Keynote speakers:

Susan Wolf (Chapel Hill)
John Martin Fischer (Riverside)
Walter Sinnott-Armstrong (Duke)

Main Program Abstracts (alphabetical order):

1. Santiago Amaya (Washington U, St. Louis): “Blame Me! It was a Slip”

   Slips of action (Norman 1981, Reason 1984) are often viewed as accidents without significance. Because of the demands of social coordination, slips are often filtered-out and patched-up to appear reasonable (Baars 1992). Some slips, however, cannot be ignored: their consequences are catastrophic. For instance, when a parent fails to drop his child at the daycare and inadvertently leaves her in the back seat of the car while he is at work (Collins 2006, Weingarten 2009).

   Catastrophic slips give rise to attributions of blame. Yet these attributions run contrary to a principle that underlies widespread views on moral responsibility. I call it the **disclosure principle**: An actor is morally responsible for her actions/omissions/attitudes only if these reveal where she stands on questions of moral value. The principle is commonly associated with so-called real-self views (Frankfurt 1971, Scanlon 1998, Smith 2005, Watson 1975). But it also plays a role on certain Kantian views (Nagel 1986, Herman 1993). The argument of this paper is that the principle needs to be revised. The practice of blaming agents for their slips is part of the data that theories of moral responsibility ought to accommodate.

   I begin the paper introducing a definition of slips that I defend elsewhere. Slips are, on my view, actions contrary to a governing intention. When an actor slips, she sets out to do something. Yet without a change of mind, she ends up doing some action that upsets the intention that first prompted her to act. Slips, thus, involve a distinctive kind of omission: a failure to do something, which the actor had reason to do. The actor who slips behaves unreasonably, even by her own lights.

   Slips are best characterized as negligent mistakes. As such, they differ from other paradigmatic examples of blameworthy behavior. The actor doesn’t intend the bad consequences of her slip; she doesn’t regard them as a mere side-effects; she doesn’t even risk the bad outcome hoping that it does not materialize. To use legal terminology: the actor does not do wrong with intent, foreknowledge or recklessness. She acts voluntarily but, due to inadvertence, her behavior falls below a standard of reasonableness.

   In the second part of the paper I present my main contention against the disclosure principle. I argue that even though the principle does not rule out negligence, accounts of negligence compatible with it cannot accommodate blameworthiness for slips. I consider various accounts. The upshot is not that these views of negligence need to be abandoned. Rather, I conclude that the **disclosure** principle cannot be an overarching principle grounding attributions of responsibility.

   One widely held view, for instance, is that negligence is blameworthy to the extent that it involves a prior ‘benighting’ act: a decision not to acquire information, or to take precautionary measures, etc. (Smith 1983, Zimmerman 1986). What connects the benighting decision with the negligent mistake, on these views, is foreseeability. The subsequent inadvertence is blameworthy because it was within the risk of the benighting decision.
It is easy to see how this kind of account can construe negligence as a mistake that discloses the normative commitments of the offending actor. Yet, as I argue in the paper, it is hard to see how they can handle slips. When a slip occurs there is typically no prior decision with which the actor’s failure can be connected, except for the decision to do the action that she in fact failed to do. Further, in normal circumstances, slips are not foreseeable: all along actors are as confident as usual about the success of their undertaking.

Not all accounts of negligence, however, think of it as resulting from a prior decision. Thus, in the paper I also discuss views according to which negligence is blameworthy because it manifests a morally deficient character (Hampton 1990, Tadros 2005), or an indifference of the actor towards the consequences of her actions (Duff 1990, Simester 2000).

First, I argue that it would be implausible to claim that slips result from a bad character. Mistakes due to bad character traits tend to be repeatable, whereas the repeatability of a slip (after a certain point) threatens its status as a slip. Also, slips have an immediate cure that is typically not available for manifestations of character deficiencies: awareness of an imminent slip is sufficient to prevent it.

Second, I argue that it would be incorrect to think that slips result from some general indifference of the actor (even if temporary) towards the consequences of her actions. Indeed sometimes slips occur to people who are being meticulous at the time. Most of the time, however, the actor is as concerned with the consequences of her behavior as much as she usually is. The problem is that she is focusing on the wrong sort of consequences.

One way to challenge my argument against the disclosure principle is to deny that slips are mistakes for which the actor can properly be blamed. In the last section of the paper, I discuss one variant of this challenge: namely, reducing responsibility for slips to strict liability. My response is that the considerations that tend to explain why some behaviors give rise to strict liability (victim compensation, future deterrence, etc.) fail to provide convincing explications for the practice of holding actors responsible for their slips. In concrete, they fail to make sense of what seems to be the most salient feature of catastrophic slips: the discovery of the slip by the actor is usually accompanied by devastating pain and remorse.

In this paper, I argue against the disclosure principle. The principle is at the core of various theories of moral responsibility, notably, real-self views. It might seem that my argument leaves rival theories in a better position, for instance, theories that take control to be a necessary condition for responsibility. Yet to conclude this would be too simplistic. Because slips involve failures of control, it is likely that these rival theories also need further revision.

• Comments: Henry Pickford (Colorado)


Suppose we believe that certain autonomy-preserving resources should be preserved by members of each generation, and made available for acquisition for members of subsequent generations. We then need to distribute acquisition-rights to these resources among members of each generation. The objective of this article is to outline and defend a novel solution to this ancient, but pressing practical problem. I will argue that we should provide all agents substantial equal opportunity, in a certain sense, to affect the initial size of their share of the unowned resources through exercise of their choices. Agents’ opportunity to exercise their choices in entitling ways should be equalized. We should not pursue material compensation for congenital disabilities. Instead, we should arrange original acquisition of unowned resources in a way, which allows us to reduce the impact of these disabilities on agents’ ability to acquire shares of these resources.
DISPOSITION

The article is structured as follows: in section 2, I motivate the project and give the problem a precise formulation. In section 3, I clarify the key concepts used in the argument. In section 4, I present a simplified case of distribution within a generation, which will allow us to isolate and analyze some widely shared, as well as some less recognized, intuitions regarding resource distribution. In section 5, I discuss some adequacy criteria for a plausible theory of distributive justice in original acquisition. In section 6, I introduce my proposal. I argue that there should be several consecutive opportunities for original acquisition of unowned resources for each generation. I also argue that there are strong, but not conclusive, arguments for the claim that the rules for original acquisition at the first occasion should be different from the rules for original acquisition at the subsequent occasions. In the course of the argument, I introduce a novel account of equality of opportunity. I argue that we should not pursue initial equality of resources or welfare, nor should we pursue initial equality of opportunity for resources, welfare or advantage. Instead, we should pursue initial equality of opportunity to make entitling choices. Agents should be given substantial equal opportunity to affect the size of their initial share of unowned resources by making certain entitling choices. The benefit of this account is that it is more sensitive to individual responsibility than contemporary accounts of initial equality of opportunity, while “levelling the playing field” in an equally plausible way. In section 7, I discuss the application of my proposal to future generations. In section 8, I state and discuss some remaining challenges to my proposal.

ASSUMPTIONS

I assume, for the sake of the argument, that agents have obligations to preserve some autonomy preserving resources for future generations. I do not here take a stand on whether these obligations correlate to rights. Initially, I will assume that autonomy-preserving resources are available for initial acquisition by each generation. Later, I will argue that if we adopt the distributive principles defended in this article, it is likely that such resources will be available for each generation. I will not discuss the non-identity problem. I will assume that the identity of future persons remains the same regardless of whether the principles defended in this article are applied or not. I only discuss rights-based accounts of justice in original acquisition. I do not defend these against other normative theories of distributive justice. Neither do I discuss whether mere justice in original acquisition, possibly followed by necessary redress, by itself preserves distributive justice in a wider sense.

SETTING THE STAGE

Suppose a population consists of agents A and B. A and B want to divide a divisible, unowned resource R among themselves. A and B are autonomous adults, whose upbringings have been identical in all morally relevant respects. They only differ regarding some of their congenital capacities, which they cannot alter. A and B are capable of realizing a certain number of opportunities each, by means of their inherent capacity, and the resource R. The relevant method for individuating and measuring the opportunities available to A and B is introduced later on in the article. Each of these opportunities is equally valuable to them. This assumption helps us focus the discussion only on increases and decreases of A’s and B’s opportunity sets, that they consider beneficial or adverse to them. The difference in ability is only caused by genetic differences, for which neither A nor B is causally or morally responsible. A and B only consider unilateralist theories of distributive justice in original acquisition. They do not consider theories, which require agents to collect permission from other agents before acquiring unowned resources. A and B have full knowledge about all the factors mentioned above. These assumptions are controversial, but necessary in order to demarcate the problem. If my account is plausible under these assumptions, its application to more realistic and complex circumstances
merits exploration. Such exploration would include discussions regarding how to implement the proposal globally.

**THE PROPOSAL OUTLINED**

1. There should be several consecutive opportunities for original acquisition of unowned resources within each generation. Unowned resources should be saved in order to be made available for acquisition at each of these occasions. For brevity, I will not discuss how many such occasions there should be. Suffice to say that the number of occasions may be morally important, and should be discussed in a full expose of the theory introduced here.

2. The size of each agent’s share of these resources should be determined by:
   - At the first occasion: the size of the opportunity set available to him/her, *iff*: he/she had exclusive access to all the resources to be distributed at this occasion, or in accordance with some prominent liberal egalitarian or left-libertarian theories,
   - And:
   - At the remaining occasions: The market value of the unowned resources he/she previously has preserved for future generations.

In summary, I am introducing a novel, non-coercive, responsibility sensitive global fund for unowned resources, which “levels the playing field” in original acquisition.

- **Comments:** Jamie Hardy (Utah)

3. **Chrisoula Andreou (Utah): “Self-Defeating Self-Governance”**

   What is it to be self-governed? This is a difficult question and one that has been the subject of a great deal of thoughtful and interesting debate. I will not attempt to directly contribute to that debate here, though I do hope that what I say will be consistent with the most influential views of self-governance or else will be plausible enough to prompt some rethinking of these views. My aim in this paper is to initiate and contribute to debate concerning the possibility of behavior that is both self-defeating and self-governed.

   In the first section of the paper, I will review a couple of points that figure in the literature as platitudes about (the relevant notion of) self-governance. As the idea of self-governance has been understood, one can lack self-governance even if one is not being controlled by some outside agent. For, according to the philosophically rich notion of self-governance at issue, one does not count as self-governed if one is governed by something with which one does not identify or which one finds alien, and this includes desires that one could honestly disown, even if they have not been implanted by another. A classic case of the relevant sort is the case of the unwilling addict, who finds herself with a desire for a drug but does not identify with this desire, and indeed sees fighting the desire as essential to expressing her true self. Different views of self-governance flesh out this skeletal description in different ways, but one point that seems to figure as common ground is that if one is self-governed, one is governed by one’s commitments and values.

   In the second section, I will explain how these points give rise to what seems to be a dilemma that suggests that informed self-defeating behavior, wherein one understands the consequences of each choice one makes, is impossible. Like self-governed behavior, self-defeating behavior—at least in the strict sense that concerns me here and that is, in my view, crucial to discussions of diachronic rationality—must be such that it can be traced to the self, rather than to something alien to the self. Relatedly, in self-defeating behavior, the one being defeated is the same as the one doing the defeating. It is easy to engage in self-governed, self-defeating behavior if one is misled by false beliefs about the consequences of some of the choices one is making. But, putting aside behavior that results from simply being misled by false beliefs, it seems like self-governed behavior cannot be self-defeating. For, consider the following line of reasoning:
If one’s behavior is self-governed, it is governed by values and commitments with which one identifies and which one does not find alien; and so, one’s behavior is true to oneself. Assuming it is based on an understanding of the consequences of each choice one makes, behavior that is true to oneself does not also defeat this same self, and so self-governed behavior is not self-defeating.

We are thus left with a dilemma that seems to undermine the possibility of informed self-defeating behavior. On the one hand, given that self-defeat involves being defeated by oneself, if some behavior is not self-governed, it cannot be self-defeating. On the other hand, given that self-governance involves being true to oneself, if some behavior is self-governed, it cannot, assuming it is informed, be self-defeating. So it seems that, if informed, neither self-governed behavior nor behavior that is not self-governed can be self-defeating.

One might be willing to loosen up the notion of “self-defeat” and count the classic case of the unwilling addict as a case of self-defeat, but here I want to see if we can get any closer to vindicating the possibility of behavior that is self-defeating in a stronger sense—a sense in which the desires and behaviors that are defeating one’s true self are ones with which one identifies. The possibility of self-defeat in this stronger sense is, I think, inextricably tied to the existence of certain diachronic rationality constraints. I will move away from the classic case of the unwilling addict and consider, in the third and fourth sections of the paper, two other types of cases in which it is tempting to say that the agent’s behavior is self-defeating. The first case, which involves discounting-induced preference reversals, gets us closer than the classic case of the unwilling addict to a case of strictly self-defeating behavior. The second case, which involves slippery-slope intransitivities, arguably get us all the way there. If my reasoning is correct, strictly self-defeating behavior, which can also be described as self-defeating self-governance, is, contrary to initial appearances, definitely possible.

I’ll end with some remarks concerning self-governance, self-defeat, and dynamic choice. As is explained in my paper, reflecting on complications associated with how choices accumulate over time allows us to see how self-defeating behavior, understood as self-governed, is possible. The challenges of choice over time, even when one is not changing over time, leave room for behavior that is faulty with respect to being true to oneself without involving false beliefs about the consequences of each choice one makes or a failure of self-governance. It follows that governance over time by a stable self is not sufficient for the avoidance of informed self-defeating behavior. It is also not necessary, since, if one is deeply fragmented, self-defeat, where the defector and defeated are one and the same, may never occur. It is clear that the avoidance of informed self-defeating behavior is rationally required and argues that fragmentation into different selves over time does not itself open one up to criticism; indeed, insofar as some amount of unity is necessary for the attribution of self-defeat, deep fragmentation may make one less susceptible to criticism from the point of view of diachronic rationality. In any case, avoiding self-defeating behavior is consistent with radical transformation over time and the call to avoid self-defeating behavior is not a call for conservatism with respect to the values that define one’s current self.

• Comments: Hannah Love (Pacific Lutheran)


An autonomous person is self-governing. Joel Feinberg expresses the idea at the heart of autonomy when he quips: “I am autonomous if I rule me, and no one else rules I.” My own judgments (evaluations, decisions) govern my actions. But are all my judgments equally my own? Different conceptions of autonomy call for different answers because they serve different purposes; no single conception of autonomy can do everything it is often pressed to do. Most contemporary ethicists are concerned with autonomy as a condition of free will and responsibility, or its role in a person’s social or political status, or its role in subjective well-being.
Accordingly, they seek to provide value-neutral accounts of autonomy in terms of certain structural or historical features of the self. On these conceptions of autonomy, it is perfectly coherent to say not only that we can autonomously acquire slavish desires, but also that we can autonomously live slavishly, so long as these desires are a part of our character, or so long as we retain the ability to periodically endorse the choice to live thus.

I argue to the contrary that a life of slavishness, servility, conformity, blind obedience, or self-effacement cannot be substantively autonomous, even if the choice to live thus is autonomously renewed periodically. The very description of the lives of these heteronomous individuals marks their difference from those whose daily life is textured by their own independent judgments. Holding otherwise implies, absurdly, that the character and actions of an independent, self-reliant individual are no more autonomous than those of an individual who willingly goes through life doing others’ bidding.

Why should we care about having substantive autonomy – autonomy as a character trait? Those few philosophers who discuss substantive autonomy typically answer that we should care because it is instrumentally useful and because it is a non-moral excellence, a condition of a personally worthwhile life. Presumably, this means that autonomy is something admirable, like scientific intelligence or musical virtuosity. I believe, however, that autonomy cannot meet the conditions for being admirable without being morally admirable. Attempts to argue otherwise seem strained and are, ultimately, unsuccessful.

On my view, a person is autonomous insofar as she is intellectually and emotionally disposed to direct herself by her own reality-oriented judgments. Further, autonomy understood thus is morally admirable. By extension, a judgment or action is autonomous and admirable just in case it is an expression of a person’s autonomous dispositions or motives. To be reality-oriented is to be disposed to track truth or understanding and act accordingly in important (non-moral) areas of one’s life and human life in general, such as relationships of love, work, and play. The requirement that an autonomous person must be self-directed in important areas is implicit in the idea of autonomy as a character trait. Both shameful and admirable trait terms implicitly carry a reference to the important.

In Section II I argue for the claim that autonomy requires reality-orientation as follows. (i) All forms of heteronomy (servility, conformity, impulsiveness, etc.) display an absence of reality-orientation; and (ii) the features usually attributed to autonomous people – independence of mind, self-reliance, self-possession, etc. – imply that the judgments by which they direct themselves must be reality-oriented. Such judgments are more “their own” than others. Further, when we consider what is required to have these qualities, we can see that autonomy is morally admirable, a virtue, in the Aristotelian sense of a disposition to think, feel, and act “at the right times and on the right occasions and towards the right persons and for the right causes and in the right manner”. We can also see why heteronomy is a vice and implies other vices. The example of literary characters such as Fyodor Karamazov (The Brothers Karamazov) and Elizabeth Bennett (Pride and Prejudice) makes these points more vivid.

Like other virtues, autonomy can vary in scope, and there is no precise formula for how self-directed a person must be to count as autonomous - other than that she must be autonomous in most important matters. An autonomous person must also care about being autonomous and, thus, about her lapses of autonomy. But how much she cares does not allow of a precise formula. Here, as elsewhere, there is a range of acceptable attitudes, and determining the range is a matter of practical wisdom.

In Section III I expand on the idea of reality-orientation, arguing that, just as autonomy entails reality-orientation, so reality-orientation entails autonomy. The main reason is that practical understanding cannot be had entirely second-hand, that is, on the basis of testimony; we must also exercise and be guided by our own powers of perceptiveness, imagination, empathy, and reasoning. But the autonomous person is not a lone wolf who never relies on others’ judgments in any matter of importance. Understanding and accepting the limits of our powers as human beings and individuals, and the consequent necessity of relying on others’ testimony, is just as much a part of
being reality-oriented as understanding the limits of reliance on others and the importance of being autonomous. Reliance on others’ testimony is compatible with autonomy when others are more likely to be right or when practical demands and time-constraints make such reliance necessary - as is frequently the case. It is incompatible with autonomy when, e.g., we are too lazy to make the effort to think for ourselves, or too neurotically self-doubtful to trust our own judgment, or too desirous of approval to care enough about the truth. All these motivations involve de-valuing self-direction.

In Section IV I consider the view that autonomy requires the ability for questioning and justifying one’s basic attitudes and values, including autonomy itself. I believe this is an overly-intellectualized conception of autonomy, an expression of philosophical parochialism. Autonomy is not limited to people with philosophical abilities.

• Comments: Elizabeth Foreman (Saint Louis)

5. Paul Baer (Georgia Tech): “Balancing Responsibility and Capacity in the Allocation of Climate Obligations”

Many of the philosophers engaged with the question of burden-sharing in the climate regime have concluded that obligations should be proportional in some fashion to responsibility (contribution to climate change) and capacity (ability to pay), a conclusion which we share. Yet justifications for particular ways of combining these principles are largely absent, with particular burden-sharing formulae being based on relatively ad-hoc criteria.

The issues here are complicated by the ambiguity of the location of obligation between individuals and states. The practical politics of the climate negotiations require obligations to be attributed to states; yet it is not obvious that obligations calculated on the basis of national characteristics and obligations calculated on the basis of individual characteristics (or aggregated individual characteristics) would be the same. In this essay, we explore some of these complexities in the justification of climate-related obligations, with an eye to identifying “reasonable” ways for combining multiple principles.

Because part of the difficulty in reaching consensus on a fair distribution of climate obligations is the multiple perspectives brought to the problem, we start by identifying several of the basic (and not mutually exclusive) paradigms that have been applied. Among these are:

1) The resource sharing paradigm. In this view, the global sinks for GHGs represent a scarce resource, and the problem to be solved is a fair allocation of this resource. Advocates of this perspective typically begin from the egalitarian argument that, because the global carbon sinks are a “global commons”, prima facie there should be equal entitlements. An important subject for dispute is whether unequal historical use requires current restitution – whether there is a “carbon” debt, and if so how it should be calculated. This raises numerous questions about fault, ignorance and choice.

2) The burden sharing paradigm. Under this view, the reduction of emissions represents a cost to be shared, and this view has driven the debates about “comparable effort” among the industrialized countries, but it is difficult to generalize because plainly some persons, and arguably some countries, are so poor that they deserve to be fully exempted from any emissions mitigation burdens; here the role of rich people in poor countries looms large.

3) The right to development paradigm. In this view, developing countries have a right not to have their development (typically, GNP growth) limited by the need to reduce their growth in energy consumption; only emissions reductions that produce greater “co-benefits” or that are subsidized by developed countries are warranted until some (unspecified) level of national welfare has been reached. Again, this brings focus to the obligations of rich people in poor countries.

4) The economic competition paradigm. In this view, all countries, whether rich or poor, are fundamentally involved in a zero-sum game of competition for markets (and, by extension, for
The competition is presumed to be fair by default, and thus the introduction of asymmetric regulations represent an unfair burden on affected industries, workers and countries. Here the focus turns in particular to affected workers in rich countries.

Our approach is to take each of these paradigms and (first) consider the direct application of these principles to individuals and countries, looking at a variety of prototypical cases, to see where they produce conflicting recommendations; (second), to try to see what disputed premises lead to these conflicting recommendations, and in particular whether they can be identified as “factual” premises or “value” premises; and (third), to see whether there are any “metaprinciples” that can be appealed to in hopes of resolving these conflicts. Our broad strategy is to seek an ethically robust solution, which could be claimed to pass a test of “not being reasonably rejected” or accepted behind a “veil of ignorance”.

As we suggested at the beginning, our assertion is that any such solution must combine aspects of responsibility and capacity, but further that it must more generally respond to concerns about fundamental interests; following Simon Caney, we hold that no climate obligations can accrue to those who do not have currently have acceptable protection of their fundamental interests to physical security (including health and nutrition), or who would be pushed below some related threshold by any such obligations.

Beyond that minimal proviso, we suggest that an acceptable solution needs to emphasize minimizing welfare losses – by imposing burdens on those with the greatest wealth – and also allowing opportunity for those who have not so far made significant use of the global commons. We experiment with some alternative formulae for a “progressive climate tax,” again using prototypical examples to see whether there are any easily defensible limits of “too progressive” or “not progressive enough.”

We conclude by assessing the relationship between individuals and nations. We suggest that nations are not merely collections of individuals who happen to be represented in the negotiations by a single government; a substantial part of the emissions profile of a nation represents policy choices, while other parts represent geographical endowments. The structure of international negotiations requires that these factors be somehow accounted for, and that national characteristics influence national obligations. Yet we hold in the end that an ethical approach to fair global policy requires fair treatment of individuals, and thus that states be presumed to apply within their own borders the principles of fairness that they would have hold globally. In brief, this requires that rich people in poor countries have comparable obligations to rich people in rich countries, with exceptions only that are of direct and demonstrable benefit to the poorest citizens.

- **Comments:** Kelly Heuer (Georgetown)

6. **Anne Baril (Arizona): “Eudaimonism and Conditionality”**

Eudaimonism addresses the question ‘how should I live?’ and finds the answer in eudaimonia- human flourishing. We ought (the eudaimonist alleges) to organize our lives so that they realize eudaimonia, and we do this (the eudaimonist virtue ethicist alleges), at least in large part, by developing and exercising the canonical moral virtues.

Recently EVE has been developed, and endorsed, as providing an alternative to contemporary normative ethical theories, such as utilitarian and deontological theories (rather than as somehow being compatible with such theories). The main addition to EVE, when it is understood as a contemporary normative ethical theory, is an account of morally right action. Rosalind Hursthouse’s account is representative. Hursthouse proposes that:

1. An action is right iff it is the action the virtuous person would characteristically perform, and
2. The virtues are just those traits needed for eudaimonia.
An objection to this account of right action is that it gives the wrong kind of explanation of *why* an action is right. If the virtues are just those traits we need to flourish, and flourishing is something that is necessarily and essentially good for the agent, then there is a troubling sense in which an agent’s action is *ultimately* right, not in virtue of the features of others (e.g. their autonomy, their sentence) but because so acting is good for the agent herself. Such an explanation of moral rightness, it is argued, seems unacceptable—‘unacceptably egoistic’, it is sometimes said in the literature (though I avoid using this term in this paper). Call this the WKE objection.

In this paper, I consider two of the most promising extant replies to this objection (Toner 2000 and LeBar 2009). I argue that these replies attempt to address the WKE objection by building virtue, or direct regard to others, into the very formal conception of eudaimonia, but that, in so doing, they give up on what is most attractive about eudaimonism.

The ‘formal’ conception of eudaimonia comprises that set of features which, together, serve as constraints on any proposed substantive conception of eudaimonia. The main such constraint is the requirement that eudaimonia be complete: worth pursuing for its own sake, while everything else I pursue, even if pursued for its own sake, will also be appropriately pursued for the sake of eudaimonia. But there are other constraints placed on a substantive conception of eudaimonia (though they are not always explicitly stated). One is that eudaimonia is a property of complete lives, not of individual moments. Another (and this is the constraint that gives rise to the WKE objection) is that the eudaimon life is something that is good for the person whose life it is.

Extant replies to the WKE objection, I argue, work by arguing that virtue (or our direct concern for others) is part of the formal conception of eudaimonia—not a part of a substantive conception of eudaimonia, but a constraint on any substantive conception. In this way, eudaimonists avoid saying that our direct concern for others is somehow ‘conditional’ on facts about our flourishing.

The problem with this kind of reply is that, to the extent that the eudaimonist ‘builds up’ this formal conception of eudaimonia, to make it into something that *would* be appropriate as the core of our narrowly-moral obligations to others, the eudaimonist gives up on what is essential to—and, I think, most appealing about—eudaimonism. Eudaimonism’s insight is that an individual must organize her life around which is “capable of organizing and focusing all the concerns and aims of my life as a whole”, of “encompass[ing] everything worthwhile into my life” (Annas 1993: 83). To make virtue part of the very notion of eudaimonia—to make it a constraint on any substantive account—would address the objection that eudaimonism offers the wrong kind of explanation of our obligations to others, but would do so at the price of no longer being a eudaimonist view. For to say that we must organize our lives around ‘attachment to’ (or direct concern for) others is no longer to say that, when it comes to the question of how, at the most basic level, to guide and organize my life—I *must live well*.

Extant replies to the WKE objection, then, are not acceptable as defenses of EVE accounts of morally right action. Yet my ultimate conclusion is not negative: I think there is an attractive route open to EVE, one that preserves what is most appealing about the view without falling prey to the WKE objection, and I close by briefly noting it. The route I suggest is that of understanding EVE as offering only an account of what how we ought, *all-things-considered*, to organize our lives, and not as offering an account of the “peculiar *moral* sense of ‘good’ and ‘duty’, which is [in modern times seen as] the proper object of ethical study”. (von Wright 1963: 1) The ‘ought’ of EVE may be understood as only an ‘all-things-considered’ ‘ought’, and not as a distinctively moral ‘ought’ (as opposed to both). So understood, EVE is in a better position with respect to the WKE objection than EVE when it is understood as a contemporary normative ethical theory. This is because it is our *distinctively moral* obligations that a eudaimonist explanation seems to get wrong, and, on the present proposal, EVE does not purport to give an account of these obligations. On the present proposal, EVE says that the appropriateness of accepting a system of morality (including requirements that we have a direct regard for others) is
conditional on the acceptance of that system being part of human flourishing. No doubt there will still be those who find this kind of conditionality objectionable—indeed, there are those who would find any kind of conditionality objectionable. But the proposed view seems to preserve what I think is an important insight of eudaimonism without giving an unacceptably egoistic explanation of our narrowly moral obligations to others.

- Comments: Nandi Theunisson (Columbia)


Recently, Michael J. Zimmerman has claimed that “When Auschwitz camp commandant Rudolf Höss had over two million people put to death, he was not to blame. When Adolf Eichmann delivered victim after victim to the concentration camps, he was not to blame. When William Calley led the massacre of hundreds of civilians at My Lai, he was not to blame.” Zimmerman argues that none of these notorious historical figures is likely to have ever satisfied the conditions of morally culpable ignorance. Since these wrongdoers were not morally culpable for the ignorance that led each to commit such atrocities, none of them can be blameworthy for the morally odious conduct each displayed.

In this paper, I argue that Zimmerman’s argument fails because his account of the conditions of morally culpable ignorance is inadequate. On his view, ignorance is morally culpable only when an agent has intentionally cultivated his ignorance by way of some prior voluntary action the agent culpably performed. This necessary condition of morally culpable ignorance incorporates two principal requirements. First, from the perspective of the agent at the time he performed this prior voluntary action, the resulting state of ignorance must have been explicitly foreseen. Second, the agent must have occurrently believed, at the time he performed this prior voluntary action, that he was acting wrongly. This account of the conditions of morally culpable ignorance is then attached to the claim that if an agent commits wrong out of ignorance, he is blameworthy for the wrong committed only if he is also morally culpable for the ignorance that produced it.

Zimmerman uses this account, together with some plausible empirical assumptions, to defend his claim that Höss, Eichmann, and Calley are not blameworthy for the horribly wrong actions they performed and the morally repellant practices in which they actively participated. It's plausible to suppose that they were all largely ignorant of the wrongness of their conduct. It's also plausible to suppose that none of them ever previously voluntarily performed actions that satisfy Zimmerman’s conditions of morally culpable ignorance. Instead, it is much more plausible to assume that none of these men voluntarily cultivated their respective states of ignorance. If Zimmerman’s account of the conditions of morally culpable ignorance is correct, then this more plausible assumption apparently entails his claim that neither Höss, Eichmann, nor Calley were blameworthy for the wrongs they committed out of ignorance.

I argue in favor of an alternative theoretical account of morally culpable ignorance. My argument is motivated by a desire to provide a plausible explanation of the conditions of morally culpable ignorance that preserves the intuition that at least some agents, especially agents who commit such atrocious moral wrongs as Höss, Eichmann, and Calley, can be blameworthy for wrongdoing committed out of ignorance that they nevertheless didn't voluntarily cultivate. I believe that Zimmerman is right to claim that most people never voluntarily engage in projects of cultivating ignorance, yet that this fact can be shown to be consistent with the claim that at least some agents who act out of ignorance that is not voluntarily cultivated are nevertheless blameworthy.

To develop my argument, I begin by inquiring why a voluntary choice to act in a way that one both foresees may lead to subsequent ignorance and occurrently believes to be wrong appears blameworthy in the first place. Such a choice appears blameworthy, I claim, because it
unequivocally reveals the agent’s faulty evaluations about the importance of certain moral reasons. Everyone has moral reason to avoid moral ignorance, and an agent who chose to cultivate moral ignorance would be a paradigmatic example of someone whose evaluations of the significance of the moral reasons he has are flawed. So this kind of choice is culpable, I contend, not merely because it is a choice, but because of the underlying faulty evaluations of moral reasons this kind of choice reflects.

But any adequate understanding of moral ignorance must surely grant that moral ignorance may sometimes be, in ordinary adults, itself constituted by certain faulty evaluations of the force of certain moral reasons. For example, Eichmann plausibly held certain horrifically mistaken evaluations of the moral importance of advancing the Nazi Final Solution. If I am right to claim that moral ignorance is sometimes constituted by certain faulty evaluations of moral reasons the agent holds, then it seems reasonable to claim that what makes a given state of moral ignorance culpable is just that it reflects these underlying evaluations of moral reasons, and not necessarily that it is indicative of any wrongful choices the agent may (or may not) have made in the past. Such wrong choices are targets of moral appraisal only because they reflect an agent’s faulty underlying evaluations of moral reasons. If things other than voluntary choices can directly reflect an agent’s faulty underlying evaluations of reasons, then these other things are also capable of opening an agent up to negative moral assessment, independently of whether these other things are owing to any particular voluntary choices the agent putatively made.

My account of the conditions of morally culpable ignorance is superior to Zimmerman’s because it more adequately explains our general intuitions about when ignorance does and does not excuse an agent from blameworthiness. Zimmerman’s account is radically revisionary, because on the very plausible assumption that most wrongdoing is committed out of unforeseen and unchosen ignorance, his account may entail that agents are never blameworthy for most of the wrongs they commit. The mistake Zimmerman has made is in thinking that there is something fundamental about the notion of a voluntary choice for blameworthiness. My argument claims that evaluative judgment is fundamental for moral appraisal, and that when moral ignorance directly reflects an agent’s underlying evaluative judgments, agents may be morally culpable for such ignorance even when such ignorance cannot be explained in terms of any prior choices the agent has made. If I am right, then Zimmerman’s claim that Höss, Eichmann, and Calley are not to blame lacks adequate justification.

- Comments: Matt Bedke (British Columbia)


Should defenders of libertarian free will think that “mere influence” mitigates moral responsibility?

Suppose a person S is subject to factors - arising, say, from her genetic endowment or early family environment - over whose existence S had no control, and which strongly incline S to decide to perform some morally wrong action A. Suppose that this influence is not so strong as to make A literally unavoidable for S, but is strong enough that without the presence of those factors, S would very likely refrain from doing A. Suppose further that these factors work by affecting what we might regard as "internal" features of S (e.g. S’s desires, or S’s ability to resist desires, or S’s values, or S’s ability to alter her desires and values, etc.), rather than by affecting merely "external" features of S’s situation (e.g. by providing an opportunity to do A, or by affecting S’s physical abilities or skills relevant to A.) Finally, suppose that S does A. Should the presence and operation of influences of this sort mitigate S’s moral responsibility for A? That is, should we say that S is less blameworthy for doing A than she would have been had she done A in the absence of such factors? Is she partly excused for doing A (at least compared to someone else who does A without being influenced), or, given that she could have refrained from doing A, is S fully responsible for doing A?
Peter van Inwagen (PVI) has defended a negative answer to these questions in his article “Genes, Statistics, and Desert.” He argues for what I shall call the “Dichotomous Attribution of Mitigation” (or DAM) view — the view that, with a few qualifications, moral responsibility is “all or nothing.” Either the influences on an action are so strong as to render someone incapable of avoiding wrongdoing (in which case she is excused from blameworthiness), or, despite being perhaps very strong, the person is still able to avoid wrongdoing (in which case her responsibility is not mitigated at all by the presence of the strong influence). Against PVI, I shall defend what I’ll call the “Matter of Degree” (or MOD) view — the view that a person’s moral responsibility for an action is a matter of degree which depends on, among other things, the degree to which he or she was able to avoid wrongdoing, and that influences of the sort in question (genetic predispositions, early moral training, etc.) can, and often do, mitigate blameworthiness.

In the first part of the paper, I shall summarize PVI’s argument for DAM. In brief, he suggests that cases involving “external” influence (factors which affect such things as skills or opportunities, but not motivational or volitional (“internal”) features of agents) do not warrant mitigation; but, he continues, there is no morally relevant difference between external influences and “internal” influences; so the latter do not warrant mitigation either.

In the second part of the paper, I shall identify two main strategies for replying to PVI’s argument. The first strategy targets the “no relevant difference” premise; I shall suggest three differences between external and internal influences that are plausibly thought to be morally relevant. The second strategy involves offering arguments for the conclusion that, even if one concedes that there is no fundamental or deep difference between external and internal influences, such influences do mitigate responsibility. I shall offer three such arguments. The first aims at producing a Sorites-based objection to the claim that only irresistible desires (or other features) reduce or remove responsibility. The second argument turns on a problematic asymmetry between praiseworthiness and blameworthiness to which PVI’s position seems committed. Finally, the third argument focuses on a set of core libertarian commitments whose conjunction stands in tension with DAM.

For instance, DAM requires the view that, so long as an influence (call it F) was not deterministic or irresistible, then its degree of strength — just how resistible it was — should play no role in assessments of moral responsibility. One might think that such a view is defensible in light of what we might call the “Ultimacy Condition” — the fundamental libertarian intuition that moral responsibility requires ultimate origination, or sole authorship, or a power that makes agents causal buck-stoppers. If F was resistible, then as we trace the causal chain backward from the action, we will not simply pass through the agent back to F; instead, the chain will simply terminate at the agent. And this (the libertarian concludes) is why we can say that the agent deserves blame.

But this picture — and the “Ultimacy Condition” — needs significant refinement: as Timothy O’Connor has put it, we are not “Unmoved Movers,” but rather “Not Wholly Moved Movers.” We are partly moved, even when acting freely, by factors beyond our control. But then how much we are moved by factors beyond our control — “our own” contribution vs. the contribution of those factors — seems importantly relevant to assessments of moral responsibility.

Another typical libertarian commitment is that character contributes to action. Indeed, on a “restrictivist” view of freedom, character can even come to wholly determine (some) actions. This latter view is often paired with “historical” or “tracing” accounts of moral responsibility which consist, in part, of the claim that one’s character is at least partly formed by prior free actions. But here again, it seems obvious that a number of features of these relations between character and action will be matters of degree. I shall argue that certain kinds of influence (by factors beyond one’s control) can make a given action far less (or more) resistible, and that it is the relative moral effort or “effort of will” required and expended that should ground judgments of responsibility and mitigation.

In the final section of the paper, I shall gesture toward a general account of moral responsibility mitigation, one that accommodates the commitments explored in the previous
section, and which grounds degree of responsibility in degree of control, rather than in the
dichotomous presence or absence of alternative possibilities.

- Comments: Georgi Gardiner (Edinburgh)

9. **Matthew Braddock (Duke): “Defusing the Demandingness Objection”**

The first section of the paper introduces the demandingness objection:

- (premise) Moral view X is too demanding.
- (premise) If a moral view is too demanding, it is mistaken.
- (conclusion) Therefore, moral view X is mistaken.

Objections of this sort have all sorts of targets. Given the wide scope of such objections, it
appears that nearly everybody—both proponents and critics of the targeted views—has a
philosophical stake, if not a practical stake, in the success or failure of the demandingness
objection. But what exactly is the demandingness objection at issue in the literature? The
schema above is underspecified—“too demanding” in what sense?—and could be filled out in a
number of ways. And philosophers have fleshed out a great variety of demandingness
objections, some of which narrowly target some form of act consequentialism and some of which
target a wider scope of views.

Given the philosophical space of objections, the question naturally presses: what is the
most formidable and philosophically interesting sort of demandingness objection? This paper
does three things: (i) it clarifies the structure of the various demandingness objections on the
market, (ii) develops a very formidable and philosophically interesting demandingness objection
that targets a wide scope of views, and (iii) challenges this formidable objection with an
argument the form of which may, interestingly, be effectively deployed in other areas of
philosophy.

The second section of the paper develops a formidable demandingness objection. It
begins by briefly sketching prominent demandingness objections in the literature in order to
distinguish them from my target set, namely a family of objections that essentially press that
some moral views are mistaken because compliance with their demands costs too much. Since
cost has various dimensions, the targeted set of objections may be distinguished along various
dimensions of cost—in terms of the sacrifice of well-being, subjective interests, projects,
and objective goods. Then I contend that since (i) implausible versions of this form of objection
mistakenly estimate costs and rely on controversial theoretical premises and (ii) since versions of
this objection that identify costs that are impossible (nomologically, psychologically, etc) to bear
only challenge a rather narrow scope of moral views, the most formidable and interesting
version of the objection runs as follows:

**Intuitive Demandingness Objection Schema**

1. Moral view X imposes obligations that require compliant agents to sacrifice G, where G is not
   impossible to sacrifice and consists in
   - our well-being
   - our well-being below some threshold of a decent life,
   - our subjective interests,
   - our projects,
   - and/or certain objective goods

2. If a moral view imposes obligations that require compliant agents to sacrifice G, it is intuitively
too demanding in the sense of intuitively mistaken because it demands too much sacrifice from
compliant agents.

3. Therefore, moral view X is intuitively too demanding.
Since such arguments get premise (1) right and are valid, I take up the critical task of challenging instances of premise (2). Even if my challenges ultimately do not persuade, I hope to have at least identified a formidable and philosophically interesting objection that moral philosophers should take seriously.

In the third section, I challenge (2) with an argument from unreliability.

**OBJECTION FROM UNRELIABILITY**

1. Our culture’s expectations, our partiality bias, and/or our socioeconomic status (partly) plausibly cause our demandingness intuitions.
2. Whether our culture’s expectations, our partiality bias, and/or our socioeconomic status cause our demandingness intuitions does not affect the moral facts about demandingness.
3. The formation of our demandingness intuitions on the basis of our culture’s expectations, our partiality bias, and/or our socioeconomic status is an unreliable process.
4. There is no good reason to think other causes of our demandingness intuitions are reliable.
5. If the known causes of our intuitions are unreliable and there is no good reason to think that other causes of those intuitions are reliable, then there is good reason to think those intuitions are unreliable.
6. Therefore, there is good reason to think our demandingness intuitions are unreliable.
7. If there is good reason to think our demandingness intuitions are unreliable, then we are not justified in holding them.
8. Therefore, we are not justified in holding our demandingness intuitions.

I support each premise but give most of the space to the defense of (1), (3), (4), and (7). My paper concludes by contending that if my argument from unreliability against premise (2) is successful (or successful upon slight modification), its form interestingly might also be deployed against many other views in philosophy that rely on intuitive premises. And intuitive premises abound in philosophy, so suitable instances of these argument forms might do some heavy-lifting. Of course, there will be objections. I address some of them in the longer version of the paper and am eager to confront more.

• **Comments:** Tom Dougherty (MIT)

10. **Danielle Bromwich (NIH) and Joseph Millum (NIH): “Fraud and Consent to Medical Research Participation”**

Valid informed consent is widely regarded as an ethical requirement for enrolling competent adults into most clinical research. The process of obtaining informed consent includes both the disclosure of information about the research to prospective participants and their understanding of information about the research. The standard view of disclosure and understanding holds, first, that a lot of information must be disclosed to prospective participants, and second, that all the information disclosed must be understood. However, a wide range of studies of people enrolled in clinical research have shown that many otherwise competent research participants do not appear to understand much of the information they have been told. The studies therefore imply that if the standard view about what research participants have to be told and understand are correct, then many research participants fail to give valid informed consent to being in medical research. And, if that is right, then we seem forced to accept that many clinical trials which we thought were ethical are in fact unethical.

In this paper we challenge the standard view of the relationship between disclosure and understanding. We argue that while it is true that a great deal of information should be disclosed to research participants, they do not need to understand everything that ought to be disclosed to them—they must just be given a reasonable opportunity to understand. We therefore conclude that while the data on participant understanding should motivate changes to how information
about research is disclosed, the number of unethical research studies is much smaller than the standard view would imply.

The two most common accounts of what people need to understand in order to give valid consent to being in clinical research are the interest and inducement views. According to the interest view, prospective research participants must understand those aspects of the research that may affect their interests or well-being. According to the inducement view, prospective participants must understand those aspects of the research that would be relevant to their decision about whether to enroll. Depending on one’s account of interests, the views may differ in what they require participants to understand; however, both require that a great deal of information should be understood. Since this information is unlikely to be common knowledge, it follows that a great deal of information must be disclosed, too.

Any plausible account of the content of the disclosure or the understanding requirement must be able to explain why not disclosing some proposition or not understanding some proposition invalidates consent. We therefore approach the question of what must be disclosed and understood for consent to be valid by considering the different ways in which it can be invalid. Within the domain of research ethics two of these ways—coercion and incompetence—have been widely discussed. We identify another important way—fraud—that has been addressed in other contexts, such as contract law, but largely neglected in clinical research. According to the account we develop, fraud occurs, roughly, when someone voluntarily fails to reveal a fact that she has good reason to believe would be relevant to the decision being made by the individual tokening consent. Like coercion, fraud renders consent involuntary.

In the context of consent to research participation, fraud can occur if an investigator fails to reveal to a prospective participant some fact about her clinical trial that she has reason to think would make a difference to his decision to enroll. For example, she might omit to mention that a known side effect of an experimental drug is that it causes unsightly blotching of the skin, a fact that would surely make a difference to some people who are concerned about their appearance. Fraud in medical research can therefore be avoided through appropriate disclosure, that is, through a full announcement of those facts that the investigator has good reason to think may be relevant to someone’s decision to enroll. This implies that a restricted version of the inducement view is the correct account of what must be disclosed.

Showing that an inducement view is the correct account of disclosure does not thereby tell us whether or not it is also the correct view of understanding. We next argue that it is not. Two possible interpretations of the inducement view can be distinguished. The first is comprehensive: prospective research participants must understand everything that could be relevant to their decision about whether to enroll. This interpretation preserves the intuitions that motivate the inducement view, but is otherwise wholly implausible: not only is it unlikely that anyone will know all the propositions relevant to someone’s decision, there may be experiential aspects to the research procedures that could not be understood prior to undergoing them. The alternative, restrictive interpretation says that participants must understand all and only those propositions relevant to their decision that the researcher knows. But this interpretation avoids the problems of the comprehensive interpretation at the cost of its intuitive support: people who give this interpretation cannot explain why a research participant who fails to understand any part of what has been disclosed to him thereby gives invalid consent. By contrast, our analysis in terms of fraud does explain why a researcher who fails to appropriately disclose information about her research that she has reason to believe would be relevant to an enrollment decision thereby undermines the validity of his consent.

We close by noting three implications of our analysis. First, how much of the information disclosed must be understood for consent to be valid is less than has previously been supposed. Second, fulfilling the disclosure requirement involves not just disclosing facts that are expected to be relevant, but also disclosing them in a manner that gives prospective participants a reasonable opportunity to understand them. The empirical data on what participants are likely to be able to understand should therefore be taken into account in designing informed consent.
processes. Finally, the analysis implies that far fewer studies have been conducted unethically than the standard view of disclosure and understanding would imply.

- **Comments:** Tait Szabo (Wisconsin-Washington County)


Many contemporary philosophers working on moral responsibility adopt a Strawsonian account of that concept. Such philosophers follow Straw in thinking that to be morally responsible is to be the appropriate target for a range of *reactive attitudes*, a class of emotional responses we have toward each other typically understood to include emotions such as anger, resentment, contempt and gratitude. On such a view, for the person who stole your wallet to be morally responsible for doing so, that person must be the appropriate target of my anger (or another reactive attitude).

In order to be successful, a Strawsonian theory must be able to characterize a special class of reasons for feeling the reactive attitudes—reasons that track moral responsibility. We can see this by noting that we commonly evaluate emotions as appropriate or inappropriate in senses that have nothing to do with whether someone is morally responsible. For example, one way in which we commonly evaluate emotional episodes is in term of their prudence. It might be thought inappropriate to be visibly angry at someone who is stealing your wallet, since that might incite him to do you further harm. If, like most people, you cannot completely suppress the expression of strongly felt emotions, you have a reason based on prudence not to feel angry with the thief. The fact that it would be folly for you to feel anger toward your assailant shows that prudence cannot be the relevant kind of appropriateness for elucidating a Strawsonian theory, for the thief’s moral responsibility for stealing your wallet is not impacted by prudential considerations.

Recently, a consensus has emerged among Strawsonians that the relevant reasons for feeling are specified by the concept of *desert*. As Derk Pereboom urges, “For an agent to be morally responsible for an action in the sense at issue in [the moral responsibility] debate is for it to belong to him in such a way that he would deserve blame if he understood that it was morally wrong [...] supposing that this desert is basic in the sense that the agent would deserve the blame [...] just because he has performed the action, given understanding of its moral status, and not by virtue of consequentialist or contractual considerations.”

However, even ruling out consequentialist or contractual considerations does not make ‘desert’ fix on a unique set of reasons for feeling. I show in this paper that claims about whether a person deserves an emotional reaction remain ambiguous between two very different kinds of reasons we have for feeling. One sort of reason-claim invoked by the language of ‘desert’ is that an emotion is *fitting*. An emotion fits its object, in the relevant sense, just in case it accurately characterizes its object as having the properties that emotion characterizes the world as having. For example, anger construes the world as containing someone who has violated a norm on proper conduct. If so, anger is fitting just in case someone has violated a norm, as when a person who steals your wallet violates the moral norm against stealing. If the person did not actually violate the norm—suppose his agency was completely bypassed and he was being controlled like a puppet—then anger would be unfitting for you to feel toward him.

‘Desert’ can also imply a different sort of claim about the appropriateness of anger that is conceptually distinct from the question of whether or not anger is fitting. To see this, consider whether the person who steals your wallet deserves your anger once you discover he is destitute and hoped to get enough money from your wallet for a few meals. It seems that he does not; because of your privileged financial situation relative to him, your assailant does not deserve your anger in what I call the *fairness* sense of ‘desert.’

By appealing to cases where the fairness and fittingness of reactive attitudes come apart, as in the above example, I show that fittingness is the aspect of emotional desert that
tracks moral responsibility. This insight has the potential to significantly impact our understanding of the debate between compatibilists and incompatibilists. To demonstrate this, I apply the distinction between fairness and fittingness to the highly influential manipulation argument against compatibilism. That argument appeals to the intuition that even someone who fulfills the conditions of the best compatibilist account of morally responsible agency would not deserve anger for doing wrong, so long as she had been manipulated (say, by neuroscientists) into satisfying the relevant compatibilist conditions. Since the person would not deserve anger, the manipulation argument holds that such a person is not morally responsible, suggesting compatibilism is flawed. I show the intuition that the manipulated person does not deserve anger is supported only by the fairness sense of desert. I also show that anger is still fitting for the manipulated person. Since in the example the manipulated person deserves anger in the fittingness sense and that sense of emotional desert is the relevant one for elucidating the Strawsonian account of moral responsibility, such examples are no threat to compatibilism. Rather, they miss their mark because they trade on the wrong sense of emotional appropriateness.

The upshot of the paper is an improved understanding of a Strawsonian account of the concept of moral responsibility, increased clarity about the reasons governing emotional responses and a new insight regarding the importance of a very influential argument.

- Comments: Evan Simpson (Memorial University of Newfoundland)


The term ‘welfarism’ is used in a variety of closely related ways in moral philosophy and axiology. In this paper, following L.W. Sumner, it designates a position about the foundations of morality—the view that “nothing but welfare matters, basically or ultimately, for ethics.” More specifically and again following Sumner, “welfarists” are committed to:

**Axiological Monism:** The list of foundational goods contains a single item: well-being. Only states of individual well-being are intrinsically valuable.

**The Priority of the Good:** Facts about which states are intrinsically good explain which duties and normative reasons we have. The fundamental point of ethics is to bring about good (or better) states of affairs.

**Agent-Neutrality:** The foundational values in an ethical theory are agent-neutral—there is a reason for each agent to promote it, or aim to bring it about.

My paper does not defend welfarism, though I open by explaining why the position is difficult to resist. Instead, I’m interested in which normative theories we should adopt if welfarism is true. I suggest that a particular kind of *caring ideal observer theory* is the most plausible expression of welfarism. Specifically, welfarists ought to use the responses of a fully-informed, instrumentally rational being who cares (or as it is put in the psychological literature—has “sympathetic concern”) for all welfare subjects to characterize various aspects of the moral point of view. For example:

**To evaluate outcomes:** A state of affairs \( S \) is **better than** a state \( S' \) iff an instrumentally rational, fully-informed being who impartially cared for welfare subjects would prefer \( S \) to \( S' \).
To characterize the morally optimal: An action, \( x \), is optimal in a circumstance \( C \), iff an instrumentally rational, fully-informed being who impartially cared for welfare subjects would perform \( x \) in \( C \).

To assign duties: One has a moral duty to do \( x \) in \( C \) iff an instrumentally rational, fully-informed being who impartially cared for welfare subjects would will (not merely hope or wish) that agents do \( x \) in \( C \).

Why?

On its face, it seems that no theory is a purer expression of welfarism than traditional varieties of maximizing aggregative utilitarianism. So what use is the device of an “idealized carer”—how could it possibly be a better expression of welfarism? The inspiration is not a nostalgic yearning for the benevolent spectators of early utilitarian history. Instead, it is inspired by the recognition of a difference between valuing welfare and valuing welfare for those to whom it accrues.

Welfarism’s allure lies, in part, in its privileged ability to capture the ideas that the moral point of view tells us to do “what is best for everyone” and that it treats each individual with equal concern. But when welfarism is expressed in terms of a demand to maximize aggregate well-being, it becomes an affront to the very welfarist ideals that motivate it. Often maximizing aggregate well-being seems to be worthless—not worth doing for anyone’s sake. Here I have in mind “Replacement”, “Mere-addition”, and “Repugnant Conclusion” type cases. For example, it seems pointless to replace (i.e. annihilate) a currently existing and happy population, to bring about a world full of many more beings whose lives are barely worth living. Is this what would be best for everyone? Of course, many feel compelled to “bite the bullet” in these cases because the alternatives have troubling implications of their own; appealing to average welfare turns out to be a non-starter, and so-called “Person-Affecting” principles which avoid “Replacement” and the “Repugnant conclusion” fall prey to Parfit’s “Non-Identity” problem.

We should be puzzled. We believe we have a moral reason to promote well-being, and the welfarist thinks that this is simply what morality is “all about”; but there is often no reason to maximize aggregate or average well-being. But we need not give up on the idea that welfare is a fundamental good, or concede that there are other independent goods or moral demands at play. Instead, if the problem is that maximizing well-being is sometimes not worth doing for our sake, then the solution should be in discerning when promoting well-being is worth realizing for our sake.

Following Elizabeth Anderson, David Velleman, and Steven Darwall I argue that to want someone’s well-being for that someone’s sake is to want it out of care for her. This is not a vacuous position. We will see that there’s a substantial difference between caring for individuals and hence their well-being, and directly caring about individual well-being—Replacement and the Repugnant Conclusion type cases are a clear illustration of that point. Thus, I think the device of the “idealized carer” is integral (though perhaps only a reliable heuristic) to understanding what an adequate welfarist morality looks like. Perhaps more importantly, I will contend it is the key to a plausible axiology for welfare—it has unique capacity to provide a well-motivated escape from both the Repugnant Conclusion and the Non-Identity Problem.

The view also has a number of independent virtues. Unlike other candidate welfarist views, it can be applied without knowing what exactly welfare consists in, or how to quantify it intra or inter personally. For I will claim that we can “read off” what’s good for a person, \( A \), by examining what someone would want or prefer insofar as they care for \( A \) and are fully informed and rational. By extension, what is best for welfare subjects can be read off from the responses of an idealized agent who cares for welfare subjects. These dispositions need not (and should not) be mediated though a substantial theory of well-being. Furthermore, unlike traditional consequentialist views, using the different dispositions of the idealized carer allows us to make
sense of the possibility, in principle, of optimal acts that it is not our duty to perform, while allowing that those categories may be extensionally equivalent.

At the close of the paper I discuss how epistemic access to the responses of the “idealized carer” is possible, though obviously imperfect. I end by replying to some salient objections.

- **Comments:** Kelly Sorensen (Ursinus)

13. **Eva Dadlez (Central Oklahoma):** “Federally Funded Elective Abortion: They Can Run, but They Can’t Hyde”

In this paper we argue in favor of federal funding of elective abortion. We address the current restrictions on public funding, and suggest that the failure to enable represented by a ban on federal funding is morally equivalent to an outright prohibition. Just as a moral equivalence can be established between killing and letting die in symmetrical cases, criteria for symmetry can be established that identify those failures to make possible that are morally indistinguishable from proscriptions. On this basis, it can be shown that restrictions on federal funding can carry the same moral liability as prohibitions.

Enacted in 1976 as an amendment to the annual appropriations bill which funds Medicaid, the Hyde Amendment initially stated that no federal funds could be used to fund abortion services. Similar amendments are still regularly attached to the annual appropriations bill. States receiving federal Medicaid dollars are not required to provide funding even for medically necessary abortions.

Ethicists have long been familiar with arguments about whether there is an inevitable distinction between killing and letting die, and with the related question of whether some failures to rescue are morally equivalent to killings. If there is no morally significant difference between killing and letting die in symmetrical cases, there is an obligation to offer aid which is just as morally pressing as the obligation to refrain from harming other agents. Opponents of this position invariably invoke a distinction between positive and negative rights, arguing (often with reference to the Kantian distinction between perfect and imperfect duties) either that there are no positive rights to speak of, or that there are very few positive rights which need concern us, or that the obligation to attend to the positive rights of agents is uniformly less pressing than our obligation to respect their negative rights.

Defenses of the Hyde Amendment seem to rest on arguments similar to those espoused by advocates of a radical moral distinction between killing and letting die. Here, however, the focus is on policies rather than actions, and the distinction in question is between forbidding and enabling. Defenders of Hyde stress that respect of so-called negative rights is never to be confused with the kinds of obligations to which acknowledging the existence of positive rights might commit us. Just as no obligation to rescue is thought to follow from a prohibition on killing, no obligation to enable a course of action is thought to follow from the contention that it is wrong to prohibit it. The wrongness of the prohibition is due to a violation of negative rights, whereas any obligation to enable would be based on some positive right to aid.

Opponents of Hyde are inclined to argue that a failure to enable can sometimes prove morally equivalent to an outright prohibition. This is not to contend that all failures to enable are morally the same as the proscriptions to which they correspond. Several factors conduce to a claim of moral equivalence. Just as Michael Tooley, Judith Lichtenberg, and others developed criteria that could help us to determine whether cases of killing and letting die were symmetrical and on that account morally equivalent, we will attempt to isolate grounds on the basis of which a failure to make possible can be regarded as morally equivalent to a proscription.

In the case of killing and letting die, where the consequences of doing and allowing are the same and where there is a clear awareness of what will follow from inaction, the agent is held accountable for a failure to intervene, provided the intervention was not prohibitively costly or burdensome. Taking our lead from the preceding, one method for identifying a moral
equivalence between proscribing and failing to enable could involve foresight on the part of policymakers. The criterion would identify policymakers who were aware that a failure to enable would almost certainly deprive those affected of the ability to choose the course of action which an outright proscription would forbid. This is just the awareness that the failure to enable would have the same liberty-limiting consequence as a proscription. It is difficult to see how policymakers could be oblivious to such a consequence. Only seventeen states offer any variety of funding for abortion, so the ban on federal funding makes abortion virtually unobtainable for the indigent in most states. The second criterion for moral equivalence would take into account the burden imposed by enablement, as balanced against the moral import of the deprivation under review.

We suspect that proponents of an inevitable moral distinction between proscribing and failing to enable will be drawn to the Doctrine of Double Effect in their response to the preceding. Presumably, the Doctrine’s distinction between foreseeing and intending would be invoked in an effort to show that intention need not accompany foresight. But what can be established on that basis in regard to something like the Hyde amendment, even if we allow that there is any morally significant difference between foreseeing and intending? We assume that our opponent would not claim outright that the good effect achieved by the Hyde amendment was a greatly reduced number of abortions. Given that the Doctrine tells us the good effect is intended, this would amount to admitting that the intention behind the amendment was an outright proscription. If the good effect were held to involve reduced tax dollars (and thereby a reduced infringement on negative rights as popularly represented by taxpayer wallets), it would still have to be shown that the arguably quite bad effect of women’s being deprived of options was not a means to the allegedly good one. It would have to be further shown that the good effect outweighed an effect which is very bad indeed. We don’t believe this could be demonstrated. In fact, the Guttmacher Institute reports that “The additional public cost of prenatal care, delivery services and welfare totals 4–5 times the amount saved by not paying for Medicaid abortions,” suggesting that any fiscal benefit the Hyde amendment produces for the taxpayer is counteracted many times over.

- Comments: Jeremy Garrett (Sacramento State)

14. Dale Dorsey (Kansas): “The Supererogatory, and How to Accommodate It”

The category of the supererogatory appears to be required to explain a number of important considered judgments about the moral status of actions. If so, we appear to have good reason to reject moral theories that fail to leave space for actions that go beyond the call of duty. Among the traditional targets of this style of critique are standard, maximizing forms of act-consequentialism. In this paper, I do not seek to defend or reject the existence of the supererogatory. Rather, I argue that the supererogatory cannot be understood in a way that is compatible with the traditional use to which it has been put in rejecting act-consequentialism. Indeed, properly understood, act-consequentialism is better placed than many non-consequentialist theories to accommodate the supererogatory.

The paper begins with an examination into the concept of the supererogatory itself. Traditionally, if an act $\phi$ is supererogatory, three things can be inferred about $\phi$:
1. Moral Goodness: $\phi$ is morally good, praiseworthy, or meritorious.
2. Permissible Failure: It is morally permissible to fail to $\phi$.
3. But for Sacrifice: $\phi$-ing would be morally required but for the amount of sacrifice $\phi$ entails on the part of the actor.

However, once these claims have been property understood and interpreted, their conjunction implies following puzzling conclusion:
**Greater:** One has greater moral reason to perform a supererogatory act than a morally required act.

This feature of the supererogatory (as traditionally understood) is puzzling, as has been noted by James Dreier. Hence to accommodate the traditional notion of the supererogatory, a moral theory must defend the claim that moral requirements are not determined by the strongest moral reason to act. In other words, one must successfully argue that *intra-* moral rationality fails to take an *optimizing* form.

But a non-optimizing form of moral rationality should be rejected. The second section of this paper addresses two important attempts to accommodate such a view. The first option is to accept the claim that moral rationality is *satisficing*. This proposal is favored by Dreier. However, I argue that no form of moral satisficing is acceptable. In particular, no form of moral satisficing can, by itself, explain *But for Sacrifice*. The second option is to accept the claim that moral rationality is not concerned *solely* with moral reasons. In other words, though moral rationality is optimizing, moral reasons are not the only reasons that are relevant from the standpoint of moral rationality. This proposal accommodates all three characteristic principles of the supererogatory, but is not otherwise plausible. In particular, this view would appear to hold that the prudential reason an individual may have to murder his wealthy neighbor in cold blood is *morally* relevant, contrary to considered judgment. Though this prudential reason may help to determine what a given person is required to do from the standpoint of all-things-considered practical reason, this prudential reason surely has very little to say about what *morality* requires us to do.

The third section of this paper offers what I take to be the best method of accounting for the supererogatory. I suggest that supererogatory actions are actions that are morally good, actions for which there may be decisive moral reason to perform, but actions for which there is not decisive *practical* reason to perform. The supererogatory becomes the class of actions that go “above and beyond” one’s *rational* requirements. If this proposal is accepted, one can accommodate the plausible claim that moral rationality is optimizing and accept the category of the supererogatory. To put my claim more precisely, my account of the supererogatory reinterprets *Permissible Failure* and *But for Sacrifice*. For any supererogatory act $$,$$

*Permissible Failure*$$:$$ It is rationally permissible to fail to $$.$$

*But for Sacrifice*$$:$$ F-ing would be rationally required but for the amount of sacrifice $$ entails on the part of the actor.

This view is subject to an obvious objection. My proposal allows that supererogatory actions can be morally required. However, by way of a response, the claim that supererogatory actions cannot also be morally required is a *conception* of the supererogatory, a *theory* of the supererogatory. It is not analytically implied by the idea of the supererogatory itself. To reject my view on the basis that it differs from other theoretical accounts of the supererogatory is to simply beg the question against my view. Of course, my view is in trouble if it grants supererogatory status to actions that are not, on reflection, supererogatory (or vice versa). But my view does not do this.

To see why, consider the fact that one cannot accept my view *and* the existence of supererogatory actions without also accepting a weak form of anti-rationalism about morality. In particular, one must reject the claim that morality is rationally overriding, i.e., that if one is morally required to $$, one has decisive all-things-considered practical reason to $$. But once we take this form of anti-rationalism seriously, we find that my approach to the supererogatory can successfully accommodate the plausible considered judgments that appeared to justify the existence of the supererogatory. In other words, once we accept the form of anti-rationalism implied by my view, there is no further reason to believe that supererogatory actions cannot be morally required.
If my argument succeeds, this is important not just for a general understanding of the nature of supererogation, but also for the use of the category of the supererogatory to reject traditional moral theories. Indeed, I argue, act-consequentialism is well-placed to accommodate the supererogatory. On my view, to accommodate the supererogatory, it must be plausible to insist that there is distance between one’s moral requirements and one’s rational requirements. But because act-consequentialism can plausibly explain this distance, act-consequentialism is better-placed to explain the supererogatory than moral theories that insist on a convergence between the demands of morality and the demands of practical reason.

- **Comments:** Derek Baker (Lignan University, Hong Kong)

15. **Kai Draper (Delaware): “Anticipating Death”**

The paper I would like to present argues that death can merit what I shall call “negative anticipation.” By “negative anticipation” I mean “anticipation with emotional distress” or, more precisely, “being troubled or disturbed by the temporal approach of some state of affairs.” Obvious examples of negative anticipation include being frightened by the fact that tomorrow one will undergo a painful medical procedure and being saddened by the recognition that one’s child will soon be leaving home. There are various forms of emotional distress that can take as an object the approach of some future state of affairs, among them fear, sadness, despair, and despondency. Thus, there are many forms of negative anticipation.

Negative anticipation is the opposite of positive anticipation. When we speak of the latter, we often speak of someone’s “looking forward to” or “eagerly awaiting” something that lies in the future. The English language doesn’t provide its users with parallel expressions for describing negative anticipation. Perhaps “dread” can be used to refer generally to negative anticipation, but its primary use is captured by Webster’s definition of “dread” as “anticipation with great fear or apprehension.” Fear and apprehension do not exhaust the ways in which one can be troubled in anticipation of a future state of affairs, and obviously not all fear or apprehension is great.

One might suppose that only (future) positive evils can merit negative anticipation and so, because death is at worst a negative evil (i.e., bad for the one who dies because it deprives that individual of the positive benefits of survival), death cannot merit (self-interested) negative anticipation. If an unexpected financial loss were to prevent me from taking a much-deserved and long-awaited vacation in August, for example, disappointment would be reasonable, but unless I would suffer some positive evil in August (a boring month at home perhaps), I would have no reason to be troubled by the approach of August or by the approach of my failure to be on vacation in August. Similarly, if I were to learn that an unexpected death would prevent me from taking such a vacation, once again it seems that disappointment, but not negative anticipation, would be reasonable. If we could generalize from examples of this sort, then we could reach the conclusion that the mere deprivation of life’s benefits cannot possibly merit negative anticipation and so death cannot merit negative anticipation.

Unfortunately, we cannot generalize from examples this sort because the loss of present benefits can merit negative anticipation even if no positive evil accompanies that loss. My thesis, then, is that a death that brings to a permanent end all of life’s benefits can merit negative anticipation. My argument to this conclusion is a simple appeal to symmetry that begins with the premise that it can be rational to look forward to escaping from misery. From there it is a short step to the conclusion that if one’s life is irredeemably miserable, then a death that will bring one’s misery to a permanent end can merit positive anticipation. If this is correct, then assuming that the value of happiness and the disvalue of unhappiness are symmetrical in the relevant respects, we can reach the conclusion that if one’s life is a happy one, and death that will bring that happiness to a permanent end, then one’s death merits negative anticipation.
It might be objected that Lucretius’s appeal to the symmetry between posthumous nonexistence and prenatal nonexistence offers a basis for rejecting my appeal to symmetry. For if it is unreasonable to be troubled by the fact that one’s happiness does not extend farther into the past than it in fact does, then given the symmetry between past and future, it might seem that it must also be unreasonable to be troubled by the fact that one’s happiness does not extend further into the future than it in fact does. Consider, however, the parallel case of unhappiness. Here there is a clear asymmetry between past and future. If I am miserable now, then it is reasonable for me to look forward to the freedom from misery that the future holds in store; but no comparable attitude towards the freedom from misery that I enjoyed in the past is rational for the simple reason that, because time’s arrow is not pointed in that direction, the past offers no escape from my current, unhappy condition. Even if, per impossible, I could somehow return to my past freedom from misery, it would not be rational for me to want to do so; for given that the direction of time remains the same, that would entail having to go through my misery all over again! Assuming once again that happiness is parallel to unhappiness, I am happy now, then even though being deprived of that happiness in the future merits negative anticipation, the direction of time’s arrow precludes the rationality of any comparable attitude towards the absence of happiness in my past. Indeed, if my whole life has been happy, I might even wish that somehow my prenatal nonexistence could return, for that would provide the opportunity to enjoy my whole life all over again.

To my knowledge, I am the first to propose this sort of response to Lucretius’s symmetry argument. Some of those who argue for asymmetry claim that unlike one’s annihilation, one’s generation at a particular time cannot possibly deprive one of any benefits. [e.g., Nagel 1979; Kaufman 1999] Others [e.g., Brueckner and Fischer 1986; McMahan 2006] concede the possibility of being deprived of benefits by being born too late, but argue that such deprivations are less serious than the deprivations that death inflicts. Unlike all of these writers, I deny neither the possibility nor the relative gravity of prenatal deprivations. Rather I deny that there is a rational analogue to negative anticipation that takes as its object a past deprivation.

- **Comments:** Chris Heathwood (Colorado)

16. **Julia Driver (Washington U, St. Louis):** “Error Theory, Fictionalism, and Imagining the Impossible”
- **Comments:** Bridget Clarke (Montana)

17. **Tim Dunn (Wisconsin, Waukesha):** “The Presumption of Egoism”
- **Comments:** Mark van Roojen (Nebraska)

18. **Robb Eason (Stony Brook/MIT):** “The Ownership Condition of Guidance Control: Meeting Empirical Challenges”

In a provocative paper Maureen Sie and Arno Wouters argue that the behavioral, cognitive, and neuro-sciences provide sufficient empirical evidence to support a new, complex picture of the ostensibly straightforward practice of ‘providing reasons’ (Sie and Wouters, 2009). This picture of the practice of providing reasons challenges a version of compatibilism reliant upon the practice of providing reasons for justifying and explaining moral responsibility. In this paper, I take issue not with the empirical findings of the behavioral, cognitive, and neuro-sciences as reported by Sie and Wouters, but rather, argue that the new compatibilist has adequate resources to meet the challenges presented by the empirical research. I argue that the challenges put forth by Sie and Wouters can be answered by clarifying the implications of the ownership conditions of guidance control as articulated by Fischer and Ravizza. (Fischer and Ravizza, 1998).
Sie and Wouters argue that most philosophers ignore developments in the behavioral, cognitive, and neurosciences (here on BCN-sciences) purporting to challenge our ideas concerning free will and responsibility. (Sie and Wouters, 2009). They claim the reason philosophers tend ignore the purported challenges is that these challenges are generally of the variety that, given findings, there is evidence to suggest that we are unable to act differently than we do act. Many philosophers believe that the ability to do otherwise is irrelevant to responsibility and free will. Citing Frankfurt’s criticism of the idea that responsibility is grounded in a Principle of Alternative Possibilities (PAP), Sie and Wouters turn their attention to “new compatibilism,” citing three paradigmatic examples (Frankfurt, 1969; Fischer and Ravizza, 1999; Wallace, 1994; Wolf, 1981). New compatibilists share in common the idea that responsibility requires “reasons-responsiveness” on the part of the agent to whom responsibility is attributed. They claim agents are responsible for what they do in virtue of acting on reasons those agents can explain and justify.

A supposed virtue of new compatibilism is that it avoids assuming the existence of a metaphysically obscure counterfactual freedom, i.e., that we could have done otherwise than we actually did. However, citing numerous empirical findings of BCN-sciences focusing on the phenomenon of confabulation (Gassaniga and LeDoux, 1978; Wegner, 2002; Wegner, 1999; Pronin, 2006; Damasio, 1994; Bargh, 1999; Haidt, 2001), Sie and Wouters argue that we do not have access to the inner processes that connect the causes of our actions with the actions themselves. In much the same way that we are susceptible to an array of optical illusions, so too we are susceptible to an array of illusions concerning our access to and the transparency of reasons supposedly guiding our actions. The evidence of BCN-sciences suggest that many of the actions for which an actor can give reasons are automatic responses, the results of sub-personal mechanisms, to external stimuli, many of which are not recognized by the actor. The findings of the BCN-sciences ultimately, therefore, complicate the practice of giving reasons. This is because, on their view, findings in the BCN-sciences promote the idea that, “‘providing reasons’ for our behavior is... a complicated process that is better described as ‘interpreting our behavior’ than ‘recalling what moved us’” (Sie and Wouters, 2009).

Viewing the practice of providing reasons as ‘interpreting our behavior’ rather than ‘recalling it’ sets up three challenges to new compatibilists’ accounts of responsibility. First, the new compatibilist must accommodate the finding that many (or most) of our actions are determined by automatic, sub-personal, processes that are triggered by external cues. Second, insofar as the practice of providing reasons is better understood as post hoc interpretation, new compatibilists are saddled with having to explain how agents have the capacity to respond to reasons in cases in which they did not in fact respond to the relevant reasons, without re-invoking the libertarian idea that freedom and responsibility are reliant up PAP. Third, new compatibilists should be able to account for the BCN-sciences’ finding that an agent’s ability to act for reasons is heavily compromised by the influence of the situation in which the agent acts.

I argue that all three challenges to the new compatibilism can be met adequately, and that it is possible to defend a position that maintains that an agent’s responsibility resides in the agent’s ability to both recognize and react to available reasons. Each of the three challenges turn on there being a gap between the reasons the agent can recognize and react to, and the true cause of the agent’s actions. That is, each challenge suggests that an agent’s reasons-responsive mechanisms are not, properly speaking, in touch with the real causes of the agent’s actions. The reasons to which an agent has access, say the challenges in effect, do not track the causal origins of the agent’s actions. However, if we look at Fischer and Ravizza’s account of responsibility and the role the practice of providing reasons plays in that account, that is, if we look carefully at their account of guidance control, it is clear that there are two conceptually distinct components (Russell, 2002). The first component of guidance control is the mechanism of reasons-responsiveness. This is the component targeted by Sie and Wouter’s challenges. The second component ignored, by Sie and Wouter’s, is an ownership condition, guaranteeing that the mechanism that issues in action is the “agent’s own” (Fischer and Ravizza, 1998: 39, 89, 170, 227,
An important feature of the ownership condition is that the agent have appropriate evidence for the agent’s belief that she is “a fair target for the reactive attitudes” resulting from the exercise of her agency in certain contexts (Fischer and Ravizza, 1998: 238). I argue that the process of gathering such evidence is a vital component to agency and is a form of “self-monitoring” (Dennett, 1984, 2005; Russell, 2002b). Furthermore, I argue that the data coming from the empirical sciences, far from delegitimizing an agent’s reasons-responsive mechanisms, serves, instead, to sharpen those mechanisms creating a second-order, more sophisticated reasons-responsiveness (Levy, 2008; Damasio, 1994, 1996, 1999; Dennett, 1984, 2003, 2005).

- **Comments:** Benjamin Mitchell-Yellin (Riverside)

19. **Mylan Engel (Northern Illinois): “The Immorality of Biomedical Animal Experimentation”**

Most arguments for the immorality of animal experimentation [AE] take one of two forms: Either they follow Peter Singer’s lead and maintain that most animal experiments are morally unjustifiable on utilitarian grounds; or they follow Tom Regan’s deontological rights-based approach and insist that the animals being experimented on possess the very same rights-conferring properties which confer rights on humans and that, therefore, experimenting on these animals is wrong because it violates their rights.

When confronted with such arguments, proponents of AE tend to casually dismiss them by rejecting the ethical theories on which they are predicated. Consider two examples. In an effort to defend AE, Carl Cohen goes to great lengths to try to show, contra Regan, that nonhuman animals lack rights and that, therefore, it’s permissible to experiment on them. Richard Vance rejects Singer’s and Regan’s arguments for the immorality of AE because he rejects the analytical ethical tradition on which they are based. As Vance sees it, Singer’s and Regan’s arguments against AE ultimately fail because of the:

“limited nature of the philosophical tools they use. Their ultimate theoretical weaknesses are extremely common among analytical ethicists. Unlike more substantive ethical traditions (for example religious or ethnic traditions), analytical ethics cannot draw on a rich array of sources—canonical texts, authoritative readings, overlapping (even contradictory) platitudes, interpretative communities, and the like. In comparison with such traditions, analytical ethics is abstract and thin... no analytical model has been able to claim adequacy.”

A moment’s reflection reveals the sophistry of such replies. Since no ethical theory to date is immune to objection, one could fashion a similar reply to “justify” virtually any behavior. One could “justify” slavery as follows: An opponent of slavery might appeal to utilitarian or deontological grounds to establish the immorality of slavery. Our fictitious slavery-proponent could point out that these ethical theories are flawed and *ipso facto* so too is any argument based on these theories. Our slavery-proponent might then assert: “Until someone can provide me with clear moral reasons for abolishing slavery, I will continue to own and exploit slaves.”

The speciousness of such a “justification” of slavery should be obvious. No one who seriously considered the brutality and inhumanity of slavery could think that it’s permissible *simply because* all current ethical theories are flawed. But such specious reasoning is often used to “justify” the equally brutal and inhumane breeding, confining, infecting, injuring, mutilating, maiming, blinding, torturing and killing of animals in animal experiments. I aim to block this spurious reply by providing an argument for the immorality of AE that does not rest on any particular highly-contentious ethical theory. Rather, it rests on commonsense moral beliefs that we all share. The significance of this argumentative strategy is this: All effective argumentation must start with premises one’s interlocutor accepts.6 The reason Singer’s and Regan’s arguments sometimes fall on deaf ears is because their arguments do not start with premises their readers share. In contrast, my argument starts with premises the reader already accepts and traces out the moral implications of those premises. Consequently, the reader is already rationally
committed to the truth of the resulting conclusion, on pain of inconsistency. Although animals are used in all sorts of scientific research, including product testing and psychological experimentation, I will focus exclusively on the use of animals in biomedical research, for if it’s wrong to use animals in experiments aimed at developing vitally important, life-saving drugs, it’s wrong to use animals to test trivial products like a new floor wax or shampoo. Having clarified the scope, significance, and rationale for my argumentative strategy, I now turn to the argument itself.

- **Comments:** Sacha Sullivan (South Florida)

20. **Lisa Fuller (Albany): “Burdened Societies and Transitional Theory”**

George Sher notes that nonideal theory “takes as its point of departure some problem that is raised either by the injustice of some past or present social arrangements or else by some limitation of what people can be expected to know or do.” Following John Rawls, nonideal theory is typically divided into two kinds: (1) “partial-compliance theory” and (2) “transitional theory.” The former is concerned with those circumstances in which individuals and political regimes do not fully comply with the requirements of justice, such as when people break the law or some individuals do not do their fair share within a distributive scheme. The latter is concerned with circumstances in which background institutions “may be unjust or may not exist at all”. This paper focuses on issues arising in transitional theory. In particular, I am concerned with what Rawls’ has called “burdened societies,” that is, those societies that “lack the political and cultural traditions, the human capital and know-how, and often, the material and technological resources needed to be well-ordered.” Rawls suggests that some such societies may find themselves in very “unfavorable conditions,” such that their “historical, social or economic circumstances” make it “difficult if not impossible” to instantiate just institutions.

The purpose of this paper is to investigate exactly how – or whether – such burdened societies can reasonably proceed towards a more just condition. As such, the paper focuses on a significantly underdeveloped area of nonideal theory, in which Rawls himself can provide very little guidance. In *A Theory of Justice* he indicates that certain political freedoms or rights of fair equality of opportunity might sometimes be forgone on a temporary basis in order to transform a burdened society into one that more closely approximates justice. However, he adds that any such policies should be designed such that they eventually bring about social conditions in which these restrictions can be lifted, and these rights and freedoms can be fully realized. Further, he suggests that we must look for “policies and courses of action that are morally permissible and politically possible as well as likely to be effective”.

These rough guidelines leave many important questions unanswered. For instance:

- Which restrictions of political freedoms or rights are morally permissible, and in what contexts?
- What if the requirements of moral permissibility, political feasibility and general expediency conflict? Are we to allow some morally questionable policies to be implemented on a temporary basis in order to arrive at a just and well-ordered society more quickly? Or should we instead accept a slower pace of change in order to ensure that no moral requirements have been violated?
- What is to be done in cases where there are morally permissible and effective policies that could be implemented, but they do not enjoy the support of many citizens or their representatives?

These are not merely abstract questions concerning the conceptual relationship between ideal and nonideal theory. They are practical questions that societies have faced in the past and will no doubt face in the future. Perhaps then, we might learn something from how such difficulties have been handled in the past? In this connection, the following three examples are instructive:
1. Affirmative action programs have been defended on the basis that the disadvantages to which certain groups have been historically subject due to their gender, race or ethnicity must be corrected in order for equal rights to be substantively realized.

2. It has been argued that paternalistic policies may be warranted when a burdened society’s cultural or political traditions have caused certain groups to undervalue themselves or others and so to oppose measures designed to instantiate greater equality.

3. Truth and reconciliation commissions have been implemented in situations where gross injustices have been perpetrated, as an alternative to the standard procedures for trying and punishing human rights violations or crimes against humanity. The argument in favor of such commissions is that they allow the society to function better in the aftermath of periods of unrest than extensive punishment of perpetrators.

These examples seem to suggest that under nonideal conditions we may sometimes override or disregard ordinary requirements of fairness, equal liberty, or criminal liability in the service of bringing about a more just society. Indeed, I will argue that the policies and procedures that count as just against a background of deeply imperfect citizens, institutions, and economic and social conditions may not be the same as those that count as just under ideal conditions. We should not take for granted that the right ways to treat people in a basically just society will resemble how they ought to be treated when various types of historical, economic or political injustices are taken into account. Further, we should not be surprised to discover that policies that are acceptable – or even morally required – as intermediate steps toward a more just condition do not always reflect our intuitions about what justice “looks like.”

In addition, I will argue that political feasibility and expediency should be taken as secondary to moral acceptability when considering policies of transition from less-just to more-just conditions, even if this means that the progress toward ideal conditions is slower. Finally, I will argue that for some burdened societies it might be impossible to move forward by implementing policies that are fully morally acceptable – since in some cases any policy solution will leave crimes unpunished, legitimate claims unmet, or rights violated. As such, sometimes it is only politically possible to move laterally between unjust circumstances, that is, from one unjust condition to another. In these cases, policies and procedures should be chosen either by choosing the “lesser-of-two evils,” or according to the likelihood that a given lateral move will open up the possibility of positive reforms at a later time.

• Comments: Steve Vanderheiden (Colorado)

21. Robyn Gaier (St. Louis): “The Second-Personal Perspective: An Other (and yet another) Argument for Externalism”

Motivational judgment internalism, or simply internalism, refers to the idea that when an agent judges that she morally ought to do some action, then she will have at least some motivation to do that act. Ethical rationalism (hereafter, rationalism) is a type of internalism. Rationalism refers to the idea that moral norms stem from the requirements of practical rationality, such that a moral agent would have a motivating reason to act according to her moral judgments insofar as she is rational. Externalism is the denial of internalism. For the purposes of this paper, I will specifically focus upon externalism as the denial of rationalism.

Because the debate between internalists and externalists is essentially a debate about the origins of moral motivation, a first-personal perspective is commonly assumed. So, for instance, according to a rationalist, moral motivation is grounded in the agent’s faculty of reason. Conversely, an externalist might proceed to argue that moral motivation depends upon the agent’s character or disposition. Stephen Darwall, however, advances the idea that moral norms are to be justified from a second-personal perspective. If we presuppose that whatever justifies
moral norms would provide a normative reason for the agent to act, then Darwall’s account may offer a new way of approaching this debate between internalists and externalists.

In this paper, I will argue that the kind of second-personal perspective which Darwall advocates cannot be grounded upon the rationalist (internalist) framework he assumes. Although I am largely sympathetic with Darwall’s second-personal account, I aim to show that his account lacks the support it needs when built upon a rationalist framework. Before explaining the criticism of Darwall’s account which will be of interest here, a brief explanation of his account is in order.

Darwall advances the idea that the normative and motivating reasons that an agent has to act morally are irreducibly second-personal reasons. To appreciate such second-personal reasons for acting, one must take up a second-personal perspective. This second-personal perspective entails more than just addressing another agent as an other. It entails the reciprocal recognition of mutual accountability. For this reason, the second-personal perspective which Darwall advocates requires the agent to have second-personal competence and to make certain presuppositions. A moral agent has second-personal competence if she has “the capacity to be guided by norms” and “the ability ... to evaluate second-personal norms one accepts, and competing alternatives, from the standpoint of one mutually accountable person among others ...” (Darwall, 2006, p. 153). Additionally, according to Darwall, to take up the second-personal perspective is to presuppose “a common competence, authority, and, therefore, responsibility as free and rational [persons], a mutual second-personality that addressee and addressee share ... that is ... recognized reciprocally” (Darwall, 2006, pp. 20-21). Taking up a second-personal perspective in this way provides the necessary perspective from which moral norms are to be justified and the perspective from which a rational, moral agent may appreciate those (normative) reasons that she has for acting morally.

The criticism against Darwall’s view that is of interest here rests upon the large role he affords to second-personal competence and to certain presuppositions which must be in place before an agent is able to take up the second-personal perspective. Taking up a second-personal perspective presupposes a first-personal perspective, but the opposite is not true. An agent may take a first-personal perspective without further assuming a second-personal perspective. Christine Korsgaard, therefore, maintains that taking up the second-personal perspective in order to justify moral norms is simply unnecessary (Korsgaard, 2007). If an agent has what Darwall refers to as second-personal competence and presupposes the other agent’s second-personal competence and authority then, according to Korsgaard, the first-personal perspective is sufficient. Because the second-personal competence of both the addressee must be presupposed in order to enter into the second-personal perspective on Darwall’s account, then it seems as though actually taking up a second-personal perspective is, indeed, unnecessary; and that moral norms may be justified from the first-personal standpoint of practical deliberation. I will further advance Korsgaard’s criticism by specifically questioning the necessity of an agent to presuppose the second-personal competence and authority of the other agent, on Darwall’s account. The purpose of these presuppositions is to establish the idea that the other (rational) agent would have the standing to hold the agent in question morally accountable. But the fact that we may be held morally accountable to those who seemingly lack second-personal competence suggests that moral accountability does not depend upon this presupposition.

If Darwall wishes to defend the idea that the second-personal perspective is, indeed, necessary, then he must afford a larger role to second-personal address. Focusing upon what is required in order to take up a second-personal perspective seems to miss what may be acquired through the second-personal address itself. The objective of this paper, therefore, is to show that affording a larger role to second-personal address may be sustained only on an externalist framework. I will not deny Darwall’s claim that normative reasons are irreducibly second-personal reasons. Rather, what I aim to question is whether we may “locate [these] second-personal reasons within an overall theory of practical reason” as Darwall’s rationalist account.
claims (Darwall, 2006, p. 147). If taking up the second-personal perspective depends more upon factors regarding the particular second-personal relationship that is in place and less upon the agent’s second-personal competence and presuppositions, then an externalist framework seems to provide the more promising foundation upon which to build the necessary second-personal perspective, or so I shall argue.

- **Comments:** David N. S. Faraci (Bowling Green)

22. **Amber Griffioen (Iowa/Marburg):** “Deceiving You, Deceiving Me: On the Nature and Morality of Self-Deception”

In much of the literature on self-deception published in the last thirty years, the debate has centered around how closely self-deception can be said to resemble interpersonal deception, and thus whether self-deception can be said to be intentional or not. More recently, so-called “non-intentionalist” accounts of self-deception have become the prevalent explanation of self-deceptive belief and behavior. Led by philosophers like Alfred Mele, these deflationary accounts claim that most, if not all, instances of self-deception are reducible to instances of some kind motivational biasing, and that that one need not appeal to an agent’s intention to deceive herself to explain her being self-deceived. However, to claim that self-deception is non-intentional may have serious consequences for the philosophical evaluation of an agent’s responsibility for her self-deception and, more broadly speaking, of the morality of self-deception in general.

In this paper, I hope to show that non-intentionalist theories fail to account for certain crucial features of self-deception, whereas intentionalist accounts are in much better shape to make sense of these features. Furthermore, I will argue that self-deception, properly speaking, is best characterized as an intentional project in which an agent actively engages, and that the behavior described by non-intentionalist accounts is better characterized as something other than self-deception. Finally, I will argue that the kind of intentionalist account I put forward leaves room for the possibility of an agent’s being not only epistemically but also morally responsible for her self-deception.

Any account of self-deception must adopt one of at least two strategies. It must either a) attempt to retain a close analogy with interpersonal deception that somehow escapes the paradoxes that appear to arise from such an analogy, or b) deny that self-deception is relevantly similar to other-deception. Intentionalists generally take the former approach, non-intentionalists the latter. However, the question remains: Is there a reason to suppose that self-deception is fundamentally intentional in nature? Non-intentionalists often appeal to the explanatory parsimony of their theory, and this is generally what drives the proponent of a deflationary account of self-deception to propose that all (or at least all paradigmatic) cases of self-deception are essentially non-intentional. The burden of proof is thus placed on the intentionalist to show a) that non-intentionalist accounts do not adequately explain the phenomena and b) that a coherent intentionalist theory can be put forward which better explains these phenomena.

This paper will focus on a few strategies the intentionalist might adopt to meet the challenge put forth by deflationary accounts of self-deception. One such strategy for the intentionalist to take is to focus on an important feature of self-deceptive irrationality that deflationary accounts tend to ignore—namely, the reflective nature of self-deception. As Dion Scott-Kakures notes, non-rational animals may process information in motivationally-biased ways that satisfy the non-intentionalist’s conditions for self-deception, but we do not generally take such animals to be capable of self-deception. Whereas the (potentially) rational agent reflects on and is moved by her reasons for acting and believing, the non-rational animal is not. However, we might think that self-deception requires a certain level of self-awareness—namely, the ability to be moved by reasons qua reasons.
Indeed, what seems to distinguish the self-deceiver from the merely epistemically-ignorant or psychologically-compelled agent (or from the non-rational animal) is not merely that the former is both motivated and in some sense “competent” to avoid her deception, but also that she actively *puts reason to use against itself* in ways that ignorant, compelled, or non-rational actors do not. Put a bit differently, the self-deceiver attempts to give herself reason to believe what she takes herself to have sufficient reason not to believe. This “pseudo-rational” feature of self-deception also provides us with a way to distinguish between mere wishful thinking and self-deception proper—a distinction often elided on deflationary accounts. Self-deceivers are at least partially *complicit* in their self-deceptions, whereas wishful thinkers are not.

Furthermore, non-intentionalist accounts do not adequately explain why self-deceived agents typically exhibit a certain level of cognitive tension. But on intentionalist accounts according to which the self-deceiver employs distinctly rational (in the sense of reasoning-giving) techniques in the service of irrationality, it is no surprise that she would experience being “at odds” with herself—for she is, in fact, undermining her own rationality by employing the very faculties that make her a rational agent in the first place. She is, in some sense, *internally* irrational—that is, she is irrational *by her own lights*.

Thus, the claim that deflationary accounts of self-deception are explanatorily more parsimonious and therefore preferable might be a bit hasty. There appear to be phenomena that non-intentionalists do not adequately address, including (but not restricted to): the reflective nature of self-deception, the cognitive tension apparent in the behavior of self-deceived agents, the distinction between ignorant, compelled, or non-rational actors and self-deceived agents, the intuitive difference in kind between wishful thinking and self-deception proper. Therefore, one might conclude that a more robust, intentionalist account of self-deception is necessary to make sense of these phenomena.

In the remainder of the paper, I will go on to discuss an intentionalist alternative that both accounts for the phenomena lacking in non-intentionalist accounts and avoids the aforementioned paradoxes. I will argue that if we view self-deception as a diachronic *project* undertaken by the agent with some awareness of what she is up to, not only can we see how and why reflective self-deceivers often exhibit significant cognitive tension, but we can also make sense of the common intuition that self-deceivers are (at least sometimes) both epistemically and morally responsible for their own deceptions. Furthermore, I will show that such an account retains a close analogy to other-deception by appealing to the importance of *persuasion* in both kinds of deception. I will conclude with a brief discussion of potential avenues for further research into the morality of self-deception.

- **Comments:** Philip Robichaud (Rice)

23. **Lori Gruen (Wesleyan): “The Ethics of Captivity”**

We generally believe that holding someone captive harms them. Insofar as captivity causes physical or psychological suffering it can certainly be considered harmful. But is there something beyond suffering that contributes to the harm of captivity? Some have argued that in denying captive individuals their freedom a captive’s interests in liberty is violated. In this paper I will explore what it means to have an interest in liberty, to evaluate who can have such an interest, and examine the conditions under which such interests may be violated. Though my analysis will be applicable to an ethical analysis of the situations of human imprisonment, in penitentiaries and Guantanamo Bay, my main focus here will be on the case of captive chimpanzees which, I will argue, poses a troubling ethical dilemma.

Liberty can be understood in a variety of ways (freedom from restraint, ability to control one’s actions, absence of arbitrary interference) and captivity involves a denial of all of these sorts of freedoms. But does the fact of being captive itself represent a violation of the captive’s interest in liberty? Recently, Alasdair Cochrane has argued that this can only be the case for
autonomous individuals who have an intrinsic interest in liberty. To be an autonomous individual with an intrinsic interest in liberty is more than having preference autonomy or freedom to satisfy one’s desires, a freedom that captivity certainly denies, but also to have the ability to frame and pursue one’s own desires. If an individual has an intrinsic interest in liberty then she is harmed in an additional way by being held captive. She is denied the opportunity to pursue her own way of life with her own conception of how that life should go. One implication of the analysis of the harm of captivity for autonomous individuals with an intrinsic interest in liberty is that no improvements in captive conditions can rectify this particular harm. While interests in physical safety, adequate nutrition and hydration, bodily integrity, psychological engagement, and the like may all be amply satisfied, the fact of the denial of liberty remains. To determine whether or not an individual is harmed by captivity, then, one must ascertain whether or not the beings held captive are autonomous beings.

I will argue that while there are a variety of different ways of understanding what it means to be an autonomous individual (e.g. internally motivated to achieve independently chosen ends; ability to self-govern; being guided by principles that one endorses; formulating one’s own conception of the good life) it makes sense to think of autonomy as coming in degrees. Further I will argue that autonomous beings do not act autonomously every time they act, but that they don’t does not undermine their status as autonomous. If this argument is correct, then it will follow that chimpanzees may be considered autonomous, even if they do not act autonomously most of the time. I will support this argument by drawing on a variety of conclusions that result from cognition research with chimpanzees in captivity as well as in the wild settings. Further, and controversially, I will argue that captive chimpanzees understand that they are captives and that we humans are their captors, and that when some individuals ask to be let out (through gestures or by escaping) they are not seeking to have some immediate desire satisfied (they are not in search of more desirable food or more interesting mates) but rather they apparently want to be free from captivity as such.

In the final section of the paper I will explore the implications of my argument for the actual conditions in which captive chimpanzees live today. There are an estimated 2400 chimpanzees living in captivity in the United States. Approximately 985 live in biomedical research laboratories, 270 live in accredited zoos, 620 live in sanctuaries, and an estimated 550 chimpanzees are living in various conditions in the entertainment industry, in roadside attractions, and as people’s “pets.” There are approximately 1400 chimpanzees in sanctuaries and zoos in the rest of the world, with the exception of Africa, where the number of chimpanzees living in captivity is increasing as a result of chimpanzees being orphaned by the bushmeat trade. In many of these settings, chimpanzees are having their most basic interests violated and are thus being harmed. But in some, particularly some of the better zoos and sanctuaries, it may be that it is just their liberty interest that is being violated. Since we cannot release captive chimpanzees from captivity without causing great harm and likely death to them, we face an ethical dilemma. I will suggest that while keeping chimpanzees in captivity, even under the best case scenarios, harms them, we can minimize that harm by creating captive conditions that, in so far as possible, promotes their “wild dignity.”

- **Comments:** Molly Gardner (Wisconsin)

**Jason Hanna (Northern Illinois): “Consent and Framing Effects”**

Deontological theories of consent face problems due to the operation of framing effects: The choices we make are often affected by the way in which options are presented. To take one classic example, people are significantly more likely to consent to a medical procedure if they are told that 90 percent of those who undergo it are alive after five years than if they are told that 10 percent are dead after five years, even though these descriptions provide the same
quantity of information. Such cases give rise to a puzzle: If a person has consented (or dissented) and we have good reason to believe that she would choose differently if the choice were framed differently, is it permissible to proceed? There seem to be two general ways of resolving this puzzle, one that appeals to actual consent and one that appeals to hypothetical consent. I will argue that neither succeeds and that the puzzle suggests that consent is less important than most deontologists believe.

The Actual Consent Proposal

According to the actual consent proposal, our treatment of the person in such cases is morally governed by his actual choice. Thus if Fred dissenting to a surgery when told that it has a 10 percent failure rate, we may not proceed, even if we are justifiably confident that Fred would have consented if told that the surgery has a 90 percent success rate. This proposal faces two problems. First, it seems to be in tension with a plausible theory about the significance of consent. Most ethicists accept the valid consent thesis, according to which a consent-token is morally valid if and only if it is sufficiently well-informed, unimpaired, and uncoerced. Each of these conditions on valid consent is presumably justified by appeal to autonomy. If so, we cannot conclude from the fact that someone actually has consented that his consent carries moral weight. Instead, the relevant question is whether the consent-token expresses his autonomy. It is very difficult to believe, however, both (a) that Fred expresses his autonomy when he dissents and (b) that Fred would have expressed his autonomy had he consented merely because the options were differently framed. The actual consent proposal either yields this implication or else, I shall argue, embraces a conception of autonomy that is too thin to support the valid consent thesis.

Second, the actual consent proposal faces an act-description problem. On one common view, consent is a psychological state or attitude. If so, it seems possible for a person simultaneously to consent to an act under one description and dissent to it under a different description. Such conflicts are most likely when the act-descriptions are complex enough that their equivalence is not immediately apparent. Further, even if this point is mistaken, the actual consent proposal must grant moral force to a person’s disposition to consent if it is to account for our intuitions about cases, especially those in which we are unable to communicate with the person. If the disposition to consent carries moral weight, however, the original problem resurfaces, since it involves a person’s consenting under one description but being disposed to dissent under a different description. (In the paper, I explain how the disposition to consent differs from traditional interpretations of hypothetical consent. Roughly, the difference is that the former does not require us to idealize away from the subject’s attitudes and thus should be acceptable to those who eschew hypothetical consent.)

Of course, one might argue that even if the actual consent proposal cannot resolve framing effects cases, actual (lack of) consent should absolutely govern our treatment of people in other contexts. This response seems unsatisfying. Given the pervasiveness of framing effects, the concession may be large enough to deprive consent of much of its practical significance. Further, if actual consent has moral force, we would surely expect it to govern our treatment of Fred in the surgery example, since this is precisely the sort of case in which consent is supposed to matter.

The Hypothetical Consent Proposal

In response to these problems, one might deal with framing effects by appealing to hypothetical consent. The hypothetical consent proposal can be developed in two further ways. According to the first, the conditions characterizing the hypothetical choice situation are merely procedural; we are to imagine that the subject is thinking clearly (e.g., is not intoxicated), has adequate time for deliberation, and is fully informed. This proposal does not seem to resolve the original puzzle, which arises because fully informed people with adequate time for deliberation sometimes choose differently depending on how the options are framed. One might reply that, in our earlier example, the inconsistency in Fred’s (hypothetical) attitudes shows that he is guilty
of some form of procedural irrationality. But even if this is correct, it provides no way to determine which of Fred's attitudes should be given priority.

It might now be claimed that the relevant hypothetical choice should meet further, substantive (as opposed to procedural) conditions. One might argue that we should implement the choice the subject would make if she were in a state conducive to good decision-making, and then specify good decision-making as, e.g., that which best reflects the subject's values or best serves her interests. This proposal may provide guidance in framing effects cases. But it is unlikely to satisfy deontologists who emphasize the importance of consent, since it allows us to appeal directly to the fact that proceeding would (or would not) protect the subject's values or interests. Consent drops out of the picture and remains relevant only as evidence for some further claim.

I conclude, first, that it is unclear how deontological theories of consent can resolve the puzzle posed by framing effects. Given the broad reach of this phenomenon, the puzzle poses a significant obstacle to the acceptability of such theories. Second, while critics of consequentialism sometimes object that it has too little regard for consent, perhaps the deontologist too must concede that consent has only evidential relevance.

- Comments: Marcus Arvan (Tampa)


Richard Posner has long argued that when considering the appropriate relationship between law and morality, it is best to leave morality aside, and in its place to focus on wealth maximization (Posner 1979 & 2001). Posner argues for maximization of wealth in large part because of the problems he believes moral theories, particularly utilitarianism, suffer from. According to Posner, utilitarianism ought to be rejected for three reasons. First, most people do not accept utilitarianism. While most of us may recognize that happiness is a value, it is not the only or even the primary object of importance in our lives. Second, Posner raises the familiar objection that utilitarianism fails to recognize the importance of, and accord sufficient respect to, individuals. Finally, utilitarianism suffers from a boundary problem. By this Posner means that utilitarianism draws the boundary of moral considerability at the wrong place. While most of us have intuitions that humans matter more than animals and members of our own community matter more than strangers, utilitarianism cannot justify these intuitions. Because of these weaknesses Posner argues that instead of maximizing utility we ought to maximize wealth.

The theory of wealth maximization tells us that we ought to seek to achieve that distribution of goods that produces the overall greatest amount of wealth. It should be noted that 'wealth' here is not limited merely to money. Instead, as Posner means it, wealth is the total of all valued things in a society “…weighted by the prices they would command if they were to be traded in markets.” So, imagine Smith owns a bike that Jones wishes to have. If we were to ask “who would be happier to have the bike?”, we may not be able to answer this question with any degree of accuracy. While Smith may be effusive about the happiness that the bike will bring him and Jones may be taciturn about the topic, this does not answer the question. Smith may be melodramatic and Jones may reserved. In other words their reports to us about the bike may be misleading. Posner claims wealth maximization avoids such epistemic difficulties. If we would like to know who values the bike more we simply need to find out what price Smith would sell the bike and how much Jones would pay for it. Imagine that Smith would sell the bike for anything over $600 and Jones would pay up to $800 for it. Should Jones offer $700 for the bike the transaction will go through, and wealth has been maximized. Smith now has $700 and Jones has something he values for $800 (for a total of $1500) whereas before the transaction Smith has something he valued for $600 and Jones has $700 in cash (for a total of $1300). The aggregate wealth has increased by $200. The advantage for Posner is that we avoid speculating about what
will make people happy and instead we focus on how much a person or persons are willing and able to pay for any good. Once we know how much you are willing and able to pay for something, then we know how much you value it. Posner assures us that the wealth maximization theory succeeds where utilitarianism fails.

In this paper I argue that Posner’s argument for wealth maximization is deeply flawed. I develop two arguments against his position. First, I will argue that Posner’s theory fares no better when held to the standards he uses to evaluate utilitarianism. Rather than explore the various replies that utilitarians have developed in response to these types of objections, I suggest we evaluate Posner’s theory by the very same standards he uses to evaluate utilitarianism. What we find, I argue, is that Posner’s theory fails for the very reasons that he suggests utilitarianism fails. That is, wealth maximization relies on a view of value most reject, it fails to respect individuals in appropriate ways, and finally, wealth maximization suffers from a significant boundary problem.

The second argument I offer against Posner’s theory of wealth maximization reveals a more fundamental flaw in his theory. Posner argues that wealth maximization avoids the epistemic difficulties that we face when we try to determine what act will make others happiest. For Posner the key to understanding how much a person values something is how much they are willing and able to pay for that thing. What’s important to see about this is that the theory of wealth maximization does not ask how much one does pay for a good to determine value but rather what they are willing to pay. Once we see that what matters for Posner is the intention to pay rather than paying we can also see that there are serious problems with this account of value. To elucidate the problem that Posner encounters I show that he has found himself in an economic version of Gregory Kavka’s ‘Toxin Puzzle’ (1983). I consider a number of ways that Posner may try to extricate himself from this problem, but ultimately I argue that none of those options will suit Posner’s purposes. What we find then is that Posner has failed to give us any compelling reason to accept his view over utilitarianism.

- **Comments:** Harald Viersen (Amsterdam)


Act-consequentialist normative theories determine rightness or wrongness of an act on the basis of the total sum value of the state(s) of affairs that will (or, would be expected to) ensue if the act is performed (relative to the total value of the state(s) of affairs that would ensue if other available acts were performed). Importantly, this does not preclude the possibility that an act itself may have some axiological valence, where by axiological valence is meant the (non-instrumental) goodness/badness of an act (or, the goodness/badness of the state of affairs of an agent performing an act), as distinguished from the rightness/wrongness of the act. Bernard Williams makes this point in *Utilitarianism: For and Against*.

An act-consequentialist may, for example, have an axiological theory according to which human flourishing is the good, and one need not be a full-fledged Aristotelian to hold that human flourishing involves actions as well as pleasurable psychological states. As such, human actions may be assessed as having positive (or negative) axiological valence independent of the pleasure or other consequences that they bring. Acts of dancing, for example, may be held by an act-consequentialist to have axiological valence, as might performance art more generally. All that act-consequentialists must deny is that whether an act is right or wrong is independent of the full range of states of affairs that will ensue if the agent performs the act. The axiological valence of the act itself may just be one of the values in the summation.

Might this ground a general act-consequentialist distinction between doing and allowing, especially pertaining to cases about which philosophers have historically been concerned, such as cases involving the distinction between killing and letting die? No one in the
literature seems to have considered this question; this paper is a defense of an affirmative answer to it.

Although it may not be too difficult to provide a theory of the good according to which acts such as dancing have axiological valence, what could ground a claim that acts such as killing have a more negative axiological valence than omissions which involve letting die? One avenue is off-limits to the act-consequentialist. Act-consequentialists cannot say that there is a difference in axiological valence between killing and letting die because killing is more wrong than letting die – for this is conceding to the deontologist that there is an assessment of the rightness/wrongness of an act which is prior to the determination of the axiological valence of the act.

To provide an axiological theory under which an act-consequentialist may have a doing/allowing distinction without conceding too much to the deontologist, this paper argues for a middle-ground position concerning metaphysical individualism. On one extreme, there is a neo-Spinozist deep ecological view according to which humans are literally One with the rest of nature. On another extreme, individual humans are metaphysically independent of anything else, even each other (in accord with what is known as Hume's Dictum, the view that there are no necessary connections between distinct existences). On a plausible middle-ground view, individual humans are in fact necessarily connected to things outside of ourselves, since we are necessarily embodied beings with a certain evolutionary history.

My claim is that this plausible view of human nature does not leave everything the same as far as axiology is concerned. I argue that, in light of the necessary connections that individual humans have to broader systems of which we are parts (and, specifically, to other humans), the goodness of a human action is affected by how the action fits into a broader system. Importantly, human actions may create systems whereas inactions (normally) do not create systems. So even if killing and letting die result in the same amount of individuals harmed, the axiological valence of an act of killing may be different from the axiological valence of an act of letting die. This is still consistent with an act-consequentialist view that whether an act (or an inaction) is right or wrong can only be determined with regard to an assessment of the total value of the state of affair which ensue from the act (or inaction).

This requires an understanding of the notion of a system; for the purpose of this paper, a system is defined as a whole which has, as parts, interacting individuals which are tokens of types which have necessary connections to each other. (Two further refinements are needed: first, the generality of types must be restricted, since any two physical objects are tokens of the type physical object, and of course the type physical object is necessarily connected to itself; second, the way in which the token individuals interact must be an instance of the necessary connection between the types and not merely an accidental interaction.) For example, when an animal eats a plant, where the animal is a member of a species which evolved to eat the plant, the animal and plant form a system.

When two humans interact with each other, the interaction forms a system, and my claim is that the system may be good or bad depending on the kind of interaction it is, in a way not simply reducible to deontological principles concerning right and wrong action. As above, even if an act of killing results in the same number of people dead as an omission, the act may be more wrong (or, in unusual cases, not as wrong) than the omission since the act may create a bad (or good) system whereas the omission does not. Although a complete axiological theory concerning the goodness of systems is well beyond the scope of this paper, the paper demonstrates a fruitful possibility for an act-consequentialist doing/allowing distinction. Hopefully, this can lead to some amount of conciliation between act-consequentialists and those who are driven by what are normally assumed to be anti-consequentialist intuitions concerning particular cases to become (highly misguidedly, in my opinion) non-consequentialists.

- **Comments:** Arthur Ward (Bowling Green)

Moral realism is the theory that there exist moral standards that are objective in a very strong sense. According to moral realism, morality is stance-independent; that is, the correct moral standards are true independently of the attitudes that any human beings have towards them, and independently of the attitudes that human beings would have towards them under idealized conditions of reflection.

Recently, some evolutionary biologists have argued that the content of human moral attitudes has been heavily shaped by natural selection. While I shall not review the empirical literature on this topic, in this paper I explore whether the truth of such a claim would have any implications for the prospects of moral realism as a metaethical theory.

I begin by laying out Sharon Street’s “Darwinian Dilemma” for realist theories of value. Street challenges the realist to explain the relationship between the stance-independent moral facts posited by realism and the evolutionary pressures which have (putatively) heavily shaped the content of our moral judgments. According to Street, realists must characterize this relationship in one of two ways. On the one hand, the realist could admit that there is no relation between the two, in the sense that the forces of natural selection have shaped our moral judgments in directions that bear no correlation to the moral truth. Call this the distortion hypothesis. Alternatively, Street holds, the realist could assert that there is a relation between the two, namely that an ability to grasp the moral truth was adaptive to our ancestors, and was therefore selected for. Call this the direct tracking hypothesis. Street argues that both of these alternatives are unattractive, and that we therefore have reason to abandon the moral realism that gives rise to the dilemma.

I agree that both of these responses raise serious difficulties for the realist. Street’s contention that she has posed a dilemma for the realist is more problematic, however, for her dichotomy of possible relationships between the selective forces and any stance-independent moral facts is not exhaustive. The most plausible realist responses to Street’s challenge neither posit that evolutionary forces have pushed us in a direction completely independent of the moral truth, nor that dispositions to form certain moral judgments were selected for because these judgments were true. I develop a third possible type of account of the relation between stance-independent moral facts and evolutionary pressures, and explain how such indirect tracking accounts avoid both horns of Street’s dilemma.

I then go on to argue that while indirect tracking accounts escape the specific critique presented by Street, they fail to adequately rescue moral realism from the epistemological challenges raised by the (putative) fact of widespread evolutionary influence. This is because the indirect tracking theorist lacks adequate grounds for rejecting the distortion hypothesis, a possibility which we must take seriously if there is good evidence for strong evolutionary influence on the content of our moral judgments. I argue that arguments for the indirect tracking view inevitably rely on first-order moral judgments, the reliability of which are exactly what is called into question by the distortion hypothesis. This renders arguments aiming to support the indirect tracking hypothesis over the distortion hypothesis viciously circular.

On the other hand, the distortion hypothesis cannot be established with any certainty either. Assuming (as I will argue) that the direct tracking hypothesis is untenable, we are faced with a choice between the indirect tracking hypothesis and the distortion hypothesis, and we lack any strong evidence in favor of either. One might be tempted to conclude from this that evolutionary challenges to moral realism fail, for as long as the indirect tracking hypothesis is in fact true, then our epistemological situation with regard to the stance-independent moral standards is pretty good after all. I argue that to so conclude would be a mistake. On the contrary, if morality is stance-independent, then the serious epistemic possibility of the truth of the distortion hypothesis is sufficient to undermine the epistemic justification of our moral judgments. I conclude that if evolutionary pressures have significantly affected the content of our moral judgments, then moral realists are saddled with the conclusion all of our moral beliefs...
are unjustified. Since such skepticism about morality is implausible, compelling evidence of strong evolutionary influences on the content of our moral judgments would give us reason to abandon moral realism.

• Comments: Eric Martin (London School of Economics)

27. Stan Husi (Rice): “A Defense of Radical Meta-Normative Skepticism”

Radical meta-normative skepticism is the view that no standard, norm, or principle has objective authority or normative force. It contests not that there are norms, standards of correctness, and principles of various kinds making possible that we succeed or fail in measuring up to their prerogatives. The public arena in which we debate how to manage our individual and joint affairs evidently is characterized by a great plurality and diversity of norms and standards directing us what to do. Undoubtedly there are plenty of oughts and shoulds according-to-norm-such-and-such, plenty of opportunities to commit mistakes according-to-norm-such-and-such, and so forth. The law, language, and games exemplify that indisputable fact best. It is unlawful in the United States to hoist any flag higher than the Stars and Stripes. Yet whether authoritative reasons, not just officials, decree that I must comply in my own disclosed yard is unclear and represents a matter wholly distinct from the recognition that I would act contrary to the law if not. The primary question, therefore, is not about norms per se but about what status norms and their implicit standards of correctness have. What radical meta-normative skepticism contests is that any norm has the status of commanding with objective authority, the status of entailing objective normative reasons to take seriously and follow their demands.

Meta-normative skepticism represents the counterpart to all forms of meta-normative realism, which hold that certain norms are authoritative, true, and correct, period, not just reflectively endorsed as correct or correct according-to-yet-another-norm; not just socially accepted and followed, but when accepted and followed then rightly so. They specify “what we ought to do simpliciter san phrase.” (Darwall 1992:156) Several philosophers have recently spent great efforts showing skepticism untenable. The problem stems from skepticism’s denial of normative authority tout court. Unlike more restricted versions, radical normative skepticism challenges not only the authority of particular norm-domains, such as morality, but finds queer the very idea that anything could ever issue authoritative demands: including rationality, instrumental reason, and epistemology. The question, however, is whether such a global form of skepticism represents a coherently defensible position that does not promptly undermine itself when seriously advocated. There are two powerful transcendental arguments challenging exactly that. Here is number one. Settling what to accept appears a reason-guided enterprise, including whether to accept skepticism. Yet skeptics cannot coherently support their view by citing authoritative reasons in their favor after they just rejected them throughout. What, then, are they doing when recommending their view? Will they not have to become silent, ushering themselves off the philosophical scene? “Skepticism that is the product of an argument cannot be total,” Thomas Nagel writes (1997:19). And here is number two. We are essentially deliberative creatures who ponder what to do. This characteristic of ours is none we are prepared to abandon. How even could we? Yet a sine qua non to the very point and purpose of deliberation appears to be that there are correct answers to our deliberative questions, answers we seek to discover rather than invent. Deliberation requires reasons, and thus our confidence in the sensibility of the former should translate into some confidence into the latter. I seek to rebut the effectiveness of both transcendental argument by demonstrating first how to support skepticism without deserting its tenets, and secondly how to deliberate in skeptical fashion.

In a nutshell, here is how the skeptic can defend his position coherently. He must acknowledge, first, that philosophical argumentation essentially is a norm-driven enterprise, and concede that he cannot appeal to authoritative norms and reasons to make their case, but show
why this concession does not compromise the employment of the requisite norms. He must show, in particular, how the norm-guided enterprise of philosophical argumentation can be carried out after authority is denied across the board: on the basis, namely, of norms void of objective authority! This strategy questions the presupposition that the issues of which norms to use and which to regard as authoritative are identical, whereas rejecting a norm’s authority does not amount to rejecting the norm itself. Skeptics contest that we must settle which norms to use in virtue of some mysterious property norms possess, some quality-seal mandating acceptance. The question is essentially a practical one and distinct from whether norms are objectively authoritative. Our fundamental practices have no internal exit strategy, no route out from within. Yet the skeptic never had any intention to exit the realm of norms. His position only challenges a particular interpretation of our commitments to norms: that we may employ them only when we can recognize them as objectively correct. For skeptics, the non-existence of authority is in fact no big deal: authority is something we neither have nor need.

David Enoch has recently developed a powerful adoption of the transcendental strategy in support of robust normative realism, which he defines as the view that there are response-independent, non-natural, irreducibly normative truths, perfectly objective and universal ones, that when successful in our normative inquires we discover rather than create or construct. (2003:21) The core concept in Enoch’s argument is deliberative indispensability which he juxtaposes to explanatorily indispensability. Enoch argues that robustly understood normative truths are indispensable for deliberation, an enterprise we have no choice but to take seriously, and that this should ground some confidence that there are such truths. In response I demonstrate how to deliberate in skeptical fashion by developing a presuppositional and norm-based approach. Deliberation is an activity that proceeds on the basis of our accepted norms and concerns. Deliberation cannot occur in a vacuum but instead requires a rich presuppositional context. We deliberate when, driven by our commitments, we actively participate in the selection and evolution of our norms and practices, searching for common ground for how to coordinate our individual and joint endeavors. The deliberative enterprise is not unconstrained and is answerable to standards of correctness as it is carried out within a tight web of norms that we do already accept, a web we continuously spin and expand.

- Comments: Mark T. Nelson (Westmont College)

28. Sam Kerstein (Maryland) and Greg Bognar (NYU): “Saving Lives and Respecting Persons”

In the allocation of resources, persons must be respected, or so many philosophers contend. Unfortunately, they often leave it unclear why a certain allocation would respect persons, while another would not. We explore what it means to respect persons in contexts in which scarce life-saving resources must be distributed.

We focus on two sorts of cases. We assume in both that we have to allocate a life-saving resource between different persons. Each person needs and wants to get the resource. But since the resource is scarce, we cannot make it available to all. Moreover, our helping these persons (or a subset of them) is not simply an act of beneficence on our part. Each person has a claim on the resource in the relatively weak sense that it would be wrong for us to refrain from giving it to her on morally arbitrary grounds (for example, because we could enrich ourselves by giving it to someone else). Finally, no person in our cases is morally responsible for her need of the resource in any way that would affect her claim on it.

In our first case, we have one indivisible life-saving drug and two patients who are identical in every relevant respect except that one of them is 20 years old and the other one is 70 years old. The patient who does not get the drug will die. If the younger person gets the drug, she will live for many years yet; if the older person gets the drug, she will die of natural causes in a few years. We call this the different age case.
One consideration in favor of giving the drug to the 20-year old patient is that it does more good: since she will live longer, saving her life creates a larger benefit. Another is that the older patient has had a longer life: she has already had her “fair innings.” It is not clear, however, that giving priority to the younger person is compatible with the idea of respect for persons. It seems that we respect the older person’s wish to go on living less than the wish of the younger person. Giving priority to the younger seems to involve not treating both patients with equal respect, although it seems to be the recommended course of action on both benefit maximizing and fairness grounds.

In our second case, we have to decide whether we save one person or five persons from certain death. Perhaps there were two traffic accidents, and one person was injured in the first and five persons were injured in the second. If we save the one person, the five will die; if we save the five, the one will die. (We assume that there are no relevant differences between the one and the five, e.g., in terms of their health or longevity.) We call this the different number case.

On benefit maximizing grounds, we should save the five. But some philosophers argue that it would be unfair to give no chance to the one: if we did that, we would fail to show proper respect for her—perhaps because her claim is not taken into consideration at all. On this view, benefit maximizing and fairness considerations point in different directions. Others argue that it is possible to save the five persons while giving no chance to the one person without failing to show proper respect for her—perhaps because her claim is taken into consideration just as much as those of others, but it is outweighed by them in one way or another.

We consider three accounts of respect for persons. An “equal worth” account holds that an action respects a person if and only if it shows proper respect for both the person’s well-being and her autonomous will concerning how her own life should go. A Kantian account contends that an action respects a person just in case it expresses respect for the person’s dignity, that is, her unconditional and incomparable value. Both the equal worth and Kantian accounts have implausible implications regarding the different age and different number cases, we contend.

We develop an alternative, “three tiered” account according to which an agent’s action expresses respect for persons if, first, in performing it he does not treat anyone merely as a means and, second, the action is of a sort which, it is reasonable for the agent to believe, would maximize persons’ preservation. In the paper, we summarize an account of treating others merely as means that we have defended elsewhere. We focus on the notion of maximizing persons’ preservation.

We can preserve persons along two dimensions. First, we can preserve a person by extending the period of time in which a being has the capacities that make up personhood. Let us call preserving persons along this dimension preserving “person years.” Second, we can preserve persons by maintaining them in existence. Let us call doing so preserving persons along the “person numbers” dimension. If we save five people for three years, then on the person years dimension we preserve 15 years, while on the person numbers dimension we preserve five people. Reflective common sense values the preservation of persons along these two dimensions.

But what might it mean to maximally preserve persons? There can, of course, be cases where maximizing preservation on one dimension fails to do so on the other. For instance, suppose we can save one person for 20 years or five persons for 2 years each. Choosing to save the one person would best preserve personhood along the person years dimension; choosing to save the five would best preserve personhood along the person numbers dimension. We develop a new way to proceed in such vexing cases, which we call the “comparative proportion procedure.”

In sum, we argue that unlike the equal worth and Kantian accounts, our three-tiered account of respect for persons has plausible implications in the different age and different number cases.

There has been recent discussion concerning the success of tracing as a way of accounting for an agent’s responsibility in particular cases. Tracing is invoked in cases in which it is intuitive that the agent is morally responsible for some occurrence yet the agent lacks responsibility grounding agency at the time of the occurrence. Responsibility grounding agency, as it is used here, is a placeholder for whatever capacities or other features of persons that distinguish us from non-moral agents, whether this involves guidance control, libertarian free will, alignment of one’s will with one’s second-order volition, or some other condition.

Tracing occurs when what is thought to explain one’s responsibility for some event that one does not express responsibility grounding agency over is one’s exercising the relevant agency over some prior event suitably related to the current one while satisfying some epistemic condition at that prior time. Consider the standard drunk driving case: Steven freely and knowingly becomes inebriated at a party and chooses to drive home. On his way home he runs over a child in the street. Surely Steven is blameworthy for the death of the child. But we can stipulate that he lacked responsibility grounding agency at the time of the accident, for he was too drunk. The reason he is responsible for the accident, the proponent of tracing holds, is that we can trace back from the untoward event, the death of the child, to his decision to get drunk and drive. And at this point he both possessed the relevant agency and could reasonably be expected to believe that his action would put others at risk.

The notion of tracing, which can be traced back to Aristotle, is thought by many to be an essential component of a theory of responsibility. But it has recently come under attack. Manuel Vargas presents cases in which we are expected to have the following three intuitions: (i) the agent is responsible for the event, (ii) the agent did not express responsibility grounding agency over the event, and (iii) there is no “suitable” prior time at which the agent did express the relevant agency and satisfied the epistemic condition. Fischer and Tognazzini come to the defense of tracing. In each of Vargas’s cases they argue that we should deny at least one of the three intuitions (i)-(iii).

They maintain that tracing is not as problematic as Vargas thinks and they “do not see how a theory of moral responsibility could adequately handle the range of drunk-driving cases, “Martin Luther cases”, and manipulation cases without some sort of tracing component; tracing just seems both highly plausible and theoretically indispensible” (553). My project here is to explore a strategy that explains the drunk driving cases without appeal to tracing. That is, I hope to show that with respect to the drunk driving cases tracing is, in fact, dispensable.

The drunk driving cases and ones relevantly like them have the following features:
(a) The agent did not exercise responsibility grounding agency over the consequence E.
(b) The agent did express responsibility grounding agency over action A.
And the proponent of tracing holds that since
(c) The agent could reasonably be expected to believe that A would result in E,
(d) The agent is blameworthy to degree d for the event E.
The strategy advocated in this paper, however, involves denying (d). This may seem counter intuitive but it seems less so when it is emphasized that while (d) is false the following is true:
(d’*) The agent is blameworthy to degree d for action A.
I argue that the tracing cases should be understood as cases involving responsibility for consequences and that the consequences of one’s actions are only epistemically relevant to responsibility. I then argue for the strong claim that agents cannot be blameworthy for the consequences of what they have done. For consider the following argument:

(1) If something is independent of one’s blameworthiness, then one cannot be
(2) The consequences of one’s actions are independent of one’s blameworthiness.
(3) Therefore, one cannot be blameworthy for the consequences of one’s actions.

The first premise has the ring of analyticity. The second premise can be shown to be true when one considers cases in which the consequences of one’s action that occur are not expectable. That is, it is easy to imagine cases in which we can hold the consequence C of some action fixed but alter the agent’s blameworthiness B. And we can also imagine cases in which C is altered but the agent’s blameworthiness B remains constant. This shows that the occurrence of some consequence is independent of one’s responsibility. Having argued that the consequences of one’s actions are not related to moral responsibility in the way that many think, the paper closes with a discussion of the ways in which consequences can matter.

• Comments: Stefan Sciaraffa (McMaster)

30. Matt King (Maryland): “Moral Responsibility and Merit”

In the recent debate regarding free will and moral responsibility, there is a tentative consensus being reached. We have begun to settle on a shared target concept to be explained. Moral responsibility is to be understood as “responsibility in the desert-entailing sense”. This is to distinguish it from causal or legal responsibility, and draws it closer to our other moral concepts. Moral responsibility and desert are natural partners: morally responsible agents can be blameworthy and praiseworthy; they can deserve blame and praise. While this is a welcome development, it is the more striking that so little has been said about what this desert relation is. One potential worry is that without saying more about desert, we risk merely replacing one difficult concept (moral responsibility) with one just as difficult (desert). At the very least, adequately evaluating extant accounts of moral responsibility seems to require explicitly engaging with the notion of desert.

This paper seeks to address this lacuna in the moral responsibility literature. I propose an understanding of desert sufficient to help explain why the blameworthy and praiseworthy deserve blame and praise, respectively. I do so by drawing upon what might seem an unusual resource. I appeal to so-called Fitting-Attitude accounts of value (hereafter: FA accounts) to help inform a conception of desert or merit, one that can be usefully applied to discussions of moral responsibility. If successful, the candidate view I sketch will achieve two goals: (1) it will fill out a bit more substantively how moral responsibility and desert are related; and, (2) it will provide a strategy for showing the compatibility of moral responsibility and various naturalistic theses about the world (including determinism).

The outline of the view connects three thoughts. First, theories of moral responsibility need some positive account of the notion of desert. Second, one natural way to talk about responsibility is in terms of blameworthiness and praiseworthiness, terms that seem to call for FA treatments. Third, FA accounts rely on there being some normative relation holding between the properties being explained and the relevant attitudes. By suggesting desert as the normative relation operative in such an FA account, we get a positive view of desert: blameworthy (or praiseworthy) agents deserve or merit blame (or praise) in the same way that the admirable deserves admiration or the fearsome merits fear. This view, while not spelling out the conditions on moral responsibility, does connect that notion with a notion of desert. Moral responsibility remains a normative notion and the relation in virtue of which agents deserve blame or praise. The concept thereby remains significantly unrevised. The explanation, in turn, is quite simple. Blameworthiness and praiseworthiness are normative properties that should be given FA analyses, those analyses employ a normative relation between the properties and relevant attitudes, and that normative notion is desert or merit.

While FA accounts of blameworthiness exist (e.g., Wallace, Darwall), this candidate view does better than extant accounts on at least three scores. First, the candidate view treats all the
relevant normative properties the same. Not only do they all get FA accounts, they get FA accounts in terms of the same normative relation: desert/merit. Extant FA accounts of blameworthiness construe the fittingness of blame in terms of fairness, which looks particularly ill-suited for the appropriateness of other normative properties. There’s nothing unfair about failing to fear a coiled copperhead at one’s feet or about admiring a despicable desperado.

Second, unlike the candidate view, extant FA accounts of blameworthiness aren’t naturally extendable to praiseworthiness. If A praises B for something despicable, this doesn’t seem to be unfair in any pressing sense, and it certainly isn’t unfair to B. Yet it seems as though blameworthiness and praiseworthiness ought to be given parallel analyses, at least insofar as relating them in the same way to their respective attitudes.

Finally, the candidate view under consideration faces no special problem regarding the wrong kinds of reason test. Relying on an independent notion like fairness can all-too easily lead to wrong kinds of reason cases when we introduce a consideration that clearly bears on fairness but not on blameworthiness (see Smith 2008). It may be unfair of me to blame Mark for stealing my car because I’m an unrepentant car thief. But this no more affects Mark’s blameworthiness than does the evil demon’s threat affect an object’s desirability. Moreover, this instance of the wrong kind of reason problem may remain even were a solution to the more general version found. In contrast, the candidate view relies only on the same normative notion present in all FA analyses, and thus is solved by whatever solution works generally.

Though the candidate view may be preferable independently, it also provides a compatibilist strategy to explaining moral responsibility. That is, it gives us a new avenue for showing responsibility to be compatible with various naturalistic theses, especially the thesis of determinism. We wouldn’t think any plausible naturalistic theses about the world undermines the claim that the fearsome deserve/merit fear or that the notable deserve/merit being noted. If the blameworthy deserve/merit blame in the same way, as the candidate view suggests, then we have made the initial case for compatibilism. Determinism would no more upset ascriptions of blameworthiness than ascriptions of fearsomeness or admirability.

- Comments: Steve Sverdlik (Southern Methodist)


Are there things we should value because they are, quite simply, good? Not because they are good for anyone. Nor because they are good things of a kind (good novels, good watches, good friends). Rather, for no reason other than their being good—period?

G. E. Moore thinks so. “The only possible reason that can justify any action,” he says in *Principia Ethica*, “is that by it the greatest possible amount of what is good absolutely should be realised.” I set aside Moore’s claim that this is the only possible way to justify an action, because I want to concentrate on what he presupposes: that there is such a property as absolute goodness, and that it is a reason-giving property.

Geach, followed by Foot and Thomson, argue that there is no such property, but they take a semantic route to this conclusion: they think that such sentences as “pleasure is good” fail to have meaning. But I doubt that this criticism of Moore can succeed.

Other philosophers are tempted by a “fitting attitude” analysis of intrinsic value. For something to be bad (period), according to this approach, is for it to be the case that one ought to oppose it in some appropriate way. But I am inclined to say, against this, that when we assert that another persons’ pain is bad, a reason is being offered for compassion. Our statement does not recommend an attitude, but serves to ground an attitude: we should be compassionate because pain is bad.

Scanlon says that when a thing has value, its being valuable does not itself constitute a reason for valuing it. He holds that whenever something is valuable, it has some feature that makes it valuable, and *that* feature is what provides a reason in favor of valuing it, not the mere
fact that it is valuable. But Scanlon’s “buck passing” analysis of value is problematic in several ways. First, he would do better to distinguish several different kinds of evaluations: (a) judgments that something is good of a kind (e.g., a good play); (b) judgments that something is good for someone; (c) judgments that something is good (period). It is implausible to suppose that in the first two types of judgment, “good” passes the buck. (That something is good for you is often a reason for you to do it. That the soup in your bowl tastes good – has a good taste – is often a reason to try it.) It is more tempting to give a buck-passing analysis of the absolute use of “good.” Second, even in this case, Moore’s way of thinking about goodness (as a reason-giving property) cannot easily be dismissed. When one has a sensation of pleasure or pain, isn’t that phenomenon apt for evaluation? Are not “good” or “bad” appropriate evaluations of them? Should we not avoid pain because it is bad?

I believe, therefore, that a longer route must be taken to reach the conclusion – if it can be reached at all – that Moore’s thesis should be rejected. What we must do, to put that thesis to the test, is to see whether we need, for purposes of sound practical reasoning, not only the concept of being good for someone, but in addition the concept of absolute goodness.

One possibility is that being good for someone consists in having two properties: first, it is partly constituted by being good (period); second, it partly consists in having some further feature (to be specified.) If that analysis were true, and if the best explanation for the reason-giving force of “good for” is that “good” (period) has reason-giving force, Moore’s thesis would be vindicated.

But I reject this compositional analysis of “good for.” If smoking is bad for George, that is not made true by, first, the badness (period) of smoking, and, second, something else having to do with George. So, I am led to the conclusion that “good” and “good for” should be treated as names of two independent primitive concepts. Accordingly, if absolute goodness and being good for someone both play reason-giving roles in our thinking, they do so independently of each other. There can be two different reasons to avoid pain, then: it is bad (period), and it is bad for someone.

These two kinds of reasons might work in cooperation or against each other. We might have a reason to do something because it is good for us, and all the more reason because it is good (period). Or we might have a reason to do what is good because it is good, even when it is bad for someone – perhaps even bad for everyone involved.

The remainder of my argument is an attempt to find convincing examples of sound practical reasoning in which absolute goodness provides a reason – either a reason that cooperates with others to strengthen the grounds for a conclusion, or a reason that defeats others. I claim that there are no such examples.

What sorts of things might plausibly counted as good (period)? Perhaps knowledge, pleasure, virtue, friendship, beauty. But these sorts of things might also be said to be good for those who have or share in these things. Do we then have two kinds of reasons to establish and sustain friendships – because doing so is good for us and others, and because doing so is good (period)?

What sorts of things should be valued because they are good (period) – even if enacting that valuation will not be good for anyone, and might be quite harmful? Several items might be proposed: persons, equality, biodiversity, beauty. I cannot explore all of these possibilities, but will briefly examine one of them: the alleged intrinsic value of persons. My thesis is that we can do justice to the commonsense idea that persons have an importance that other animals lack by attending to the higher degree to which a human life can be good – not absolutely – but for the one who is living it.

- Comments: Gwen Bradford (Rice)
The Doctrine of Double Effect (DDE) states roughly that there are stronger moral reasons against inflicting harm as a means to an end than there are against causing harm as a byproduct. On one reading this is a claim about reasons not to act with certain intentions, while on another it speaks against performing acts with a certain objective causal structure. While there are good reasons to prefer the objective reading, some have dismissed it as absurd. In this paper I defend the objective reading against this dismissal, which seems to depend largely upon the conflation of criteria of objective wrongness with other criteria. I present an objective formulation of the DDE according to which a harm’s bringing about a particular good weakens the extent to which that good counts in favor of bringing about the harm. This solves several problems and helps explain an intuitive injunction against benefitting at others’ expense.

The DDE offers us a plausible way to unify a wide variety of non-consequentialist intuitions that cannot be explained by other deontic distinctions like that between doing and allowing harm. These include the intuitive moral differences between diverting threats and using people as shields, collateral damage and terrorism, and failing to aid someone due to a resource shortage and failing to aid her so as to benefit from her misfortune. But the DDE has often been sub-optimally formulated and applied in implausible ways. I thus begin with some ground clearing, arguing that we should interpret the DDE as attributing heightened reasons against bringing about harms to some beings as a means of preventing harms to others when all else is held equal. So interpreted the DDE does not speak against non-harmful killing or harming someone for his own greater good. The Doctrine also applies to non-lethal harms, making absolutist versions non-starters. Finally, the DDE speaks against harming as a means only when ceteris is paribus – plausible defeaters include consent, culpability, and obligation to incur harm.

The DDE is often interpreted as saying that there is stronger moral reason against intending harm as a means than doing things that one merely foresees will result in harm. But, as authors like Bennett, Gibbard, Thomson, and Scanlon have argued, what we have moral reason to do does not depend upon our intentions in doing it (at least in the cases alleged to support the DDE). You can do the right thing for the wrong reasons, for instance if you were to divert a trolley from hitting five to hitting one out of spite for the one. While our blameworthiness after the fact might depend upon our intentions in acting, what we should do in the first place does not.

But, I argue, the intuitions in favor of the DDE are in the first instance forward-looking intuitions about what we should do. As such, they most directly support an interpretation of the Doctrine according to which there are stronger reasons against performing acts with a certain intention-independent causal structure, namely those that bring about good as a result of bringing about harm. While this objective DDE makes our moral reasons independent of our intentions, it can explain why it seems blameworthy to intend harm as a means. Failing to respond to genuine moral reasons contributes to one’s blameworthiness and disestimability, and intending harm as a means reflects a failure to respond to the moral reasons that exist according to the objective DDE.
Several authors have, however, dismissed the objective DDE as absurd. The best case for such dismissal is given by Alistair Norcross, who presents a scenario in which Moe does what his evidence suggests will save five people but kill Homer as a side effect. Having the same evidence as Moe, the Pope awards Moe a Seal of Approval. But Moe’s evidence about a triggering device was misleading, and his act actually saved the five as a causal consequence of killing Homer. When the Pope learns of this, she indignantly retracts her Seal. Norcross concludes from this that it is just too implausible to suppose that the mechanics of a triggering device could make “a moral difference” without making a difference to either Moe’s mental states or who lives or dies.

I argue that Norcross’s conclusion would be unsubstantiated in the absence of his story about Moe, but that this story is simply an invitation to conflate criteria of objective wrongness with criteria of subjective wrongness, blameworthiness, and estimability. The objective DDE, like actual-consequence formulations of consequentialism, offers a (partial) criterion of objective rightness, which we are to try to approximate by using our evidence to determine the likelihood that our acts will have the properties of which the theory speaks. We merit praise and blame to the extent that we responsibly succeed or fail to do (for the right reasons) what our evidence suggests is objectively right. The Pope’s failure to recognize this is no evidence against the objective DDE.

The objective DDE might be interpreted as making the implausible claim that it’s worse to bring about harm if the harm does any good. But a better formulation attends to Jonathan Dancy’s distinction between considerations that oppose acts and considerations that weaken other reasons. While the fact that one has promised to do A is a reason to do A, the fact that the promise was given under duress weakens this reason without itself counting against doing A. I argue that we should understand the DDE as claiming that a harm’s bringing about a good weakens the extent to which that good counts in favor of bringing about the harm. This interpretation solves problems in the literature about whether the good consequences of an act otherwise permitted by the DDE could make it forbidden, or change its status from supererogatory to obligatory. It also helps explain the intuitive plausibility of significant moral barriers against benefitting some at the expense of others in terms of such benefits losing some of their status as things that we have intrinsic reason to confer.

- **Comments:** Alastair Norcross (Colorado)


Error theorists (most prominently Mackie, Joyce) are committed to two claims. As Mackie articulates them: (1) “Ordinary moral judgments include a claim to objectivity, an assumption that there are objective values,” (Ethics 35) and (2) “There are no objective values” (alternately put, “moral scepticism”; 15 and passim).

The climate of philosophical opinion has changed markedly since Ethics was published. In 1977, non-naturalism was not a popular, widely advocated, or even (Parfit plausibly suggests) widely understood position in metaethics. But in recent years there has been a resurgence of non-naturalist thinking, the development of a new non-naturalism (Nagel, Parfit, Scanlon, Dancy, Shaffer-Landau). My project in this paper is to explore the implications of the new non-naturalism for error theory: In the light of the new non-naturalism, how is error theory to be understood, and how plausible a case can be made for it?

The obvious first thought is that the new non-naturalism makes the error theorist’s job in one way easier and in another way harder: it is easier to argue for (1), because it is a commitment the non-naturalist and the error theorist share, but harder to argue for (2). I argue that this first thought turns out to be roughly right, but that, though it is in some ways harder to argue for (2) in the light of the new non-naturalism, a promising case for (2) and thus for error theory can still be made.
I begin with (1). It is extremely tempting to understand the (putatively problematic) claim to objectivity attributed to ordinary moral judgments in (1) as the claim that there are genuine normative reasons as non-naturalists understand them. For Mackie himself appeals to the older non-naturalist tradition, to Sidgwick and Moore, in articulating the idea of an objective claim to objectivity. Thus Sidgwick is, along with Plato and Kant, one of his three exemplars of the claim to objectivity in “the main tradition of European moral philosophy” (30). A little later he remarks:

No doubt it was an extravagance for Moore to say that ‘good’ is the name of a non-natural quality, but it would not be so far wrong to say that in moral contexts it is used as if it were the name of a supposed non-natural quality, where the description ‘non-natural’ leaves room for the peculiar evaluative, prescriptive, intrinsically action-guiding aspects of this supposed quality. (31-2)

But the evidence here is complicated. Derek Parfit diagnoses much of the moral philosophy of the 70s, including Mackie’s work, as betraying a failure even to understand the non-naturalist concept of a normative reason. If Mackie doesn’t understand the concept of normativity, he presumably does not take it to be part of ordinary moral judgments. I argue first that Parfit is here uncharitable to Mackie: that though Mackie is not immune to the influence of the climate of philosophical opinion in which he wrote, and so is sometimes disposed to understand normativity as just motivational force, his argument from queerness is best reconstructed as having as its target the queerness of genuine normativity as non-naturalists understand it. So understood, Mackie and non-naturalists share a commitment to (1); the new non-naturalism helps understand and articulate the first component of error theory. This has important implications not just for Mackie but for Joyce: it means that the argument for error theory Joyce offers in The Myth of Morality, relying as it does on a commitment to Williams’s views about normativity, is in important ways misconceived.

I then turn to the arguments for (2), moral scepticism. Mackie provides the following brief and illuminating summary of the case as he sees it:

The considerations that favor moral scepticism are: first, the relativity or variability of some important starting points of moral thinking and their apparent dependence on actual ways of life; secondly, the metaphysical peculiarity of the supposed objective values, in that they would have to be intrinsically action-guiding and motivating; thirdly, the problem of how such values could be consequential or supervenient upon natural features; fourthly, the corresponding epistemological difficulty of accounting for our knowledge of value entities or features and of their links with the features on which they would be consequential; fifthly the possibility of explaining...how even if there were no such objective values people not only might have come to suppose that there are but also might persist firmly in that belief. (49)

Of these, the first is the argument from relativity; the second, third, and fourth are elements of the argument from queerness. In the light of the new non-naturalism, only the second retains real force. What is problematic about objective values is the idea that they are intrinsically action-guiding. In seeing this intrinsic action-guidingness as metaphysically problematic, the error theorist is committed to a kind of metaphysical naturalism typically shared by proponents of other prominent views in metaethics: non-cognitivism (e.g. Blackburn, Gibbard) and moral naturalism (Brink, Railton, Sturgeon). The argument from queerness turns out really to be an argument against intrinsic action-guidingness by appeal to metaphysical naturalism.

Non-naturalists sometimes respond to this argument by, in effect, rejecting metaphysical naturalism (Nagel, Scanlon). Error theorists can here largely rely on an appeal to companions in guilt: they can be comfortable in allowing that the argument from queerness is no more plausible than the metaphysical naturalism widely accepted in metaethics and elsewhere. The more interesting and specific question error theorists face is whether it is specially problematic both to allow, as I argued error theorists should in articulating (1), that there is a coherent (non-naturalistic) concept of a normative reason, and to deny that there are in fact any normative reasons. I argue that this combination is not specially problematic, but that it does
have important implications for the status error theorists should attribute to (2). Contrary to what some interpreters have held, error theorists should claim that (2) is only contingently true.

- **Comments**: Ryan Hay (USC)

37. **Matjaž Potrč and Vojko Strahovnik (U Ljubljana, Slovenia)**: “Moral Dilemmas and Vagueness”

- **Comments**: Scott Forschler (Northland Community and Technical College)

38. **Mark Rosner (Queen’s University)**: “The Irrationality of Aktratic Action”

Recent debates concerning aktratic action have questioned their traditional construal as a paradigm instance of practical irrationality. According to the traditional view, aktratic action, or intentionally acting against one’s own better judgement, is always a paradigm instance of irrational action. Insofar as it is possible for an agent to flout her own judgement as to what is the best course of action, such an agent is, by necessity, practically irrational.

The irrationality of aktratic action stems, it seems, from the fact that an agent flouts her own evaluative judgement as to what is best in the circumstances when forming an intention regarding how to act. That is, there is the general assumption that the rationality of action entails that the formation of our intentions normally tracks what we take to be good; acting intentionally means, in part, acting in light of what an agent took to be good or have value in her situation. These claims regarding the nature of acting intentionally, understood as acting for a reason, and their connection to what an agent takes to be good in her situation seem to entail that acting aktratically just is irrational from the agent’s perspective. An agent ignores what she herself claims to be the reason for which she should act.

Writers such as Alison MacIntyre and Nomy Arpaly have questioned this thesis, claiming aktratic action need not always be irrational. There can, they hold, be cases of non-irrational or arational aktratic action. What’s more, because of this fact, the possibility arises that aktratic action can sometimes be the rational option to pursue or the most rational thing to do in a particular situation. Far from being the paradigm of practical irrationality, acting against our own better judgement turns out to be a kind of rational response, on occasion, when deciding how to act.

Drawing on cases of so-called ‘inverse akrasia’, they hold that when agents are sensitive to features or facts of a situation which outstrip their more conscious deliberations (e.g. their judgement as to what is best in the situation) acting aktratically on the basis of such sensitivities can be the most rational thing to do. The stock example of such arguments, the case of Huckleberry Finn, seems to emphasize that there are agents who can be sensitive to considerations they themselves would recognize as the most rational course of action, but of which they are only dimly aware at the time of acting. Relying on what we might term their more standing judgements as to what is best at the time of acting, agents such as Huck Finn are conflicted just because they sense that what they think they ought to do is incorrect, by their own lights. Inasmuch as Huck aktratically acts on that dimly felt reason, his action displays a form of rationality as he flouts his own best judgement in order to do what he would recognize to be the best thing to do in the situation, had he deliberated in a more adequate manner.

I argue in this paper that such a conception of the rationality of akrasia is in tension with our most plausible theory of the nature and explanation of action. Drawing on Joseph Raz’s writings on the nature of agency, I claim that the account forwarded by MacIntyre and Arpaly undermines their own commitment to the explanation of intentional action via the idea of acting on a reason. Raz’s work focuses on the features that make action intelligible from the point of view of the agent as the key to explaining action. Now this explanation is not merely a collection of the causal set of events which led to the movement of the agent but rather an explanation of the facts the agent took to make the action an intelligible object of choice. What made the action an intelligible object of choice was the fact that there were considerations which showed the action
to be good in the eyes of the agent. The agent’s reasons, then, for which she acted, were some seeming ‘good-making’ characteristic which made the option they were deliberating about the one to choose. Those reasons may not have required one action or another but they did render the agent’s options eligible in her eyes – one’s that might be worth pursuing.

This story about the nature and explanation of action commits us to the irrationality of akrasia due to the requirement that action be intelligible from the agent’s own perspective in order for it to qualify as intentional action. I argue that the commitment that both Arpaly and MacIntyre both express to the idea of intentional action as acting on a reason cannot vindicate the rationality of akrasia and would rather require jettisoning our most plausible theory of rational agency.

In addition, borrowing a distinction from T.M. Scanlon, I argue that we can make better sense of the cases of ‘inverse akrasia’ by making a distinction between irrationality narrowly construed and the most rational thing to do. Such a distinction preserves Raz’s story of rational action while also explaining why those cases seemed to be counter-examples. What’s more, the distinction helps identify the specific nature of the irrationality associated with akrasia and allows for a more adequate treatment of this form of practical irrationality.

- Comments: Matt Kisner (South Carolina)

39. Abe Roth (Ohio State): “Intention, Shared Activity, and Team Reasoning”

A central concern in the theory of collective intentionality is understanding how joint or shared activity is related to the agency of participating individuals. But when we try to make use of an important insight about human agency – namely, that intentions are commitments that settle practical issues – our theory of shared activity is threatened with inconsistency.

On the one hand, when we do something together I am not just engaged in and committed to what I do. That would be the case of doing something on one's own. Rather, I am in the relevant sense committed to the entirety of the activity, to what we’re all doing. As Searle puts it, I am pushing the car only as a part of our pushing. It is natural to try to understand this participatory commitment in terms of intentions: I intend for us to be dining together or I intend to dine with you or, to use Bratman’s locution, I intend that we dine together. Appeal to such intentions would go a long way toward accounting for the characteristic coordination of participants in shared activity, and of the responsiveness of each to the other(s).

On the other hand, adopting an intention is something I do to settle a deliberative issue: weighing several options, I decide on A-ing, and thereby intend to A. My intentions provide fixed points to structure subsequent practical reasoning and action: what means to A-ing should I take? What else must I intend to secure those means? How may I pursue other ends in ways that are compatible with A-ing? This suggests the so-called Settling Condition that I can only intend what I take to be up to me to decide or settle. It is a violation of a rational requirement on intending if I were to intend something I don’t think I can settle, and thus regard my ensuing plans and actions as likely coming to grief. Applying the point to collective action, to say that I intend for us to be dining together presumes that whether we’re dining together is something for me to settle. But the idea behind shared activity is precisely that it's not entirely up to me what we do. Our problem, then, is that shared activity would seem both to demand and to disallow one and the same intention on the part of each participant.

One strategy for solving this problem suggests that participatory commitment does not require an individual participant to intend the entire activity. It will suffice for the participant simply to intend her part of the activity. My critical remarks will focus on a sophisticated version of this strategy, which makes use of a distinctive form of strategic reasoning. Team reasoning has been invoked to suggest how an individual might count as rational in opting for more cooperative behavior in scenarios discussed in game theory, such as the Prisoner’s Dilemma and Hi-Lo. Some have proposed the theoretical redeployment of team reasoning to characterize the
intention to do one’s part in shared activity. On this approach, the intention is understood in terms of the team reasoning that is thought to lead to it. I will argue that this does not account for participatory commitment essential to shared activity. That’s because team reasoning fails to individuate any appropriate attitude of intending that could underlie participatory commitment; the thought that it does is illusory.

Recall that the original problem started with the thought that participatory commitment entails an intention regarding the entire activity. This in turn entails the exercise of some authority over fellow participants, settling what they will do. This is hard to reconcile with the activity being shared. The team reasoning approach is just one of several proposed solutions to this problem. Others involve appeal to conditional intentions, yet others to intentions based on predictions. What’s common to all these approaches is that they are incompatibilist in the sense that they accept that this authority to settle what others do cannot be reconciled with shared activity, and that an adequate account of the latter (in particular, of participatory commitment) can be had without any form of authority.

But should the incompatibility thesis be accepted? I argue that authority, properly understood, is compatible with shared activity. More specifically, suppose that the authority is symmetrical in that each participant can issue intentions to settle what fellow participants will do. Suppose further that the locus of responsibility for the defeat and/or revision of any intention issued on the basis of such an authority extends beyond the individual exercising the authority. I think that compatibilism in those circumstances is very plausible. I also make the case for thinking that that we often do indeed have this authority by arguing from a revised and broadened understanding of what it is to act for a reason.

• Comments: David Killoren (Wisconsin)

40. Brook Sadler (South Florida): “Acting Out of Character”

We sometimes say that a person’s actions were “out of character,” meaning, roughly, that her actions were inconsistent with her character. But is it really possible to act out of character? If so, what view of character will allow us to make sense of this idea? In this paper, I address these questions, retaining the ordinary observation that sometimes people act in ways that appear to be uncharacteristic of them without conceding to situationism that character is, at best, merely a matter of "local" traits (Doris, 2002). Ultimately, I defend the idea that the practice of attributing to people character traits such as "honest," "kind," or "cruel" is a legitimate practice for a limited set of purposes, but that a richer notion of character that is not reducible to such terms, even to a set of them, provides a more accurate understanding of human personality and best accounts for the phenomenon of acting out of character. Finally, I suggest that this richer notion of character is compatible with key features of both virtue ethics and Kant’s ethics.

In some cases, it might be thought uncontroversial that people can and do act out of character. A person acting under the influence of powerful drugs or hypnosis may behave in ways starkly at odds with her typical behavior or her values. In such cases, where some of the hallmarks of agency—autonomous self-control, choice, reflective endorsement, volition—are absent, we might be reluctant to call the behavior "action" at all, so little does it seem to belong to, or arise from, the agent herself. Since such cases already display a highly attenuated connection between the agent and her behavior, I set them aside, focusing discussion instead on cases where the agent appears to meet the usual standards for autonomous action. For example, she acts intentionally, perhaps after some deliberation; she is aware of alternative courses of action; she possesses a reasonable understanding of the nature of her acts and the consequences of them; etc. Thus, by the usual measures, the agent appears to be the author of her actions. Nonetheless, she acts in a way that appears anomalous for her. It is this sort of case about which I want to ask whether it is really possible for an agent to act out of character.
I consider first a negative answer—it is not possible for an agent to act out of character—and discuss the assumptions about character assessment that seem to underwrite it. The negative answer suggests what I call an All-In view of character: A person’s character is constituted by all of her character trait-relevant behavior. Thus, even anomalous behavior goes toward establishing the agent’s character traits. One motivation for adopting the All-In view may be that there is no way to draw a line between trait-relevant behaviors that count toward trait-attribution and those that do not because, for the sort of cases we are considering, people appear to meet the hallmark conditions for autonomous agency that usually make us confident that their behavior emerges from or contributes to their character. Another motivation for the All-In view may be a concern with assigning praise and blame or holding persons responsible for their actions. Accordingly, the worry is that agents may be let off the hook for wrongdoing that, while fully autonomous, was not deemed characteristic. Finally, one might subscribe to the All-In view on the basis of a simple behaviorism: You are whatever you do, without exception.

Does the All-In view of character make sense? I will not argue that it doesn’t. Rather, I reveal some of the assumptions that the view involves and point to the limitations of this way of thinking about character. One might hold the All-In view on the assumption that certain acts are of such moment (seriousness, gravity, consequence) that even a single occurrence is sufficient to warrant character-trait attribution. High-impact actions contribute to character regardless of their frequency. For example, it might be argued that a single act of deliberate cruelty warrents the attribution of cruelty, no matter how anomalous for the agent who committed it. One problem with the act-impact assumption is that it might plausibly apply to certain types of act but not to others; the account would need to provide an explanation of which act types are high-impact and which are low-impact. For reasons I offer in the paper, I think we cannot successfully make a general distinction of this sort. Determinations of high- versus low-impact acts will require a more context-sensitive and detailed account of character than that which focuses on discrete traits.

Independent of the act-impact assumption, the All-In view may rely upon the idea that character-trait attribution is justified by the frequency with which the agent performs trait-relevant behavior relative to countervailing trait-relevant behavior (e.g., the frequency of honest acts relative to the frequency of dishonest acts). In order to make accurate trait attributions, we must include in our assessment of an agent all of her acts; otherwise, we get a distorted or misrepresentative ratio. Thus, even acts that appear to be out of character must be seen as constitutive of character, always contributing to the diachronic determination of traits. To be sure, the frequency assumption does fit well with important claims about the diachronic nature of character formation. However, I argue that there are serious problems with determining the frequency appropriate to trait-attribution. These problems reveal the need for a richer account of character.

I argue that the motivations for the All-In view of character can be accommodated without the problematic act-impact and frequency assumptions. A psychologically complex, narrative view of character that makes only limited use of character trait terms can best explain cases of acting out of character. This view allows us to see what is apparently anomalous in an agent’s action, without losing sight of the motivational, cognitive, and deliberative structures that give rise to it—and also without thinking that character is fundamentally fragmented or merely situational.

• Comments: Michael Bukoski (Arizona)

41. Jeff Sebo (NYU): “Constructivism about the Self”

What is the self? Does it exist prior to our having a self-conception? Or is it a product of our having a self-conception? In this paper, I argue that the latter answer, combined with empirical information about human psychology, has radical, but ultimately intuitively plausible,
implications: for example, it implies that a person can have many selves at once, and it also implies that two or more people can share a self.

First, I review Daniel Dennett’s argument that the self is the “center of narrative gravity”. Whereas a center of gravity is an abstract object that we posit to explain the behavior of material objects, a self is an abstract object that we posit to explain the behavior of persons.

Second, I present psychological evidence that many people have multiple personalities, in the non-pathological sense that we have different thoughts, feelings, and habits in different contexts. Specifically, we develop a single, unified personality early in life. Then we learn that we have to act differently in different social situations. For example, we have to act one way in front of our parents, another way in front of our teachers, another way in front of our friends, and so on. And in the process of mastering each of these roles, we become the character that we play in that role, in a method acting kind of way. But since these characters have different, and sometimes incompatible, thoughts, feelings, and habits, we also compartmentalize them from each other to a degree. And the upshot is that we split our single, unified personality into multiple, psychologically continuous personalities. Then, later in life, we settle into a smaller number of social roles, and so we either abandon some of our personalities, or fuse them back together again.

Third, I argue that this empirical information both supports, and finds support in, constructivism about the self. Specifically: this empirical information suggests that our thoughts, feelings and habits are much less stable than they seem to be. So why do they seem to be so stable? It must be because we take a (weighted) average of our thoughts, feelings, and habits over time, and then attribute this average to ourselves as a whole.

Fourth, I argue that if constructivism about the self is right, then we can have many selves at once. Compare: How many centers of gravity are in an object? Technically speaking, there are many. But how many does it make sense for us to posit for practical purposes? That depends on the object. For example, if an object has a single, unified structure (like a sword), then it makes sense for us to posit only one. But insofar as an object has distinct parts (like nunchucks), it makes sense for us to posit many: one for each part, and then another for the object as a whole.

Similarly: How many selves are in a person? Technically speaking, there are many. But how many does it make sense for us to posit for practical purposes? That depends on the person. If a person has a single, unified personality, then it makes sense for us to posit only one. But insofar as a person has multiple, psychologically continuous personalities, it makes sense for us to posit many: one for each personality, and then another for the person as a whole.

Fifth, and relatedly, if constructivism about the self is right, then two or more people can share a self. Compare: How many centers of gravity should we posit for a group of objects? That depends on the group. If the objects in the group have no gravitational pull on each other (like a group of marbles), then it makes sense for us to posit one for each member of the group, and leave it at that. But insofar as the objects in the group do have a gravitational pull on each other (like a group of magnets), it makes sense for us to posit many: one for each member of the group, and then another for the group as a whole.

Similarly: How many selves should we posit for a group of people? That depends on the group. If the people in the group are acting independently of each other (like two strangers walking next to each other), then it makes sense for us to posit one (or more) for each person, and leave it at that. But insofar as the people are acting together (like two friends taking a walk together), it makes sense for us to posit one (or more) for each member of the group, and then another for the group as a whole.

Finally, I argue that if all this is right, then constructivism about the self can provide a unified explanation for many features of our self-conceptions. First, it explains why selves seem to have such strange metaphysical features, e.g., they seem to come in degrees, survive fission and fusion, and survive temporal gaps. (It would be strange if “real” objects had these features, but it makes sense that abstract objects can have them.) Second, it explains why we seem to have many selves at once, and therefore it justifies a literal reading of claims like: “I’m sorry I lost my
temper last night. I wasn’t myself”; “He turns into someone else entirely around his friends”; “They became the same person when they started dating”; and so on. Third, it explains why we seem to have many selves but not a steady stream of momentary selves, as Galen Strawson implausibly suggests. Fourth, it implies that individualists and communitarians are both right: Persons have selves and social groups have selves – and we construct all of these selves in the same exact way. Finally – and I can only tease this conclusion here, although I defend it in detail elsewhere – constructivism about the self grounds a powerful theory of agency, according to which each self is a separate agent, with his own reasons, duties, and rights.

- **Comments:** Carl Hammer (CUNY-Baruch)

42. **Andrew Sepielli (Toronto): “Normative Uncertainty for Non-Cognitivists”**

It seems that normative judgments are doubly gradable. First, we can be more or less confident regarding normative matters. I'm more confident that the balance of reasons weighs against murder than that the balance of reasons weighs against abortion. Second, there are greater and lesser degrees of reason strength represented in judgment. I'm close to certain that the balance of reasons weighs very strongly against murder, and I'm also close to certain that the balance of reasons weigh weakly against being late for social appointments.

This shows up as a banal truth for the cognitivist about normative judgment, but as a problem for the non-cognitivist. For Michael Smith has argued that, if the non-cognitivist is right, then normative judgments don't have enough structure to be doubly gradable. His argument is essentially: The evaluative "oomph" of normative judgment that cognitivism assigns to the content of a mental state – i.e., belief – non-cognitivism assigns to the mental state itself. And so the strength of reasons that the cognitivist can account for in terms of gradable properties represented in the content of a judgment, the non-cognitivist can account for only in terms of the judgment itself's being gradable. But the gradability of the judgment should correspond to the agent's degree of confidence, as indeed it does on the cognitivist picture. So cognitivism is a more satisfactory view than non-cognitivism when it comes to representing gradable judgments about gradable features. Non-cognitivism can accommodate either degrees of judgment or degrees of reason strength represented in judgment, but not both. From this Smith concludes that we must reject non-cognitivism.

There are two camps with an interest in challenging Smith's argument. Recognizing that there seem to be two gradable dimensions, non-cognitivists will want to show how they can capture them both. And those who are concerned to defend the existence of normative uncertainty (i.e. credences of less than 1 in normative propositions), recognizing the plausibility of non-cognitivism, will want to show how this phenomenon can be captured by it. (It is beyond dispute, I take it, that our judgments sometimes represent reason strength as gradable.) Yet Smith's argument has been met with silence from both camps. There have been responses from some expressivists, but expressivism is a meta-semantic theory; non-cognitivism is a psychological theory. And indeed, all of the serious responses have come from expressivists who are not non-cognitivists, but rather "ecumenical expressivists", who claim that normative claims express both cognitive and non-cognitive states. What's more, these expressivists have tried to represent normative uncertainty using, and sometimes only using, cognitive components of normative judgment; they seem no more confident than Smith that pure non-cognitivism can capture the phenomenon. So non-cognitivists are still left wanting. So too, I think, are ecumenical expressivists. For not everyone has been satisfied with their responses to Smith. It's recently been argued that, precisely because they try to shoehorn gradability of judgments into the cognitive component, their view cannot accommodate thoroughly normative uncertainty. If this argument is right, then they need the same sort of help that the pure non-cognitivist does.

In this paper, I'm going to show how non-cognitivism can give us something like normative uncertainty. But I'm going to do this in an oblique way. I'm going to show that non-
cognitivism is plausible only if it is coupled with a meta-semantic theory, and that that theory must be expressivism. Non-cognitivist expressivism is plausible only if it is able to solve the Frege-Geach Problem, which it can do only if the non-cognitive state expressed in normative language meets certain structural constraints. The Frege-Geach Problem is but one member of a family of closely-related problems about normativity and public language. And I'll claim that expressivism can solve the problems in this family only if we put even further constraints on the sort of state expressed. A non-cognitive state that meets all of these constraints, I'll argue, can also be used to represent both gradable dimensions. So while not all non-cognitivisms can meet Smith's challenge, all otherwise plausible non-cognitivisms can, and this is about all we can ask for from a theory of normative judgment. I'll conclude the paper by noting some methodological issues raised by the solution I've proposed.

- **Comments**: Mike Huemer (Colorado)

43. Saul Smilansky (Haifa): “Should We Be Sorry that We Exist?”
- **Comments**: Melinda Roberts (College of New Jersey)

44. David Sobel (Nebraska): “Parfit’s Case against Subjectivism”

Derek Parfit, in *On What Matters*, argues that all subjective accounts of normative reasons for action are false. Parfit has three arguments to this conclusion. In this paper I will focus on his “Agony Argument.” I take this to be his favorite argument against subjectivism as only this argument is called “decisive.” The first premise of the Agony Argument is that we have current reasons to avoid future agony. Its second premise is that subjective accounts cannot vindicate this fact. So, the argument concludes, subjective accounts must be rejected. I will accept the first premise of this argument and that it is valid. The main thesis of this paper is that subjectivists can account for our reasons to get pleasure and avoid agony. I conclude that the Agony Argument does not justify the rejection of subjective accounts. Independently of that, I will examine Parfit’s understanding of the distinction between objective and subjective theories. I claim Parfit offers a surprisingly narrow understanding of subjectivism such that even if his critique were successful, this would be bad news for fewer theories than we might have thought. Finally I reply to some possible worries about my arguments.

To evaluate The Agony Argument we need to understand what Parfit means by agony and pleasure. Parfit accepts that current phenomenological states only count as agony or pleasure if they are liked or disliked. And this liking or disliking itself is not a reaction we have any reason to have. We simply have such reactions to some phenomenology and not to others. That is, he rejects Benthamite Hedonism which claims that the intrinsic nature of some phenomenology, regardless of our response to it, provides reasons. But given that we like or dislike certain phenomenology, he claims that we have reasons to get what we like and to avoid what we dislike.

But subjective accounts cannot vindicate this truth, Parfit claims, for two key reasons which will concern us at length below. First, he claims, likings are importantly different from desires and so subjectivists cannot appeal to the reason-givingness of likings. Second, subjective accounts can only grant reason-giving authority to desires I currently have or would currently have after informed deliberation. But a person might lack any current desire, even after informed deliberation, which would give weight to future likings or desires. And so subjective accounts cannot capture the thought that one’s future agony necessarily provides one with reasons now to take steps to avoid it.

Against this first argument I have two responses. First, I take issue with Parfit’s reasons for thinking that likings are different from desires. Secondly, I argue that even if likings are not a kind of desire, still they properly belong in the set of attitudes which the subjectivist can appeal to.
They are contingent, not rationally mandated pro-attitudes. To the extent that such attitudes provide reasons, to that extent subjectivism is vindicated.

Against Parfit’s second argument that the subjectivist cannot account for our reason to avoid future agony I argue that Parfit has no good grounds for attributing such a view to subjectivists. The best reconstruction of an argument for this attribution to subjectivism involves it seeming to follow from Williams’ requirement that reasons be capable of motivating after sound deliberation. One might think that since it is a contingent matter whether a person is moved by her future concerns, it must be a contingent matter whether a subjectivist can say that one has reason to take steps today to put themselves in a position to act in conformity with reasons they will have tomorrow. But Williams’ requirement only claims that reasons must motivate after one has engaged in sound deliberation. And I argue that it is quite plausible that sound deliberation, conceived along subjectivist-friendly grounds, requires one to now care about what one will later care about.

The subjectivist who claims that ideal procedural deliberation involves caring about one’s future concerns is not assessing the content of one’s future concerns and whether the objects of such concerns are worthy of being desired. Rather such a subjectivist principle is only concerned with whether one comes to care about the option as a result of an accurate understanding of it. The principle that one should now care about what one will later care about gives one no guidance until one starts to care about this rather than that for no good reason. It is in this sense that I am saying that such a view borrows no objectivist principles about what is worth caring about in the first instance. It is quite different from claiming that a person has a desire-independent reason to be moral or eat chocolate. The claim involves only the thought that if one will care about something later, one should now care about that fact. This seems continuous with the idea that one’s passions set the ultimate goals and further reasons are hostage to what promotes our ultimate goals. Reason is still the scout or slave to the passions. Can it really be said that it is a distinctively objectivist principle that one should act so as to maximally comply with one’s subjectively determined reasons over one’s life?

- **Comments:** Trent Dougherty (Baylor)

**45. Edward Song (Louisiana State): “Legitimacy as Affirmation”**
- **Comments:** Peter Higgins (Eastern Michigan)

**46. Roy Sorensen (Washington U, St. Louis): “Knowledge-Lies”**
- **Comments:** Julia Staffel (USC)

**47. Gopal Sreenivasan (Duke): “Equality, Opportunity, Ambiguity”**

In this essay, I expose an apparently unnoticed ambiguity in the notion of an ‘opportunity,’ distinguishing [*substantive*](#) conceptions of ‘opportunity’ from [*non-substantive*](#) conceptions. A salient context in which this distinction is significant is that of articulating the ideal of ‘equality of opportunity.’

What is especially notable is the way in which the unnoticed ambiguity in ‘opportunity’ interacts fatefully with a well-known ambiguity in the notion of ‘equality,’ namely, that marked by Parfit’s (1991) distinction between ‘equality’ and ‘priority.’ Since Parfit’s distinction is orthogonal to mine, we end up with a four-fold matrix for interpreting the ideal of equality of opportunity. Clarity obviously recommends keeping the four interpretations distinct from one another. But I shall also argue that one of them is not well defined.
1. Discussions of distributive justice often distinguish the question ‘what to distribute?’ from the question ‘how to distribute?’ Borrowing (but extending) some terminology from Cohen (1989), we can call the ‘what?’ question the currency question, and let advantage function as a generic place-holder for the right answer to it. Both my distinction and Parfit’s lie on the plane defined by the currency question, despite some appearance to the contrary in Parfit’s case.

One might begin with the perfectly general question, what is an ‘opportunity’? However, it is more profitable to jump ahead and consider how having an ‘opportunity for some advantage’ contrasts with simply ‘having the advantage’ itself. More specifically, we should ask whether someone can have the former independently of the latter.

According to non-substantive conceptions of opportunity, the answer depends on whether the person has ‘refused’ the advantage, i.e. chosen in some sense not to have it. For anyone who has not refused advantage X, there is no difference between ‘having X’ and having ‘an opportunity for X.’ That is because, on the simplest non-substantive conception, ‘opportunity’ merely functions in ‘opportunity for some advantage’ as a device for expressing acceptance of the following thought: when someone has refused an advantage, her not having the advantage does not count against the justice of its distribution. There is nothing else to it. Hence, for anyone who has not refused advantage X, the truth condition for ‘has the opportunity for X’ just is ‘has X.’ (The paper also illustrates some less simple non-substantive conceptions).

To illustrate, recall one of Sen’s favourite examples (1985: 200-202): A fasting monk lacks the advantage of ‘being nourished.’ Yet this does not impugn the justice of any distribution of nourishment, since the monk has freely chosen not to eat. To accommodate this point, we can refine our specification of the relevant advantage, substituting ‘opportunity for nourishment’ for ‘nourishment.’ For the same reason, Sen prefers the language of ‘capability’ to that of ‘functioning,’ where capabilities are defined as ‘opportunities to function.’ Capabilities, that is, are defined in terms of non-substantive opportunities.

2. By contrast, on substantive conceptions of opportunity, an ‘opportunity for some advantage’ is something anyone can have independently of ‘having the advantage’ itself. In particular, someone who lacks advantage X and has not refused X can still have an ‘opportunity’ for X. For there is something to the opportunity, which stands free from the associated advantage. It falls to advocates of substantive conceptions to spell out what this something is.

To confirm that at least one substantive conception is coherent, let us turn to the traditional understanding of equality of opportunity as a political ideal. I shall take Williams’ (1962) classic analysis as representative. On his analysis, the appeal to ‘opportunity’ substitutes an alternative equilisandum in the case of scarce advantages, for which equal distributions are impossible. Thus, to use Williams’ example, even if the distribution of grammar school education [education*] is perforce unequal, equality can still be served by equalising the ‘opportunity for education*’ instead. Notice, however, that an ‘opportunity for education*’ can only substitute for education* here if it can be enjoyed by those who lack education* and have not refused it.

Now a crucial feature of ‘equality of opportunity,’ so understood, is that the distributions it governs are competitive. Hence the traditional metaphor of the ‘[level] playing field.’ This feature allows us to identify the substance of the relevant opportunities with the terms on which the competition is conducted. The opportunities are ‘equal’ just in case the competitive terms are ‘fair.’ (My formulation leaves it open what is required to make these terms fair, i.e. to level the field.)

Arneson (1989) defends ‘equality of opportunity for welfare’ and Cohen (1989) defends ‘equal opportunity for advantage.’ But each invokes ‘opportunity’ only non-substantively (each explicitly makes an analogue of Sen’s move). Their ‘equality of opportunity’ slogans therefore
obscure important differences between their positions—both versions of ‘luck egalitarianism’— and traditional equality of opportunity, which appeals to substantive opportunity.

Following Parfit (1991), we can define ‘advantage’ in either relative (i.e., comparative) or absolute (i.e., non-comparative) terms. Egalitarians speaking strictly are concerned with the distribution of relative advantage, whereas prioritarians are concerned with the distribution of absolute advantage. Arneson (2000) now defends ‘responsibility catering prioritarianism,’ amending his earlier position to clarify that his concern lies exclusively with absolute levels of advantage. Schematically, he has moved from (2) to (4):

<table>
<thead>
<tr>
<th>‘Opportunity’</th>
<th>Substantive</th>
<th>Non-substantive</th>
</tr>
</thead>
<tbody>
<tr>
<td>Equality</td>
<td>1. equality of opportunity</td>
<td>2. Luck egalitarianism</td>
</tr>
<tr>
<td>Priority</td>
<td>3. priority of opportunity</td>
<td>4. responsibility catering prioritarianism</td>
</tr>
</tbody>
</table>

Since the distinction between absolute and relative levels applies unproblematically to goods, and therefore to ‘advantage,’ there is no conceptual difficulty in Arneson’s move. However, the same cannot be said of the move from (1) to (3). When its substance is identified with the terms of a competition, it makes sense (indeed, it is compulsory) to understand ‘opportunity’ in relative terms. However, it is at best obscure to speak of absolute levels of substantive opportunity, as (3) requires.

- **Comments:** Roksana Alavi (South Texas)

48. Sarah Stroud (McGill): “Partiality and Plural Agency”

Many of the philosophers engaged with the question of burden-sharing in the climate regime have concluded that obligations should be proportional in some fashion to responsibility (contribution to climate change) and capacity (ability to pay), a conclusion which we share. Yet justifications for particular ways of combining these principles are largely absent, with particular burden-sharing formulae being based on relatively ad-hoc criteria.

The issues here are complicated by the ambiguity of the location of obligation between individuals and states. The practical politics of the climate negotiations require obligations to be attributed to states; yet it is not obvious that obligations calculated on the basis of national characteristics and obligations calculated on the basis of individual characteristics (or aggregated individual characteristics) would be the same. In this essay, we explore some of these complexities in the justification of climate-related obligations, with an eye to identifying “reasonable” ways for combining multiple principles.

Because part of the difficulty in reaching consensus on a fair distribution of climate obligations is the multiple perspectives brought to the problem, we start by identifying several of the basic (and not mutually exclusive) paradigms that have been applied. Among these are:

1) The resource sharing paradigm. In this view, the global sinks for GHGs represent a scarce resource, and the problem to be solved is a fair allocation of this resource. Advocates of this perspective typically begin from the egalitarian argument that, because the global carbon sinks are a “global commons”, prima facie there should be equal entitlements. An important subject for dispute is whether unequal historical use requires current restitution – whether there is a “carbon” debt, and if so how it should be calculated. This raises numerous questions about fault,
ignorance and choice.

2) The burden sharing paradigm. Under this view, the reduction of emissions represents a cost to be shared, and this view has driven the debates about “comparable effort” among the industrialized countries, but it is difficult to generalize because plainly some persons, and arguably some countries, are so poor that they deserve to be fully exempted from any emissions mitigation burdens; here the role of rich people in poor countries looms large.

3) The right to development paradigm. In this view, developing countries have a right not to have their development (typically, GNP growth) limited by the need to reduce their growth in energy consumption; only emissions reductions that produce greater “co-benefits” or that are subsidized by developed countries are warranted until some (unspecified) level of national welfare has been reached. Again, this brings focus to the obligations of rich people in poor countries.

4) The economic competition paradigm. In this view, all countries, whether rich or poor, are fundamentally involved in a zero-sum game of competition for markets (and, by extension, for profits and employment). The competition is presumed to be fair by default, and thus the introduction of asymmetric regulations represent an unfair burden on affected industries, workers and countries. Here the focus turns in particular to affected workers in rich countries.

Our approach is to take each of these paradigms and (first) consider the direct application of these principles to individuals and countries, looking at a variety of prototypical cases, to see where they produce conflicting recommendations; (second), to try to see what disputed premises lead to these conflicting recommendations, and in particular whether they can be identified as “factual” premises or “value” premises; and (third), to see whether there are any “metaprinicples” that can be appealed to in hopes of resolving these conflicts. Our broad strategy is to seek an ethically robust solution, which could be claimed to pass a test of “not being reasonably rejected” or accepted behind a “veil of ignorance”.

As we suggested at the beginning, our assertion is that any such solution must combine aspects of responsibility and capacity, but further that it must more generally respond to concerns about fundamental interests; following Simon Caney, we hold that no climate obligations can accrue to those who do not have currently have acceptable protection of their fundamental interests to physical security (including health and nutrition), or who would be pushed below some related threshold by any such obligations.

Beyond that minimal proviso, we suggest that an acceptable solution needs to emphasize minimizing welfare losses – by imposing burdens on those with the greatest wealth – and also allowing opportunity for those who have not so far made significant use of the global commons. We experiment with some alternative formulae for a “progressive climate tax,” again using prototypical examples to see whether there are any easily defensible limits of “too progressive” or “not progressive enough."

We conclude by assessing the relationship between individuals and nations. We suggest that nations are not merely collections of individuals who happen to be represented in the negotiations by a single government; a substantial part of the emissions profile of a nation represents policy choices, while other parts represent geographical endowments. The structure of international negotiations requires that these factors be somehow accounted for, and that national characteristics influence national obligations. Yet we hold in the end that an ethical approach to fair global policy requires fair treatment of individuals, and thus that states be presumed to apply within their own borders the principles of fairness that they would have hold globally. In brief, this requires that rich people in poor countries have comparable obligations to rich people in rich countries, with exceptions only that are of direct and demonstrable benefit to the poorest citizens.

- Comments: Kendy Hess (Holy Cross)
Addiction is a social problem and a neurobiological one. I want to suggest that it is also a philosophical problem. Indeed, I want to suggest that the apparently clinical question “What is addiction?” turns on a more philosophical question, “What, if anything, is wrong with addiction?”

My argument starts with clinical diagnostic criteria of common addictions. Unfortunately, they cannot distinguish addictions from non-addictive passions. To make this distinction between addictions and passions myself, I first look to the way that the diagnostic criteria exclude non-addictive physical dependences, such as our dependences on food and water. To exclude those non-addictive dependences, the criteria claim that an addiction must be “maladaptive”: our use of food and water certainly is not maladaptive, whatever “maladaptive” may mean. I try to apply that same formula to our strong passions. Unfortunately, it is not clear what it means to have a “maladaptive” passion, so it’s not clear whether “maladaptive” will distinguish addictions from non-addictions.

I take up the challenge of figuring out more specifically what defect “maladaptive” points to. I present counterexamples to some possible suggestions, such as that “maladaptive” simply refers to cultural or biological norms. I suggest that what “maladaptive” was getting at can be more profitably considered by asking a more general question: what is bad about, or wrong with, addictions, which is not wrong with non-addictive passions?

I first rule out the effects of the objects of addiction, even the relatively common effects like intoxication and elimination of withdrawal. Effects vary significantly, and no effect is common to all addictions, let alone an effect whose wrongness would distinguish addictions from non-addictive passions.

I work instead to answer the question of what is wrong with addiction from two observations. First, addicts make bad decisions because of their addictions, choices to pursue the object of the addiction instead of more worthwhile goals, and to arrange other parts of their lives to accommodate that pursuit. Secondly, addicts have a tendency to relapse. This is enough to distinguish addictions from non-addictive compulsions, which I do as follows.

I consider whether there is a flaw unique to the addict’s thinking that causes bad choices. The alternatives I canvass, e.g., that the addict fails to understand her deepest motivations, fail to account for any flaw unique to the addict. Instead, I note that what seems unique to the addict is that her misvaluation of those things relevant to her addiction would chronically and stubbornly resist contrary evidence. This is unlike a non-addictive passion, which would respond to contrary evidence—perhaps not quickly, but eventually. Even our strongest non-addictive dependences—those for food, water, and sex, for example—can be resisted for long periods when the benefits seem worth the cost.

The question, then, is why the addict’s misvaluations chronically resist contrary evidence. I consider and reject several suggestions that fail to distinguish the addictive from the non-addictive desire. Finally, to explain why the addict’s misvaluations chronically resist contrary evidence, I take seriously the thought that the addict is unable to keep her resolution to stop. This is not because the addictive desires overpower or seduce, but because they are simply unaffected by decisions. Addictive actions needn’t be on the basis of thought, explicit or not, about one’s reasons to act.

This is the key to my account. I suggest that many addictive actions issue from no significant thought about what one should do at all. There is simply action. For example, I’ve been trying to quit smoking, but my smoking friends don’t know that, or they don’t care. One day, I run into a friend who is just getting ready to smoke and who offers me a cigarette. I unthinkingly, habitually say “sure” and reach for the cigarette. I may realize that I’m accepting, but I don’t debate with myself in any way what I’m doing. I simply act (although an impulsive action does not preclude further thought).

On this proposal, what makes addiction’s misvaluation chronically resistant to contrary evidence is that addicts are unable to carry through on any resolution to stop because, regardless
of any resolution to quit, they may act impulsively. Acting impulsively in this way can reinforce that misvaluation. Such impulsive actions can reinforce the misvaluation either by prompting rationalization of the action or through certain unconscious mechanisms, which are not coincidentally common to many objects of addiction.

Impulsive actions, together with such reinforcement, maintain the addict’s misvaluation in the face of contrary evidence. Contrary evidence can only influence a person’s thoughts about what to do. If impulsive action bypasses one’s thoughts and leads to rationalization and/or reinforcement of the initial misvaluation, then contrary evidence will have a hard time changing a person’s misvaluations.

This position on addiction accounts for what is wrong with addiction. What is wrong is that the addict misvalues addiction-related objects and actions, which is unlikely true of non-addictive passions; moreover, she misvalues in a way that chronically resists correction, which is even worse. Further, the addict regularly acts without thinking, which may show some additional defect.

Most significantly, this position accounts for how we would naturally debate about whether a person is addicted to something or just passionate about it. Specifically, it shows that the debate about whether someone is, say, a caffeine addict will turn on whether caffeine is as valuable as the person thinks and whether her misvaluation chronically resists contrary evidence—particularly via impulsive actions.

• Comments: Abigail Gosselin (Regis)

50. David Sussman (Illinois, Urbana-Champaign): “Pain for Constructivists”

In the final lecture of The Sources of Normativity, Christine Korsgaard notes that pain presents a serious challenge to the constructivist account of value she has offered. Korsgaard contrasts her constructivism with the “substantive realism” of Moore, Prichard, and Nagel. Substantive realism holds that some concerns are worth caring about or acting on just because of their intrinsic qualities. For the realist, there can be no non-question-begging arguments that might properly convince someone that a supposed value is really worth caring about. At the end of the day, the most the realist can do is to ask the skeptic to reconsider a putative value in a more careful and clear-eyed way. Hopefully, with focussed attention and imaginative engagement, the skeptic will come to see the importance of what is at stake.

Substantive realism is often attacked for invoking obscure non-natural properties in a way that makes a mystery of how we could come to know or care about the supposed normative facts. Yet when it comes to physical pain, realism seems to get things exactly right. Pain appears to be bad just because of what it is, so that the only way to appreciate this disvalue is to experience pain first-hand and carefully attend to its character. Our aversion to pain makes perfect sense in light of pain’s intrinsic character, a character that shows itself in such a clear and immediate way that no real doubt can insinuate itself.

Against realism Korsgaard argues that some fact can become a normative reason only through a person’s “reflective endorsement” or willingness to accept that consideration as a reason to act. For Korsgaard, practical reasons must be grounded in our commitment to them, rather than being prior facts that might serve to explain and justify our making of such commitments. Pain, Korsgaard argues, is not some kind of sensation with a quality that gives us a reason to avoid it. Instead, pain is itself a “perception of a reason”: that is, an immediate awareness that some commitment that is partially constitutive of our most basic practical identity as animals is in jeopardy. Our perception of such reasons is importantly different from the way we might recognize some other sort of fact. To perceive that I need to change my condition is, in part, to already be making up my mind to do so. Korsgaard contends that in this case, there can be no sharp distinction between cognition and motivation, between grasping what is so and beginning to respond to it appropriately.
I argue that Korsgaard ultimately fails to make sense of why pain is so relentlessly awful. The account she offers cannot avoid the bizarre conclusion that, when physical pain is recognized as being independent of any underlying disease or injury, we have no reason to take it seriously. However, this failure need not commit us to any kind of substantive realism about pain. I argue that Korsgaard’s constructivism ultimately has the resources to give a far more plausible account of the badness of pain than the one she actually presents in Sources. To see this, we need to move away from the perceptual account of pain that Korsgaard adopts, a model that is overly focused on the “first-personal” aspects of rational agency. To respond to the realist’s challenge, Korsgaard should instead consider pain’s essentially “second-personal” aspects, a shift that would still be consistent with her overall constructivist commitments.

I argue that pain involves more than just a felt need to protect or nurse some part of one’s body, as a response to a perceived threat. At least in creatures like us, pain also has an essential expressive dimension. Part of what it is to be in pain is to be disposed toward such “natural expressions” as wincing or crying out. I argue that such behavior involves a kind of inherent directedness towards others who might share our practical identity. Prior to any decisions or intentions we might make, to cry out in pain is to address a kind of plea toward others for help, a plea that carries an implicit hope that it might be recognized and accepted in the spirit in which it is given. If so, then the possibility of sympathetic response is already built into our experience of pain from the start, rather than something to be added to that experience through further psychological mechanisms. If we do in fact share the identity in question, we may well find ourselves unable to turn a deaf ear to such pleas, even though the sufferer lacks the standing to make any specific claims on us. Such pleas are less than a demand, but more than just a psychological stimulus.

What’s crucial here is that sympathetic response does not presuppose that the sufferer’s cry is based on a correct awareness of his situation. My daughter may cry out in fear of the monster in her closest, and my knowledge that there is no such monster does not negate the urgency of her pleas for me. My love for my child is such that I recognize her as having a kind of standing to make such appeals to me, appeals which have an importance that goes beyond whatever conditions might justify her making them in the first place. I argue that Korsgaard can and should understand pain in this way: not as a perception of a reason, but as something more like a plea that some part of oneself address to another, where the connection between these various aspects of the self has the form of something like a relationship of love or friendship between different people. In this way Korsgaard can take even groundless pain seriously, without having to treat the pain as just an intrinsic quality of a sensation that we apprehend through a kind of immediate intuition of value.

- Comments: Robert Jones (CSU-Chico)

51. Frans Svensson (Uppsala): “Virtue Ethics and Right Conduct: A Reconsideration”

How should we understand the aspirations of virtue ethics (VE) with regard to the topic of right conduct? In this paper, I will consider four different proposals for how to answer this question. Of these, I will argue that the first three are all subject to serious objections and should therefore be rejected (§§I-III). The fourth proposal, on the other hand, seems to me more promising and deserving of greater attention than it has hitherto been awarded. This proposal is developed in §IV, and I will respond to various objections to it in §V. I end by indicating some questions for future research.

The three defective proposals are: (i) VE aims to produce an account of what makes right conduct right, couched exclusively in terms of the virtues or the virtuous agent. (ii) VE aims to produce an account of what makes right conduct right, but not, strictly speaking, in terms of the virtues or the virtuous agent, but rather in terms of virtuous action. On this view, virtues are, at least in part, defined as character traits disposing their possessors to perform what is
independently characterized as virtuous actions, and it is the latter that determines whether a person’s conduct is right. (iii) VE does not aim specifically to produce an account of what makes right conduct right, but instead to provide a decision-procedure, couched exclusively in terms of the virtues or the virtuous agent (or perhaps in terms of virtuous action) that could be used to figure out what we should do in cases of uncertainty.

The main problem (though not the only one) with (i) and (ii) is that they get the order of explanation the wrong way around. While it seems plausible to think that a virtuous agent reliably get things right in her conduct, her conduct is not right because she would do it. Rather, she would behave in the relevant ways because they are right. Similarly, I will argue that, even if we were to grant that the virtues could be understood as dispositions to perform what is independently characterized as virtuous actions, insofar as a certain action is really virtuous, this is because it is right; not the other way around. With regard to (iii), I believe that (a slightly qualified version of) the old objection that appealing to the virtues or the virtuous agent in our moral thinking is either superfluous or unhelpful, is correct: the fully virtuous, at least, would be able to see what the circumstances demand of her for herself and she therefore does not need any help from a decision-procedure, whereas those of us who lack the practical intelligence or wisdom characteristic of the fully virtuous will not be much helped by thinking about what someone possessing such intelligence or wisdom would do.

Now the rejection of (i)-(iii) might prompt someone to suggest that VE should simply abandon any ambition to say anything about right conduct; that it should leave that topic for consequentialists and deontologists to fight over and instead restrict itself to questions about moral character. This is too quick, however. There is (at least) one more positive proposal to consider, one that involves ascribing to VE a radically different approach to the topic of right conduct than we usually find in other moral theories.

According to this fourth proposal, VE indeed seeks an account or conception of right conduct, but not in the form of “a finite and manageable set … of moral principles” (whether they be of an all-things-considered kind or merely pro tanto) that together specify the properties that make all right conduct right, or in the form of something that each person could always put to use as a method of decision-making when trying to solve specific moral quandaries. It is rather part of the background assumptions, on this view, that neither of these things can plausibly be had, and the idea of pursuing them is thus never seriously entertained. Instead, VE approaches right conduct via the virtues or the virtuous agent, aiming to explain what we must be like in order to reliably get things right in all relevant aspects of our behaviour. This includes inquiring into such things as the kind of knowledge or intelligence that is needed to appreciate the morally salient features of one’s circumstances, and the conditions for how one must choose to bring about what one should, which also need to be fulfilled in order for one’s conduct to count as perfectly right.

The result is an account of right conduct in the form of an ideal that we should strive for in our behaviour. In the light of this ideal, we can distinguish between perfectly right or fully virtuous conduct (conduct that is right in all relevant respects) and, for example, conduct that is right, but only in the sense of being decent or ok (right in some respects, but not all). We should furthermore be able to apply such distinctions to ourselves, and on that basis determine whether our conduct is characteristically closer to being ideal than merely decent, or the other way around. In both cases, there are aspects of our conduct that we still need to work on or improve, even though the work that remains to be done is obviously greater in the latter case than in the former.

• **Comments:** Bryan Reece (Oxford)
Matthew Talbert (West Virginia): “Moral Competence, Moral Blame, and Protest”

A common assumption about moral responsibility is that wrongdoers are open to moral blame only if they had the capacity to respond to the moral considerations that counted against their behavior. This assumption imposes a moral competence requirement on blame. Agents who fulfill this requirement can discern and respond to morally relevant features of their environments; agents who do not fulfill this requirement are impaired in some way that limits their responsiveness to moral considerations. This paper provides a counterpoint to compatibilist theories that emphasize the importance of moral competence for blameworthiness; I argue that it is sometimes appropriate to blame wrongdoers who could not have responded to the moral considerations that counted against their wrongdoing.

Like many compatibilists, I assume that a person is blameworthy if she is a proper target for negative reactive attitudes like resentment, and that a person is a proper target for these attitudes if she treats others with unjustified ill will or contempt. In the first section of this paper, I argue that even if it was not possible for a wrongdoer to respond to certain moral considerations, her actions may still express the sort of ill will to which blame responds. I provide initial support for this conclusion by developing an analogy with Harry Frankfurt’s argument against the Principle of Alternate Possibilities, which says that a person is morally responsible for her behavior only if she could have done otherwise.

Frankfurt argues that even if an agent could not have done otherwise, she is responsible for her behavior if she acts for her own reasons and not simply because she could not have done otherwise. I make a similar claim about morally impaired agents. A morally impaired wrongdoer may be incapable of responding to certain moral considerations, but when we seek to explain the wrongdoer’s actions we might naturally cite the considerations she counted in favor of so acting. Thus, while a deeply committed racist may be incapable (for all practical purposes) of recognizing the moral claims of members of a certain group, our explanation of his mistreatment of these people might still refer to his judgment that their interests lack normative weight. I would argue that, given the judgment on which he acts, the committed racist’s actions express the sort of contempt to which blame properly responds even if his errors in moral judgment were unavoidable.

In the second section of this paper, I consider cases in which morally competent wrongdoers are blameworthy, but their moral competence does not explain their blameworthiness. We can imagine cases in which a morally competent agent and a morally impaired agent commit similar wrong acts because of similarly offensive normative judgments. In such cases, it will be psychologically open to the morally competent agent to avoid her offensive judgments in a way in which this is not open to the impaired wrongdoer. However, I argue that this does not mean that the morally competent agent’s actions express a more serious form of contempt for those affected by his actions than do the actions of the impaired agent. This provides some additional, indirect support for the conclusion that moral competence is not necessary for blameworthiness.

The third section of this paper addresses the view that morally impaired agents are not proper targets for blame because blame communicates moral demands and thus presupposes the competence to respond to these demands. Gary Watson, for example, says that the “negative reactive attitudes express a moral demand, a demand for reasonable regard.” Moreover, according to Watson, “[t]o be intelligible, demanding presumes understanding on the part of the object of the demand.”

In response to Watson, I argue that while the negative reactive attitudes express a demand for reasonable regard, they do not do this primarily by issuing the demand in question. This is because these negative attitudes arise after an injury has been done, at which point it is too late to issue the demand that a person show us due regard. Watson suggests that a demand for regard is unintelligible if it is imposed on an agent who cannot conform to this demand, but there is also something problematic about presenting this demand to an agent who has already flouted it. I suggest, then, that reactive attitudes express a demand for regard by expressing the offense
we take when this demand is not observed. However, when we focus on the responsive, offended nature of the reactive attitudes, it becomes less clear why these attitudes should be intelligible only when they are directed at competent agents. I argue that, since the actions of both competent and impaired agents can express contempt, the negative reactive attitudes are fitting responses to both sorts of agents.

Finally, I argue that there is a moral point to blame even when we cannot hope to bring about contrition or moral reform in the person we blame. In such cases, blame may be founded on the interest we have in protesting the offensive judgments of impaired wrongdoers. For example, the person victimized by the committed racist mentioned above may wish to protest the racist’s (perhaps unavoidable) judgment that the victim’s interests do not count as reasons to refraining from certain actions.

Whereas Watson characterizes blame as a demand, I suggest that blame can take the form of a moral protest in which a person who is powerless to convince others of her moral standing seeks to reaffirm her own commitment to this standing. Because it can involve recrimination, resentment, and the insistence on moral facts, I suggest that such protest is a form of moral blame, but it is not a form whose intelligibility depends on whether its target is morally competent.

• Comments: Johann Klaassen (Independent Scholar)

53. Michael Tiboris (UCSD): “Childhood and Status in Kantian Non-Ideal Theory”

There’s a popular intuition that moral responsibility is, almost by definition, an “adult” concept. One reason child soldiers, criminals, drug addicts, and teen mothers are especially upsetting because they are children engaging in characteristically adult behavior. The adult/child distinction is more than biological, it is a normative distinction meant to pick out what is owed to each group in terms of justice. In this essay, I challenge a recent attempt to defend the adult/child distinction within the auspices of a major moral theory. Tamar Schapiro argues that Kantianism can offer us both an explanation of why we are justified in dismissing children’s claims to belief and self-determination. Her view is based on the idea that children lack the kind of principled self-conception that allows them to make the kinds of authoritative and stable judgments which demand respect. Her view is compatible with developmental pictures of agency insofar as the process of developing a self-constitution takes time. Schapiro argues that, in order to make sense of this transition from non-agent to moral person, we must adopt that she calls a “non-ideal Kantianism,” which she compares to Rawls’ distinction between ideal and non-ideal political theory. Schapiro draws her notion of the non-ideal from Rawls’ concern that we must find ways to make principles justified under idealized conditions retain their force under real-world conditions which threaten to make the principles unrealizable or irrelevant. The same must be true, she argues, when we use an idealized conception of autonomy to ground moral principles for real live agents.

The turn to the non-ideal allows her to support her view with empirical observations about development. For example, she makes the intriguing suggestion that children develop these self-conceptions through play, which requires that they “try-on” different principled perspectives.

I argue against this view in two ways, both of which are related to her claim that the Kantian version of the adult/child distinction can only be understood in a non-ideal way. My first objection argues that actual evidence from developmental psychology shows that children have the marks of self-constitution at a very young age. On her account, prohibitions on paternalism kick in when an agent has a principled deliberative perspective, a self-constitution which is both “stable” and “authoritative.” While this characterization of self-constitution has a chance at explaining why paternalism is unjustified from an idealized perspective, there is strong evidence from research in developmental psychology that children are both stable and authoritative with
respect to their self-constitutions in the way she requires. Moreover, moving the bar for stability and authority in order to exclude children won’t fix the problem. Aside from being a question-begging defense of the adult/child distinction, it would likely extend justified paternalism over a large number of uncontroversially adult agents.

My second objection begins by pointing out that, on the non-ideal view, facts about the developmental environment directly impact the kind and quality of self-constitution one can have. If development of a constitution is susceptible to influence from bad or indifferent parents, deprivation, isolation, and fanaticism, then it isn’t obvious that all self-constitutions are worthy of respect, protected from paternalistic intervention. In the ideal world, agents are rational and self-reflective about their principled deliberative perspectives. It is partly this feature of agents which makes their constitutions respect-worthy and protected from paternalism. Schapiro, following Korsgaard, would probably respond to this by claiming that possession of any constitution makes it worthy of respect. The fact that a self-constitution is has the formal characteristics of stability and authority, this is enough to trigger the demand for respect.

This tactic, I argue, only works in the ideal case. If we accept it, then in the non-ideal there will be little reason to deny children’s play respect and protection from paternalism. Here, as in the previous objection, I think the real issue is that we have a tendency to mistake the adults’ lack of interest in children’s self-representations with their lack of value.

After developing these two objections I consider an alternative tactic for approaching the adult/child distinction. Schapiro calls hers an “attributability” view of the adult/child distinction; one which claims that “a child’s lack of reason consists in her being incapable of making her own choices, whether good or bad.” This is to be contrasted with what she calls “proficiency” views which defend the idea that “a child’s lack of reason consists in her being incapable of making good choices.” Schapiro thinks that only the attributability view can account for the strong normative division between adults and children. I think this is incorrect, and I conclude by pointing to a way in which proficiency can do the job all by itself.

- **Comments:** Heather Salazar (Spring Hill)


Discussions of the distinction between “thick” and “thin” evaluative concepts commonly note that what is distinctive of thick concepts (such as ‘cruel,’ ‘honest,’ ‘greedy,’ or ‘generous’) is that they somehow \*hold together\* evaluation and non-evaluative description. A crucial question about how thick concepts do this concerns the theoretical location of their evaluative contents, and hence the sense in which they can be used to communicate evaluations (e.g., semantically or pragmatically).

This paper argues that the evaluative contents of thick concepts are presuppositional implications of a certain generally recognized kind. In other work, I analyze how thick concepts behave when one speaker regards the very concept of another as \*objectionable\* on account of the evaluative outlook invoked in the latter’s usage of the concept. There I argued that these data cannot be explained by treating the evaluative contents of thick concepts as part of their semantically expressed meaning. Here I offer a positive defence of the presuppositional account.

The descriptive stage of my argument is that thick concepts behave in various types of denials just as the presuppositional account predicts. Imagine you don’t accept Catholic sexual morality and so are unwilling to join the Pope in condemning people or their sexual conduct as ‘lustful’ or ‘lewd,’ praising them as ‘chaste,’ etc. If you regard ‘chaste’ (as used by the Pope) objectionable, what you deny is that things are in any way good qua, or for, satisfying the kind of non-evaluative conditions concerning sexual reservedness which are entailed by ‘chaste.’ (The same goes if you are the sort of fundamentalist who denies that attitudes called ‘tolerant’ are in any way good for displaying willingness to share community, etc., with people with different outlooks.) As I’ll describe in more detail, your available replies to (1) include (2)-(4), plus more, but not (5):
(1) Abstinence from extramarital sex is chaste.

(2) Abstinence from extramarital sex isn't chaste, although it signals dedication to not being sexually provocative, because it is in no way good for doing so.

(3) Abstinence from extramarital sex might be chaste, but that doesn't mean that it would be in any way good for exhibiting sexual reserve.

(4) Whether or not abstinence from extramarital sex is chaste, it is in no way good for signaling dedication to not being sexually provocative.

(5) Abstinence from extramarital sex isn't chaste, since chastity requires abstinence not only in action but also in thought.

As (2) indicates, in expressing your disagreement you will want to direct your objection specifically at the sort of evaluation that typically goes with your interlocutor’s ‘chaste’ talk. But as (3) and (4) indicate, this can be independent of whether or not you think the thing in question satisfies ‘chaste’ (so far as its non-evaluative features go). That would be unexpected if the evaluative content denied in (2)-(4) were part of the meaning of ‘chaste,’ but predictable if (2) uses negation metalinguistically to plug a presupposition of (1). (5) is di_erent: it shares with (1) a commitment to the evaluative content denied in (2)-(4). These and other patterns in how the evaluative contents of thick concepts project, or get plugged, in di_erent contexts of negation (and other embeddings) are just what one would expect if they were presuppositions.

The theoretical stage of my argument is to address the foundational issues with any presuppositional account, namely what the relevant evaluative presuppositions are like and how they arise.

Contexts of objectionable concepts are especially challenging despite the account's fit with the data. It seems that (a) the Pope's utterance of (1) can be conversationally appropriate even when he knows that his interlocutor regards his use of ‘chaste’ objectionable and will typically refuse to accommodate his evaluative presupposition, but (b) intuitively, his utterance needn't fail and (c) the presupposition can still be a commitment of his as a speaker. This seems to conict with the predictions of standard (Stalnaker/Lewis) stories about presupposition accommodation.

My solution is to think of the evaluative contents of thick concepts still broadly along the lines of a pragmatic account, as the presuppositions of speakers which impose constraints on the common ground, but to show how hearers can in the right circumstances infer the relevant evaluative presuppositions, thereby accommodating them, but refuse to admit a commitment to them into the conversational record. (Models of how this type of presupposition works can be found in linguistics.)

My account will appeal in part to the common view that thick concepts derive their evaluative point from the evaluative outlooks, practices, traditions, and the like, in which their use is embedded. This can be used partly to explain how thick concepts come to carry evaluative presuppositions as a class, and how the evaluative presuppositions of utterances deploying a particular thick concept can be conversationally derivable even without detailed knowledge of the particular context of utterance. Such inferences can be facilitated by knowledge of a common conversational association of linguistic expressions (at least in the idiolects of those who use them evaluatively) with evaluative presuppositions. I'll consider, e.g., how hearers' knowledge of Catholic sexual morality interacts with the interpretation of ‘chaste,’ ‘lustful,’ and ‘lewd' in the mouths of those who accept it.

Both stages of my argument aim to appeal, as far as possible, only to independently plausible semantic and pragmatic principles. The upshot of the paper is a systematic and explanatorily powerful alternative to the standard view that the evaluative contents communicated by thick concepts are parts of their semantically expressed meaning.

- Comments: Jamin Asay (Chapel Hill)

56. Justin Weinberg (South Carolina): “Moral Hazards”

In this paper I offer a conception of moral hazard and an assessment of the idea’s use in moral argument.

Moral hazard, in its general form, is a phenomenon in which morally good agency directed at some party incentivizes morally bad agency by that party. An example of morally good agency is making cars safer for drivers. An example of morally bad agency is driving recklessly. When making cars safer for drivers incentivizes recklessness by drivers, as some studies suggest, we have moral hazard.

Sometimes forms of good agency are objected to because they create moral hazard. Call this the moral hazard objection, or MHO. For example, a policy of unconditional free rescue of stranded hikers (good agency) is sometimes objected to because it will encourage people to go on adventures they are not qualified for, endangering themselves while knowing that others will bear the costs of their lack of preparation (bad agency).

While the term “moral hazard” is often identified with insurance issues, the phenomenon arises in various contexts, for example:

the personal—
should I politely downplay an insult if doing so will embolden a friend’s insensitivity?

the societal—
should our social safety net offer assistance to those who are badly off owing to their own choices, rather than luck, if doing so will encourage them to continue to make imprudent choices?

the global—
should we engage in vast geo-engineering projects to protect us from climate change, if doing so lets each of us continue to treat the environment poorly?

We hear the MHO in these contexts and others, ranging from worries about parents spoiling their children to questions about whether governments should help rebuild homes ravaged by predictably regular natural disasters.

Despite how widespread moral hazard is, the concept has received surprisingly little treatment by philosophers. In this paper, I have two aims: first, to defend a conception of moral hazard that makes sense of what is common to the cases mentioned in the previous paragraphs yet also designates a sufficiently specific and morally interesting phenomenon; and second, to clarify the moral hazard objection and determine under which conditions it has normative force.

Regarding the first aim, my treatment of moral hazard expands the term beyond its traditional narrow application in insurance contexts, and I explain why in the paper. The general account offered above—the phenomenon in which morally good agency directed at some party incentivizes morally bad agency by that party—requires clarification and defense, which I provide, particularly in regards to the ideas of good and bad agency, and the proximity and scope of these kinds of agency.

To make progress on the second aim, it helps to have names for the agents involved. To keep things simple, call the party who would be performing the good agency the good agent and the party whom this good agency would spur towards bad agency the bad agent. The most sustained philosophical treatment of the moral issues surrounding moral hazard to date, Benjamin Hale’s “What’s So Moral about the Moral Hazard?”, attends to the morality of the decisions of the bad agent. Hale argues that, in insurance contexts, the bad agent does not necessarily act wrongly by increasing her risk-taking in response to the fact that the good agent will bear the costs of her doing so. While I agree with Hale’s main point, which is largely about what (if anything) makes the agency of the bad agent bad (or wrong), that is only part of the story.
My approach differs by shifting our focus to the decision of the good agent. It is to the good agent that the MHO potentially applies. Here, two sets of questions come up.

One set is about the broader normative framework required to advance or be moved by the MHO. One issue (of several) here is responsibility. The MHO says the good agent should refrain from acting so as to prevent the bad agent from acting badly. It thus has the good agent taking responsibility for the agency of the bad agent (and in this respect is similar to paternalism). This prompts two concerns, both of which are associated with the core deontological idea that we are responsible for our own moral agency, not simply what happens as a result of the exercise of our agency. The first is that it involves wrongly “taking a decision out of someone else’s hands” (Korsgaard, “The Right to Lie”). The second is that it overburdens the good agent with responsibility for states of affairs directly brought about by others. Those deontologists with these concerns will not be able to advance the MHO.

The other set regards applying the MHO. Once we have cleared the issue regarding the normative commitments consistent with the MHO, there is still the question of when it applies, and how strong is it when it does apply. My view is that the objection is often used carelessly, and that identifying the conditions under which the objection is relevant can help us assess its strength.

To that end, I identify four individually sufficient conditions for the relevance of the MHO. Each of these rides atop the background condition that what is being objected to is a course of action in which what I have called the “good agency” does in fact increase the incidence of the “bad agency.” They are: (A) the good agency fails to prevent all of the bad agency’s bad effects, and thus increases the incidence of the non-prevented bad effects; (B) the good agency discourages the substitution of the bad agency with behavior that would have produced more beneficial effects; (C) the bad agency reflects poorly on the character of the agent; or (D) the acts constituting the bad agency are disrespectful or otherwise intrinsically wrong.

I then look at the issue of climate-altering geo-engineering to see how existing discussions of moral hazard in that context are affected by the foregoing analysis.

- Comments: Ben Hale (Colorado)

57. Brynn Welch (Madison): “Narrowing the Realm of Legitimate Parental Partiality”

In “Legitimate Parental Partiality,” Harry Brighouse and Adam Swift argue for controversial limits on the realm of legitimate parental partiality. For example, they argue that parent-child interactions, such as bequeathing large sums of money to one’s child or over-investing in one’s child’s education, are not justified by appeal to the family, even though they may be justified all things considered. In this paper, I offer a defense of their conclusion and argue that the realm of legitimate parental partiality is even narrower than Brighouse and Swift maintain. Specifically, I argue that they are willing to allow for compensatory parental partiality when their argument precludes it.

Brighouse and Swift argue that we ought to allow and support those parent-child interactions that are necessary for the flourishing of the relationship, even when those interactions give the child an unfair advantage in terms of life prospects. The parent-child relationship is one that provides a number of valuable goods that are unique to the relationship; that is to say, one cannot get these familial relationship goods from other relationships. Children require a healthy parent-child relationship to realize these goods and, furthermore, the realization of these goods is essential to a child’s development and flourishing. Many, though not all, adults also require the realization of familial relationship goods in order to lead flourishing lives. As such, we ought to allow and support those interactions that are essential for the realization of familial relationship goods, even when such interactions undermine principles of distributive justice, such as fair equality of opportunity.
For example, a child whose parent reads bedtime stories to her is likely to have a competitive advantage over children whose parents do not read them bedtime stories. Even if the competitive advantage is not always present and, when it is, any advantage associated with this interaction is accidental, the interaction can nonetheless undermine fair equality of opportunity. However, the interaction seems essential to the realization of familial relationship goods. As such, appeal to these very important and unique goods justifies the interaction. In other words, the interaction is justified by appeal to the family.

However, they argue that many parent-child interactions are not essential to the realization of familial relationship goods. For example, it is generally the case that a parent does not contribute something essential to the realization of familial relationship goods when she over-invests in her child’s education. Similarly, the relationship does not suffer if she does not bequeath large sums of money to her child. This claim is even more plausible if society is arranged such that no parents leave large trust funds for their children. These interactions can and often do undermine fair equality of opportunity and, as such, require justification. However, because the interactions are not essential to the realization of familial relationship goods, they are not justified by appeal to the family. They may be justified, all things considered, but they are not justified simply in virtue of the fact that they are parent-child interactions.

Brighouse and Swift allow for the possibility of compensatory parental partiality, or partiality that aims to ensure that the child have the same – not better – life prospects as other similarly talented and motivated children will have. For example, compensatory parental partiality might be permissible in cases in which a failing school system is likely to leave the child at a competitive disadvantage. A plausible defense of such partiality is that these interactions help achieve fair equality of opportunity rather than undermine it.

However, there are at least three reasons for doubting that Brighouse and Swift’s argument has the resources to justify such an allowance. First, none of the premises in the argument appears dependent on the social context. Brighouse and Swift offer reasons for thinking that certain interactions are essential to the realization of familial relationship goods while others are not. Additionally, they argue that when such interactions make others worse off than they would otherwise be, those interactions require justification other than appeal to the value of the family. There is no reason to think this is not true in cases of compensatory partiality. However, even if one argues that certain interactions become essential to the realization of familial relationship goods in certain social contexts, there are still other reasons for thinking that compensatory parental partiality is problematic. Second, Brighouse and Swift stress that other distributive goals are important, and these goals might be undermined even when fair equality of opportunity is not. As such, showing that a particular interaction does not undermine fair equality of opportunity does not necessarily provide sufficient justification for the interaction. Third, it is not clear that such interactions do not in fact undermine fair equality of opportunity; in fact, the individuals who are made worse off by instances of compensatory parental partiality are already among the worst off and so the interaction might require more – not less – justification.

Once we make Brighouse and Swift’s argument explicit, it becomes clear that the controversial conclusions they reach are much more intuitively plausible than they often strike readers as being. Furthermore, it becomes clear that the realm of legitimate parental partiality might in fact be much more narrow than Brighouse and Swift claim.

- **Comments:** Claudia Mills (Colorado)
58. Shay Welch (Williams): “Democratic Equality, the Freedom Threshold, and Strong Sufficientarianism”

Recently ethicists have begun to switch the question of equality simpliciter to a matter of priority and/or sufficiency. They reject strict equality and instead argue for distributive frameworks that prioritize the worse off, which separates them from egalitarianism proper. Though social justice concerns usually fall under the egalitarian umbrella, I advocate a sufficientarian approach to social justice problems. Sufficientarianism posits a priority threshold over which differences among agents warrant no moral consideration but those differences below the threshold are all highly significant. I posit the Freedom Threshold as a way of addressing problems of social injustice and I draw from Elizabeth Anderson’s conception of democratic equality to flesh out this threshold. In this article I have three aims. First, I demonstrate that the Freedom Threshold handles social injustice in a manner consistent with Anderson’s approach; however, for the freedom threshold to achieve the goals of democratic equality, I argue that the social group, rather than the individual, must be the primary target of equality claims. Addressing the social group first in matters of injustice avoids many problems that arise when considering equality at the individual level. Therefore, I also show that this approach yields a unique understanding of individual (in)equality. Third, I examine how a specifically deontological, ethicopolitical threshold impacts the characterization of a strong sufficientarian model.

In section one I provide a description of the forms of egalitarianisms in play in this discussion: traditional strict egalitarianism, sufficientarianism, and democratic equality. Larry Temkin characterizes egalitarianism as concerned with the misfortunes of individuals who suffer through no fault of their own and holds that society or some agent is morally required to offset the bad consequences of moral luck for reasons of fairness. Sufficientarianism holds that priority must be given to the worse off but also that there is a limit to this consideration. For this reason, sufficientarians posit a threshold, a point that once surpassed individuals can be said to have “enough”; any further distributive discrepancies carry no moral significance. However, Elizabeth Anderson argues that the proper objective of egalitarianism is the elimination of social injustice and oppression rather than bad luck. Thus egalitarians ought to utilize a notion of equality that reflects social and relational values such as respect between individuals and capabilities needed to self-determine. She terms this understanding “democratic equality”.

In section two I explain the importance of considering social groups when making equality claims for the purpose of social justice and democratic equality. As Nancy Hirschmann explains, choices themselves are discursively constructed and so what counts as a barrier to choice and action depends on the values that particular societies hold and so varies among social groups. By looking to the social group, theorists can more accurately determine if individuals face structural or systemic inequalities. When theorists ignore equality claims at the social group level sufferers of oppression are held to the same degree of personal responsibility as those in the dominant group, which contributes to the recreation of their condition. This is why some luck egalitarians propose to improve everyone’s position. Yet this results in an adverse dichotomy between accounts that enforce blind personal responsibility and accounts that assume none. For that reason, the only way to circumvent this problem is to first proffer an account of equality and social justice for social groups that intervenes in the social conditions.

In section three I lay out how the social group approach to democratic equality can be used to fill out a sufficientarians framework and explain why this is the best approach for democratic equality purposes. When thinking of democratic equality in terms of a threshold, inequalities that exist below the threshold qua conditions of oppression need to be remedied with social resources that create access to the capabilities that Anderson lists before personal responsibility can be the criteria for evaluating individual equality. Yet material resources eliminate only the most obvious of external barriers. If the normative goal is to attain democratic equality, social institutions must emphasize positive liberty via intangible, social resources and values social
Institutions to more fully eradicate barriers to individual freedom. The need for positive liberties that correct conditions of social group oppression is morally urgent up to the point that the agent reaches the freedom threshold. The threshold, then, is defined in terms of full access to the material and immaterial resources of freedom. Once the person hits the threshold, she has a complete set of positive liberties that allows for free action and robust choice sets immediately above the threshold. In effect, then, there will be no morally relevant “unit difference” between individuals above the threshold insofar as each can engage in unhindered pursuits and use their agency as they see fit.

Finally, in section four I demonstrate the ways a threshold cashed out in terms of democratic equality changes the underlying structure of sufficientarianism, which brings it closer to traditional egalitarianism. I highlight three significant changes: individuals can choose to live below the threshold, the continuity/discontinuity debate between weak and strong sufficientarians is avoided, and the deontological freedom threshold employs a notion of equality that carries inherent value. Why does it matter that the Freedom Threshold turns a teleological account into a deontological one? My response is that it likely doesn’t matter much for the framework itself but it most matters for social justice purposes. Democratic equality, unlike other teleological models, does not make the outcome the focus, but rather the action. Given this assumption, it is crucial that sufficientarianism can maintain a deontological threshold if social justice aims for better understandings of and conditions for both freedom and personal responsibility.

- Comments: Adam Hosein (Colorado)

59. Stephen White (UCLA): “What’s Wrong With Coercion?”

This paper takes up the question of what makes coercion wrong. I begin by arguing that coercion – or, more specifically, making coercive threats – is a distinctive wrong, not reducible either to the wrong of the threatened conduct or to the wrong of issuing threats more generally. I then turn my attention to the familiar view that coercion is wrong because it constitutes an invasion of autonomy.

Although this line of thought is a natural one, I argue that it is very difficult to develop the idea that coercion invades a person’s autonomy -- or in some way renders a person’s choice less free -- in a way that is theoretically satisfying. Theories that adopt this autonomy-based approach either fail to account for the full range of cases in which issuing coercive threats is wrong, or they fail to provide a non-trivial explanation of why such threats are wrong.

If an autonomy-based theory is to succeed, it has to do two things. First, it must supply a conception of autonomous action that does not have, built into it, the condition that the action is not coerced. Second it must explain how coercion undermines autonomy conceived of in the specified way. In the paper, I take up and criticize an account developed by Harry Frankfurt, which is representative of the autonomy-based approach. According to Frankfurt, coercion violates a person’s autonomy or freedom because a person who succumbs to a threat either acts from a motive she is unable to resist or one that she at least fails to identify with or endorse. Frankfurt’s theory does, I think, provide a sufficiently independent understanding of autonomy to be explanatorily powerful. However, he fails to show that, in general, a coercive threat -- along the lines of, “Give me the money or I’ll shoot” -- undermines the victim’s autonomy as Frankfurt understands it.

The basic problem that Frankfurt faces does not hinge on the details of his account and I conclude the discussion of his view with some lessons that apply to autonomy-based approaches more generally. Any view which, like Frankfurt’s, appeals to some sense in which the coereee’s choice is defective -- e.g., unfree or non-rational or involuntary -- will run up against the same difficulty. A coercive threat primarily works by altering the reasons the recipient has for pursuing certain options over others. But responding to changes in one’s situation that make certain
alternatives more attractive or reasonable is just what acting rationally and autonomously normally involves. It’s hard to see how coercion renders the victim’s choice defective in the sense required by the theory.

I go on, in the second half of the paper, to offer a different account of why coercion is wrong. I begin by noting that the recipient of a threat, at least when the conduct being threatened is impermissible (as in standard cases of wrongful coercion), faces a special deliberative problem. She must take into account the fact that opting for certain courses of action will likely lead another person to treat her in ways that are morally prohibited. I argue, next, that one generally has a claim against having to consider, in deliberating about what to do, the possibility that others will behave impermissibly as a result of the decision one makes.

Ultimately, on my view, the problem with coercion is that it amounts to an illegitimate transfer of responsibility from the coercer to the victim. To see this, suppose that a mugger threatens to shoot you if you don’t give her your wallet. Now, effectively, it is up to you to decide whether to do one thing and be shot, or do something else and walk away unharmed. From your point of view, it is as though you are now responsible for ensuring that your right not to be shot is respected. But this is a responsibility that properly falls to others — including, relevantly, the one who is now pointing a gun at you and demanding your money. In making her demand, however, the mugger forces you into a position where it is now essentially up to you to make sure your right will not be violated. This is because the mugger makes you aware, via her threat, that whether you will be shot turns on the decision you now have to make. Your life depends on handing over your wallet, and the choice is yours.

The wrong of coercion, I conclude, lies in this. It undermines the victim’s freedom to rationally deliberate about what to do on the assumption that others will respect her rights no matter what she decides. In undermining this freedom, the coercer wrongly transfers to the victim the responsibility for ensuring her own rights are not violated by the coercer. And this is mode of influence, I claim, that is not morally permissible.

• ** Comments:** Eric Chwang (Colorado)
Poster Presentations:

1. Roksana Alavi (South Texas): “Sex Industry and Moral Culpability”
3. Jamin Asay (Chapel Hill): “Creeping Minimalism Resolved”
5. Michael Bukoski (Arizona): “Consequentialism and Special Obligations: An Implication of an Agent-Centered Prerogative”
7. Tom Dougherty (MIT): “Chancy Charity”
8. Trent Dougherty (Baylor): “Reducing Responsibility: The Ethics of Belief is Ethics”
9. Christopher Franklin (Riverside): “The Value Account of Blame”
10. Abigail Gosselin (Regis): “Drugs and the Self”
11. Carl Hammer (CUNY-Baruch College): “Social Relations and Constructivist Publicity of Reasons”
13. Fred Harrington (Wisconsin): “Rights and Agent-Relative Consequentialism”
15. Ryan Hay (USC): “Hybrid Expressivism, Slurs and Jerks”
17. Peter Higgins (Eastern Michigan): “Is there a Right to Freedom of International Movement?”
18. David Killoren (Wisconsin): “Group Agents, Obligations, and Feasibility”
23. Melinda Roberts (The College of New Jersey): “Does the Moral Status of Merely Possible Persons Imply that Early Abortion is Wrong?”
24. Stefan Sciaraffa (McMaster): “Law’s Soft Claim to Authority”
25. Sam Shpall (USC): “Rational Requirements, Strictness, and Commitment”
28. Kelly Sorensen (Ursinus): “Dirty Hands and Morality’s Two Subject Matters”