Consequentialism and the

Agent's Point of View

Abstract: I propose and defend a novel view called '*de se* consequentialism', which is noteworthy for two reasons. First, it demonstrates — *contra* Doug Portmore, Mark Schroeder, Campbell Brown, and Michael Smith, among others — that agent-neutral consequentialism is consistent with agent-centered constraints. Second, it clarifies the nature of agent-centered constraints, thereby meriting attention from even dedicated non-consequentialists. Scrutiny reveals that moral theories in general, whether consequentialist or not, incorporate constraints by assessing states in a first-personal guise. Consequently, *de se* consequentialism enacts constraints through the very same feature that non-consequentialist theories do.

1. Introduction

Budding philosophers learn early that act consequentialism (hereafter, just *consequentialism*) has some surprising implications. For example, its defining claim that, roughly, you may always do whatever has the best outcome appears to permit killing an innocent person to save others. However, 'New School' consequentialists argue that surprising implications like these follow only from optional interpretations of the basic consequentialist framework. According to them, consequentialism can mimic, for example, a so-called 'agent-centered' *constraint* that forbids a killing no matter its outcome, forbidding killing someone even to prevent another from killing that very same person and several more. If New Schoolers are right, very little about consequentialism's reliance on *good* shapes theorizing since it is consistent with agent-centered constraints on acts. If that's so, then we might understand all of morality through an evaluative lens.[1]

But the New School view has a catch. Only if we understand value in a particular way, it seems, can we reconcile consequentialism and constraints. In particular, New Schoolers assume that, "in order to [mimic constraints], [they] must deny that agents ought always to produce the most good and accept instead that each agent ought always to produce the most of what is good-relative-to-her".[2] This assumption is costly. Critics allege that consequentialism's compelling claim that you may always produce the *best* outcome cannot mean anything like 'best-outcome-relative-to-you' since the latter is agent-relative but the former is not — it is 'agent-neutral'. Critics insist that New School consequentialism is, therefore, not recognizably consequentialist.

As I argue below, the New Schoolers' assumption is false. Agent-neutral consequentialism and constraints are consistent. Standard-issue *subjective* consequentialism, which holds, roughly, that an agent must do what will have the best *expected* outcome, is consistent with constraints. Expected outcomes reconcile consequentialism and constraints because they are individuated by the agent's expectations or, more broadly, by her 'point of view'. They are thus already agent-relative, in a sense explained below. As a result, mimicking constraints does not require both agent-relative conceptions of the good and expected outcomes. Expected outcomes alone suffice.

This paper has two main aims. The first is to show that agent-neutral consequentialism is compatible with constraints. It turns out that the best way to achieve this aim is to hold that two outcomes that represent the same event can differ in value. In particular, consequentialists who wish to mimic a constraint on killing should hold that, roughly, presenting a killing in the *de se* or first-personal mode makes it worse.[3] The second aim is to show that this position is coherent.

Pursuing these aims reveals a view that merits attention from even dedicated non-consequentialists for two reasons. First, it illuminates how constraints are enacted in a moral theory more generally. Second it shows how some normative concepts, like *good*, are more plastic than is often assumed. As a result, the common method of dividing ethical theories by their privileged normative concept — that is, by whether they understand rightness through rules, reasons,

rationality, value, etc. — does not carve at the deepest joint in ethics. Rather, a more fundamental fault line lies between theories that take the 'point of view of the universe' and those that take the agent's point of view.

2. The Standard Recipe

Consequentialists characteristically claim that, roughly, you may always do what produces the best outcome. No action is forbidden independently of its consequences. This claim conflicts with the idea that some actions, like killing, are intrinsically forbidden. I'll call a rule that forbids you from acting some way regardless of the consequences a *constraint*.[4]

To illustrate, suppose that a constraint on killing exists and that, as a result, you are never permitted to kill, even in cases like the following:[5]

> *Mafia*: The Mafia are certain to kill five innocent strangers unless you kill a sixth, which is
> certain to stop the Mafia from killing the five.

The constraint implies that you may not kill even as the only means to prevent more killing. Consequentialism, by contrast, appears to permit you to kill the sixth if doing so produces a better outcome. Many find this permission heinous. As a result, consequentialists have reason to look for interpretations of their view that forbid agents from killing in cases like *Mafia*. New Schoolers allege that a 'standard recipe' offers just this. It has two ingredients: a 'broadened' conception of outcomes and an 'agent-relative' notion of value.

Consequentialists often hold that an act's outcome is the bounded sum of its downstream effects.[6] New Schoolers, in contrast, conceive of outcomes more broadly. 'Broad outcomes' include not only the act's downstream effects but also the act itself. For example, when I kill, the broad outcome involves not only the fact that someone has died but also the fact that I've killed them.[7]

Broadened outcomes do not, by themselves, reconcile consequentialism and constraints. If consequentialists wish to forbid killing innocents, they must assume that it always produces the worst outcome.[8] However, Schroeder (2006) describes a case where this assume produces a contradiction.[9] Suppose that each of Hans, Franz, and Jens must choose whether to kill in order to prevent the other two from killing. I'll represent the outcome where Hans kills as 'H', and similarly for F and J. If killing is forbidden, then for any X and Y $\in$ {H, J, F} such that X≠Y, X>X&Y, where '>' expresses agent-neutral 'is better than' and 'X&Y' names the outcome where X and Y kill.[10] To mimic a constraint on killing, consequentialists must claim that, from Hans's perspective, J&F>H. That's because, according to them, Hans ought not kill only if the outcome of his killing is worse than the outcome of both Jens and Franz killing. But the parallel fact from Jens's perspective, H&F>J, proves a contradiction.[11]

Therefore, it seems that unless consequentialists also employ 'agent-relative' value, where an outcome has a value only relative to an agent, their view is inconsistent under plausible assumptions. I'll represent the asymmetric, transitive, irreflexive agent-relative better-than relation using '$>_X$'. For example, rather than claiming that Hans's perspective implies that J&F>H and that Jens's perspective implies H&F>J, New Schoolers instead claim that the agents' differing perspectives imply different but consistent rankings: J&F$>_H$H and H&F$>_J$J.

A central challenge for New Schoolers is interpreting the concept expressed by '$>_X$'. It can be taken to mean *better-for-X*, such that 'J&F$>_H$H' means that outcomes where Jens and Franz each kill someone are better *for Hans* than ones where Hans kills someone. The notion of something being better for someone than an alternative is simple and familiar. Eating two bowls of ice cream is better for me than eating three. However, a moral theory that requires each to promote what's good for them seems like a kind of teleological egoism, not consequentialism.[12] As Schroeder claims, "ethical egoists do *not* believe that it is always permissible for you to bring about the most good. They believe that it is always permissible for you to bring about the most of what is good *for you*. But since *good*

and *good for* express different concepts, the idea that egoists find compelling turns out not to be at all the same idea that consequentialists find [c]ompelling."[13] Using *better-for-X* appears to forbid New Schoolers from drawing on intuitive support for consequentialism.

New Schoolers can instead claim that '>ₓ' expresses a different concept, *better-relative-to-X*.[14] However, critics argue that agents lack a pre-theoretic grasp of *better-relative-to-X*, so that talk of *better-relative-to-X* makes as little sense as talking about *orange-relative-to-X*.[15] Or it can be interpreted in terms of what one should prefer, such that A >ₓ B just when and because the reasons for X to desire that A are stronger than the reasons for X to desire that B.[16] But it's unclear whether I always have most reason to prefer what's best.[17]

Perhaps independent reasons favour embracing one of these broadly agent-relative conceptions of value. I don't know. My thesis is simply that consequentialists don't need one to mimic constraints. So I'll bracket discussion of agent-relative value for the remainder to focus on a novel, alternative approach.

3. *De Se* Consequentialism

3.1 Presentational Differences Can Be Moral Differences

Objective consequentialists characteristically evaluate, roughly, an action's future effects.[18] Whether you should press the button depends only on what happens once you do. This view has counterintuitive implications. Even if all of your evidence indicates that pressing the button will result in an otherwise avoidable fiery catastrophe and even if there is simply *no* reason to suspect that your pressing the button has *any* moral upside, you should still press the button if doing so provides the best possible outcome. This implication results from evaluating only the actual effects of an agent's action, which obscures the potentially morally significant effects of how the action appears from the agent's point of view, such as whether it is risky or reckless.

To take account of risk, other forms of consequentialism attend to an act's intended or foreseen consequences, rather than to its actual consequences; these forms of consequentialism take the agent's point of view. Very roughly, what matters for some *subjective* consequentialists are an action's *expected* effects not its actual effects. Crucially, this shifts from thinking of outcomes extensionally, as actual future states, to thinking of them from the agent's point of view, as objects of attitudes like expectation, belief, credence, anticipation, etc. As others note, particularly Ana Mahtani, this grants subjective consequentialism unexpected nuance.[19]

To illustrate:

> *Lois Lane.* Lois Lane has just learned the bad news: aliens are about to destroy the world. Worse still, her cell phone has enough battery to make only one call. She can either call her friend Clark Kent to come save the day or she can call her friend Superman. She knows Clark Kent as the nerdy guy in the office. She knows Superman as someone who's faster than a speeding bullet, more powerful than a locomotive, and able to leap tall buildings in a single bound. Unbeknownst to Lois, Clark Kent is numerically identical to Superman.

By adding some numerical values to this scenario representing Lois's evaluations and expectations, the table below describes the features of Lois's choice relevant to some subjective consequentialists:[20]

[Table 1]

From Lois's point of view, while each friend is equally likely to answer the call, having Superman arrive to save the day is better than having Clark Kent arrive. What should Lois do? Since no matter what number Lois dials, the same person will show up, *objective* consequentialism implies that Lois can dial either number. But that's counterintuitive. Given that Lois thinks of Clark as a mere office worker, calling Clark would be a mistake when she could call someone who's more powerful than a

locomotive. *Subjective* consequentialism ratifies this intuition. Since by Lois's lights Clark is a desk jockey and Superman is faster than a speeding bullet, it's better to call Superman than to call Clark.[21]

Clark arriving to save the day is metaphysically identical to Superman arriving. But they are distinct outcomes for Lois because she regards them as distinct possibilities. These outcomes differ only in how a particular presents itself to Lois — as an office worker in one case and as able to leap tall buildings in a single bound in the other. Nevertheless, these merely *presentational differences* — differences in how a particular presents itself to an agent — clearly have moral significance for many subjective consequentialists since they explain why Lois should call Superman and not Clark Kent.

A second presentational difference is particularly significant. Just as Clark and Superman are identical, the referent of 'I', in my use, and the referent of 'the actual author of this article' (@AU) are identical. And just as (merely epistemic) possibilities involving Superman can misleadingly present themselves as involving only an office worker and not a superhuman alien, possibilities involving @AU can misleadingly present themselves as involving only him and not me. For example, I can entertain what thrilling turns my life might have taken had I not been the actual author of this article. Likewise, I can imagine someone else, apparently distinct from me, writing it.

Now, if the presentational differences between Lois's Superman possibilities and her Clark Kent possibilities can ground evaluative differences, we cannot be surprised when the presentational difference between me myself and someone else (who happens to be me) also makes an evaluative difference. To illustrate, suppose that the local grocery store has selected you and several others to win a month's worth of groceries. The prize merely compounds your luck: you are extremely well off, having benefited from a generous inheritance. You line up in one of the two queues formed to distribute the prizes. As you progress, only you and a woman in the adjacent queue remain. You are then informed that, owing to a clerical error, only one prize is left. You wonder whether you or the woman would benefit more from the prize. Because you are likely better off than her, you reason that she would probably benefit more from receiving the prize. So you judge it better that she receives it

and that you should, therefore, decline the prize. But, twist of fate, you failed to recognize yourself in the mirror: you were her all along.

Prior to recognizing that you are the woman next to you, the kind of subjective consequentialism that we are considering implies that the outcome where you receive the prize is worse than the outcome where the woman receives it. These two outcomes are distinguished only presentationally. In one, you are "presented to [your]self in a special and primitive way, in which [you are] presented to no one else", the so-called *de se*.[22] In the other, you are presented to yourself third-personally in the impersonal mode: you are presented as a stranger to yourself. Outcomes that involve *de se* thought — "*de se* outcomes" — can therefore differ in value from metaphysically identical outcomes presented differently, just as Lois's Clark Kent and Superman outcomes can differ.

Summing up, enacting agent-centered constraints in a consequentialist framework requires representing differences in agents' points of view. The standard recipe used by New School consequentialists to mimic constraints represents an agent's point of view in *her* distinctive evaluative standards or in the reasons that *she* has to prefer certain outcomes to others. In contrast, forms of subjective consequentialism like the one just presented represent an agent's point of view in how the potential effects of her actions present themselves to her, not in her evaluative standards. Presentational differences, as we will see, reconcile agent-neutral consequentialism and constraints.

3.2 Private Outcomes

A view that I will now develop, called *de se* consequentialism, employs *de se* outcomes to reconcile constraints and agent-neutral consequentialism, demonstrating that they are consistent, *contra* Schroeder, Smith, Portmore, Brown, etc. It requires two assumptions. First, it requires assuming that some outcomes involve *de se* thought, following discussion above. For example, when I drink in order to quench my thirst, the outcome that I intend by drinking — namely, that *I myself* quench my thirst — is necessarily *de se* or essentially indexical. Likewise, when deciding whether you will get

the prize or whether the woman in the other line (who happens to be you) will get it, the first outcome is necessarily presented in the *de se* mode since that's what distinguishes it from the second.[23]

The second assumption is that *de se* thoughts are private. In *de se* thought, you are "presented to [your]self in a special and primitive way, in which [you are] presented to no one else". This claim is also widely assumed.[24]

If some outcomes involve *de se* thoughts and if all *de se* thoughts are private, then some outcomes are private. For example, when Franz is deciding whether to kill or whether Hans and Jens will kill, the outcome entertained by Franz where he himself kills is private. I'll use 'I_F' to represent that outcome, distinguishing it from other presentationally-distinct outcomes where Franz kills, such as F, the publicly-accessible third-personal counterpart of I_F. Since an agent's outcomes, as we've seen, are partly a function of her thoughts and since only Franz is in a position to think his *de se* thoughts, necessarily, I_F is an outcome only for Franz. This makes I_F private.

I_F is essentially *de se* or first-personal; it is an outcome that only Franz can entertain. I_F contrasts with F which, as I'll use it for the rest of the discussion, is an essentially *third-personal* outcome. Just as entertaining I_F requires thinking of Franz in a "special and primitive way" that only Franz can, entertaining F requires entertaining an outcome not only where Franz kills but also where Franz is represented as distinct from oneself. As a result, Franz cannot have F as an outcome unless he is suffering from some kind of *de se* confusion. The moment that confusion lifts, he loses F as an outcome since, necessarily, that outcome represents Franz as a distinct agent or, as I put it earlier, "as a stranger to himself".[25]

Outcomes' privacy is weaker than it may seem in at least two ways. First, though your outcomes are private, they nevertheless permit publicly accessible explanations of your actions. This claim follows from a more general observation about first-personal thought. Even if it is essentially private, I can still grasp the *de se* thought expressed by your utterance of "I am thirsty". Perhaps I

grasp it by implicitly grasping the rule that takes your utterance and its context to a Kaplanian content; perhaps I grasp it through some other explanation. But whatever that explanation is, *de se* thoughts are not totally inscrutable simply because they are private. By the same token, although *de se* outcomes are private, they contribute to explanations of intentional action. For example, when we're trying to explain why Mark is drinking and I claim that it's because he wants to quench his thirst, then my claim does not fully capture the outcome that Mark desires. But it's close enough that we can fill in what's missing using pragmatic means.

Second, that outcomes are private does not imply the absurd proposition that an action affects only its agent or anything of the sort. That confusion rests on misunderstanding the meaning of 'outcome' in this context. 'Outcome', as I'm using it, is a consequentialist term of art, denoting the bearer of value on which an act's rightness allegedly depends. As such, even if an action's intended outcome is private, its effects need not be.

Private outcomes reconcile agent-neutral consequentialism and agent-centered constraints because their imagined conflict relies on certain outcomes being public. For example, on the standard view, the total set of outcomes in Schroeder's scenario is {J, F, H, J&H, J&F, H&F}. Jens's perspective implies that H&F>J and Hans's perspective implies that J&F>H. Because we're assuming that the outcome where Jens kills (J) in Jens's choice set (that is, {J, H&F}) overlaps with one that appears in Hans's choice set (that is, {H, J&F}), two instances of the principle that more killing is always worse — J>J&F and H>H&F — combine to yield a contradiction.[26]

We can derive a similar result more informally. Jens absolutely must not kill only if H&F>J. A constraint on killing therefore implies that J is extremely bad. But if J is extremely bad, then, plausibly, others should sometimes kill in order to avoid J, which straightforwardly conflicts with a constraint on killing. Because J is part of at least two different agents' outcomes, assigning J the value necessary for securing a deontic verdict for one agent may ensure the wrong deontic verdict for the

other agent. The contradiction falsely thought to demonstrate the inconsistency of agent-neutral consequentialism and constraints thus depends on public outcomes.

$De\ se$ consequentialism's private outcomes, however, avoid this implication. Privacy limits their deontic import to only the single individual whose $de\ se$ thoughts constitute the outcome. For example, because I_F is only Franz's outcome, only his actions are affected by I_F's value (or disvalue). Because the deontic significance of $de\ se$ outcomes is limited to a single agent, we can claim that I_F is very bad indeed without implying anything at all about what Jens or Hans should do (and similarly for I_H and I_J).

Accordingly, private outcomes let consequentialists mimic constraints. $De\ se$ consequentialism implies that Hans's choice set is {I_H, J&F}, Franz's choice set is {I_F, H&J}, and Jens's choice set is {I_J, F&H}, yielding a different total set from the standard interpretation of Schroeder's scenario: {I_H, I_J, I_F, H&J, H&F, J&F}. Critically, since $de\ se$ outcomes are private, no $de\ se$ outcome where an agent herself kills is part of a different agent's choice set. As a result, the total set of outcomes {I_H, I_J, I_F, H&J, H&F, J&F} *can* be ordered to mimic a non-consequentialist prohibition on killing just so long as the $de\ se$ outcomes are the worst possible outcomes. This assumption yields the following ranking in terms of the agent-neutral better-than relation: H&J=H&F=J&F > I_H=I_J=I_F.[27] According to this ranking, Hans's choice set {I_H, J&F} implies that Hans should not kill since J&F > I_H and similarly for the rest.

To be clear, I have not yet *justified* ordering outcomes in this way. I've merely shown that $de\ se$ consequentialism lets us do so, thereby squaring constraints and agent-neutral value, which is the paper's main aim. Whether we can justify this ranking on independent grounds and whether $de\ se$ consequentialism is an independently attractive view are *further* questions, which I begin to address below. But I want to be explicit that articulating $de\ se$ consequentialism as I've just done, rebuts Portmore, Schroeder, Brown, Smith, Nair, *et al.* since it establishes that agent-neutral

consequentialism can be reconciled with constraints so it achieves the two aims set out in the introduction.

Agent-centered constraints follow from *de se* consequentialism only given a certain axiology. So the framework itself, apart from that axiology, implies nothing about our moral obligations and permission. Nevertheless, the framework provides indirect support for agent-centered constraints by suggesting how they amount to more than mere 'rule-worship'. For example, if we concede that it's better that one is killed than that five are, we are faced with a familiar challenge: why *not* kill the one? Attempting to answer this question "leaves an ugly stigma."[28] *De se* consequentialists, however, can dismiss the challenge by denying that it is better that I (myself) kill one than that the five are killed by others. Furthermore, rather than countervailing evidence against the truth of consequentialism, consequentialists can assert that the intuitive wrongness of killing one in order to save five tracks the badness of you yourself killing. This allows them to conscript constraint-supporting intuitions as *evidence* of *de se* value. Philosophers attracted to consequentializing, the project of reconciling consequentialism and constraints, feel the pull of both consequentialism's compellingness and of their constraint-supporting intuitions. *De se* consequentialism matters because it shows that these two claims do not pull in opposite directions. Rather, they pull towards accepting an axiology that places special importance on the first-person perspective.

4. *De Se* Consequentialism and its Competitors

4.1 Currying the Standard Recipe[29]

Contrasting *de se* consequentialism with agent-relative consequentialism highlights the former's appeal. Thomas Hurka alludes to a subtle ambiguity when he describes the "concept of agent-relative goodness, or what is good from a person's point of view."[30] What's good relative to an agent and what is good from their point of view differ but can easily be confused. For example, if we meet for lunch,

and I order a soup and you order a sandwich, it might appear that we disagree about what's good to order. Or it might appear that we're using an agent-relative concept, *good-to-order-relative-to-X*. However, if it turns out that the waiter has handed me a menu with only soups on it and that he has handed you one with only sandwiches on it, then what's good to order from our differing points of view differs, but that doesn't entail that we have different conceptions of what's good to order. We needn't be employing the concept *good-to-order-relative-to-X*. Rather, we can be thinking lucidly and consonantly about what's good to order relative to different menus.

The very same mechanism that explains our different lunch orders also allows agent-neutral consequentialists to enact agent-centered constraints. New Schoolers assume that incorporating some measure of agent-centering in consequentialism requires an agent-centered notion of value. This assumption is natural. After all, consequentialists characteristically understand rightness through goodness. As a result, it's natural to assume that mimicking an agent-centered notion of rightness requires an agent-centered notion of goodness. But *de se* value relies on the inherent differences between agents' points of view, particularly with respect to *de se* thought, to enact constraints.

Its mechanics are subtle. This is so partly because, at a certain level of abstraction, *de se* consequentialism and agent-relative consequentialism involve closely related formalisms. The standard recipe avoids contradiction in Schroeder's scenario by distinguishing the evaluative relations that H&F and J&F stand in, relative to Hans and Jens. Rather than implying the contradiction that H&F>J&F and J&F>H&F, the standard recipe implies that H&F>$_J$J&F and J&F>$_H$H&F. A process known alternately as Currying or Schönfinkelization entails that every distinction between relations, like the standard recipe's distinction in evaluative relations, is logically equivalent to a distinction in their *relata,* like *de se* consequentialism's distinction between outcomes like I$_F$ and F. For example, we might say that the silver medalist exemplifies two, distinct monadic properties: being-next-to-the-gold-medalist and being-next-to-the-bronze-medalist. Currying offers an equivalent description. We could instead claim that the silver medalist stands in two instances of the same binary property *being-*

next-to: once by being next to the gold medalist and again by being next to the bronze medalist. Currying shows that these two descriptions are logically equivalent: necessarily, you stand in the 'next to' relation to the gold medalist iff you satisfy the monadic property of being-next-to-the-gold-medalist. Currying the standard recipe distinguishes a particular killing as it appears as Jens's outcome from how it appears as Hans's outcome rather than distinguishing Jens and Hans's evaluative standards.

To mimic constraints, consequentialists must hold that "the amount of value a state of affairs has [varies] from agent to agent."[31] Currying limits what we can vary. In particular, since what an agent should do is a strict function of which of her available *outcomes* is *best*, to mimic constraints we either have to make available outcomes agent-centered or we have to make what's best agent-centered. While the standard recipe makes what's best agent-centered, *de se* consequentialism offers a natural and independently motivated account of outcomes that are, in a certain sense, agent-centered. In other words, agent-relative consequentialism varies the value associated with a state of affairs by varying the function from states of affairs to values from agent to agent. *De se* consequentialism, by contrast, varies the outcome associated with that state of affairs from agent to agent. Thus, the main difference between the two is that the former makes the evaluation of an outcome agent-relative whereas the latter makes the outcome itself agent-relative.

While perhaps formally trivial, this difference is philosophically substantive. Agent-relative consequentialism's main challenge, as we've seen, is squaring agent-relative goodness with the 'compelling' consequentialist idea that it's never wrong to do what's best. *De se* consequentialism, in contrast, does not make evaluation agent-relative. It makes what's evaluated agent-relative. So it uses 'good' in the sense expressed by the compelling consequentialist thought that it's never wrong to do what's best.

As a result, *de se* consequentialism does not involve a novel *kind of value*. It involves only a novel value *bearer*. My invitation to consider the value associated with the first-person perspective

thus resembles an egalitarian's invitation to consider the (agent-neutral) value of equality. Utilitarians may decline the invitation, but not because the value allegedly exhibited by equality is *unintelligible*. Utilitarians simply doubt that equality has familiar, intelligible agent-neutral value. Similarly, when I claim that *de se* killings are worse than their third-personal counterparts, I mean that the difference makes the former worse in the familiar way that other bad-making features do. Worries about intelligibility, such as those that afflict $>_X$ do not apply to *de se* consequentialism.

Moreover, *de se* consequentialism's sensitivity to presentational differences is a considerable advantage over agent-relative consequentialism in cases of *de se* or indexical confusion, such as the grocery case above. For example, forms of consequentialism imply that you should give away the grocery prize only if they recognize a difference between you getting the prize and the woman in the other queue (who happens to be you) getting the prize. But once we recognize that difference, we no longer need agent-relative value to reconcile consequentialism and constraints, as I've shown above.

Nevertheless, *de se* consequentialism's advantage in vindicating the compelling consequentialist thought that you may always do what's best can be doubted.[32] That thought is compelling partly because when something is agent-neutrally best, then it's best in a public or non-parochial way. However, if some outcomes are private, then their value, it might be thought, is private. So even if *de se* consequentialism complies with the letter of the compelling thought, it might seem to violate its spirit.

This impression is misleading. For example, your qualitative experience of your own happiness may be private or inaccessible to me. But that does not diminish its authority over what I should do in a utilitarian account of moral obligation. Though your experience of happiness is private, the authority of its value is public. Likewise, though a *de se* outcome is accessible only to its agent, the outcome is not good or bad *only for her*. It is good or bad in a public, agent-neutral way.

Outcomes' privacy, however, is incompatible with a second appealing feature of consequentialism, beyond the compelling thought.[33] Consequentialism is appealing for social policy

work because it offers a simple way for value to be added up across persons. However, if my *de se* outcomes are outcomes only for me, then their value cannot be straightforwardly aggregated from the policy-maker's point of view: from that perspective, your *de se* outcome is a third-personal outcome. As a result, *de se* value 'disappears' when considered from a third-personal point of view. The special value of the first person perspective cannot, therefore, be part of policy or public decision-making.

This concern is a feature of constraints themselves, not of *de se* consequentialism *per se*. On a simple deontological framework, the wrongness of *my* killing matters to my decision making in a way that it doesn't matter to your decision making and *vice versa*. Thus, this shortcoming appears to be inherent to constraints themselves, afflicting any way of enacting them, not just *de se* consequentialism's particular method of doing so.

Finally, *de se* consequentialism might seem to violate the spirit of the compelling thought in a different sense. Following Scanlon buck-passing analysis of value, suppose that, necessarily, something is agent-neutrally bad just when there is sufficient agent-neutral reason to disvalue it.[34] Following Schroeder, suppose that, necessarily, there is sufficient agent-neutral reason to disvalue something just when there is sufficient reason for *anyone* to disvalue it.[35] Thus, if the outcome where I kill has *de se* (dis)value, then everyone has a reason to disvalue me myself killing. And that's very odd if only I can think the thoughts involved in that outcome.

While odd, this implication is not fatal. Although the *de se* (dis)value of me myself killing gives a reason for anyone to disvalue that outcome, only I am in a position to act on it.[36] The outcome falls under the sphere of my agency and no one else's. Insofar as the implication is odd, it is benign.

4.2 *De Se* Thoughts, Centered Worlds, and Properties

There are, in principle, as many ways to Curry the standard recipe as there are ways of individuating outcomes by agents. For example, Campbell Brown and Jamie Dreier each explore mimicking constraints by distributing value over centered worlds or properties, which David Lewis uses to model *de se* content.[37] Yet I've chosen to Curry the standard recipe using *de se* thoughts, without mention of centered worlds or properties. It's natural to wonder whether that choice is sound.

While centered worlds or properties help make *de se* thought formally tractable, they obscure *de se* consequentialism's most important feature: private outcomes. As we've just seen, *de se* consequentialism can mimic constraints only because some of its outcomes are private. Their privacy follows from the assumption that some outcomes are *de se* thoughts and that all *de se* thoughts are private. Once we understand how some outcomes involve *de se* thoughts, private outcomes follow naturally.

Rather than distinguishing directly between first- and third-personal thoughts, we could instead distinguish between ordered pairs taken to represent those thoughts, as Jamie Dreier does, following Lewis. However, while this paper owes much to Dreier's work, applying this approach to the project of consequentializing is more hindrance than help. Some (principally Herman Cappelen & Josh Dever and Ofra Magidor) have offered potent criticisms of that approach. These criticisms suggest that Lewis's account relies not only on centered worlds or properties but also on an allegedly cryptic *and unreduced* notion of *self-predication*.[38]

To see this, suppose for *reductio* that if agent *x* predicates property *F* of *x*, she *self-predicates F*. Now suppose that Heimson is deceived about his own beliefs. Although he believes that he's Hume, the cognitive dissonance of that false belief has put Heimson in denial. As a result, Heimson does not believe that he himself believes that he's Hume. If H is Heimson's set of beliefs then a Lewisian centered worlds approach seems to imply that:

1. $\{<w,x>: w \in$ [Heimson believes that he is Hume] and x=Heimson$\} \nsubseteq$ H

One day he sees a handsome stranger in line at the grocery store wearing a dashing pink velvet turban. Heimson recognizes the turban from the well-known portrait of Hume. The man must have chosen to don the turban, Heimson reasons, because the man thinks that he's Hume. Hence:

2. {<*w,x*>: *w* ∈ [He believes that he is Hume] and *x*=Heimson} ⊆ H

However, Heimson fails to notice that he's looking at his own turbaned reflection. The stranger is Heimson himself. Since 'He' in (2) refers to Heimson, (1) and (2) are inconsistent. So we must reject one *or* recognize a difference between merely predicating a property of oneself and *self*-predicating it.[39] As Liao puts it, "there seems to be a fundamental conceptual distinction between ascribing properties to *oneself* and ascribing properties to an individual possessing a unique and exhaustive list of non-trivial properties."[40]

This challenge suggests that working *only* with centered contents, at least as they are conventionally understood, fails to capture the nuances of *de se* belief. And since private outcomes depend on the nuances of *de se* belief, employing centered worlds to model the outcomes involved in consequentialism at best obfuscates how to reconcile consequentialism and constraints. At worst, it fails to reconcile them altogether.

Indeed, seemingly because of this obfuscation, Campbell Brown is misled into denying that constraints and consequentialism are consistent. He begins by assuming that the outcome of an action is an entire world.[41] Then, to explore agent-relativity, he pairs worlds with individuals. He writes, "If the centered worlds <*x, i*> and <*y, j*> are centered on different people (i.e., if *i≠j*) then they make no difference to whether [a theory of what's right maximizes value]".[42] That's true: individuating outcomes by individuals does not, all by itself, reconcile consequentialism and constraints. After all, regardless of how we evaluate things, some set of centered worlds will have maximal disvalue and, other things equal, the consequentialist theory will permit agents to kill in order to avoid them. Only if the very worst outcomes are *private*, as *de se* thoughts are, can we be

sure that we'll never have to kill in order to avoid them. But thinking of outcomes as centered worlds obscures their privacy, which is what misleads Brown.

He concludes his discussion of centered worlds with the following reflection:

This way of trying to accommodate agent-relative theories within consequentialism therefore fails. To be clear, I do not claim that agent-relative views are incoherent. I claim only that they are incompatible with the view that there is a single thing, the good, that we all ought to maximize. Such views are better thought of as saying that there are many things that ought to be maximized, a different one for each of us.[43]

Brown is, as I've just argued, mistaken to claim that views like *de se* consequentialism are "incompatible with the view that there is a single thing, the good, that we all ought to maximize". Moreover, his claim that "such views are better thought of as saying that there are many things that ought to be maximized, a different one for each of us" is ambiguous. *De se* consequentialism is *not* "better thought of as saying that there are many [kinds of value] that ought to be maximized, a different [value] for each of us." Its key claim is "that there are many [sets of outcomes whose value] ought to be maximized, a different [set] for each of us."

In sum, *de se* consequentialism achieves constraints by varying outcomes between individuals, not by varying kinds of value. *De se* consequentialism is thus best understood as the view that the absolute, agent-neutrally worst thing, to which all other outcomes are preferable, is a *de se* killing. But your *de se* killings are never my outcomes nor *vice versa*.

5. *De Se* Value

5.1 For all agents X, X>I$_X$

De se consequentialism relies on what I'll call '*de se* value' to enact a constraint on killing, which is ordinary, agent-neutral value (or disvalue), grounded in what's distinctive about first-personal thoughts about killing. We can make its role vivid by adding the perspective of a bystander to Schroeder's case of Hans, Franz, and Jens. Suppose that Klaus is watching the other three bicker about who's going to get their hands dirty. We're assuming that every killing is bad. Accordingly, from Klaus's perspective, Jens killing one is better than Hans and Franz killing two, and *mutatis mutandis* for the rest. Klaus's perspective suggests, quite intuitively, that H=J=F > H&J=H&F=J&F. However, if we assume that H&J=H&F=J&F > I$_H$=I$_J$=I$_F$ to mimic constraint on killing, transitivity implies that H=J=F > I$_H$=I$_J$=I$_F$, from which we can extrapolate that for all agents X, X>I$_X$ — that is, *de se* killings are worse than their third-personal counterparts. I'll focus on this implication for the rest of the paper.[44]

   De se value might initially seem puzzling: that an outcome's presentation can make it worse is odd. But this oddness offers no objection to *de se* consequentialism. That's because I'm not relying on pre-theoretically plausible axiological claims to justify an absolute constraint on killing in the way that we might justify a particular policy by the number of people it helps. Rather, the order of explanation is the opposite: I'm showing what a constraint on killing entails about the structure of value in an agent-neutral consequentialist framework. Consequently, *de se* consequentialists might agree that *de se* value is odd but insist that the evidence for agent-neutral consequentialism is so great and that the intuitive force of a constraint on killing is so strong that *de se* value is worth embracing as a genuine moral discovery, regardless of how puzzling it may seem.

   As a result, I don't need to show that *de se* value is an independently plausible axiological posit. I need only to show that it is coherent. However, one could reasonably suspect otherwise,

particularly given an initially compelling argument that I call *the Supervenience Argument*. According to it:

1. X and I_X represent the same metaphysical possibility.

2. The normative supervenes on the non-normative.

3. If X and I_X represent the same metaphysical possibility, they are non-normatively alike.

4. (MP1-3, 2) Therefore, X and I_X are normatively alike, so *de se* value violates supervenience.

I've stipulated (1) and I assume (2), so I'll rebut the Supervenience Argument by showing how (3) mischaracterizes outcomes, at least as the *de se* consequentialist understands them.

5.2 Representation and Supervenience

It's natural to assume that evaluative differences between outcomes wholly supervene on differences between the metaphysical possibilities that they represent. But that's false: Lois Lane's Superman outcomes differ from her Clark Kent outcomes while being metaphysically identical to them. Hence, it's a mistake to treat outcomes simply as proxies for metaphysical possibilities. Treating them as proxies collapses subjective consequentialism into objective consequentialism.

An analogy helps to diagnose the mistake behind the Supervenience Argument. It is natural to think of, for example, @AU and $I_{@AU}$ as two different vistas on the same landscape. One vista might provide a better view than the other, but that doesn't make the landscape any more beautiful. Likewise, it's tempting to think that though the moral repugnance of a killing may be more vivid from a first-personal standpoint, that doesn't make the killing itself any more repugnant.

But *de se* consequentialists don't conceive of outcomes in this way. Outcomes — an act's foreseeable effects — themselves are the bearers of deontically-relevant value, not the states of affairs that they represent. Normative supervenience therefore implies that evaluative differences between

outcomes necessitate non-evaluative differences between outcomes, not in the *states* that those outcomes represent. Thus, because outcomes are evaluatively significant in their own right, a more appropriate analogy is thinking of @AU and I$_{@AU}$ as two *paintings* of the same landscape that differ in value because one is beautiful and one is ugly. Evaluative differences between paintings do not supervene on what those paintings represent. It is clearly a mistake to think that two paintings differ in their degree of beauty only if they differ in what they depict. Although paintings are representations, their value is at least partly independent from what they represent. If expected outcomes are like paintings in this respect, it's also a mistake to think that outcomes' evaluative properties supervene on what they represent.

Thus, premise (3) of the Supervenience Argument is false. Normative supervenience is consistent with *de se* value. For example, the state of affairs where Clark Kent arrives on the scene is identical to the one where Superman arrives. Nevertheless, the outcome corresponding to the second state of affairs exemplifies a property that the first lacks: the second is an outcome where it's *likely* that the world is saved, relative to Lois's information. This epistemic property — being such that it's likely that the world is saved — is *not* a property of the state of affairs that both outcomes represent since the outcomes differ relative to this epistemic property while representing the same state of affairs. Rather, it's a descriptive property of the outcome in its own right, *qua* representation, that makes a considerable difference to its value. The property of being *de se* is similar: a descriptive property of an outcome that affects its value.

The analogy between paintings and *de se* outcomes has limits. There are ugly paintings of beautiful things and beautiful paintings of ugly things. The analogy thus suggests an unconstrained relationship between an outcome and the corresponding event or state-of-affairs. But an analogous decoupling of the value of an outcome from the value of the corresponding event is unappealing; surely the value of the outcome is at least partly a function of the value of the event.

A second analogy is helpful here. Nomy Arpaly and Timothy Schroeder argue that co-referring representations can differ evaluatively because a particular kind of goodness, *good will*, depends on presentational differences like those on which *de se* consequentialism relies.[45] According to Arpaly and Schroeder, "complete good will is an intrinsic desire for the right or good, *correctly conceptualized*."[46] This account straightforwardly implies that two representational states (in this case, desires) can vary in value (particularly, in their quality of will) while representing the same state. For example, the desires of an alien who finds the firing of c-fibers very unpleasant, like fingernails scraping over a chalkboard, will involve the same state of affairs as the desires of a pure-hearted altruist who abhors seeing people in pain. However, only the altruist's actions exemplify good will: when the alien is moved to alleviate someone's suffering, the alien's deployment of the concept *c-fiber firing* does not 'correctly conceptualize' the suffering. The quality of the alien's will is thus worse than the altruist's, though both are moved by desires that differ only in how they present a given state. However, the alien's actions do not merit praise because of this difference.

De se consequentialism is similar: simply swap 'good will' with '*de se* (dis)value' and swap 'merits praise' with 'is right'. Arpaly and Schroeder hold that some acts differ in their praiseworthiness based only on differences in good will grounded in presentational differences between desires. *De se* consequentialism holds the analogous position that some acts differ in their rightness based only on differences in *de se* value grounded in presentational differences between outcomes. The value of an outcome is, therefore, still closely related to what it represents: only outcomes associated with constraints can exhibit *de se* value, which can only intensify the badness of a killing. The link between an outcome's value and of the corresponding state is far more closely linked than the beauty of a painting and of what it represents.

This analogy between good will and *de se* value is no coincidence. Julia Driver also finds a structural similarity between subjective consequentialism and Kantian ethics. She writes,

Subjective forms of consequentialism are also internalist in the same way the Kantian system is, since on one version of subjective consequentialism, for example, right action is defined in terms of expected consequences, and on this version of the subjective view a virtue would be a disposition to act so as to maximize expected good or to try to maximize the good. Across this parameter, then, subjective consequentialism and Kantian ethics are the same sort of ethical theory, whereas subjective and objective consequentialism are radically different, as they locate the source of value very differently.[47]

Driver's observation bears repeating. Kantianism and various subjective consequentialisms, such as *de se* consequentialism, share a deep and relatively neglected similarity; they explain facts about moral rightness by facts about the moral significance of certain internal mental states.[48] Given this resemblance, it is no wonder that *de se* consequentialism exhibits features characteristic of non-consequentialist views like Kantianism, such as constraints. The main difference between subjective consequentialism and Kantianism lies in consequentialism's *telic* nature. Whether an agent does what's right, according to subjective consequentialists, it is associated with the right outcome — or, equivalently, whether the agent has the right telic state — not on other states such as the 'maxim' of their action.

6 Fault Lines in Ethical Theory

6.1 Companions in Guilt

De se consequentialism is noteworthy not just for consequentialists sympathetic to constraints, but also for anyone interested in what we can call, following Shyam Nair, the fault lines that divide various positions in ethics.[49] For example, in his influential "Structures of Normative Theories" and in subsequent work, Jamie Dreier defends the claim, which I have defended here, that the agent-

neutral/agent-relative distinction is orthogonal to the consequentialist/non-consequentialist distinction: accepting agent-relativity in ethics does not require rejecting consequentialism. However, this claim might be taken to diminish the importance of the agent-neutral/agent-relative distinction, which, philosophers largely agree, is one of the "greatest contributions of recent ethics."[50] So it's worth asking whether constraints trace a genuine fault line in ethics. I will argue that they do, just not the fault line between consequentialism and non-consequentialism.

I'll focus in this final section on identifying and defending one claim in particular: that claims expressing agent-centered constraints typically — perhaps necessarily — involve indexical terms, particularly 'I'. Doing so shows that *de* se value is noteworthy not only because it grounds a consequentialist account of constraints, explaining why they might amount to more than mere 'rule-worship'. It also illuminates how any moral theory that posits constraints thereby attributes moral significance to differences like the one between first- and third-personal thought. As we will see, drawing out the essential indexicality of many agent-centered constraints supports Stephen Darwall distinction between ethical theories that start from the 'point of view of the universe' and those that start from the agent's point of view — or, as he puts it, theories that start from the 'inside-out' and those that start from the 'outside-in'.

This fault line merits attention not only for its own sake but also because it provides a 'companions in guilt' defense of *de se* value. *De se* value is at least *prima facie* puzzling. But it is merely the consequentialist aspect of the puzzle that agent-centered constraints pose for *any* ethical theory, whether consequentialist or not. Moral theories, as we'll now see, seem to require attributing moral significance to what I've been calling presentational differences to enact constraints. Insofar as *de se* value is odd, therefore, that's only because agent-centered constraints are odd. The oddity of *de se* value offers an argument against constraints themselves, not their expression in *de se* consequentialism.

Deontological and reasons-based ethical frameworks, which are often taken as competitors to consequentialism, express constraints as follows:[51]

1. It is always wrong to kill.

2. There is never most reason to kill.

Neither claim explicitly mentions an agent. This is essential, for one important role of constraints is to offer practical or deliberative guidance about how to act. However, stating (1) with an agent, as in (3) below, undercuts its practical guidance:

3. It is always wrong *that anyone* kills.

For example, in cases like *Mafia,* it is certain that someone will kill. Consequently, (3) offers no practical guidance on whether to kill in *Mafia* since killing is unavoidable. In contrast, (1) tells us what to do: not kill, even if someone else will. Similarly for (2). As a special case of (3), (4) also fails to play constraints' practical role:

4. It is always wrong for @AU to kill/ that @AU kills.

I've introduced '@AU' as a name for me, the actual author of the article. As such, it's natural to suppose that a constraint on killing, as it applies to me, is captured by something like (4). But (4) omits the first-person, making it susceptible to *de se* confusion, such as when I doubt that I'm @AU, undercutting (4)'s practical guidance.[52] For example, suppose that I accept (4), that it is always wrong for @AU to kill. But suppose that I don't know that I am @AU. When faced with an apparent choice between killing someone or having @AU kill them and I doubt that I am @AU, I will infer that I may kill since it is always wrong that @AU kills, which is inconsistent with a constraint on killing.

The difference is that (1) and (2), in contrast to (3) and (4), use a non-finite clause, namely, 'to kill'.[53] Non-finite clauses in environments like (1) and (2) tacitly involve first-personal thought by covertly expressing a generic first-personal pronoun tied to the *de se*, making them equivalent to 'it is always wrong *that you yourself* kill' and 'there is always most reason *that you yourself* not kill'.[54] This is

a general feature of *to*-clauses embedded under null grammatical subjects.[55] For example, 'It is good to see you again', as uttered by me, expresses that it is good that *I myself* see you again. Consequently, as others have noted, deontological and reasons-based theories don't incorporate agent-centered restrictions simply because of their apparent focus on actions.[56] Rather, (1) and (2) are agent-centered because they incorporate a *covert* indexical element tied to the first person. That is, what makes these restrictions agent-centered is that they're essentially indexical.

With (1) and (2)'s covert *de se* sense made explicit, we're in a better position to appreciate a deep similarity between the three accounts of agent-centered constraints:

5. De Se *Consequentialist*: It is always the worst outcome [that *you yourself* kill].

6. *Conventional Deontology*: It is always wrong [that *you yourself* kill].

7. *Reasons-Based*: There is never most reason [that *you yourself* kill].

The three approaches are distinguished only by the normative property or concept that each privileges: value for consequentialists, norms or rules for deontologists, and reasons for the third. Each enacts the agent-centered constraint on killing by associating maximal negative moral status (*e.g.*, 'always worst', 'always wrong', 'always most reason to not') with a clause that requires a first-personal reading.

As a result, each faces the uncomfortable questions that may make *de se* value seem odd. In particular, if morality applies to each agent impartially, why should an agent attribute special moral significance to *her* killings? For example, (6) attributes wrongness only to killings that you yourself perform. Likewise, (7) implies that *your* killings are uniquely disfavored by the normative reasons that matter. These implications are puzzling because they indicate a profound and obscure moral asymmetry between oneself and others, which conflicts with morality's apparent impartiality. But the conflict is rooted in the nature of constraints, not in the moral theories themselves.

De se value is, therefore, the species of a genus. A wide range of moral theories attribute special moral significance to first-personal thoughts in order to enact action-guiding constraints.

Consequentialists argue that value exhausts moral significance. As a result, consequentialists attribute a novel kind of value to these thoughts, *de se* value. But *de se* value is not clearly more implausible than these other ways of interpreting the moral significance of the *de se* through special reasons or special rules.

So it seems that regardless of whether we enact constraints through reasons, rules or value, we must understand them indexically which helps explain observations that others have made about their logical structure. Agent-centered constraints seem to possess a distinctive distribution of agent-referring variables.[57] The idea that agent-centered constraints are essentially indexical builds on this variable-based account in at least two ways. First, it shows that having the appropriate agent-referring variables is necessary but not sufficient for being agent-centered; the referents of those variables must be understood first-personally or be undermined by cases of *de se* confusion. Second, the essential indexicality of some agent-centered constraints explains *why* such rules are unified by that particular distribution of variables: that distribution is necessary for the rule to be understood first-personally, which is necessary for interpreting it as an agent-centered constraint.

6.2 Ethics from the Inside-Out

Two points from the preceding discussion merit particular attention. First, enacting constraints requires the kind of information that appears only from the agent's point of view; it is therefore the universe's point of view, often assumed by consequentialists, that is incompatible with constraints, not consequentialism itself.

Second, more complex (but no less consequentialist) forms of consequentialism, like some subjective ones, allow constraints when they take the agent's point of view. Normative concepts like *good* are, therefore, more plastic than theorists typically suppose. A given ethical theory's first-order permissions and prohibitions are shaped more by its point of view than by its preferred flavour of

normativity. Thus, the traditional fault lines between ethical theories traced by their privileged normative concept (rights, goodness, virtuous character, reasons, etc.) may not carve at the most central ethical joints. Class ethical theories by whether they adopt the agent's point of view or whether they adopt what has been called, variously, Sidgwick's "point of view of the universe", Nagel's "point of view of nowhere", or Williams' "absolute conception of the world" cuts at a deeper joint.

A distinct route leads Stephen Darwall to a similar conclusion. He traces a fault line between ethical theories that start from the 'outside-in' and those that start from the 'inside-out', writing:

> To put it in a rough and preliminary way, moral integrity involves a person's guiding his life by his own moral judgment, properly understood, and the fundamental responsibility of the moral life is the maintenance of integrity, so conceived. Instead of beginning outside the moral agent with a view about states of affairs that are intrinsically worthy of promotion, the alternative begins inside the moral agent with a view about moral character and integrity. The rationale for agent-centered restrictions is itself agent-centered.[58]

Darwall's point is this. Some values don't appear in states of affairs, considered apart from the point of view of particular agents.[59] If we start theorizing about morality using only those states of affairs, and we are willing to recognize no values apart from those that appear in them — if we theorize about morality "from the outside-in", as Darwall puts it — then values that are visible only to agents, like integrity, seem puzzling. For example, if we start theorizing about a painting's beauty without the painting itself or a photo of it, but rather with a (very long) numerical readout specifying the location and velocity of each particle that makes up the painting, it's going to be very difficult, perhaps impossible, to judge whether the painting is beautiful using only the readout. The painting's beauty won't appear when approaching it from the 'outside-in'. Rather, we must behold the painting

with our own eyes, from our own perspective, to apprehend its beauty. The same holds true for *de se* value as a kind of moral ugliness. *De se* value does not appear when we examine what *de se* thoughts represent. It emerges only when we think the thoughts themselves. So it makes a great difference whether ethical theorizing starts from the point of view of the universe, which excludes or prescinds from indexical thought, or from the point of view of a particular agent, which is constituted by it.

Objective consequentialism is a paradigmatically 'outside-in' approach to ethics. It limits itself to only those evaluative features that belong intrinsically to states of affairs. If Darwall is right, those features cannot provide a rationale for agent-centered constraints; we cannot justify constraints if we begin theorizing from the point of view of the universe and begin thinking about value from the 'outside-in'.

Consequently, to vindicate constraints we must approach them from the 'inside-out', and exploit indexical features that appear only from an agent's point of view. As I've argued, it should be unsurprising, then, that capturing constraints using outcomes requires positing evaluative features of outcomes that are not evaluative features of the states that those outcomes represent. Since representations mediate agents' engagement with the world, constituting their point of view on it, it is natural to exploit, for example, beauty, an evaluative feature characteristic of representations, to explain agent-centered constraints from the 'inside-out'.

Recognizing this fault line is especially important because it explains the mistaken insistence that consequentialism is incompatible with constraints. If we assume that every consequentialist theory is an outside-in theory, then constraints and consequentialism are indeed incompatible; constraints appear only from the inside-out. But consequentialism is not essentially an outside-in approach. If we begin consequentialist theorizing from the agent's point of view, in the manner of subjective consequentialism and of *de se* consequentialism in particular, then a consequentialist justification for something like agent-centered constraints is achievable through *de se* value.

Nevertheless, an apparent tension divides *de se* value and the axiology that standardly underpins consequentialism, which often regards value as, roughly, what's desirable 'from the point of view of the universe' or from the outside-in.[60] 'The point of view of the universe' offers an appealing perspective on value because it fixes a standard of correctness or objectivity for evaluation. Objectivity might thus seem essential to agent-neutral value. Correspondingly, because *de se* value is invisible from that point of view, its objectivity might be doubted. And if *de se* value lacks objectivity, it might seem like a merely covert form of agent-relative value.

De se consequentialists can respond by appealing to an ancient source of objectivity in ethics: the *phronimos* or practically wise agent. One standard formulation of virtue ethics holds, very roughly, that you should act in some way just when and because the *phronimos* would act that way in your position. Practical wisdom provides the objective standard against which an act's rightness is judged. We can extend that standard to ground *de se* value's objectivity by holding that one outcome is agent-neutrally better than another just when and because the *phronimos* would prefer the first to the second, were she (*per impossibile*) in your shoes and knew only as much as you do now. To be sure, response-dependent accounts of value like this face well-known challenges, such as the Euthyphro. But these challenges are no more daunting than the challenge of explicating what 'the point of view of the universe' is. Consequently, *de se* consequentialists can vindicate the objectivity of *de se* value by grounding that value in the point of view of a privileged agent, rather than that of a universe.

7. Conclusion

De se consequentialism employs agent-neutral value and agent-centered outcomes to enact constraints. Consequently, it provides a counterexample to the claim that "rankings of outcomes will need to be agent-relative […] in order to consequentialize certain non-consequentialist theories." To enact constraints, *de se* consequentialism implies that some *de se* outcomes are worse than their third-

personal counterparts. This may seem puzzling. Nevertheless, I have argued that it is no more surprising than parallel implications from competing non-consequentialist theories and that it has precedent in a familiar evaluative property, beauty.

Table 1

Expected Effects ▷ ⎯ ⎯ ⎯ ⎯ ⎯ ⎯ ⎯ ⎯ Actions▽	Superman arrives to save the day.	Clark arrives to save the day.
Lois dials Superman's number	(1) 10 utility at 99% likelihood = 9.9 value.	(3) 1 utility at 1% likelihood = .01 value.
Lois dials Clark Kent's number	(2) 10 utility at 1% likelihood = .1 value.	(4) 1 utility at 99% likelihood =.99 value.

Caption: Lois Lane's Choice

[1] See James Dreier, "Structures of Normative Theories." *The Monist* 76 (1993): 22-40; Jennie Louise, "Relativity of Value and the Consequentialist Umbrella." *The Philosophical Quarterly* 54 (2003): 518-536; and Martin Peterson, "A Royal Road to Consequentialism?" *Ethical Theory and Moral Practice* 13 (2010): 153-169 for remarks in this vein. For critical discussion, see David McNaughton and Piers Rawling, "Honoring and Promoting Values." *Ethics* 102 (1992):835-843; Campbell Brown, "Consequentialize This." *Ethics* 121 (2011): 749-771; Paul Hurley, "Consequentializing and Deontologizing: Clogging the Consequentialist Vaccuum," *Oxford Studies in Normative Ethics* Vol. 3 (2013): 123-153; and S. A. Schroeder, "Consequentializing and Its Consequences," *Philosophical Studies* 174 (2017):1475-1497.

2 Douglas Portmore, "Consequentializing Moral Theories," *Pacific Philosophical Quarterly* 88 (2007): 39-73. See also Douglas Portmore, "Consequentializing," *Philosophy Compass* 4 (2009): 330: "Of course, the rankings of outcomes will need to be agent-relative (and, perhaps, also time-relative) in order to consequentialize certain non-consequentialist theories." Brown offers a similar thought in "Consequentialize This": "Some views of right and wrong are incompatible with [agent-neutral] consequentialism, no matter what theory of the good one adopts."

3 Perhaps, as Louise, "Relativity of Value" and Michael Smith, "Two kinds of Consequentialism," *Philosophical Issues* 19 (2009): 257-272 observe, temporal considerations force a parallel treatment of *de nunc* outcomes. Since the treatment of the *de se* and the *de nunc* will be extremely similar, I bracket the issue.

4 I'm using 'constraint' to mean what Samuel Scheffler means by 'agent-centered restriction'. He writes that "an agent-centered restriction is, roughly, a restriction which it is at least sometimes impermissible to violate in circumstances where a violation would serve to minimize total overall violations of the very same restriction, and would have no other morally relevant consequences" in Samuel Sheffler, "Agent-Centred Restrictions, Rationality, and the Virtues," *Mind* 375 (1985): 409-419.

5 I'm assuming that moral dilemmas where, for example, you should kill and you should not kill, are impossible. Jussi Suikkanen, "Consequentializing Moral Dilemmas," *Journal of Moral Philosophy* 17 (2020): 261-289 argues that moral dilemmas pose a special challenge to accommodating constraints in a consequentialist theory. His arguments' lessons are orthogonal to my aims here.

6 If outcomes extend across the future without limit, knowing whether an act is permissible implausibly requires knowing elusive facts about the distant future — see James Lenman, "Consequentialism and Cluelessness," *Philosophy & Public Affairs* 29 (2000): 342-370 and Elinor Mason. "Consequentialism and the Principle of Indifference," *Utilitas* 16 (2004): 316-321 for more. I assume that the outcome of, for example, refraining to kill someone can include the killing of two by others.

[7] Paul Hurley offers reasons to be suspicious of this claim in "Why Consequentialism's 'Compelling Idea' is Not," *Social Theory and Practice* 43 (2017):29-54; "New Consequentialism and the New Doing-Allowing Distinction" in *Consequentialism: New Directions, New Problems,* Ed. Christian Seidel, Oxford University Press, (2019):176-97; and "Exiting the Consequentialist Circle: Two Senses of Bring it About," *Analytic Philosophy* 60 (2019):130-163.

[8] I'll drop the explicit 'innocent' qualification going forward and implicitly assume that all killings discussed below are killings of innocent people.

[9] See Mark Schroeder, "Not So Promising After All: Evaluator-Relative Teleology and Common-Sense Morality," *Pacific Philosophical Quarterly* 87 (2006): 348-356.

[10] '&' does not express conjunction and is syncategorematic, saving me from inserting parentheses throughout.

[11] Given Hans's perspective, J&F>H. But H>H&F by the assumption. So J&F>H&F by transitivity. Likewise, given Jens's perspective, H&F>J. But J>J&F by the assumption. So H&F>J&F by transitivity. However, > is anti-symmetric, resulting in contradiction.

[12] Portmore, "Consequentializing Moral Theories" disputes this claim.

[13] Schroeder, "Not So Promising, p.354. Of course, not everyone finds consequentialism compelling — see Hurley, "Why Consequentialism's 'Compelling Idea' Is Not."

[14] Discussions of this concept include Thomas Hurka. "Moore in the Middle," *Ethics* 113 (2003): 599-628; Michael Smith, "Neutral and Relative Value after Moore," *Ethics* 113 (2003): 576-598; Portmore, "Consequentializing Moral Theories"; Jussi Suikkanen, "Consequentialism, Constraints, and the Good-relative-to: a Reply to Mark Schroeder," *Journal of Ethics and Social Philosophy* 3 (2008): 1-8; and James Dreier, "In Defense of Consequentializing," *Oxford Studies in Normative Ethics* 1: 97-119. For arguments that *some* axiology must mimic constraints, see Peterson, "A Royal Road to Consequentialism?".

[15] Mark Schroeder, "Teleology, Agent-Relative Value, and 'Good'," *Ethics* 115: 265-295.

[16] See Smith, "Two Kinds of Consequentialism," and Douglas Portmore, *Commonsense Consequentialism (*Oxford University Press, 2011) for views defended along these lines.

[17] Christopher Howard, "Consequentialists Must Kill," *Ethics* 131 (2021): 727-753.

[18] I'm using 'subjective' and 'objective' consequentialism in roughly the manner that Peter Railton, "Alienation, Consequentialism, and the Demands of Morality," *Philosophy and Public Affairs* 13 (1984): 1134-171 does. I'm not the first to exploit the distinctive features of probability and expectation to defend consequentialism. Elements of my approach are anticipated by Frank Jackson, "Decision-Theoretic Consequentialism and the Nearest and Dearest Objection," *Ethics* 101 (1991): 461-482 and Graham Oddie and Peter Milne, "Act and Value: Expectation and the Representability of Moral Theories," *Theoria* 57: 42-76.

[19] Ana Mahtani, "The Ex Ante Pareto Priniciple," *Journal of Philosophy* 114 (2017):303-323.

[20] I'm using 'utility' as a catch-all term for the features of the possible state-of-affairs represented by an outcome that contribute to the outcome's value. An outcome's value is (partly) a function of its expected utility.

[21] Like Mahtani, I reject "naive referentialism about credence". On the view that we favour, it's possible to have distinct credal attitudes towards distinct epistemic possibilities that are intrinsically metaphysically identical. See also David Chalmers, "Frege's Puzzle and the Objects of Credence," *Mind* 120 (2011): 587-635 and Jesse Fitts, "Chalmers on the Objects of Credence," *Philosophical Studies* 170 (2013): 343-358.

[22] Gottlob Frege, "The Thought: A Logical Inquiry," *Mind* 65 (1956): 289-311.

[23] *De se* consequentialists therefore follow many philosophers in assuming that outcomes figure in the psychological explanation of action and that, as José Bermudez, "Yes, Essential Indexicals Really Are Essential," *Analysis* 77 (2017): 690-4 puts it, "psychological explanations are essentially indexical". The phrase 'essential indexical' of course traces back to John Perry, "The Problem of the Essential Indexical," *Philosophical Review* 86 (1979): 874-97, but defenders of the thesis also include Hector-Neri Castañeda, "Intentions and the Structure of Intending," *Journal of Philosophy* 68 (1971): 453-66; Tyler Burge, "Reason and the First Person," in *Knowing Our Own Minds,* C. Wright, B. Smith and C. Macdonald (Eds.), (Oxford University Press, 2000); and José Bermudez, *Understanding "I"* (Oxford University Press, 2016). Herman Cappelen and Josh Dever, *The Inessential Indexical* (Oxford University Press, 2013) criticizes the thesis. Matthew Babb, "The Essential Indexicality of Intentional Action," *Philosophical Quarterly* 66 (2016): 439-456 and Bermudez, "Yes, Essential Indexicals Really Are Essential" respond.

[24] Bermudez *Understanding "I"* argues that *de se* thought is not private. According to him, "an account of the sense of "I" must allow tokens of "I" to have the same sense as tokens of other personal pronouns such as "you" in appropriate contexts" (p.79). However, Bermudez's package of views offers no succor to opponents of *de se* consequentialism. José Bermudez *Frame It Again* (Cambridge University Press, 2020) defends a controversial *ultraintensionality* according to which even if tokens of "I" can share a hyperintension with tokens of "you", they nevertheless differ in ways that allow those tokens to be evaluated differently, which is all that *de se* consequentialism requires. This shows that *de se* consequentialism relies on weaker assumptions than those that I adopt above. I adopt them only to simplify exposition.

[25] One might worry that known identities pose a problem for *de se* consequentialism. When Lois knows that Clark is Superman, then she cannot rationally assign different values to her Clark and Superman outcomes. Likewise, when someone knows that F and I_F represent the same state, they must assign them the same value. But, as we'll see, mimicking constraints requires assigning those outcomes different values.

This worry is less forceful than it appears. Since only I am in the position to think my *de se* thoughts, only I am in a position to assign values to $I_{@AU}$. When I know that @AU and $I_{@AU}$ represent the same state, @AU is *not* one of my outcomes, since it is essentially a third-personal representation of myself. Just as Franz can have F as an outcome only if he is suffering *de se* identity confusion, I can have @AU as an outcome only if I'm similarly confused and I can rationally assign @AU and $I_{@AU}$ different values when confused.

Some might insist that I can have @AU as an outcome even if I'm not confused because I can imagine myself from a third-personal perspective while identifying with the imagined person. First, such thoughts have long puzzled philosophers. It seems plausible that when we imagine ourselves third-personally, we are engaging in the mere pretense that we are distinct from the imagined person. Or perhaps our identity is distributed across two distinct perspectives. In any case, neither description seems to fit outcomes like @AU, which essentially present their agents as distinct from oneself. Second, even if I can imagine myself third-personally while maintaining the integrity of my identity, *imagining* @AU and *having* @AU as an outcome are two different conditions. I'm concerned only with the latter. Observations about the former are orthogonal.

26 Recall that, given Hans's perspective, J&F > H. But H > H&F by the principle. So J&F > H&F by transitivity. Likewise, given Jens's perspective, H&F > J. But J > J&F by the principle. So H&F > J&F by transitivity. However, > is anti-symmetric, resulting in contradiction.

27 To avoid proliferating symbols, I'm not using '=' to express identity but rather 'is equal in value to', the binary relation between outcomes that holds when their value is identical. I should note that the ordering above violates a stronger principle than the principle that killing is always bad, according to which two killings are always worse than one: for any X, Y, and Z $\in$ {H, J, F, I$_H$, I$_J$, I$_F$} such that Y≠Z, X>Y&Z. But this principle is straightforwardly inconsistent with a constraint on killing, since it entails that it is always better to kill one in order to save two.

We might reasonably deny this stronger principle on independent grounds even if we are drawn to the weaker one set out in section two. The weaker principle does not make comparisons between the disvalue of various killings. It holds only that killing is always bad. The stronger principle, by contrast, holds that two killings are always worse than one. If we think that killings are not comparable in disvalue or that killings can differ sufficiently in value, we should reject the stronger principle regardless of our sympathy for consequentializing.

28 Dreier, "Structures of Normative Theories".

29 With apologies to the referee who objected to this pun.

30 Hurka, "Moore in the Middle", p.611.

31 Oddie and Milne, "Act and Value", p.44.

32 I thank a referee for *Ethics* for pressing me to clarify my position here.

33 I am grateful to José Bermudez for discussion of this point.

34 Thomas Scanlon, *What We Owe to Each Other* (Harvard University Press, 1998).

35 Mark Schroeder, "Reasons and Agent-Neutrality," *Philosophical Studies* 135 (2007): 279-306.

36 My position echoes those taken by Ralph Wedgwood and John Broome on a similar issue, as reported in Mark Schroeder, "Oughts, Agents, and Actions," *Philosophical Review* 120 (2011): 1-42.

37 David Lewis, "Attitudes *De Dicto* and *De Se,*" *Philosophical Review* 88 (1979): 513-543.

38 Ofra Magidor, "The Myth of the De Se," *Philosophical Perspectives* 29 (2015): 249-283.

[39] Lewisians might pursue the first approach by choosing to represent the property that Heimson (self-)predicates when he first-personally believes that he is Hume as $\lambda x.\ x$ is Hume, distinguishing this property from the one (self-)predicated in the third-personal that *that guy* (who is Heimson) believes that he is Hume, *viz.*, $\lambda x.\ x$ is such that *that guy* believes that he is Hume. If these properties differ, then it looks like Lewisians can use that difference to account for first-personal thought at the level of the properties ascribed in belief, not using a controversial and distinctive kind of *self*-ascription. Mark Schroeder and Michael Milona (2019) "Desiring Under the Proper Guise," *Oxford Studies in Metaethics* 14: 121-143 explores this strategy for desire ascriptions.

[40] Sam Liao, "What Are Centered Worlds?" *Philosophical Quarterly* 242 (2012): 314.

[41] Brown, "Consequentialize This," p.759. Brown explicitly excludes consequentialisms that rely on "subjective probabilities", such as *de se* consequentialism, on the grounds that he is interested in objective, not subjective, rightness. But theories of objective rightness that incorporate subjective probabilities are perfectly consistent. So Brown's interest in objective rightness does not diminish *de se* consequentialism's relevance to his project. I am neutral on whether *de se* consequentialism explains objective rather than subjective rightness.

[42] Brown assumes that a theory of what's right is consequentialized by a theory of value just when what's right (R) for an individual (i) given a set of options (A) — that is, when $R_i(A)$ — is the most valuable ($grt_{\leq}$) option in A — that is, when $R_i(A) = grt_{\leq}(A)$. However, if ought implies can in at least one sense of can, then the only options that can be right for an agent are ones choosable by that agent. So we should relativize A to individuals, yielding A_i, and instead hold that a theory of value $grt_{\leq}$ consequentializes a theory of right (R) just when $R_i(A_i) = grt_{\leq}(A_i)$. If options vary between individuals in the way that *de se* consequentialism suggests, then it's possible that $R_i(A_i) = grt_{\leq}(A_i)$.

[43] Brown, "Consequentialize This," p.763.

[44] Daniel Muñoz, "The Rejection of Consequentializing," *Journal of Philosophy* 118 (2021): 79-86 argues that intrinsic differences between actions cannot ground moral differences between outcomes for consequentialists. One might worry that this threatens *de se* value. However, since, for example, J and I_J are caused by *token identical actions*, moral differences between those outcomes are not the result of intrinsic differences between actions. As a result, *de se* value is not a 'broad' feature of outcomes, so appealing to that feature complies with the stipulation that only narrow differences between outcomes are moral differences. Muñoz's critique of consequentializing therefore does not apply to *de se* consequentialism.

[45] A referee for *Ethics* astutely observes that *de se* value is unusual because it seems to be a kind of *non-instrumental* but *derivative* value. The value of a good will, on some analyses, is similar in this respect.

[46] See Nomy Arpaly and Timothy Schroeder, *In Praise of Desire* (2013, Oxford University Press) p.163, emphasis mine.

[47] See Julia Driver, *Uneasy Virtue* (Cambridge University Press, 2001), p.70.

[48] 'Internal' may have to be adjusted to account for factive mental states.

[49] Shyam Nair, "A Fault Line in Ethical Theory," *Philosophical Perspectives* 28 (2014): 173-200; Shyam Nair, "Fault Lines in Ethical Theory," in *The Oxford Handboook of Consequentialism,* Douglas Portmore (Ed.) (Oxford University Press, 2020).

[50] Hurka, "Moore in the Middle," p.628.

[51] Matthew Hammerton, "Deontic Constraints are Maximizing Rules," *The Journal of Value Inquiry* 54 (2020):371-388 discusses various logically distinct formulations of constraints. Although grammatically distinct, (1) and (2) above are roughly analogous to his formulation #4, with the *caveat* that I have assumed that the correct treatment of *de nunc* outcomes in a consequentialist framework parallels *de se* consequentialism's treatment of *de se* outcomes.

[52] James Dreier, "Two Models of Agent-Centered Value," *Res Philosophica* 97 (2020): 345-362 discusses similar cases of 'indexical uncertainty'. He aims to show that views like *de se* consequentialism encode strictly more information than views like agent-relative consequentialism and that such information is not "junk information". I believe the arguments above support his claims.

53 In Nathan Howard, "Sentimentalism about Moral Understanding" *Ethical Theory and Moral Practice* 21 (2018): 1065-1078 and Nathan Howard, "Primary Reasons as Normative Reasons," *The Journal of Philosophy* 118 (2021):97-111, I defend a conception of normative reasons for action according to which they comprise goals. For example, the moral reason to save Bill Gates from drowning is not merely the *fact* that he's in mortal danger but the goal pursued by the desire to save his life. This moral reason contrasts with, for example, a second merely *self-interested* reason *also* given by that fact that Bill is in mortal danger which comprises the different goal pursued by the desire to get rich by saving Bill. It's natural to wonder what makes the first goal moral and the second goal merely prudential or self-interested. A natural answer is that the first one involves a (predicatively) *better* goal than the second. But analyzing moral goals in terms of agent-neutral value might seem to make moral reasons incompatible with agent-centered constraints in just the manner that some mistakenly think that agent-neutral consequentialism is incompatible with agent-centered constraints. However, since the goals I have in mind are denoted by *to*-clauses, the arguments above show that the conception of normative reasons for action that appeals to me is compatible with agent-centered constraints in just the same way that *de se* consequentialism is. I became interested in this whole project because I was worried --- needlessly, it turns out --- that they were not compatible. In retrospect, it feels a little like I created a sledgehammer to kill a fly but, happily, the sledgehammer seems to have other uses.

54 It is widely acknowledged that *to*-clauses in environments like (1) and (2) express a covert anaphoric or deictic pronoun, PRO. Jason Stanley and Timothy Williamson, "Knowing How," *Journal of Philosophy* 98 (2001): 411-444, for just one example, exploit this observation to defend intellectualism about *know-how*.

55 For a comprehensive defense, see Frederike Moltmann, "Generic One, Arbitrary PRO, and the First Person," *Natural Language Semantics* 14 (2005): 257-281, with important antecedents in Gennaro Chierchia "Anaphora and Attitudes," in *Language in Action*, R. Bartsch et al. (Eds.)(Dordrecht: Foris Publications, 1990). PRO generally inherits its referent from the sentence's grammatical subject, but in sentences with null subjects like (1) and (2) it takes on an *arbitrary* referent, which, Moltmann argues, is necessarily interpreted *de se*.

[56] For example, Tom Dougherty, "Agent-Neutral Deontology," *Philosophical Studies* 170 (2013): 327-357 articulates an ostensibly agent-neutral deontology. I doubt that such a view is possible, largely for the arguments given above and for those advanced by Matthew Hammerton, "Is Agent-Neutral Deontology Possible," *Journal of Ethics and Social Philosophy* 12 (2017): 319-324.

[57] David McNaughton and Piers Rawling, "Value and Agent-Relative Reasons," *Utilitas* 7 (1995): 31-47 and "Agent Relativity and Terminology Inexactitudes," *Utilitas* 7 (1995): 319-25; Douglas Portmore, "McNaughton and Rawling on the Agent-relative/Agent-neutral Distinction" *Utilitas* 13 (2001):350-356; and Hammerton, "Deontic Constraints are Maximizing Rules."

[58] See Stephen Darwall, "Agent-Centered Restrictions from the Inside-Out," *Philosophical Studies* 50 (1986): 291-319, p.305.

[59] Darwall's claims echo some of Locke's claims about secondary qualities. Secondary qualities are, roughly, dispositions of particulars to cause certain perceptual reactions in creatures like us. For example, properties like being loud or being crimson are secondary qualities but the sound and light wavelengths associated with them are not. Whether secondary qualities exist is controversial. But many philosophers, such as David Wiggins, *Needs, Values, Truth* (Oxford: Blackwell, 1987) p.107; John McDowell, *Mind, Value and Reality* (Cambridge University Press, 1998), pp.112-150; Crispin Wright, "Moral Values, Projection and Secondary Qualities," *Proceedings of the Aristotelian Society, Supplementary Volumes 62* (1988): 1-26; and Hallvard Lillehammer, *Companions in Guilt: Arguments for Ethical Objectivity* (Palgrave McMillan, 2007), pp.89-110 have found it fruitful to analogize moral value with them. I also pursue this analogy above.

[60] I thank a referee for *Ethics* for pressing me to clarify my position here.