Consequentialism and the Evaluation of Action *qua* Action

Andrew Sepielli
University of Toronto

Forthcoming in
Suikkanen and Kauppinen, *Methodology and Moral Philosophy* (Routledge)
Please ask if you'd like to cite this version.

1. Introduction

It's typically thought that consequentialists and deontologists are offering competing theories of the same thing. But what is that thing? Is it goodness or value '*qua* action' (Korsgaard 1996, p. 62; Schapiro 2001, p. 99; Katsafanas 2013, p. 61)? 'Acting well' (Muller 2004; Hanser 2008)? The 'excellent exercise' of 'agency' (Strom 2011, p. 232) or 'practical reason' (Lear 2004, p. 3; Vogler 2002, p. 29)? One sometimes sees these put forward as the central subject matter of normative ethics, but I have trouble imagining many consequentialists going in for this sort of talk. Certainly it sticks in my craw.

Could it be that this language strikes the contemporary ear as forced or high-flown, and as such, offends against the aesthetic tastes of your typical self-styled "no-nonsense" consequentialist? That's probably part of it. But for me it goes deeper than that. I abjure this kind of talk because I think it smuggles in a harmful kind of fetishism. I got into normative ethics because I was bothered that, for example, some people were dying of malnourishment while others had *foie gras* to waste, and that the rich and powerful were effectively bribing politicians through campaign contributions. I wanted to figure out how to deal with these problems in the best way possible, and I wanted arguments that could bring others around to my side.

And yet I noticed that from some quarters there issued a discourse that was seemingly more focused on action and less on the extra-agential world — on 'excellence in the exercise of the capacity for rational human agency'…or whatever; on the structure of action; on the *explanation* of action; on the distinction between action and mere behaviour. But what about the whole 'people dying' thing? Doesn't that seem more important than winning a gold medal in the Action Olympics?

Look, I'm not completely obtuse. I know that when I write 'deal with these problems', I'm talking about acting, and that the kinds of 'arguments' I was searching for concerned the evaluation of action. Nobody who wants to make a dent in the world's problems can claim not to care about action. Action is the only channel for influencing the world that we can control. And yet, I've long had this nagging thought that many philosophers are devoting attention to the category of action

1

that is, somehow, excessive, inappropriate, *fetishistic*. And it has seemed to me that correcting for this would put us in a position to see the deep appeal of consequentialism.

This paper is an attempt to make these nagging thoughts more precise.

Here's the gist: The standards by which we evaluate things are not all on equal footing. The standard of what we ought, overall, to do stands in a certain kind of evaluative relation to the standard of what would be the best football play, or the standard of what would count as the best pedagogy. Philosophers have coined a few different labels for this relation. They say the 'all-things-considered (hereafter: 'ATC') "ought"' standard is more 'robust' (McPherson 2011, p. 233; Cuneo 2014, p. 155; Finlay 2018) or 'authoritative' (Husi 2011; McPherson 2018) than the 'best football play' standard, and that the latter is 'lean' (Cuneo 2014, p. 155) or 'merely formal' (McPherson 2011, p. 226; Finlay 2018). We would think that someone who guided his conduct by the latter standard rather than the former in the event of a conflict would be making a certain sort of mistake, over and above the one he'd be making just by falling short of *some standard or other*. We could rightly accuse him of *fetishizing* the standards of football.

Well, I want to claim that there's an action-guiding standard that stands to the standard of the ATC 'ought' as the latter, in turn, stands to the standard of the 'best football play', or to standards of etiquette, legality, and the like. It is more authoritative than the standard of the ATC 'ought', or indeed any other standard that evaluates action *as such*, *qua action*. The standard I have in mind is the one we apply when we say it's good or bad, fortunate or unfortunate, that something exists or occurs. It's the one we're applying when we say that it would be bad were the Earth to be hit by a gamma ray burst, or good if my cold went away. It is way of evaluating *things*, generally. An action is, like anything, a thing, and so we can evaluate actions by this standard, too. We can say, for instance, that it was unfortunate that Fred and Mary Trump had intercourse, or that it would be good if the action of my calling my wife happened now. It's just that we are not evaluating them specifically *as* actions when we apply this standard.

I am going to say that actions fare well by this most-authoritative standard insofar as they yield good consequences. In this sense, consequentialism is correct. Whether or not consequentialism is right account of what we practically ought to do, it's the right theory regarding a standard that supersedes or overrides the standard of what we practically ought to do. Someone tho disregarded this superseding standard in favour of the standard of the ATC practical 'ought' or that of right action one would be, again, guilty of a kind of *fetishism*. Only here, it's fetishism of the proprietary standards of action, rather than those of football.

I take it that a successful defence of this view will have to show, first, that consequentialism really is the correct theory of what meets this standard; second, that this standard really can be action-guiding; and third, and I think most importantly, that this standard is more authoritative than specifically practical standards like that of the ATC practical 'ought'. I'll turn to the third one now.

2. What Is 'Authoritativeness'?: In Search of An Answer

Let me begin by warning you off of an error. To say that one standard is more authoritative than another is *not* to say that it is more fundamental metaphysically, or conceptually, or in the order of normative-ethical explanation. For example, one might accept Susan Hurley's (1989) view that concepts like 'just' and 'unkind' are prior to those like 'reason', all while thinking that an agent who guides her action by an application of 'just' rather than 'reason' in the event of a conflict has made an error of practical reasoning.

Derek Parfit cashes out the distinction between the a 'ought'-standard and those of football and law by saying that the former is 'reason-implying' while those like the latter two are merely 'rule-involving' (2011, p. 144-5). A standard is reason-implying just in case one necessarily has reason to do whatever meets that standard.

Should we think of authoritativeness fundamentally in terms of being reason-implying? I think we should not. This is partly because of what I want to say later on about which standards are most authoritative — specifically, that a standard that is not obviously reason-implying is nonetheless more authoritative than all of the standards that *are*. But there are other reasons for rejecting a Parfit-inspired account. First, it seems to offer no illuminating story of why the standard of what I have (some/most/sufficient) practical reason to do is itself authoritative or authority-conferring. That standard is reason-implying, of course, but then again the standard of legality is legality-implying. It seems that we've done very little to explain why one is more authoritative than the other. Moreover, this account is too coarse-grained. As I'll argue in a moment, there are many grades of authoritativeness, and it seems possible for one standard to be more authoritative than another despite *neither* reaching the level of being reason-implying.

Generally, the strategy of defining the elusive idea of authoritativeness in terms of other evaluative or normative concepts strikes me as a mistake. For what we seek is an informative account of why standards involving some of these concepts are more authoritative than standards involving others. A definition in terms of one or more concepts in this family will be explanatorily shallow.

So what are the other possibilities?

We might try to characterize authoritativeness not *via* definition — in other words, not by the *inferential role* of its concept — but *via* the concept's *non-inferential role*. As a starting point, we might want to say something like this: The standard of 'ought' or right action is authoritative, while the 'good football play' standard is not, because only the former is characteristically linked with action. If I think something is right, then under normal conditions, I'll be motivated to at least some degree to do it; not so for my thought that something is a good football play.

But I have some worries about this. One it that the 'good football play' standard *does* seem to have a special link with motivation. A quarterback in the huddle of a football game might come to think 'Red 84 would be the best play right now'; that, *sans* any other evaluation, would be enough to make him call out 'Red 84!'. Now, an opponent might think that this appearance is deceiving — that either the quarterback also thinks, at some level, that he *ought* to call Red 84, or else that he has some purely contingent desire to call out Red 84. Either way, there is no constitutive link between 'best football play' and motivation of the sort that obtains between 'ought' and motivation.

But appearances are surely worth *something*, and so I'd want to know what countervailing reason this opponent has for denying the constitutive link between thoughts about the best football play and motivation to act. Additionally, such an opponent will have trouble explaining what makes such a thought a normative or evaluative one at all. If 'best football play' or 'illegal' doesn't have a constitutive link with motivation, or a similar non-inferential role, it's not clear what puts it on the same side of the evaluative/non-evaluative line as 'right' or 'best', rather than on the same side as 'will create a loud rumble'.

The non-inferentialist might fix this problem by adopting a more finely-structured account of authority and motivation. He can say that both 'ought' and 'best football play' have constitutive links with motivation, but they differ in two respects: First, in the event of an apparent conflict between what we ATC ought to do and what would be the best football play, it's the role of the former that will manifest itself; in other words, we'll tend to do what we think we ought to do. Second, thoughts about the best football play will be motivating only in certain contexts — for example, that of a football game in which the agent is a player or coach with the expected aims.

But then what makes a 'best football play' thought evaluative and a 'loud rumble' thought not?[1] Both will motivate sometimes and not others, and both will generally fail to exhibit their motivational potential when up against a conflicting 'ought'-thought. Let me float an answer that I find plausible, recognizing that it needs further development: The belief that doing A will create a loud rumble can play a motivational role, but what it can't do is perform the same kind of 'silencing' function *vis a vis* other motivations that I described 'ought'-judgments as doing. By contrast, a belief that A is the best football play can perform this function, even if its motivational

---

[1] Some may think that we can draw all these distinctions by adverting to desires: Non-evaluative beliefs can move you only if they are matched up with desires, which may be contingently-held, while evaluative beliefs either: (a) don't require desires; (b) generate desires themselves; or else (c) move us by matching up with a necessarily-held desire (e.g. the desire to do the right thing). I'm resistant to going that way. This is partly because it doesn't seem to help at all with the distinction between more-authoritative and less-authoritative evaluative standards. It's also partly because I generally don't think that talking about desires does much philosophical-classificatory work. We can make all the distinctions we ought to care about just by talking about when someone is moved this way or that; the vocabulary of desires functions more as a convenient theoretical overlay. (This suggests that the concern about so-called 'moral fetishism' (Smith 1994) is misguided.)

force is, in turn, silenced by the presence of an 'ought'-judgment. A football player might in some circumstances be moved to do A by her thought that this would likely get the team a first down or put them in the 'red zone', or that this would be most likely to put points on the board, or that this would 'establish the run' and thereby make the 'play action' more effective in this game or in future games. But these thoughts would not generally move the player who thought 'but overall, this is not the best football play.' The picture lurking behind this distinction is that evaluative thought has the function of reconciling, and thereby subsuming, all of the myriad aims and aversions that we might have in some context; it's a clearinghouse' for them. By contrast, non-evaluative thoughts don't have this kind of function.

But there is a another problem for the non-inferentialist strategy as stated. It doesn't seem to allow for *mistakes* in which standards we take to be authoritative. It holds that standard 1 is more authoritative than standard 2 if we would generally be motivated in accordance with the former in the event of a conflict. But this seems to work better as an account of what it is for us to *take* standard 1 to be more authoritative than standard 2. We should allow for the possibility that someone might take standard 1 to be more authoritative than standard 2 when in fact 2 is more authoritative than 1. I'm especially concerned to allow for such mistakes, because one of my main claims in this paper is that they're not uncommon. Many of us treat standards like those of the practical 'ought' and practical reasons as maximally authoritative, but do so mistakenly.

The natural fix is to bring in the notion of rational action or the rational agent. We might say that our applications of the more authoritative standards are linked up in the aforementioned ways motivation *only* in agents who are practically rational. (See Smith 1994, p. 61)

Now, I have no quarrel with this characterization. It seems right. But I'll need to go beyond it for the purposes of this essay. For I want to convince you that the standard of the good or fortunate action is more authoritative than that of right action or the practical 'ought'. I won't be able to do this by relying on the premise that the agent who guides her behaviour by the latter at the expense of the former is *irrational*. No one who does not already accept the conclusions of this paper will assent to this.

So I want a deeper account, which does not *compete* with this rational-motivation-based view but rather *undergirds* it — from which it follows that it really is more rational to go with the apparently good action rather than the apparently right one, and therefore that the goodness standard really is more authoritative than the rightness or 'ought' standard.

3. What Is 'Authoritativeness'?: A Deeper Account
Here's a kind of off-the-cuff take on the difference between the ATC 'ought'-standard, and something like the 'best football play' standard: 'The former evaluates action in terms of a broader category — action — while the latter evaluates it in terms of a narrower one — football play. The "ought"-standard is general, then, while the "football play'" standard is specific.'

'Broad' vs. 'Narrow'; 'General' vs. 'Specific': I contend that this is the right way of thinking about the fundamental difference between more- and less-authoritative evaluative standards. Morality seems more general than justice, and justice more general than *corrective* justice. And the 'more authoritative than' relation seems to track that. But as stated, it's imprecise, and easy to shrug off. Since I'm going to be putting a lot of argumentative weight on some heterodox views about which standards are maximally authoritative, we'll need a more theoretically respectable way of supporting these views. Think of what follows as an explication of this hazy, suggestive take on authoritativeness.

The phrase 'all things considered' is a clue. As a first pass, we can say that a standard of evaluation is maximally general/broad, and as such, authoritative, if it *considers all things*; a standard is less authoritative if it does *not* consider all things. An evaluative standard does not consider all things when it's constitutive of the concept of that standard that some things are irrelevant to its proper application. For example, it seems to be ruled out by the concept 'best football play' that considerations unrelated to winning football games bear which football play is best. We can reasonably argue about whether 'Red 84' is the better play because it's more likely to put points on the board now, or where 'Green 29' is because it will take more time off the clock. But only someone who did not understand the concept would sincerely claim: '"Red 84" is the better football play because it's more beautiful' or '…because my wife bet money that we would run it' or '…because it would contribute to world peace'. Nor does a standard consider all things when it's part of the concept that some things *are* relevant to its proper application — i.e. that some things are, of conceptual necessity, ruled *in*. It's not an 'open question' whether world peace is relevant to which is the best football play; but neither is it an 'open question' whether contribution to winning the game is.

By contrast, a standard considers all things if nothing is ruled in or ruled out as relevant by the concepts involved. *All* things are considered, but they're *merely* considered.

But now recall that I want to argue that some other standard is more authoritative than even that of the ATC 'ought'. It may seem difficult to reach this conclusion, though, if authority is cashed out in terms of generality, which is in turn cashed out in terms of how many things are (merely) considered. For we're talking about the *all-things-considered* 'ought'. It would appear to consider *all things*.

But here I think appearances are deceiving. For the standard of the ATC practical 'ought' is a specifically or proprietarily practical standard. So are the standard of right action, the 'practical reasons' standard, and perhaps some others. They are not merely ways of evaluating action; they are ways of evaluating it *qua* action, or to borrow a phrase from James, as 'good in the way of' action or not. (See Korsgaard, Schapiro, and Katsafanas *infra*; also Wald 2017, p. 133.) They assess

6

actions as fulfilling or failing to fulfill the function that is associated with the kind: *action*, and with no other kind.

The meaning of this '*qua*' and 'in the way of' are clearer, perhaps, in the case of belief. On the one hand, we may evaluate belief *qua* belief, as good or bad in the way of belief, in terms of the function associated with the kind: belief. There's disagreement about how to do this, of course, but the consensus among epistemologists is that goodness *qua* belief is exclusively a matter of things like truth, knowledge, evidence, and reliability (and whatever else grounds or conduces to these). These are pertinent to whether a belief is fulfilling the function of its particular kind. The fact that a holding a belief will make me happier, by contrast, is irrelevant to how it fares by the proprietary standards of belief.

But that's most certainly *not* because happiness doesn't matter, or because it in no way tells in favour of something that it yields happiness. For we may also evaluate a belief as good or bad *simpliciter* in the same way we might evaluate a gamma ray burst or a cold's going away. That holding belief makes me happy is obviously relevant to whether it is good in this general way.

That the proprietary standards of belief evaluate it *qua* belief explains why certain considerations are ruled in as relevant and others ruled out as irrelevant. Similarly, I want to claim, the fact that the proprietary standards of action evaluate it *qua* action, rather than simply as good or not *simpliciter*, goes towards explaining why some considerations are ruled in or ruled out. **Considerations are ruled in if they are germane to the quality of an action *as an action* — to its fulfillment (or not) of the function of the kind: action; they are ruled out if they are not, even if they pertain to its quality more generally.**

Now, that in and of itself wont tell us much about right action. But that in combination with other theses just might. For suppose we also accepted, with Korsgaard (2009), the claim that action is self-constitution, and so action that succeeds *qua* action is just action that succeeds *qua* self-constitution. If we can then show, as Korsgaard believes she can, that only actions that accord with the categorical imperative go toward constituting oneself properly, then we'll be led to the conclusion that actions are right if and only if they accord with the categorical imperative. Or suppose we accept Paul Katsafanas's 'Nietzschean constitutivist' claim that action aims at 'seeking and overcoming resistance', and that only those actions that constitute 'a life of continuous attainment' will achieve this (2013). Then only such actions will succeed *qua* action — such that they are the right actions, or the ones we practically ought to do.

There are other ways of using the specifically practical nature of concepts like 'right' to explain their application that does not so obviously rely on theories about what constitutes action. Indeed, I think that the deontologist *implicitly* employs them all the time. And it's not hard to imagine a situation in which she would make them explicit. Let's say some consequentialist argues as follows: '100 people are about to die as the result of a drought. A flood miraculously occurs which

ends up saving all of those people, but killing 50 people. Would it be good, or fortunate, overall, for that flood to happen? Of course it would! So if an action kills 50 and saves 100, then that action is the right action. What's wrong with that?'

The deontologist would no doubt say — and reasonably so — that evaluations in terms of goodness or fortunateness of something's occurring are unlike evaluations in terms of rightness, or the practical 'ought, or anything like that: 'We're not asking whether it would be *good* or *fortunate* were an action to occur that killed 50 and saved 100. We're asking whether that's the right thing to *do*, whether that's what she practically *ought* to do it.'

'"Good", "right"…What's the difference? They're all just ways of praising action,' the consequentialist might reply.

'The difference,' the deontologist will say, 'is that when we talk about an action's being right, we're evaluating it *as* an action, not just as just one more kind of event, alongside floods and gamma ray bursts. We're not asking whether it's a good thing that's also an action; we're asking whether it's a good action.' The more thoughtful deontologist will have some further story about why this difference matters; the point here is that she thinks it *does* matter, for some reason or other. She is willing to say one thing about whether it's good that the action occurs, and another about whether the action is right, and this is explained by the fact that the 'rightness' standard evaluates action *qua* action, while the goodness standard does not.

Now, while the senses of 'right' and 'best football play' are doing some work in grounding (im)propriety of certain applications of each, respectively, it seems that further claims are 'doing more of the work' in determining the proper application of 'right'. After all, it's constitutive of the concept 'best football play' that only what's conducive to winning football games is relevant to its application. By contrast, almost no one would say that it's part of the sense of 'right' that, e.g., the 'doing/allowing' distinction is relevant to its proper application. To get the moral significance of doing vs. allowing, we'd need some further claims that are not 'contained' in this concept. But as we saw above, people do call upon the concept to do *some* work in delimiting its application: A consideration bears on which action is right if and only if it bears on its quality *qua* action; then we need some ethical theory, plus maybe some metaphysical or other philosophical theory, to fill in an account of goodness *qua* action.

In light of all of this, here's what I'd want to say about authority generally: One standard is more authoritative than another just in case the sense of the concept of the former delimits the concept's application *more* than does the sense of the concept of the latter. This is my way of making precise the idea that more authoritative standards are somehow broader, more general — that they consider, but *merely* consider, more things. As we saw above, it's intuitive that the concept of the best football play delimits its own application more than does that of rightness or the ATC 'ought'.

But as I suggested earlier, the concept of goodness does not delimit its own application at all, since it does not evaluate actions even *qua* actions.

You might ask: 'But doesn't it evaluate actions, and other things, *qua* things? How does that delimit the application of "good" any more than the "*qua* action" bit delimits the application of "right"?' But it's misleading to say '*qua* things'. The goodness standard does not evaluate actions not *qua* anything at all. For it is very implausible that anyone, even the most reflective philosopher, is asking what makes something good *as a thing*, makes it exhibit excellence in the way of thinghood or thingness, when they ask what would be good to exist or occur.[2]

What's to be said, then, in favour of my view about authoritativeness?

First, it seems to give us the right answers in the obvious cases. It implies that the 'rightness' standard is more authoritative than the 'best football play' or 'best pedagogy' standard. It implies that the standard of morality generally is more authoritative than the standard of justice specifically. It implies that the standard of ATC practical reasons enjoys a good deal of authority relative to many other standards — and unlike a story on which being authoritative just *is* being reason-implying, it *explains why* that is.

Second, it accounts for the conclusion of section 2, for which we were seeking some explanation — that in the event of an apparent conflict between more authoritative evaluations and less authoritative ones, rationality requires that we guide our conduct by the former. The story goes like this: The sense of the concept of a less authoritative standard rules in or rules out certain considerations. And so it motivates only it's tokened in a context in which I have taken for granted certain ends or aversions — namely, only when I have ends or aversions which are linked up in the right way with what the concept's sense has ruled in as relevant, and has not ruled out as irrelevant. For example, since the concept 'the best football play' rules in anything related to winning football games as relevant and rules out anything else, it gets a rational-motivational purchase on me in contexts in which I aim to win a football game, and not elsewhere. But now think of what happens when we so much consider a more authoritative standard — e.g. when we wonder: 'Yes, A is the best football play, but what *should* I do?' This seems to have the effect of wiping out 'best football plays''s motivational niche. I stop taking for granted those ends and aversions that are relevantly related to what's constitutively ruled *in* by that concept, and may adopt ends and aversions that *had* been ruled *out*. Now motivation by this standard looks irrational because, to put it simply, I don't care about anything that the standard had designated as important, and now care about many things it had designated as unimportant. The rational thing to

---

[2] Some artists and art critics do speak of *thing*ness, but it's clear they have a different notion in mind than I do. Johanna Drucker writes:'[T]the concept of 'thingness' links sculpture to objects in and of the world in a combination of [art] and mass culture production. Thingness embraces the specificity and allusional properties of material…' (1997, p. 157).

do is to act on the more authoritative standard, the few rulings-in of which correspond to my *new* ends and aversions, and the few rulings-out of which do not.

Third, my theory of authoritativeness explains why following less-authoritative standards in these conflict cases is not just irrational, but is mistaken in this deeper, 'fetishising' way, too. The thought is that the way we get around the world should be dictated, to the maximum possible extent, *by the world*. Maybe it's my pleasures and pains and those of others that, ideally, would direct my actions; maybe it's my will and the wills of others; maybe my attachments play a role. But the purest way of responding to the world is to favour some of these and disfavour others, just on account of the features themselves. Less authoritative standards put something in the way — the extra constitutive content of the concepts of those standards. When I do an action because it is optimal *qua* football play or pedagogy, I am moved to do it not only by the action, its consequences, my preferences, others' wills, and so forth, but by the fact that the concept I'm employing in evaluating this action has a certain constitutive structure — one that rules in some things as relevant to its applications, and rules out other things as irrelevant. That's not to deny that it's often *useful* or *enlivening* to guide one's behaviour by less-authoritative concepts. (Part of the story here would be akin to the one Michael Bratman tells about why it's worthwhile to make and follow plans.) But there is some kind of shortcoming we're exhibiting when we do that; we're falling short of an ideal of guiding ourselves by unblinkered, *pure* evaluation.

4. Why Consequentialism is the Correct Theory of Good or Fortunate Action

Reasonable people disagree about whether consequentialism is the correct theory of right action. It's not obvious. But I do think it's obvious that consequentialism of a certain sort is the correct account of when it's good for a thing to exist, for an event to occur, and for a fact to obtain. It follows that consequentialism of this sort is the correct theory of which actions are good, of when it's fortunate that some action occurs, since an action is — whatever *else* it is — a kind of thing or event. According to this sort of consequentialism, what makes a thing good are the past, present, and future consequences of its existing, occurring, or obtaining. These include the consequences that inhere in its very existing, occurring, or obtaining.

This sort of consequentialism holds that, all else being equal and in conjunction with other plausible claims, it would be fortunate for a weather event to occur that kills a few people, but saves more from dying from a drought-induced famine. Correspondingly, it would seem to hold that it would be fortunate for an *action* to occur that had the same basic causal profile.

But perhaps that's too hasty. 'If all you can show is that consequentialism is the correct account of the "good action" standard,' some might respond, 'then you haven't gotten very far. There are good reasons to think that any plausible moral theory can be represented as a form of consequentialism (Portmore 2007). To get the results you seem to want, you'd need to show that the correct form of consequentialism about goodness/fortunateness is the agent- and time-neutral version that the untutored think of when they think "consequentialism".'

This is a pretty big-picture paper, so I don't pretend to have a decisive reply. I will say this, though: However plausible deontology is as a theory of the *right*, and however interesting it is to try to see if we can 'consequentialize' deontology, the theory of the good that forms part of consequentialized deontology is just not a plausible theory of the good. Whether or not you favour a 'fitting-attitude' *analysis* of 'good', we can take the fittingness of attitudes as, at the very least, *evidence* of what's good. And to that end, it just seems strange, *unfitting*, for me to *prefer*, *ceteris paribus*, a world in which six people are killed to a world in which five are killed. (See also Schroeder 2007)

Less drastic than full-fledged 'consequentialized deontology' is something like what Robert Nozick (1974, p. 30) termed a 'utilitarianism of rights'. An adherent of this view might argue: 'You seem to be counting only the causal profile of an action in its "consequences". But surely one of the consequences of my doing an action is that that very action is done. But the state of affairs in which some people die *because of a killing* is worse than the state in which the same number of people die due to something non-agentive — a gamma ray burst, a spasm, and so on.'

Well, I'm not so attracted to the idea that, *ceteris paribus*, a killing adds any disvalue over and above that produced by the death it brings about. But even if it does, even if a 'utilitarianism of rights' is correct and value of 'the act itself' figures into the value of its own consequences, this is a far cry from deontology as most people understand it. Of course this was Nozick's point in introducing this theory. Most obviously, this view will permit, *ceteris paribus*, killing N people as a means of preventing the killing any number greater than N. And so my argument in this paper still cuts serious ice in normative ethical debates.

On top of its other merits, my version of consequentialism helps us to revitalize a kind of criticism of non-consequentialism that seems very compelling at first but typically withers upon examination. Some consequentialists like to say that deontologists fetishize the purity of their own agency. They're willing to accept the worse outcome so long as it means that they themselves can avoid acting wrongly (See Williams 1981, p. 50). The reasonable response from the deontologists is to say, 'Look, we're all talking about right action here; so insofar as I'm guilty of fetishizing my own agential purity or some such, so are you. You also care about acting well, about excellent rational agency. We just disagree about what that consists in.' But someone who guides her behaviour by the kind of consequentialism I'm proposing here is *not* exhibiting such a concern. She can regard the deontologist — and for that matter, the textbook consequentialist — as fetishizing agency and 'acting well' insofar as he insists on guiding his behaviour by less-authoritative, specifically practical standards. This criticism of the deontologist would be like the criticisms that most of us would level at someone who couldn't help but make decisions from the narrow standpoint of what would make him a good soldier, a good teacher, a good party host.

5. Some Precedent for My View

I am not the first person to have observed that some ways of looking at action are more amenable to consequentialism and some to deontology. Korsgaard and Talbot Brewer, both opponents of consequentialism, have sought to locate its appeal in a certain conception of action — a conception that they then attack in the expectation that this will undercut consequentialism's appeal. Korsgaard writes:

> 'According to Mill, action is essentially production, and accordingly its function is to bring something about. Whether an action is good depends on whether what it brings about is good, or as good as it can be. The influence of this conception of action on contemporary Anglo–American moral philosophy has been profound. Nowadays even moral philosophers who are not utilitarians appear to be comfortable only if they can explain moral value in terms of the production of various goods and harms. Deontological considerations are sometimes characterized as "side–constraints," as if they were essentially restrictions on ways to realize ends. As such, they have been found mysterious by many philosophers. If the whole point of action is to produce the good, how then can it be good to restrict that production?' (2009, p. 8)

And here's Brewer, discussing the 'dogmas of desire' that he regards as part of a general "world-making" conception of agency: 'On close inspection…it turns out that the dogmas of desire are integral elements of a substantive vision of ethics — a vision at once worldly, progressive, and anti-contemplative.' (2009, p. 8) Later, he suggests that the 'production-based' view of action seems to lend support to 'the prevailing idea that the proper practical response to the good is always to promote it, and where possible to bring about more rather than less of it'" (2009, p. 235)

Obviously, I think Korsgaard and Brewer are onto something, and one can't help but applaud any attempt to get at the deep roots of a view one opposes, rather than play about on the surface. But it's important to see that these critics are saddling the consequentialist with a particularly *metaphysical* spin on the broad idea that we think of action in terms of production, or bringing things about. On their telling, the consequentialist believes that action *is* production, just as one might believe that the statue *is* the clay or pain *is* C-fibres firing. The consequentialist is presented as engaging in an action-theoretic project of the sort I mentioned in the introduction.

But I, for one, don't think action is production. I don't have any view about what action is, other than to say that it is a *thing* — as everything is — and that it can be product*ive*; it can bring things about. Mine is an value-theoretic precisificatoin of the same broad outlook. Whether or not action is production, it's mostly what an action produces that makes it good or not if it occurs, because it's mostly what things in general produce that makes it good or not if they occur. And evaluation in terms of this general goodness or value — a kind of evaluation that we can make of any thing whatsoever — is more authoritative than the kind of evaluation that applies to actions specifically, *qua* actions.

I discuss Korsgaard and Brewer here partly to emphasize this difference between the metaphysical and value-theoretic spins on the picture of action as crucially involving production. But I also raise it to quell the suspicion that the 'nagging thoughts' I expressed in the introduction were just me imagining things. Two of the most searching philosophers on 'the other side' have also noticed that consequentialists and deontologists, far from simply holding different views about the very same thing, seem to have different conceptions of the very subject matter of ethics. And both of them seem to agree with me that a satisfactory resolution of the debate about consequentialism will rest largely on a determination of which of these conceptions is correct.

6. Objections and Responses

Right Answers in Ethics?
It be be objected: 'You say that a standard is more authoritative the less the concept of it delimits its application. But then it becomes unclear how there can be correct or incorrect ways of applying these more-authoritative standards. Consider that Korsgaard and Katsafanas can claim to derive their normative conclusions only because they're applying a standard that judges actions *qua* actions. For they can claim without too much implausibility that action has a constitutive aim, or a *telos*. You can't say the same thing about things *simpliciter*. Nothing's constitutive of being a thing — other than existing, I suppose. But even if we put aside these constitutivists, there just seems to be a deficit of objectivity when you're starting with such thin evaluative concepts. Sense, after all, determines reference. Concepts like "best football play" have lots of sense, so it's no trouble to see how they refer. And the clever theorist can eek some sense out of "right" or the practical "ought". But "good" *simpliciter* seems to have virtually *none*.'

This is worth taking seriously. It presents us with a dilemma of a sort that we see over and over in ethics: If a mode of evaluation is too thick, then it is merely parochial, unimportant, readily ignorable. But if a mode of evaluation is too thin, then it is unclear how it can have conditions of correctness or incorrectness. Many philosophers are quick to declare of Korsgaard, Katsafanas, Velleman, and other constitutivists that they're packing much too much into action or the concept thereof — that action as such is really just an exercise of will, and that's it; nothing more to say. But the constitutivists are putting their robust conceptions of action to a certain use — they're trying to provide what we might loosely term a *foundation* for ethics. Their critics had better either offer a foundation of their own, accept the skeptical consequences of not having one, or show how we we can have right answers in the absence of one.

Without belabouring the point, let me simply say that I opt for the third strategy. In other work, I offer a quietist meta-ethical view on which ethics can be objective even in the absence of 'foundations' derived from metaphysics, semantics (to which would belong claims about a concept's sense or constitutive structure), or the theory of practical rationality. The basic idea is that the debates we care about in normative ethics are much like what get called 'merely verbal' or 'non-substantive' disputes, like James's case of the two hikers arguing about whether the man is

going ''round the squirrel' (1907). They afford none of the values that inhere in correct prediction, in action that tends to bring about the general satisfaction of our desires, or in accurate representation of the world in any robust sense of the phrase. But the results of metaphysics, semantics, or the theory of rationality could bear on these debates, I claim, only by bearing on which beliefs have or promote these values. And so these fields are irrelevant to the question of which, if any, ethical views to adopt or which to jettison. The only grounds for going one way or other in ethics are specifically ethical grounds, which one can apprehend only from within an ethical-evaluative standpoint. (Sepielli *ms*)

Is The "Most Authoritative" Standard Practical?
Perhaps you've been struck by this thought: 'Your suggestion is that we guide our actions by this more authoritative "goodness" standard, rather than by a less authoritative "rightness" or "practical reasons" standard. But perhaps you are asking for the impossible. After all, how can a standard that evaluates everything from sneezes to spasms, earthquakes to gamma ray bursts, the existence of God to the non-existence of the perfect mousetrap, serve as a direct basis for action, given that the evaluation of these other things can't lead to action?'

I have two responses. First, I reject the idea lurking in the background that we can only be moved to action by evaluations *of action*. Suppose a waiter's question 'Soup or salad?' prompts you to evaluate the two items, and conclude that the soup is better. It seems that this conclusion is sufficient to rationalize your speech-act of replying 'Soup, please!'. I see no good reason to think that such an evaluation can render action intelligible only *via* an intermediate evaluation of the speech acts: *saying 'soup'* and *saying 'salad'*. This seems to me to exemplify a more general phenomenon — that we can be moved to actions by our choices even when the *objects* of those choices are not themselves actions.

Why might someone deny this? Perhaps it's because an exercise of the will can't by itself bring about the appearance of either the soup or the salad. I need some help from the outside world, including from other people. Insofar as we acquire the soup or salad, we can only get them *by* engaging in something that is entirely controllable by the will — here, *asking* for one of the items. But the idea that we can only be moved to action by choosing among things that the will can control all by itself, with no help from the rest of the world, has implausible consequences. It implies that even *action*-evaluations can move us to act only if they are evaluations of so-called 'basic' actions — actions that you don't do *by* doing anything else, or that you can be practically certain you'll manage to pull off (Danto 1965). But most things we'd think of as ordinary actions are not basic; they can't happen without help from the unruly world to supplement the input from our wills. However, it does seem that by deciding on some non-basic action, I can impel myself to perform more proximate actions. I can conclude that I ought to open the front door, and start acting based on that conclusion; I'm not limited to choosing among actions like *rotate my hand slightly to the right while squeezing the doorknob*, and praying that this gets me what I want.

With that said, I grant that evaluations of outcomes and objects can't *always* impel me to act. Notably, they can't do it in cases where other considerations are salient. For example, an ordinary person contemplating what he practically ought to do in the 'fat man' version of the trolley problem will not be impelled to rational action simply by the thought that five dead is worse than one dead. It does seem that, in some cases, the objects of our evaluation must be actions themselves, not just the products of action.

But does that spell doom for the action-guidingness of the 'goodness' standard? Hardly. We need to distinguish sharply between the matter of *what's evaluated* in some situation, and that of *which standard* is being used to evaluate those things in that situation. Suppose it's right that I can reach an action-guiding conclusion in some situation only by evaluating all and only the actions that are possible for me in that situation. It does not follow from this that the *kind* of evaluation to which I subject these actions must be an evaluation of them *qua* actions.

We can put this point another way. Elijah Millgram has coined a clever name, 'Intendo', for the practice of agency generally: 'Intendo is the game you are playing whenever you do anything at all; "agent" is thus the generic role in the generic game (the analog of "player" in chess or baseball).' (Millgram 2005). Insofar as the exercise of agency is, in a certain sense, inescapable, this game of Intendo is inescapable; it's one we all must play, all the time. It doesn't follow from this alone that we have to assess moves in Intendo by the standards of Intendo, any more than we must assess moves in *Nin*tendo by the standards of Nintendo. (Surely I'm not the only one who often placed more value on exploring the strange world of *Super Mario Brothers 2* than on winning the game.)

I think the failure to appreciate this distinction between what's evaluated and how it's evaluated is one reason why the *uber*-standard of goodness or fortunateness, which exhibits such supersessory power in my own mind, seems not to do so in the minds of most other people. Ordinarily, we're accustomed to 'reading off' the appropriate standard of evaluation in a context from our conception of the things we're evaluating in that context. A coach in a football game is evaluating prospective football plays, and is thinking of them *as* football plays. It is very natural, then, to evaluate them *qua* football plays — that is, by criteria the relevance of which is explained by kind-concept that figures in the mode of evaluation. Similarly, then, I think it is very natural when the things one is called upon to evaluate are actions, and only actions, to read off the standard of evaluation from this — and so to evaluate them by standards that are proprietary to action. So in a way, I'm asking you to do something that always feels a bit funny — i.e. to evaluate X's, and only X's, by standards that apply to a much broader range of things.

6. Concluding Remarks

I have tried to present at least a presumptive case for a certain form of consequentialism. If the major moves are right — i.e. that this standard being both maximally authoritative and practical —

then I think the kind of consequentialism I'm pushing is on firmer ground than is consequentialism as a theory about what we practically ought to do.

Is it a *new* form of consequentialism? I'm not so sure. Certainly most present-day consequentialists do frame their theory as a response to the question of 'What ought I to do?' But perhaps they're using the language of 'ought' to talk about whichever standard for evaluating action is thinnest and most authoritative — which, if I'm right, is the standard of goodness/fortunateness. It would be revealing to learn how other consequentialists respond to explications of the practical 'ought' concept in which its status as an evaluation of action *qua* action is made specific. My own response to talk of 'acting well' and 'excellent rational agency' has always been to think, 'Who cares? I'm a lot of things — a thinker, a feeler, *Dasein*, a nexus of causal forces; what is the special significance of being a good, or "excellent", *agent*?' This suggests that I, at least, hadn't been taking action-specific evaluation to be maximally authoritative in the first place.

REFERENCES

Brewer, Talbot (2009) *The Retrieval of Ethics* (Oxford: Oxford University Press).

Cuneo, Terence (2014) *Speech and Morality: On the Metaethical Implications of Speaking* (Oxford: Oxford University Press.

Danto, Arthur (1965) 'Basic Actions', *American Philosophical Quarterly* 2:2, pp. 141-8.

Drucker, Johanna (1997) 'Thingness and Objecthood', in *Sweet Dreams: Contemporary Art and Complicity* (2005) (Chicago: University of Chicago Press).

Finlay, Stephen (2018) 'Defining Normativity', forthcoming in *Dimensions of Normativity: New Essays in Metaethics and Jurisprudence*, David Plunkett, Scott Shapiro, and Kevin Toh (eds.) (Oxford: Oxford University Press.

Hanser, Matthew (2008) 'Actions, Acting, and Acting Well', in *Oxford Studies in Metaethics*, Volume 3, Russ Shafer-Landau (ed.) (Oxford: Oxford University Press), pp. 171-198.

Hurley, Susan (1989) *Natural Reasons* (Oxford: Oxford University Press).

Husi, Stan (2011) *Building Reasons without Authority* (Rice University, dissertation).

James, William (1907) *Pragmatism* (1995) (Mineola, NY: Dover Books).

Katsafanas, Paul (2013) *Agency and the Foundations of Ethics* (Oxford: Oxford University Press).

Korsgaard, Christine (1996) *Creating the Kingdom of Ends* (Cambridge: Cambridge University Press).

Korsgaard, Christine (2009) *Self-Constitution: Agency, Identity, and Integrity* (Oxford: Oxford University Press).

Lear, Gabriel Richardson (2004) *Happy Lives and the Highest Good: An Essay on Aristotle's Nicomachean Ethics* (Princeton, NJ: Princeton University Press).

McPherson, Tristram (2011) 'Against Quietist Normative Realism', *Philosophical Studies* 154:2, pp. 223-40.

McPherson, Tristram (2018) 'Authoritatively Normative Concepts', in *Oxford Studies in Normative Ethics*, Volum 13, Russ Shafer-Landau (ed.) (Oxford: Oxford University Press).

Millgram, Elijah (2005) 'Practical Reason and the Structure of Action', in *Stanford Encyclopedia of Philosophy*.

Müller, Anselm (2004) 'Acting Well', *Royal Institute of Philosophy Supplements* 54, pp. 15-46.

Nozick, Robert (1974) *Anarchy, State, and Utopia* (New York: Basic Books).

Parfit, Derek (2011) *On What Matters, Vol. I* (Oxford: Oxford University Press)

Portmore, Douglas (2007) 'Consequentialising Moral Theories', *Pacific Philosophical Quarterly* 88, pp. 39-73.

Schapiro, Tamar (2001) 'Three Conceptions of Action in Moral Theory', *Nous* 35:1, pp. 93-117.

Schroeder, Mark (2007) 'Teleology, Agent-Relative Value, and "Good"', *Ethics* 117:2, pp. 265-300.

Sepielli, Andrew (ms) *Significance without Substance: A Defence of Pragmatist Quietism in Metaethics*

Smith, Michael (1994) *The Moral Problem* (Hoboken, NJ: Wiley-Blackwell).

Strom, Gregory (2011) *Multitasking, Consequentialism, and Practical Imagination* (University of Pittsburgh, dissertation).

Vogler, Candace (2002) *Reasonably Vicious* (Cambridge, MA: Harvard University Press).

Wald, Benjamin (2017) *Judging the Guise of the Good by its Fruit* (University of Toronto, dissertation).

Williams, Bernard (1981) 'Utilitarianism and Moral Self-Indulgence', in *Moral Luck* (Cambridge: Cambridge University Press).