral cases, but in the paradigmatic possibilist-actualist dilemmas which were supposed to motivate that theory in the first place.

A final point: practical rationality is generally taken to be a matter of doing what one has most reason to do. On the Jackson-Pargetter account, however, reasons are decisive only relative to a set of options. So, Bob has more reason to X than Y, and more reason to Y than Z, but there is no fact of the matter about which option he has *most* reason to perform. But we started with the dilemma: what would it be most reasonable for Bob, in his situation, to do? According to Jackson and Pargetter, this question is unanswerable. Given the structure of ought-claims, asking it is a mistake.

Now, the question seems well-formed. This appearance could turn out to be misleading. But it gives us a reason to prefer an account which gives an answer to an account which calls the problem unanswerable. The next section will consider other accounts: that they answer the question, while the Jackson-Pargetter view does not, is a point in their favour.

2″, at this point, appears incorrect as an account of the way ought-judgments function, and its key notion of relativized sets of options is obscure. So until the theory is further clarified, we're stuck with the fact that 1, 2′, 3, and 4 all look unassailable, except that one of them must be false.

7. Theoretical Payoffs

The arguments of this paper, if successful, seem to have demonstrated our commitment to an inescapable inconsistency. But the point has not been to lead the reader to a state of Pyrrhonian *ataraxia*. There are theoretical payoffs to framing this debate in terms of an inconsistent tetrad.

7.1. Methodological Points

The first payoff is methodological. Jackson and Pargetter [ibid.: 239] point out that according to possibilism:

> Even if I know that I ought to arrange for a taxi given I will drink too much tonight, it is wrong for me to ask myself whether I will in fact drink too much. … We submit this as a reductio.

Humberstone [1991: 154] points out using facts about what one will do to move from conditional oughts to full-fledged oughts results in unacceptable claims about what one ought to do:

> …[S]uppose that you are in fact going to stab the person sitting on your left, your reason being, let us imagine, simply that you are irritated… and you are now considering which knife you ought to use—a short one, or a long one. Given that you are going to stab him, you ought to use the shorter knife… Thus, if we were to endorse the detachment inference, we should have to draw the conclusion that you should stab the person on your left with the short knife.

Jackson and Pargetter are pointing out that possibilism is inconsistent with principle 1. From this they conclude actualism. Humberstone responds by pointing out that actualism is inconsistent with 3; ergo, possibilism. If the thesis of this paper is correct, this is a bad way to argue for either position. Both actualism and possibilism are clearly inconsistent with strongly held moral principles. But that's because our principles themselves are inconsistent. So, unless a principled denial of principle 1 or 3 can be offered, it is illegitimate to use a reductio against the rival view as evidence for one's preferred theory—the reductios are simply too easy to produce.

Moderate actualists such as Carlson and Louise have been more sensitive to the need to account for both practices of self-management and accountability. But it is unclear which principle, 1, 2′, 3, and 4, they would revise or deny. This points to an advantage of stating the problem as an inconsistent tetrad: by forcing theories to be explicit about which claim they would revise, we can clarify the dispute and the nature of rival solutions.[10]

A solution to the possibilism–actualism debate needs a clear and motivated denial of at least one of the four claims. We must either be able to square the thought that Bob had

[10] Along these lines, Louise [2009: 331] complains that most defenders of possibilism defend a moderate version of the thesis, so that it is hard to determine what exactly about actualism that they find objectionable. With the inconsistent tetrad, philosophers can pinpoint how their proposed solutions differ.

most reason to walk away with the thought that he was never permitted to act rudely, or we must give up one of the two thoughts.[11] Otherwise our commitments remain inconsistent.

Framing the debate in terms of an inconsistent tetrad allows us to become clearer on how the rival solutions differ (which premises would they deny or revise). It clarifies what types of arguments are needed to solve the problem (denials of one premise or another must be motivated). And it shows that a standard form of argument in this debate so far, the reductio, tells us very little about which position is correct.

7.2. Radical Solutions

It's a commonplace that philosophers should defer to common sense, when possible. But if our common-sense commitments are inconsistent that sort of deference is impossible. We are justified, then, in looking to more radical theories in order to resolve Bob and Procrastinate's dilemmas.

We may need to adopt a revisionary picture of agency. For example, I have treated the freedom to choose an option as an all or nothing matter. But maybe freedom comes in degrees. R. Jay Wallace [1999 :654], for instance, argues that addiction 'impairs one's capacities for reflective agency,' even though the agent is still 'equipped *in a basic degree* with the powers of reflective self control' (*italics mine*). Louise [2009: 340] says that agency is 'something that comes in degrees.'

Neither philosopher acknowledges this as a potential revision of the concept of freedom, but it is. We do not generally think of someone as partially free to do something, and if we were to start, we would need to answer several questions. How should decision theory represent one's partial options? How do reasons favour and disfavour partial alternatives, as opposed to full-fledged ones? How does the 'ought implies can' principle apply to obligations I am partially able to fulfil? Do partial obligations result? 'You kind of must tell the truth' sounds bad.

The most natural way of representing a partial ability to X denies the agent is genuinely free to X at all. Instead, we treat the agent as completely free to Y, which has some probability of resulting in the agent's Xing. For instance, I can 'sort-of' hit the bull's-eye because I am completely free to perform some more basic action, which results in my hitting the bull's-eye with imperfect reliability. But if this is what my partial ability amounts to, the most I could be obligated to do is try my hardest to hit the bull's-eye. Ought implies can, and increasing the likelihood that the bull's-eye is hit is the extent of my can.

On this model, if Bob is partially able to control his temper, he would not really be free to refrain from punching. He would only be free to perform some mental exercise and take deep breaths, which has some chance of causing his pathology to subside. So his options really are *walk away* or *risk disaster.*

This sort of picture fits naturally with a quasi-Platonic model, for example the model put forward by Gary Watson [1977], which decomposes the person into his rational capacities plus a set of non-rational dispositions. The genuine agent here is identical with the rational capacities, and good agency is a matter of successfully shepherding one's non-rational dispositions. The agent is obligated to shepherd as skilfully as is within his power, since that

[11] Christopher Woodward [2009: 225-6] claims that we should think of the possibilism-actualism debate solely in terms of reasons; obligations have been a red-herring that have introduced confusion into the debate. This isn't necessarily wrong, but this approach risks glossing over the reason these cases trouble us: reasons don't simply favor one alternative or the other, they also justify that alternative, and in these cases they seem to come apart; moreover, the obvious explanations for why they come apart (discussed in **Section 5)** fail.

is what he is free to do. This picture grants an unquestionably legitimate reason against shaking: it would be too risky. Bob is stuck with an awful *thumos*, and it sometimes slips its leash.

So we can resolve the problem by declaring that there are no non-pathological vices; whenever a person succumbs to temptation, it is because the genuine agent was overpowered by a non-rational psychological state.[12] Then we can deny 1 while allowing that self-management is justified. Or, alternately, we can say that the difference between a pathology and a vice is that the merely vicious successfully resist the vice some reasonable percentage of the time, though they lack complete control. Then we can deny 4: walking away is impermissible because rude; shaking hands is impermissible because *genuinely* risky.

Another revisionary option involves a different sort of decomposition of the agent. We might interpret Derek Parfit's [1984] arguments to show that Bob at *t-1* is, at the fundamental level, a different agent from Bob at *t-2*: the person over time is a category we use because of our parochial interests, or because it usually coincides with psychological continuity, but the real locus of psychological activity is the person at a time. In this case, Bob at *t-1* really should treat his future self as this other, unpleasant fellow, whom he should take steps to keep away from his tennis partner. (For similar suggestions, see [Jackson and Althman 1984; Carlson 1999: 265–6].)

This solution allows us to say that 1 and 3 only appear to be in conflict thanks to a fallacy of equivocation. 1 is a claim about how what one's future self—a separate agent—will freely do. 3 is a claim about how the fact that an agent would freely choose some greater wrong cannot justify the very same agent choosing the lesser wrong.

Both pictures, with their serious revision of the scope of an agent's powers, imply similarly severe revisions of our obligations, and of when agents can justly be held responsible.

A third revisionary picture would involve decomposing the notion of 'freedom' rather than agents. I may be free to clean my office, and Procrastinate to write the review, but there is a stricter sense of 'freedom' which includes only those actions immediately performable. There is a sense in which one ought to do the best thing one is loosely free to do, but that ought is not action-guiding. Ought in the action-guiding sense implies the strict sense of can.[13]

This story may be able to resolve the inconsistency. 1, 3, and 4 use a univocal notion of 'free to,' and so would presumably only appear to be inconsistent, again thanks to fallacy of equivocation. The story must explain, however, why we use a single concept of 'freedom' to pick out two different relations; it must also explain the function of the impractical ought associated with one of the relations.

A conclusive argument for any of these solutions is beyond the scope of this paper. It would require comparison of the arguments for and against each. But I would like to point to several attractions of my preferred solution, a quasi-Platonic model of agency.

With both Bob and Procrastinate we feel that there is a sense in which they are free to meet their more demanding obligations, but also a sense in which they aren't. This is clearly at work in the idea of partial freedom, but also seems to be at work in the Jackson-Pargetter account of an agent's options shifting with conversational context, and in the suggestion that

[12] Again, Watson [ibid.] would provide the model for this sort of solution. He argues that weakness of will is an incoherent category—agents act badly because they judge they have most reason to, or else they suffer from a pathological motive.

[13] Thanks to Michael Smith for this suggestion; also see Jackson's [1998: 44-5] treatment of freedom as a folk-concept to be revised.

'freedom' might have a strict and loose sense. The Platonic model can explain and clarify these intuitions in a principled and appealing manner.

As has already been pointed out, the Platonic model can easily account for partial freedom by treating self-control as an imperfect skill. The same device can be used to explain and clarify a disjunctive notion of freedom. Strict freedom is complete freedom: the agent is strictly free to X just in case she would be guaranteed to X if she tried to X. Loose freedom is partial freedom: the agent is loosely free to X just in case she would have a reasonable chance of Xing if she tried. (Or, since we're assuming deterministic worlds, an agent is strictly free to X just in case she Xs in all nearby worlds in which she tries; and loosely free if she Xs in a reasonable portion of nearby worlds in which she tries.)[14]

The Platonic story can further explain why we would identify these two different relations (strict freedom and loose freedom) with a single term. As Watson [1977: 333–6] points out, even if we concluded that akratic agents were not, strictly speaking, free to resist their temptation, the folk concept of weakness of will would have a justification: its role in our practices of accountability. Blaming those with extraordinarily hard to resist compulsions for succumbing would be pointlessly cruel. But holding those with more manageable compulsions accountable encourages people to try to resist temptation, and it incentivizes taking steps to improve one's powers of self-control. If, on the other hand, we were to start treating it as a genuinely open question whether, in the majority of cases, the wrong-doer were really culpable, we would be inviting mass irresponsibility.

So, freedom in the strict sense might be the only genuine freedom. Freedom in the loose sense depends too much on the cooperation of things outside one's rational control. Yet we still comprehend both relations under the term 'freedom,' to mark out the range of actions society could reasonably demand of us.

From this, we also get an obvious story of the two corresponding senses of 'ought.' Strict freedom corresponds to the ought of advice, of most reasons, of rational self-control; loose freedom to the ought of moral (or societal) expectation, the things you ought to do to avoid accountability. (This picture lets us explain some of the intuitions behind the Jackson-Pargetter view as well.)

Finally, the Platonic model provides an explanation of why Possibilism (or denying 1) has some appeal in moral cases, but none in purely prudential cases. Our intuitions are asymmetrical because society can legitimately blame or sanction agents for lacking the powers of self-control necessary to achieving certain minimal moral goods, essentially as a way of coercing them to augment those powers. This same coercion in the name of the agent's own good would be intolerable, however. We make the mistake of thinking that Bob can't have decisive reason to walk away, because he would remain blameworthy, according to reasonable conventions. In prudential cases, there are no conventions of blame to skew our intuitions.

---

[14] Honoré [1964] analyzes a *general* ability to X in terms of normally succeeding at X when one tries. Smith [2003] develops this idea into an account of *particular* abilities in terms of counterfactual success. It should be noted, however, that Smith argues convincingly that the counterfactual account of abilities provided above is a serious oversimplification: abilities shouldn't be analyzed in terms of success in the nearest possible worlds, but rather success in a raft of privileged possible worlds. The Platonic model can accommodate this: An agent is strictly free to X if she Xs in all privileged scenarios in which she tries, and loosely free to X if she Xs in a reasonable portion of the privileged scenarios in which she tries.

It's also worth noting here that Smith [ibid.] treats agents with imperfect counterfactual success at Xing in the privileged scenarios as completely free to X. He may be correct—that is, the Platonic picture might be wrong—but then we will need either an argument for denying one of the premises of the tetrad, or another way of characterizing the distinction between strict and loose freedom, in order to escape the paradox.

This is also the key strength the Platonic model seems to possess over the Parfitian. Parfit [1984: 318–20] argues from his picture of identity over time to the conclusion that prudential requirements, which tell us to look after the interests of our future selves, are a species of moral requirements, which tell us to look after the welfare of others. The future self is, after all, another. But then the Parfitian solution conflicts with our intuition that choosing the lesser evil might be unjustified in the moral case, but is perfectly respectable in the prudential. The Platonic model more successfully respects this intuition, and explains it.

The Platonic model, then, has these attractions: it provides a simple picture of self-control as the exercise of an imperfect skill; it can use the picture to explain a number of our intuitions; and it provides a rationale for standard moral practice. It does this by denying that one's agency can ever be part of one's situation, but then going on to deny that anything is part of one's agency so long as one's rational control over it is imperfect.

8. Conclusion

Is a revisionary theory necessary? It's hard to say for sure—tinkering with the four premises may solve the problem. But I suspect revision is more likely the right answer. Recall our initial way of framing the problem: how does agency itself fit into the agent's situation?

The problem is that we think of ourselves as both free and as part of the causal order. The tension between these two pictures is normally taken to be resolved if we can show our freedom to be compatible with determinism. But the problem also asserts itself at the practical level. When determining what morality requires of us, we must either see the components of our psychology as part of the causal order, or as elements in our free agency. We cannot sensibly regard them as both, because each picture comes with a unique picture of our responsibility for these psychological states.

If they are part of the causal order, we are responsible for manipulating and managing these states just as we are responsible for manipulating and managing objects in the physical environment. But our responsibility extends no further. We can be faulted for recklessly leaving them about, or intentionally using them to cause harm, but we are not at fault when they harm others in unforeseeable ways. 'How was I to know my anger would throw my fist your way when you said that? It had never happened before.' Nor can we be blamed for leaving some other good unrealized, because we were occupied with managing an unruly vice.

On the other hand, if they are elements of our free agency then our responsibility for them is unlimited. At the same time, managing them must be a self-deceptive sort of task, because it is both unnecessary and impossible. Unnecessary, because a self-determining will cannot *succumb* to temptation if it chooses not to succumb. Impossible, because if the desire to strike my opponent is part of that self-determining will, the one making the choice, then it has power to overrule any of my prescriptions against it, because I have that power and cannot give it up.[15]

Lingnan University

References

[15] Thanks to Arthur Chen, Colin Klein, John Maier, Tristram McPherson, Michael Smith, Kelly Trogdon, my audiences at the University of Hong Kong's February 2010 colloquium and the Australasian Association of Philosophy's 2010 Conference at the University of New South Wales, and my anonymous reviewers for criticisms, discussion, and advice.

Carlson, Erik 1999. Consequentialism, Alternatives, and Actualism, *Philosophical Studies* 96/3: 253–68.

Goldman, Holly S. 1977. Dated Rightness and Imperfection, *The Philosophical Review* 85/6: 449–87.

Humberstone, I. L. 1991. Two Kinds of Agent-Relativity, *The Philosophical Quarterly* 41/163**:** 144–66.

Hursthouse, Rosalind 1999. *On Virtue Ethics,* Oxford and New York: Oxford University Press.

Honoré, A. M. 1964. Can and Can't, *Mind* 73/292: 463–79.

Jackson, Frank 1998. *From Metaphysics to Ethics: A Defense of Conceptual Analysis,* Oxford and New York: Oxford University Press.

Jackson, Frank, and Robert Pargetter 1986. Oughts, Options, and Actualism, *The Philosophical Review* 95/2: 233–55.

Jackson, Frank, and J. E. J. Altham 1988. Understanding the Logic of Obligation, *Proceedings of the Aristotelian Society, Supplementary Volumes* 62**:** 255–83.

Korsgaard, Christine 1986. Skepticism about Practical Reason, *Journal of Philosophy* 83/1: 5–25.

Louise, Jennie 2009. I Won't Do It! Self Prediction, Moral Obligation and Moral Deliberation, *Philosophical Studies* 146: 327–48.

Nagel, Thomas 1972. War and Massacre, *Philosophy and Public Affairs* 1/2: 123–44.

Parfit, Derek 1984. *Reasons and Persons,* Oxford and New York: Oxford University Press).

Smith, Michael 1995. Internal Reasons, *Philosophy and Phenomenological Research* 55/1: 109–31.

Smith, Michael 2003. Rational Capacities (or, How to Distinguish Recklessness, Weakness, and Compulsion), in *Weakness of Will and Varieties of Practical Irrationality,* ed. Christine Tappolet and Sarah Stroud, Cambridge: Cambridge University Press: 17–38.

Thomason, Richmond H. 1981. Deontic Logic and the Role of Freedom in Moral Deliberation, in *New Studies in Deontic Logic,* ed. R. Hilpinen*,* Holland: Reidel: 177–86.

Wallace, R. Jay 1999. Addiction as Defect of the Will: Some Philosophical Reflections, *Law and Philosophy* 18/6: 621–54.

Watson, Gary 1975. Free Agency, *The Journal of Philosophy* 72/8: 205–20.

Watson, Gary 1977. Skepticism about Weakness of Will, *The Philosophical Review,* 83/3: 316–339.

Williams, Bernard 1965. Ethical Consistency, *Proceedings of the Aristotelian Society, Supplementary Volumes* 39: 103–24; reprinted in *Problems of the Self,* ed. Bernard Williams 1973, Cambridge and New York: Cambridge University Press: 166–86.

Woodward, Christopher 2009. What's Wrong with Possibilism, *Analysis* 69/2: 219–26.