Northwestern University
Evanston, IL
Society for the Theory of Ethics and Politics
7th Annual Conference
May 16–18, 2013
Keynote addresses:

Talbot Brewer (University of Virginia): "What Good Are the Humanities? A Heartfelt and Impolitic Defense"

Sarah Buss (Michigan): "Personal Ideals, Rational Agency, and Moral Requirements"

Conference program | Call for papers

The conference is free and open to the public.
Previous conferences:

2012 Harry Frankfurt, T.M. Scanlon
2011 Philip Pettit, R. Jay Wallace
2010 Elizabeth Anderson, Christine Korsgaard
2009 Samuel Scheffler, Seana Shiffrin
2008 Susan Wolf, J. David Velleman
2007 Barbara Herman, Stephen Darwall

Northwestern University Society for the Theory of Ethics and Politics

7th Annual Conference

Program

Thursday, May 16 (John Evans Alumni Center)

Morning Session

9:00 Coffee and Refreshments Available

9:15–10:40 "Love, Benevolence and How to Share a Beloved's Ends"
Michelle Mason, University of Minnesota
Commentator: Kyla Ebels–Duggan, Northwestern University

10:50– 12:15 "Love and Agency"
Adrienne Martin, University of Pennsylvania
Commentator: Jennifer Lockhart, Auburn University

Lunch

Afternoon Session

2:15–3:40 "Functional Explanations for Constructivists"
John Mumm, Fordham University
Commentator: Amy Flowerree, Northwestern University

3:50–5:15 "Carving a Niche for Immoderate Moral Realism"
Patrick Grafton–Cardwell, Purdue University
Commentator: Raff Donelson, Northwestern University

Dinner

Friday, May 17 (John Evans Alumni Center)

Morning Session

9:15–10:40 "Aristotle on Choosing Virtuous Action For Its Own Sake"
Yannig Luthra, UCLA
Commentator: Cristina Carrillo, Northwestern University

10:50–12:15 "The Salience of Moral Character"
Jon Garthoff, University of Tennessee
Commentator: Karen Stohr, Georgetown University

Lunch

Afternoon Session

2:15–3:40 "Autonomy: Incoherent or Unimportant"
Mikhail J. Valdman, Virginia Commonwealth University
Commentator: Gwen Bradford, Rice University

3:50–5:45 Keynote address: "Personal Ideals, Rational Agency, and Moral Requirements"
Sarah Buss, the University of Michigan
Commentator: Andrea Westlund, University of Wisconsin–Milwaukee

Reception at the John Evans Alumni Center– Everybody is invited

Saturday, May 18 (John Evans Alumni Center)

Morning Session

10:15 Refreshments

10:35–12:00 "On the Justness of Defensive Wars"

Lee-Ann Chae, UCLA
Commentator: Chelsea Egbert, Northwestern University

Lunch

Afternoon Session

2:00–3:25 "Responsibility and the Value of Intelligible Beings"
Nandi Theunissen, Johns Hopkins University
Commentator: Kristina Gehrman, Miami Ohio University

3:35–5:30 Keynote address: "What Good Are the Humanities? A Heartfelt and Impolitic Defense"
Talbot Brewer, University of Virginia
Commentator: William Bristow, University of Wisconsin–Milwaukee

Dinner

LOVE, BENEVOLENCE, AND HOW TO SHARE A BELOVED'S ENDS

Michelle Mason[1]

Draft: Please do not cite or quote

ABSTRACT (124 words)

How should we understand the nature and content of the normative reasons to which love gives rise? According to the so-called *benefactor view*, a lover should act toward the beloved in accordance with a norm of beneficence. I agree with recent criticism that the benefactor view provides an uncompelling normative ideal of love. According to an alternative *shared-ends view*, love directs us to share our beloved's ends in intimate adult relationships. The shared-ends view, I argue, suffers problems of its own. In response, I sketch a third, *shared-goods view* of the reasons to which love gives rise. On the shared-goods view, love directs lovers to pursue a shared good to which the lovers, qua lovers, are jointly committed.

INTRODUCTION (4461 words)

A human life devoid of personal love, were such a life to exist, would have at least this much going for it: it would simplify the question of how we should conduct ourselves. When we love someone, considerations that we otherwise need never have contemplated not only purport to make a claim on us, they stake an especially strong claim. If I didn't love my spouse, for example, considerations having to do with his needs, interests, or desires wouldn't enter into my deliberations about how to manage my affections, my projects, or my time in the way that they do. Relationships of personal love, in short, press us to take account of considerations that need never surface in love's absence. This fact alone ensures that a lover navigates deliberative territory made more turbulent in love's wake. The question I address here concerns the nature of this turbulence: How should we understand the nature and content of the normative reasons to which love gives rise?

[1]Associate Professor of Philosophy, University of Minnesota. I thank members of the MPLS Theory workshop, especially Sandra Marshall and Zach Hoskins, for their questions and comments. I also acknowledge the NEH for the award of a summer stipend that supported research and writing that informed the present work.

In what follows, I propose to take love to be a form of regard that is at once an affective appreciation of, and practical stance toward, the concrete particular who is its object. In providing an account of the nature and content of the reasons to which love, thus understood, gives rise, I mean to specify and defend as a normative ideal the practical stance that best answers to the appreciation of particular persons in which love partly consists. I hone my normative ideal against two competing accounts of the reasons to which love gives rise, the so-called *benefactor view* and the *shared-ends view*.[2] According to the benefactor view, a lover should act toward the beloved in accordance with a norm of beneficence. I agree with recent criticisms that the benefactor view provides an uncompelling normative ideal of love. The alternative shared-ends view, I argue, suffers problems of its own. In response, I defend a third, *shared-goods view* of the reasons to which love gives rise. On the shared-goods view, love directs lovers to pursue a shared good to which the lovers, qua lovers, jointly commit themselves.

I. DAISY'S DESIRE

In the final chapter of Henry James's *Daisy Miller: A Study*, an angry Winterbourne demands from Giovanelli an account of the circumstances that led to Daisy's death: "'Why the devil,' Winterbourne asked, 'did you take her to that fatal place?'" "That fatal place" is the Colosseum, a reputed breeding ground for malaria, and Winterbourne has spied Giovanelli and Daisy there just days before. Of the ensuing encounter between Winterbourne and Giovanelli, James writes:

> *Mr. Giovanelli's urbanity was apparently imperturbable. He looked on the ground a moment, and then he said, "For myself I had no fear; and she wanted to go."*

[2] Here, I follow Kyla Ebels-Duggan's in distinguishing the two views. See her "Against Beneficence: A Normativ Account of Love," *Ethics* 119 (2008): 142-170. Hereafter, in-text page references are to Duggan (2008).

"That was no reason!" Winterbourne declared.

Winterbourne serves as my foil for bringing into focus the benefactor view of the reasons to which love gives rise. Giovanelli, in contrast, will serve as a character against which to hone and assess a version of the competing shared-ends view.

II. LOVE AND BENEVOLENT CONCERN

Winterbourne believes, reasonably, that Daisy's visit to the Colosseum caused her to contract the illness that leads to her death. Given his belief that satisfying Daisy's desire facilitated her death, he refuses to recognize it as providing any reason to aid her plan. On what I'll call the optimistic reading of the novella, this refusal is motivated by his love for Daisy. The optimistic reading is plausible because we typically count a concern for the health of the beloved as a characteristic concern of a lover. This suggests we take love to direct a lover to recognize a normative reason to concern himself with the beloved's health. More precisely, a lover takes the consideration that Φ-ing protects the health of the beloved to provide a *pro tanto* reason to Φ.

In attempting to account for this reason, we might note that protecting Daisy's health is good for her – it is a central constituent of her well-being – and that lovers, as such, have reason to protect and promote the well-being of those they love. This suggests that what I will call a norm of benevolence comprises at least part of a compelling normative ideal of the practical side of love:

Love's norm of benevolence

(l, Φ) (If Φ will protect or promote the well-being of l's beloved, then l has a *pro tanto* reason to Φ)

Interpreting Winterbourne's refusal as motivated by such benevolence fits well with an account of love according to which it is a form of valuing a person. We typically take ourselves to have reasons to protect and preserve things we value – a prized book collection, an endangered species, Venice. Indeed, were I to profess to value a particular book collection while allowing the volumes to decay into disrepair, or to value Minnesota's dwarf trout lily while instructing my landscaper to eradicate every last one from my yard, or to value Venice while campaigning to have its canals turned into parking lots – well, in those cases my actions would properly cast doubt on either my sincerity or my facility with the concept of value. To value something just is, *ceteris paribus*, to regard oneself to have reasons to protect and promote its well-being or otherwise preserve it in its valuable state. If taking oneself to have reasons to protect, promote, and preserve what we value in its valuable state is an appropriate orientation to the value of inanimate objects and non-human living things, then – absent some special explanation – a compelling account of love as a form of valuing a person would need to recognize such reasons as being among those a lover, as lover, has with respect to his beloved.

Suppose, to the contrary, that we deny a compelling account of love must recognize the lover's norm of benevolence as among those that properly guide the lover qua lover. In that case, when our book "lover" with the grossly unkempt library also professes to love his wife while similarly disregarding her welfare, his disregard evidences neither a lack of sincerity nor confusion about love. This is a conclusion we do well to resist. It should come as no surprise, then, that some of the most eloquent philosophical treatments of the reasons of love give considerations pertaining to the well-being of the beloved priority in the practical thought of the lover.

III. THE BENEFACTOR VIEW[3]

Whatever initial appeal the benevolent concern I've ascribed to Winterbourne has as part of a normative ideal of love, that appeal quickly gives way to some worries.

Consider, for example, Kyla Ebels-Duggan's criticism of so-called benefactor views of the reasons of love. Duggan identifies Harry Frankfurt as providing "a perfect statement" of the benefactor view, a view that characterizes love as "a concern for the well-being or flourishing of the beloved" (144) and a lover's reasons with respect to the beloved as "reasons to do things for his beloved" (145). So, described, lovers on the benefactor view comply with what I've introduced as love's norm of benevolence. As Duggan argues, however, such compliance is consistent with a stance more properly viewed a form of disrespect for the supposed beloved.

Duggan marshals her own literary example of Scobie, the protagonist in Graham Greene's *The Heart of the Matter*. The relevant facts of Scobie's case are that he takes himself to be responsible for his wife, Louise's, happiness. In meeting what he takes to be his charge, he acts both unilaterally and deceptively in his attempts to please Louise. At the novel's end, Louise claims of her now dead husband that he didn't love her. Accepting Louise's complaint, Duggan proceeds to diagnose Scobie's purported love as not merely failing to provide a compelling ideal of love but, worse, manifesting disrespect for Louise in totally ignoring her status as an agent owed the status of deliberative equal. As Duggan sums up her case: "The benefactor view then runs afoul of the risk of disrespect by making treatment that is appropriate for someone with impaired agency the standard for all relationships" (148).

Now, I have no interest in defending the way that Scobie relates to Louise as part of a

[3] I ignore here certain revised versions of the benefactor view (such as what Ebels-Duggan calls the "specified benefactor view") because any norm that places the beloved in the position of passive recipient of the lover's care (as does even the specified benefactor view) is an unfruitful candidate normative ideal for adult love between equals.

normative ideal of love between equals.[4] However, I depart from Duggan in the lesson I would have us draw from Scobie's case. Duggan takes the benefactor view's primary fault to be that it understands love to direct us to act in ways that in fact fail to properly value the beloved because they fail to properly value her agency. Benefactors fail to properly value this agency because they fail to recognize the power of a beloved's choices to provide them reasons to act. Duggan proceeds to argue that not only is the norm of benevolence insufficient to capture the practical side of love, acting on the norm of benevolence is incompatible with love because incompatible with a form of respect partially constitutive of love.

I agree, of course, that someone prepared to disregard another as an equal partner in the conduct of their life together thereby fails to present a compelling normative ideal of love between equals. But that point holds independently of whether we take the beloved's welfare or her choices to be reason-providing; that point follows immediately from the fact that what we are considering is meant to be a relationship between persons. To relate to a person as a person one must, at a minimum, recognize a norm of justifiability to him or her.[5]

While I have no interest in defending Scobie, then, I am concerned to vindicate Winterbourne's benevolence as a loving response. One difficulty in doing so is that it is easy to interpret Winterbourne as disregarding Daisy in much the way Scobie disregards Louise – call this the pessimistic reading of James' novella. Taking literally Winterbourne's insistence that Daisy's expressed desire provides *no* reason at all for assisting her in visiting the Colosseum lends itself to the pessimistic interpretation. But we needn't interpret Winterbourne as believing that Daisy's desires and choices require no hearing at all; it may be that he finds this particular desire of Daisy, on reflection, to provide *insufficient* reason for facilitating her visit – a

[4] As previously noted, I reject it as a normative ideal even in the case of parental love for a child.
[5] Hence, my refusal to endorse the benefactor's norm as appropriate even in cases of love for a child.

conclusion he comes to because his love focuses his attention on the risks to her well-being. Wishes need not be heeded to be heard. Although denying a hearing to a beloved's wishes *would* mark a failure of love, refusing those wishes need not.

To see this, imagine a revision of James's work in which Daisy approaches Winterbourne with her desire to see the Colosseum. A debate about the relative merits versus risks of such an outing ensues, with Winterbourne ultimately refusing to assist Daisy in her adventure. Whatever there is to be said in favor of seeing the Colosseum by the light of the moon, those considerations are in his estimation outweighed by the risk that the trip will end in the demise of a young, charmed life.[6] Although Winterbourne ultimately acts contrary to Daisy's desire in order to preserve and protect her welfare, he is not properly charged – as is Scobie – with treating his beloved "as a passive object of care rather than as a full-fledged agent" (Duggan, 148). Winterbournes proceeds neither unilaterally nor deceptively. "I simply cannot help you, Daisy," we can imagine him saying as he attempts to justify himself to her, "I love you too much to have a hand in your ruin." In so concluding, Winterbourne fails Daisy neither in love nor respect.

Although it is insufficient to account for the reasons of love, then, the lover's norm of benevolence is not a complete nonstarter. Before pursuing the role it might play in a more compelling normative ideal, let us return to Giovanelli for a cautionary note with which to approach Ebels-Duggan's alternative.

Perhaps I have been unfair to Giovanelli. Helping others realize their desires is, after all, often motivated by love. Notice, however, that the role that Giovanelli affords Daisy's desire in his deliberations appears to be completely independent of the desire's content. Although the risks

---

[6] I do not mean to deny that certain desires that do not admit of further justification can provide reasons for their fulfillment. If an end is intrinsically valuable – as I take viewing the Colosseum by the light of the moon to be – then one's attempt to provide another reasons to join in its pursuit will involve articulating/specifying the intrinsic value in question. (The Collosseum by moonlight is, I assure you, stunningly beautiful.)

to a foreigner visiting a reputed breeding ground for malaria were available to Giovanelli in advance, he appears unready to afford such a risk any relevance.[7] Having assessed that there was no risk to himself, recall, Giovanelli decides that Daisy's wanting to go is sufficient reason for him to aid her in her goal. In doing so, he remains blithely unconcerned with what the content of Daisy's desire bodes for her welfare. Without such concern, talk of Giovanelli's honoring Daisy's desires or choices rings hollow as an expression of love.

IV. THE SHARED-ENDS VIEW

On Ebels-Duggan's own favored alternative to the benefactor view, love directs us not to promote our beloved's welfare but to share his or her ends.[8] As the example of Giovanelli illustrates, sharing another's ends can be risky business. All the more important, then, to be clear about what qualifies as sharing a beloved's ends in the way appropriate to reciprocal love between adults treated as equals.

As my imagined scenario between Winterbourne and Daisy highlights, a natural way to avoid the result that one party is left a passive object of another's care is to articulate a normative ideal that prescribes how lovers should come to adopt an end as a shared object of *joint* endorsement and pursuit, that is, by prescribing compelling norms for joint deliberation.[9] The upshot of such processes, when successful, is that the lovers come to share ends in the sense of recognizing reasons for *them*, i.e *as a couple*, to pursue the end through a joint commitment to realizing the end together. How, then, are we to characterize the norms compliance with which

---

[7] Giovanelli never attempts, for example, to plead ignorance of a risk to Daisy's life in order to deflect blame for her demise.

[8] "Rather than contributing to each other's welfare, doing things for each other, I hold that love directs us to share in each other's ends, doing things with each other" (156)

[9] The relevant forms of deliberation may range from informal conversation and invitations to imagine "what-if," to more formal analyses and weightings of the pros and cons of potential pursuits.

culminates in such sharing of ends?

On Ebels-Duggan's proposal, as I understand it, sharing ends is the normal outcome of two persons reciprocally following the norms she refers to as selection authority and authority in judgment. The norm of selection authority directs the lover to respect the reason-giving force of a beloved's provisional ends (156). The norm of authority in judgment directs the lover to proceed on the presumption that the beloved's ends are good ones (158).

At a first pass, a beloved's selection authority amounts to this:

"by choosing from among the set of [morally] permissible projects, [the beloved] gives you reason to pursue the chosen ends with her rather than concentrate your efforts on some other worthwhile pursuits" (156).[10]

In order to escape worries about unilateral adoption of ends, any individual choice must be understood as a conditional judgment: the beloved chooses to adopt the end, pending the lover's approval (157). This is the sense in which the ends are "provisional." More formally, we have

Love's norm of selection authority[11]

(l, Φ) (If Φ is a morally permissible provisional end of l's beloved, then l has a *pro tanto* reason to pursue Φ with the beloved in favor of pursuing some other worthwhile end)

Although I want to return to the restriction of reason-giving choices to those among morally permissible projects, I first want to consider a complication introduced by regarding the lover's reason to arise from the beloved's agency as exercised in her choice of provisional ends. Ebels-Duggan suggests that in matters properly of joint concern – as would be any matter that would place "significant demands" on one's partner – a lover makes the pursuit of her end

[10] Citing the example of a spouse deciding whether to take a new job in a distant city far from other family, Duggan suggests that: "What she may do is make a conditional choice: she will accept the job offer, pending your approval. We might think of this conditional choice as setting a provisional end. This provisional end, in turn, gives you reason to grant the needed approval. The claim that her choice, although conditional, makes on you is a way of capturing her individual authority" (157).

[11] Although I continue to formulate these norms as providing reasons for individuals, the simplifying procedure should not at this point matter; moreover, continuing to formulate them this way highlights the fact that each lover must be able to endorse the shared reason in his or her own voice. For more on this point, see Westlund (2009), p. 10.

conditional on its acceptance by the other partner. There is an ambiguity, however, concerning how the provisional end is supposed to generate a reason for the lover to "grant the needed approval" (157).[12] With respect to any end, either there is something to be said in favor of it, in virtue of which I propose it for us to take up or there is not.[13] If so, then the relevant substantive considerations must play a role in providing whatever reasons my partner has to grant the needed approval that signals his commitment to join or otherwise assist me in its pursuit. If not, then recalling that we are meant to be considering matters generating "significant demands," a partner has a reason to stop short of endorsing my end. Indeed, the relevant substantive consideration might instead provide a reason to disabuse the beloved from her own attraction to the end. Such is the case, for example, with blatantly immoral ends. Were one's spouse to propose a joint venture of supplying meth to the neighborhood youth, just in case you approved, we would not be tempted to the view that this provisional end gives you a reason to grant the needed approval.

Ebels-Duggan admits as much in the case of morally impermissible ends but her reasoning supports a broader conclusion I would have us draw. That an end is morally wrong is but one consideration that speaks against it. End's might also be stupid, dangerous, vulgar, or otherwise a waste of time. These substantive considerations that speak in favor, or against, an end must carry normative weight, independently of the fact of a beloved's provisionally choosing them, if a lover's approval is to be anything other than a rubber stamp. Admitting as much is compatible with allowing that the fact of our beloved's provisional choice to pursue intelligible categories of value as they are instantiated in this particular end provides an *additional*

[12] Summing up her understanding of the first norm, Duggan also writes: "First, you acknowledge his selection authority: by choosing, your partner gives you reason to act toward the accomplishment of his ends, in preference to other worthwhile goals" (162). On one parsing of the cited passages (on p. 156 and 162), the reason-giving power of the provisional end relates to its being worthwhile; on another parsing, the reason-giving power of the choice obtains regardless of the worthwhileness of the end (at least, so long as it is not a morally impermissible one). On my view, the former parsing yields a much more compelling normative ideal, for the reasons I provide here.

[13] I allow that one of the things that may be said in favor of an end is that it is appropriately regarded as intrinsically valuable and, so, properly valued as an end-in-itself (for no further reason).

consideration in favor of our pursuing together these values as instantiated in this end – as opposed, for example, to my favoring some other end that has similarly valuable features but which does not equally lend itself to our joint pursuit.[14] This, it seems to me, is how we should understand our beloved's power to make it the case that we should enjoy together the respite, fresh air, and challenge of birdwatching, say, in favor of geocaching.[15] Finally, admitting the relevance for the lover of the substantive considerations that favor (or not) an end is also compatible with love's directing us to remain receptive to those of our beloved's ends that may initially appear trivial.

Morally impermissible ends are on Ebels-Duggan's account a special case. Contemplating a move that would deny their practical significance by allowing such choices to be reason-providing but always overridden by the lover's reasons for not participating, she argues:

> [S]ince such a reason [the purported reason generated by the choice of a morally impermissible end] could have no practical upshot, I am more inclined to say that here we run up against a limit on the beloved's [selection] authority: the provisional adoption of an impermissible end simply can't generate reasons for you (162)[16]

If a failure of practical upshot is sufficient to mark such a limit on selection authority, however, then it would appear to likewise reach a limit in the case of otherwise ill-considered ends.

---

[14] Note that the beloved's selection authority thus is not wholly/merely epistemic in nature.

[15] At one point, discussing authority in judgment, Ebels-Duggan considers an example where one lover enjoys birdwatching whereas the partner sees nothing to be said for it (though nothing in particular to be said against it, either). This kind of apprently trivial but otherwise unobjectionable case strikes me as the strongest one for claiming that the mere setting of a provisional end provides one's partner with a reason to approve it and aid in its pursuit. Presumably, there are reasons for caring about birdwatching, that is, recognizably attractive features of the experience. It is not as if birdwatching is an idiosyncracy and my beloved's an unmotivated desire. Thus, there will be intelligible categories of value to which my beloved can appeal as we decide whether to embark on ornithological endeavors together. What my beloved's choice of pursuing these categories of value via birdwatching *does* succeed in adding to the deliberative context is an additional consideration that renders the pursuit of those values by means of birdwatching preferable to other activities that enjoy the same features: birdwatching is something that we can come to enjoy together whereas, given our lack of skills, mountainclimbing is not.

[16] The quote continues: "You may still owe it to your beloved to consider her view that it would not be impermissible to undertake the project in question, if indeed this is her view. But in the end, you will have to rely on your own considered judgment about this" (162).

Were no favorable substantive considerations waiting in the wings in the case of otherwise bad or apparently trivial ends, they would fall victim to the same argument Ebels-Duggan presses against immoral ends. If all one's beloved could offer in attempting to convey the value of the provisional end is "I *just* want to," the reason she thereby provides you could be overriden by any competing claim grounded in substantive considerations (that is, any consideration in favor of pursuing another end above and beyond a beloved's having chosen it).[17] Consider: my spouse wants to run counterclockwise circles around the backyard oak in the light of the moon; there is homework to be supervised, children's laundry to be washed, not to mention the items to collect for the food pantry, calls to be made getting out the vote, and whatever other good might need to be done in our little part of the world. In the context of everyday life, the reasons love purportedly creates simply in virtue of my spouse's setting provisional ends begin to appear very anemic very quickly. In the current state of the world, I'm inclined to think these reasons could have no practical upshot. If the norm of selection authority fails to apply in the case of a beloved's morally impermissible ends because otherwise we are forced to posit reasons with no practical upshot, then it likewise fails to apply in the case of a beloved's otherwise unworthwhile ends. In short, Ebels-Duggan's reasoning fails to justify their asymmetrical treatment.

There is a second point worth noting about a beloved's provisional morally impermissible ends. If an end truly is morally impermissible, then of course a lover has a reason to avoid sharing it. But in such a context love directs me not merely to refuse to participate; it also directs me to do my part in attempting to set my beloved straight.[18] Ebels-Duggan's conclusion

[17]Cite Korsgaard, "The Reasons We Can Share," .

[18] If immoral choices provide me a reason to dissuade my beloved from its pursuit, so too, does my beloved's provisional choice of otherwise objectionable ends provide a reason to dissuade my beloved from their pursuit – at least insofar, that is, as I love her.

regarding my beloved's morally impermissible ends is too weak: not only does my beloved's provisional choice of a morally impermissible end fail to provide me a reason to participate in it, the provisional choice gives me a reason to dissuade my beloved of its value.[19]

Denying these two points leaves us with a norm of selection authority that, if reciprocally complied with, yields lovers not engaged in the lofty-sounding sharing of ends but in the more sinister sounding enabling, aiding, or abetting of bad plans.[20]

Although the shared-ends view's second norm, that of authority in judgment, seems directed to blunting the force of my concern with an end's content, it instead sharpens its point. Sharing the worry that once can comply with the first norm and come to share an end "while regarding it as foolish or worthless," Ebels-Duggan proposes to address it by appeal to the second norm of authority in judgment:

"In granting another [authority in judgment], you treat her choice of an end as if it were evidence that the end is worthwhile. This doesn't require you to treat her judgment as infallible, but you must operate under the presumption that her choices are good ones" (158-59)

That is,

Love's norm of authority in judgment

(l, Φ) (If Φ is a provisional end of l's beloved, then l should presume that the beloved's choice of Φ is evidence that the end is worthwhile)

Now, this norm cannot plausibly direct the lover to *defer* to the beloved in matters concerning what is worthwhile. Rather, the norm directs the lover to treat the beloved's choice as defeasible evidence that there is something to be said in favor of the end – something other than the mere fact that the beloved has chosen it – that speaks to the end being a worthwhile one to pursue.

---

[19] Should I fail, and should the values to which my beloved would commit diverge too far from those I can reasonably endorse, I likely will find my love undermined as its grounds – those features in virtue of which I love the beloved – are thrown into doubt.

[20] Duggan appears too preoccupied with the potential of the benefactor view to counsel "derailing" a beloved's end in objectionably paternalistic ways to appreciate this point. Refusing to aid and abet a beloved in a moral wrong is not a morally objectionable form of derailing their proposed joint project. Neither, it seems to me, are attempts to educate a beloved concerning the risks or unseen pitfalls of the otherwise unchoiceworthy ends he proposes.

Adopting this stance, the lover recognizes the beloved's choice as *prima facie* evidence that *bona fide* reasons to share the end exist. Any subsequent discussion or deliberation concerning its joint pursuit must concern what, if any, these *bona fide* reasons are and their relevance to an all-things-considered judgment about what the lovers should do. To refer to this norm as one that confers on the beloved some form of authority thus is at best misleading because it reinforces what I have suggested is a mistaken view of the power of a lover's desires or choices, as such, to provide reasons for the lover.

V. The Shared-goods view: A Sketch

The shared-ends view of the reasons love creates, recall, is intended to correct for the benefactor view's failure to respond appropriately to the beloved's agency. I have argued that the shared-ends view risks erring in the other direction by suggesting that the beloved's agency – as exercised in the choice of provisional ends – has reason-giving power largely independent of the content of the ends toward which it is directed. As a corrective, I've emphasized the importance of attending to the content of those ends whose pursuit lovers propose to share. I believe that a more compelling account of the reasons love creates thus lies with a third alternative, one that combines the insights of the benefactor view by acknowledging that a lover has reason to concern herself with the welfare of the beloved with the insights of the shared-ends view by acknowledging that a lover has reason to structure her relationship with the beloved in a way that respects her agency.

In sketching this alternative, let me begin with a word about what I take to be the reasons *for* love. Briefly, I subscribe to a version of a property theory of love, according to which love responds to – and is warranted by – certain features of the beloved that the lover values. The

object of love, the beloved, is a concrete particular who possesses an incredibly complex array of interrelated features, some subset of which provide the lover's grounds for loving the beloved. It is in virtue of those features of the beloved to which my love is a response that the beloved is capable of generating reasons for me that I would not have in our love's absence. Although the beloved's agency is of course among those features of the beloved that a lover has reason to value, the beloved's agency need not occupy any privileged status in the lover's scheme of values, let alone be regarded as the only feature of the beloved that makes a practical claim on the lover.

Given that the beloved would cease to exist were her welfare not protected and preserved to at least some minimal degree, considerations pertaining to a beloved's welfare yield *pro tanto* reasons for her lover to protect and preserve her. So, too, considerations related to the rational agency of the beloved typically will be among those that provide *pro tanto* reasons for a lover. Thus, of course a beloved's choice of provisional ends may create reasons for a lover to endorse them – but only insofar as they relate to goods that the lover can endorse as such in her own voice.

On the shared-goods view of the reasons love creates, these reasons are all ultimately grounded in values of the beloved that likewise sustain the lover's love.[21] Reciprocal love between adults is, as a practical stance, one that we may for good reasons renounce. Should a

---

[21] Ideally, then, Winterbourne's love for Daisy likewise directs him in a manner that is sensitive to the fact that she is, as James described her, "a child of nature and freedom" (James, Preface).

In virtue of this feature, a fully developed account of the shared-goods view of love's reasons would distinguish it from the account of lovers' joint deliberation that Andrea Westlund defends. On Westlund's view, as I understand it, although "individual concerns, commitments, preferences, and the like" provide "non-arbitrary starting points" for joint deliberation, they do so simply in virtue of belonging to the beloved and irrespective of their connection, if any, to the grounds of the lover's love. On the shared-goods view, those reasons a lover can recognize as "reasons-for-us" are mediated by their connection to features of the beloved in virtue of which the lover properly values her. For Westlund's view, see "Deciding Together," *Philosophers' Imprint* 9:10 (2009).

beloved's choice of provisional ends prove too far removed from values the lover can appreciate, this may in fact signal the relationship is best brought to an end.[22]

My suggestion, in short, is that the kind of normative authority a beloved has with respect to her lover – that is, the ability she has to provide the lover reasons that the lover would not otherwise have – depends both on the value of those ends that she proposes for joint pursuit and those features in virtue of which the lover properly values her.

The unity in what might otherwise appear an *ad hoc* hybrid of an ideal emerges once we recognize a common root in the lover's concern to protect and preserve the beloved in the valuable state to which her love responds. The successful result of lovers who deliberate with such mutual concern is the sharing of ends capable of inspiring their joint commitment and yielding a common vision of how they are to structure a life well lived together.

[22] If one of the reasons I love my spouse is his dedication to creative work, then among the reasons our love creates are reasons for me to join him in that pursuit by helping facilitate his next screenplay (and reasons against structuring our joint endeavors in ways that threaten it). If another of the features in virtue of which I love my spouse is his childlike enthusiasm about certain sports, then our love creates reasons for us to schedule our time to accommodate the occasional game (despite, perhaps, the havoc it wreaks on his aging knees). Finally, if I adore him for his whimsy, love indeed may direct me to recognize a reason to run counterclockwise circles with him around the backyard oak in the light of the moon. However, should it come to pass that he proposes pursuing the joint venture to sell meth to the neighborhood youth, and should I prove unable to persuade him otherwise, then we may well have reached a point where his provisional ends provide me reasons to abandon the practical stance of love toward him altogether.

Note to commentator and readers: This short paper is a small slice of a much larger project. In order to provide some context and motivation for what I do here, I begin with a brief sketch of the larger project's central ideas. At NUSTEP, I will talk rather than read the paper, but everything below is fair game for discussion.

Love and Agency[1]

The context: A moral psychology with three interrelated ideas at its core

1. **The right theory of human motivation is a dualist one**. We have two motivational representations: rational and subrational. Subrational motivational representations are what Kant called "inclinations." They are representations of outcomes as *to-be-attained*, or *to-be-avoided* (or food as *to-be-eaten,* the predator as *to-be-fled*, etc.). When we are moved directly by subrational motives, we do not perform actions, but behaviors for which it would be a category mistake to demand justification in terms of reasons. Rational motivational representations are what Kant called "maxims." They are representations of outcomes as providing reasons to take measures to bring them about. *Go to the kitchen in order to satisfy my desire [inclination] for more coffee* represents the outcome, *satisfaction of my desire for more coffee*, as a justifying reason for going to the kitchen. When we act on the basis of rational motives, we perform intentional actions that we are implicitly committed to defending on the basis of reasons. As the example reveals, the satisfaction of desires/inclinations—that is, subrational motives—is one kind of outcome we can incorporate into our maxims—that is, rational motives.

[1] Acknowledgments redacted.

2. **Some emotional attitudes involve incorporating subrational motives into rational motives**. Put another way, some of the states we pre-theoretically class as "emotions" are best analyzed as syndromes of subrational and rational motives, unified by the incorporation of the former into the latter. In this paper, I aim to make this moral psychology concrete and to show some of its advantages by examining what it would mean to conceive of love as such a syndrome.
3. **A norm of respect constitutes interpersonal relationships and therefore also constitutes emotional attitudes that involve rational motives regarding the treatment of others**. "Constitutes" here means "by definition governs." Thus: a relationship is successful as an interpersonal relationship only insofar as the parties abide by (or at least strive to abide by) a norm of respect; and an emotional attitude that involves an other-regarding rational motive is successful as such only insofar as the person feeling/engaged in it respects (strives to respect) that other. Again, in this paper I aim to make this moral psychology concrete and to show some of its advantages by examining what it would mean to conceive of love as constituted by a norm of respect.

Love as a syndrome—the incorporation conception

Drawing on a dualist theory of human motivation, love emerges as simultaneously a passivity and an activity. Its passive aspects are both the feelings that give rise to certain subrational motives and those motives. When we love a

person, we (paradigmatically) find both her proximity and her flourishing pleasurable, her absence and her suffering painful. Because of these feelings, we become attracted to having her near and to contributing to her flourishing, and we find the counterparts of these outcomes aversive—these attractions and aversions are subrational motives. Then, when this passive, pathological side of love develops into love in its fullest sense, we adopt the maxims of being near to the beloved, and of contributing to her flourishing. These rational motives are active projects, exercises of our agency.

The lover treats her "desire" for the beloved—the pleasurable feelings associated with the beloved and the resulting subrational motives—as practical reasons. She also, of course, treats the features of the beloved that produce these feelings and attractions as reasons: love is not exclusively self-centered, on this account. But a crucial element of love is *endorsing* the way one feels about the beloved, and how appealing and attractive one finds her and her flourishing. I'll call this the "incorporation conception" of love.

The norm of respect

Here is one way to understand Kant's view that the Categorical Imperative—the Formula of Humanity in particular—is among the principles of rationality: to successfully engage in an interpersonal activity requires responding appropriately to the nature of a person, which means treating her as an end in herself. That is, the

requirement to treat the members of humanity as ends in themselves, to *respect* them, is a constitutive norm of interpersonal engagement.[2]

One way to engage with a person is to love her: as a friend, as a partner, as a sister, daughter, mother, etc. Each of these is a mode of *love* because each involves the feelings, subrational motives, and rational motives discussed above. These are *distinguishable modes* of love because they are defined in part by what it takes to respect the beloved. For example, the physical liberties one can respectfully take with a lover or partner would be failures of respect if taken with a friend; adult siblings need not coordinate their daily routines in the way that domestic partners do; and children are permitted to make unexpected demands on their parents that others are not. To fail to show respect in the particular way demanded by a particular kind of love relationship is to fail to relate to the beloved as a person, which is to fail to love her—or at least to love her *well*.

First advantage of the incorporation conception of love: it explicates both paradigmatic and ambiguous cases

We should not think of the above conception of love as providing a set of necessary and sufficient conditions for love. Instead, it tells us about *love in the fullest sense*—the most complete love happens at three levels: feeling, subrational motivation, and rational motivation. Thus, lacking engagement at any of these three levels—or having only minimal engagement at any of them—means falling

[2] For a detailed and compelling discussion of how norms can be constitutive of activities, see Christine M. Korsgaard, *Self-Constitution: Agency, Identity, and Integrity*. (Oxford: Oxford University Press, 2009).

short of love in the fullest sense. One may nevertheless count as loving another, given sufficient engagement at the other levels.

Consider, for example, an objection Velleman raises to syndromic conceptions of love where part of the syndrome is the desire to be close to the beloved. He points out that there are many genuine cases of love where one loves someone "whom one cannot stand to be with," such as an ex-spouse, a smothering parent, or an overcompetitive sibling. "In the presence of such everyday examples," Velleman writes, "the notion that loving someone entails wanting to be with him seems fantastic indeed."[3]

Let's begin with the divorced couple, and how we might think about them pre-theoretically. In what way do they still love each other? Not all romantic love that ends does so badly, so one possibility is that, as the lovers have aged and changed, their *romantic* love has transformed into a *friend's* love. I will discuss this possibility in the next section. It is likely Velleman has a more complicated situation in mind. Some couples separate in spite of continuing to feel passion for each other, because conflicting feelings and commitments prevent them from loving each other well. If things really go off the rails, some passionate couples even come to hate and deliberately seek to hurt each other. In the right moment, such lovers may see they are unable or unwilling to treat each other as they deserve—to satisfy the requirement to respect each other—and decide it would be better to go their separate ways. Here, it is, as Velleman says, a "dark truth" that they love each other, because their demons have shouted down their better angels.

[3] David Velleman, *op. cit.* p. 353.

Any syndromic conception of love allows its proponent to say that this is a case of imperfect or less-than-full love. There are even better stories to tell, if we adopt the incorporation conception. This conception allows us to recognize and explain a deep dissimilarity between two versions of this case. In one version—the version available to any syndromic conception—the couple parts simply because their hateful feelings and attractions overwhelm their loving ones. In another version, where they part out of love, they part because they have each adopted the maxim of promoting the other's flourishing, and so they choose not to pursue closeness, recognizing that their conflicting attractions will lead them to hurt each other. This choice happens because they see themselves as having *sufficient reason* to refrain from hurting each other. Thus, the dualist theory of motivation, by including both subrational and rational motives, has more explanatory power than alternative, monist (either Humean or rationalist) theories of motivation, and gives us the resources to account for the possibility of love in the absence or paucity of paradigmatic elements of love.

The other resource the incorporation conception has available is the distinction between love that succeeds in following the constitutive norm of respect and love that tries but fails. We can see the value of this resource in relation to Velleman's case of the difficult relative. A difficult relative is the object of many loving feelings, subrational motives, and rational motives, even if one finds her company painful. The absence of the attraction to spending time with difficult relative isn't the absence of some necessary condition for loving that person—it is rather the absence of an element that is characteristic of most paradigmatic forms of love. The love one has for such a person is not love in its fullest sense, and one

knows this: "I love my mother, *but* I wish she would butt out of my life." Moreover, an important part of one's complaint, and part of the reason one is unable to love such a person in the fullest sense, is that *she* falls far short of the ideal of familial love, by failing to respect one—being in her presence means enduring her efforts to interfere with one's choices, or being on the receiving end of disrespectful behavior. Nevertheless, one empathizes with her, cares about her flourishing, follows maxims of promoting her well-being and trying to be near her as much as one can tolerate and as much as is compatible with proper self-respect. That is, one tries to love as best one can, even while the object of love is herself pretty bad at it.

Generally speaking, the incorporation conception makes good sense of our ambivalence about a range of cases. Consider a man who abuses his wife, but is also passionate about her and easily distraught at the thought of losing her. There is *some* way in which he loves her, but another in which he absolutely does not. His abuse, let us imagine, takes the form of both subrational behavior such as physical violence brought on by rages and maxim-based action such as plotting to control her in a misbegotten effort to ensure her fidelity. Some of this behavior is the effect of feelings and attractions alien to love, such as anger and an urge to hurt. But some of it is also probably due to loving feelings and attractions. Unchecked or under the wrong influences, the attraction to intimacy can evolve into a compulsion to possess that can readily cause maltreatment of its object. The same is true of his abusive maxims: some are based on attractions alien to love, but some may very well arise out of loving attractions. That compulsion to possess could be the basis for a maxim of controlling the beloved to ensure her fidelity. Thus, while part of our ambivalence about the case comes from the fact that he

doesn't seem fully engaged by the feelings, attractions, and maxims of love, that cannot be the whole story. The problem isn't just that his love is inadequate to overcome his violent and possessive urges; the problem is that his love takes a form that feeds on and reinforces these urges. It is *bad love.*

Second advantage: Captures continuities and differences between various forms of love

In discussing the idea that a norm of respect is constitutive of love, I gestured at the fact that this idea sheds light on how different forms of love are related to each other. Consider the couple who parts not because they are unable to love each other well, but because their love has changed, lost its passion, and become a strictly *friendly* love. Their story might go something like this: Once, the feelings of pleasure they got from various forms of closeness—physical, emotional, intellectual—gave rise to subrational attractions to these things. On the basis of these attractions, they adopted such closeness as an end. (They also found they cared about each others' well-being and adopted related ends, but I will not address this part of the story.) As time went on, and they and their situation changed, these feelings and attractions faded, and eventually went cold. Perhaps this is the point where they decided to part ways or, more likely, since they still loved each other, perhaps they made an effort to rekindle the passion of their relationship.

In what way did they "still love each other" and why would this love urge them to try to feel passion once more? For one thing, many of the other feelings, attractions, and maxims of love were still present—especially those having to do

with their beloved's flourishing. Even more importantly, though, this couple still had the ends of closeness, even while lacking the attractions that were the original impetus for adopting those ends. We can continue to have such ends simply out of habit, but more often we do so out of a sense of what is owed to the beloved and to oneself. We have expectations and needs that we come to rely on each other to fulfill; we may, for example, want to be wanted, even if we find we do not want the other very much—and we may understand that our beloved has the very same need. If our story concludes with the couple parting ways, it is because their efforts to reacquire passionate feelings and attractions failed, and they found that their dispassionate love did not provide sufficient basis for continuing their lives together in the same way as before. They nevertheless continued to love each other, but as friends rather than lovers.[4]

More generally, the incorporation conception can appeal to not only different varieties of loving feeling and attraction, but the different ways that two people's ends can intertwine, and resulting differences in the norm of respect governing a loving relationship. All kinds of love involve some version of the feelings, attractions, and ends I have been discussing. For example, among the primary differences between friendly and romantic love are the degree and kind of intimacy that the friend or lover finds pleasurable. These differences encompass, of course,

[4] Velleman also argues that it is a mistake to conceive of love as a syndrome of feelings and motives because there clear cases of love where the lover lacks the desire to promote the beloved's flourishing, and because such conceptions make love into an unappealing pathology of overvaluation and transference. I leave it to the reader to imagine how to extend the responses I have developed to these arguments.

sexual intimacy, but also emotional and intellectual. Consider the fact that, when two friends share extremely deep emotional and intellectual intimacy, as Dora Carrington and Lytton Strachey seemingly did, we are inclined to say they are "in love" with each other in an atypical fashion, rather than that they are unusually close friends. These differences at the level of feeling in turn generate differences at the levels of attraction and maxims: one is attracted to enjoying these different intimacies with the friend and lover, and sets the ends of doing so. The result is that friends and romantic lovers intertwine their lives and their ends in different, though related, ways. Because of these differences, there are also differences in what it takes for friends and lovers to satisfy the requirement of respect.

The incorporation conception can even account for love for non-persons, such as the love for a pet, a work of art, a city, or even a cause. Each of these things is *valuable* in its own way, and so there are norms—analogues of the requirement to respect persons—governing our relations to them. It wouldn't make sense to say I should respect my dog, in the technical sense I am using, since she doesn't have any rational ends or the capacity to agree to or share my ends without rational conflict. Nevertheless, she is a sentient creature with needs, pleasures, pains, quirks, amusements, and anxieties. To love her is thus to enjoy her company and be attracted to the idea of contributing to her doggy flourishing; to make it a project to enjoy her company and to promote her flourishing; *and to strive to respond appropriately to the kind of creature she is, the value she embodies*.[5] I'll

[5] This ability to extend to not only different kinds of interpersonal love, but also love for animals and things is a major advantage the incorporation account has over, say, Velleman's or Jeanette Kennett's. (Velleman *op. cit*. and "Beyond Price," *Ethics* 118:

not venture a theory of animal value here, but any decent person has a sense of it, at least with regard to some animals.

Third advantage: Makes sense of love with and without reason

We love the people, animals, and things we love for reasons. Or do we? On the one hand, we think love is, as Niko Kolodny puts it, "an appropriate or fitting response to something independent of itself. Love for one's parent, child, or friend is fitting, one wants to say, if anything is."[6] On the other hand, we rebel at the idea that we should be able to justify our love; any reasons we might offer seem too trivial, too local, to underwrite our emotion. The incorporation conception makes sense of this ambivalence, because it says the roots of our love—our feelings and subrational motives—are unreasoned, while our loving maxims treat these and other considerations as reasons. Thus, there are genuine justifying reasons for which we love, but love in the fullest sense outstrips our reasons. The

---

2 (2008), 191-212. Kennett, "True and Proper Selves: Velleman on Love," *Ethics* 118: 2 (2008), 213-227.) Velleman argues that the love we have for persons is a response to their rational nature. No matter how compelling his account is for other reasons, it is troubling that he makes interpersonal love radically discontinuous with the love we have for non-rational entities. Similarly, Kennett argues that love is a response to the beloved's ability to value, where valuing is not an exercise of rational capacities, but an emotional and aesthetic awareness, perceptiveness, and responsiveness. Love as Kennett conceives it can take a broader range of objects than it can under Velleman's conception, but still not broad enough.

[6] Niko Kolodny, "Love as Valuing a Relationship," *Philosophical Review* 112:2 (2003), p. 135.

incorporation conception is thereby invulnerable to the objections Kolodny raises against both "quality" and "no-reasons" views of love.

According to the quality view, love arises in response to some of the beloved's (non-relational) traits, and those traits are the reasons for loving her. According to the no-reasons view, there are no reasons for loving—love is a "basic" or "unmotivated" desire.[7] On the incorporation conception, love is a syndrome of feelings, subrational motives, and rational motives, and it is correct to ask for reasons in the full justificatory sense in relation to and only in relation to the rational motives (i.e. maxims). Feelings and subrational motives are not the sort of phenomena that admit of justification.[8] When we come to the maxims of love, by

[7] This is Frankfurt's view.

[8] There is another sense in which we might ask for "reasons" for feelings and attractions, however—feelings and attractions are the sort of thing that can *make sense* or *fail to make sense*, even while they are not for justification through an appeal to reasons. We can make sense of a person's feelings and attractions not just by relaying a strictly causal account of where they come from, but through narrative and interpretation. If I describe certain scenes from my childhood, you might wind up thinking, "Well, no *wonder* you hate the cold so much (find it so painful)" or "Now I see why you want to live in the country." I haven't justified this propensity to pain or this attraction. Nor have I failed to justify these things; the point is that asking for justification is a category mistake. Nevertheless, I have made it possible for you to understand them in a way you couldn't before. The same is possible with regard to the feelings and attractions of love. We can, though

contrast, we are in justification territory. We can ask people to justify the ends they adopt and the means they are willing to take for the sake of their ends. The reasons for love are that the beloved has traits that make the lover *want* to be close to her, *want* to promote her well-being, and so on, and it is not disrespectful for her to have these ends.[9]

So the incorporation analysis isn't exactly a no-reasons view of love nor is it exactly a quality view. It does, however, have elements of both kind of view Kolodny finds objectionable, because the feelings and attractions of love do not admit of rational justification, and the maxims of love are justified in part by the fact that the beloved has traits that make the lover want to be near her and to promote her well-being. It is worth considering, then, whether Kolodny's objections to either kind of view apply to the incorporation conception.

Kolodny first raises a general objection to quality views that, at least for many cases of love, they mistake an expression of love for the reason for love. This is especially true for family members; the reason one loves family members is that they are family, though one way of expressing love is by appreciating their

---

a sort of hermeneutical activity, come to understand why people love whom they love.

[9] That is, these features cause the lover to take pleasure in the beloved's nearness and in contributing to her flourishing, and thereby cause the lover to be attracted to these prospects.

traits. This point is consistent with the incorporation conception. Relational qualities can give rise to subrational motives just as well as non-relational traits.[10]

Kolodny also raises three more specific objections to quality views of love: the problems of constancy, non-substitutability, and amnesia. For the sake of time, I'll address only the long-standing problem of non-substitutability, but I am happy to talk about the other two problems during q&a.

The problem of non-substitutabilityis this: If the reason to love A, the problem goes, is that she has qualities *qrs*, then doesn't one have just as much reason to love B, who also has those qualities? Here is where I think it is very important that, on the incorporation conception, the reasons for love are based in the lover's attractions. For whether one has reason to love B depends not on whether she has qualities *qrs*, but whether those qualities in her give rise to the feelings and attractions of love. It wouldn't be irrational or unreasonable for the lover of A to fail to have the feelings and attractions of love regarding B, because feelings and attractions are not the sort of thing that can be irrational or unreasonable. It may, further, *make sense* for the lover of A to fail to have the feelings and desires of love for B (even if B is a quality-"clone" of A)—the

[10] Moreover, understanding and appreciating the beloved's personal qualities is an essential part of the means to pursuing the ends of closeness and promoting her flourishing. To make the point more concrete, on the incorporation conception, it makes perfect sense for a parent to love her child—to have the feelings, attractions, and maxims of love—*because he is her child*. Then, in striving to promote the child's flourishing, the parent needs to seek to understand and appreciate his other personal traits (for example, it takes something different to promote a shy child's flourishing than it does to an outgoing child's flourishing). Also, in seeking to maintain and modify and/or develop the appropriate intimacy with one's child as he matures, it is important to attend to who he is and what he is like. This is how appreciating the beloved's qualities can be an expression of (the maxims of) love.

circumstances where a set of traits gives rise to these responses can be pretty specific. Now, if the lover of A does come to have the feelings and attractions of love in response to B's qualities *qrs*, then she has reason to love B—i.e. to adopt the maxims of love—assuming it is permissible for her to do so and that she wouldn't, for example, be betraying A by doing so.

The problems Kolodny raises for no-reasons views stem from the fact that, from both the first person and the third person perspectives, love or its absence can seem appropriate or inappropriate.[11] It is inappropriate for the parent to fail to love the child, or for the abused wife to love her abuser. Now, my intuitions about these cases are actually not so straightforward. Regarding abused partners, it seems to me unjustified that they should continue to care about and concern themselves with their abusers in certain ways—I want to argue with them that they are being *unreasonable*. However, given the history of many abusive relationships, it also seems to me *perfectly understandable* that many abused partners still love their abusers in some way. Regarding parental love, I think that once one has adopted the social role of parent, one has a moral obligation to cultivate a loving relationship with the child—this means adopting various ends, and also seeking to cultivate the feelings and subrational motives of love. Again, though, it is *perfectly understandable* that some parents lack these feelings and motives, and cultivate the loving relationship primarily out of duty rather than out of "natural" love. Anyone who shares my ambivalence about these claims of inappropriate or appropriate love should like the incorporation conception, because it explicates this ambivalence nicely.

[11] Kolodny also presses the problem of amnesia against the no-reasons view.

Finally, Kolodny argues that, unless we have the right account of the reasons for love, we cannot distinguish the psychological states constitutive of love from their occurrence outside of love. His example is awaking with a sudden and persistent urge to promote the flourishing of his daughter's classmate, whom he knows only distantly. This desire is one of the desires constitutive of love, but its occurrence here is not a case of love. This is a real problem for an account like Frankfurt's, since he conceives of love as nothing more than a second-order desire. But it is no trouble for the incorporation conception, because what are missing in the case of Kolodny's urge to benefit his daughter's classmate are the feelings and rational motives of love. It's true, if he woke up suddenly with full-blown feelings and subrational motives of love for this child, and then went on to adopt the rational motives of love on this basis, he would qualify as loving the child. This would be an odd case of love, one we cannot really make sense of, because we cannot understand how the feelings and subrational motives of love could just spontaneously generate in this way, nor can we see how such nonsensical considerations could provide a plausible basis for adopting a set of ends that would profoundly affect one's life. It is one thing to adopt the end of getting to the ice cream parlor because one suddenly has a whim for ice cream, and another thing entirely to take on the project of coming closer to and benefitting a new person because one has a sudden urge. Odd and mysterious though this love might be, however, I cannot see why we shouldn't call it just that.

Third advantage: Makes love a volitional form of valuing

The incorporation conception is what we might call a "volitional" analysis of love—that is, it makes loving someone essentially a matter of being motivated in certain ways. The appeal of a volitional analysis, generally speaking, is that it is intuitively true in paradigmatic cases that loving someone involves being motivated to seek intimacy with them, to promote their flourishing, and the like. Furthermore, the feelings we associate with paradigmatic cases—pleasure in the beloved's company, sorrow at their suffering—are readily explained by the success or frustration of such motives. However, Velleman, Kolodny, and others argue that at the heart of love is a psychological state distinct from being motivated: valuing.[12] And the notion that to love is to value certainly rings true. The incorporation conception of love allows both views to be true; volition and valuing are not distinct activities or attitudes. The notion that valuing should not be understood as a form of motivation probably gets some of its plausibility from the assumption that the moving force in motivation is always a subrational or arational state ("desire"), and it is fairly ordinary to value something without having any particular desires regarding it. If we instead assume the dualist theory of motivation that underwrites the incorporation conception, according to which there are both rational and subrational forms of motivation, it becomes eminently more plausible that valuing is a form of motivation. In particular, adopting maxims of protecting, benefiting, publicizing, securing, or understanding are all excellent candidates for valuing, depending on the thing being valued. We value artworks by seeking to secure, publicize, and understand them; we value people by seeking to respectfully protect, benefit, share intimacy with them; and so on.

[12] Kennett op cit agrees with Velleman and Kolodny about this.

The incorporation conception also entails an important aspect of valuing someone by loving them, an aspect that might seem distinct from motivation, and that is being emotionally vulnerable to them. Both Velleman and Kolodny emphasize this aspect of love. Kolodny writes, "Love is a kind of valuing. Valuing X, in general, involves (i) *being vulnerable to certain emotions regarding X*, and (ii) believing that one has reasons both for this vulnerability to X and for actions regarding X (150, emphasis added)."[13] Similarly, Velleman holds that loving someone is recognizing a value in her (rational autonomy) that gives one reason to allow oneself to be emotionally vulnerable to her. And of course the incorporation conception includes this, that loving someone involves being vulnerable to painful feelings on her behalf when she suffers, and on one's own behalf when she does not return one's love, or does not treat one well; and it involves being susceptible to positive feelings when she flourishes, and when she loves one back and treat one well. These feelings are the psychologically most basic element of love—they give rise to subrational motives to pursue closeness with the beloved and to promote her flourishing, and these motives are in turn the basis for adopting these outcomes as ends.[14]

---

[13] From the perspective of the moral psychology developed in section 1, Kolodny's view actually is a volitional account of love, because he claims love involves motivating judgments about practical reasons. I assume he thinks his view is not a volitional one because it does not put subrational desire at the heart of love.

[14] What should we say about the bad/abusive lover? Does he "value" his beloved? It seems to me we should say he values her in exactly the way he loves her: badly. His way of "valuing" her fails to recognize her nature, and what that nature requires of people

Conclusion: A general model for (some) emotional attitudes?

The incorporation conception of love explicates a great many of our intuitions and pre-theoretical beliefs about love. At the same time, it has several theoretical virtues, including providing a unified account of love as a syndrome of feelings and motives that is nevertheless a mode of valuing; resolving some long-standing puzzles about love, such as the problems of non-substitutability and constancy; applying to both paradigmatic and ambigous cases, while explaining why the ambigous cases are ambigous; and explaining the continuities and differences between various forms of love for persons and love for non-persons. All this is an argument not just for this conception of love, but also for the moral psychology that underpins it and makes it possible. I hope this discussion convinces some to consider the possibility that other emotional attitudes are ways of incorporating subrational feelings and attractions into our rational agency. Interpersonal anger, for example, may be best analyzed as a syndrome of feelings, attractions (and aversions), and maxims: when a person is angry with another, she feels pain at the person's presence and flourishing, has aversions to both of these things, and also treats those aversions as reasons to avoid the person and contribute to the person's suffering. This analysis requires refinement, of course; as it stands, it is true of an entire family of emotional attitudes, of which interpersonal anger is only one member. But the general strategy should be clear. There have been great

---

who value her. The case is analogous to someone who "values" a Fabergé egg by always carrying it in her pocket, or the Shroud of Turin by carefully laundering it with organic detergent.

strides in the last few decades in developing a compelling Kantian moral psychology and demonstrating its normative and explanatory power with regard to moral motivation. It is, however, a largely untapped resource when it comes to understanding emotions.

Functional Explanations for Constructivists

J. Thomas Mumm

One way of framing the realist-anti-realist debate in metaethics is in terms of two basic explanatory projects. On the one hand, metaethicists ought to accommodate, as far as possible, the "moral appearances": the basic intuitions that we take to be central to moral practice and our self-understanding as moral agents. On the other hand, we ought to provide an account of moral properties and moral truth that is consistent with our best scientific understanding of the world. Mark Timmons has called these the internal and external accommodation projects, respectively (Timmons, 1999). The dialectical situation between realists and anti-realists ordinarily takes the following form. Realists claim that only they can provide an account consistent with the moral appearances while anti-realists claim that only they can provide a metaphysically and epistemologically respectable one. Generally speaking, realists are strong on internal accommodation, weak on external, and anti-realists have it the other way around.

One of the appealing features of constructivism is that it promises to offer a third way, making sense of a robust form of moral objectivity without positing metaphysically or epistemologically mysterious moral facts or properties. Where non-reductive realists must dig in and appeal to brute, inexplicable moral facts, constructivists claim to explain these facts without explaining them away. But according to a line of attack developed by Chris Heathwood (2012), this is to claim the impossible. No metaethical theory can avoid positing brute moral facts. If this is so, it appears that anti-realists, constructivist and otherwise, lose one of their primary dialectical advantages.

In this paper, I will argue that there is a solution to this problem, one that is uniquely available to constructivists. I will contend that constructivists can provide a particular kind of *functional explanation* that makes sense of moral facts in terms of the nonmoral without threatening what I will call the *autonomy* of moral discourse. According to this line, the nature of moral discourse is to be explained in terms of the function of that discourse. If this function is to be achieved, moral claims must be oriented toward certain success conditions. And it is from these success conditions that we can construct a suitable bridge principle. This bridge principle gives shape to moral truth but is not itself a first-order moral claim. In fact, I will argue, given the kind of explanation I aim to provide, it is not *a priori* knowable that this principle would be ratified from *within* the practice of moral evaluation.

This paper will proceed in three stages. In the first section, I will present Heathwood's attack against anti-realism. In the second section, I will outline the strategy I think constructivists should follow, proposing a form of functional explanation that can ground a constructivist bridge principle. In the third section, I will consider a toy constructivist model to illustrate how such an explanation is supposed to work. Finally, in the fourth section, I will argue that, armed with the right kind of functional explanation, constructivists can retain their dialectical advantage over realists and consistently reject the existence of brute moral facts.

1 Bridge Principles and Brute Moral Facts

In response to this supposed dialectical advantage, some realists, such as Shafer-Landau and Parfit, have developed a partners-in-guilt reply, arguing that we also encounter brute facts in mathematics and physics. But this move rests on a questionable analogy between ethics and mathematics. Chris

Heathwood presents a different kind of partners-in-guilt reply, one that strikes closer to the heart. According to Heathwood, it's not only non-reductive realists who posit brute, inexplicable moral facts. *Any* metaethical theory must do the same, including reductionist and constructivist theories (Heathwood, 2012).

Heathwood's central strategy is to show that any moral theory must either explicitly posit bedrock moral truths or must explain moral truths in terms of what he calls a "bridge principle", so-called because it supposedly bridges the realms of moral and nonmoral facts. Divine Command Theory, for example, explains moral facts in terms of the commands of God. The problem, Heathwood argues, is that any such bridge principle amounts to a moral fact, and one for which no further explanation is possible. It is, in other words, just another brute moral fact.

To see how this is supposed to work, we need to look at an analogy Heathwood draws between bridge principles and those paradigmatic brute moral facts, Ross' *prima facie* duties. Heathwood suggests that we can formulate these duties in the form of "Rossian principles". Take the *prima facie* duty to keep one's promises:

> **Rossian Principle:** If a person has made a promise to perform some act then the person has, in virtue of that, a prima facie moral obligation to perform that act.

According to Ross, facts of this kind have no source, and cannot be explained in terms of anything more basic.

Heathwood's contention is that bridge principles of any kind are analogous in structure. Take the Social Contract Theory. According to Heathwood's line, its central claim is, like the Rossian principle, a moral claim. Start with the ordinary formulation:

> **SCT:** An action is wrong (or right) if and only if rational, self-interested contractors would agree, on the condition that others do so as well, to rule it out (or allow it) in deliberation about what to do.

Part of the biconditional is the following claim:

> **SCT*:** If rational contractors would agree, on the condition that others do so as well, to rule out φ-ing in deliberation about what to do, then, in virtue of that, φ-ing is wrong.

Heathwood points out that this principle shares the same structure as the Rossian principle:

> **Rossian Structure:** If such-and-such nonmoral condition holds, then such-and-such moral condition holds in virtue of that.

He argues that since the Rossian principle is undeniably a moral claim, bridge principles like SCT must be as well (Heathwood, 2012).

Heathwood takes himself to have established the following. Constructivists attempt to explain moral facts in terms of nonmoral facts. But any explanatory principle they posit will, at least in part, bear a Rossian structure, and will therefore be a moral claim. Constructivist theories, therefore, must always posit brute, inexplicable moral facts. The dialectical advantage is lost. Call this the Bridging Problem.

I will argue that constructivists can overcome this problem by employing a functional explanation of moral discourse. Following this strategy, they can give a principled account of why the correct bridge principle need not be a principle internal to morality, and yet can still serve to help explain and give shape to moral facts. One of Rawls' influential constructivist suggestions was that moral truths are solutions to practical problems. The strategy I am proposing retains the spirit of Rawls' idea while making it more precise.

2 Functional Explanations

I'll begin with some definitions. Let D be a domain of discourse. The domain includes all the beliefs, true and false, that count as part of the relevant discourse. "Murder is good", for example, is a moral belief, even if obviously false. Call the set of all true beliefs in D the **shape** of D. And call the set of all

warranted beliefs in D the **internal view** of D.

There are at least two tasks we are faced with in providing a philosophical account of a domain of discourse. First, we must say something about the shape of the discourse. What does it mean to say that some beliefs in that domain can be true? This is the problem discussed in the last section. But we must also say something about the internal view of the discourse. What is it that warrants inferences and basic beliefs in that discourse? This is the epistemology of the domain. It is only of secondary concern for our purposes, but to anticipate a conclusion I will be drawing down the line, it is useful to point out that this distinction between shape and internal view can serve as one way to pin down what it means for a domain of discourse to be *autonomous*.

If a domain is autonomous, then one can only address and justify claims about warrant *from the inside*, as engaged practictioners. Morality might well turn out to be autonomous in this sense, but an artificial discourse like that of particle physics might not. For though it could be that from the inside, beliefs about the existence of particles are justified by appeal to their explanatory power, it is also true that we can ask, from the external standpoint of ordinary empirical discourse, whether such beliefs actually represent physical entities. If they do not, such beliefs might only be *weakly* warranted, where "weak" warrant could be cashed out in terms of a fictionalist account, or something along those lines. If they do, then such beliefs can be strongly warranted, but this normative status would be conferred by the standards of warrant of an external standpoint.

By definition, a belief in domain D is true if and only if it is contained in the shape of D. Part of what we must do in accounting for the objectivity of some target domain of discourse is to figure out how its shape is determined. Notice that the intuitions that morality is objective and that we can be mistaken in our moral beliefs together amount to the intuition that morality has a definite shape. If we can provide an external explanation of why it has the shape it does, and not some other, then we would have made significant progress in advancing the external accommodation project while at the same time remaining true to our strong intuitions about moral objectivity.

2.1 A Schema for Functional Explanation

My thesis is that the shape of a domain of discourse depends on the function of that discourse. As a toy model, consider the Hobbesian Social Contract Theory (SCT) described above. On one possible interpretation, this theory posits a particular function for moral discourse: namely, that of solving collective bargaining problems. Grant for the moment that moral discourse emerged to play this role. A natural question is *why*? What features of moral discourse enable us to successfully solve such problems? To answer this question is to provide what I will be calling the **success conditions** for the achievement of the function.

A Hobbesian answer might go as follows: in order to solve collective bargaining problems, we must fix on and regulate ourselves in accordance with principles that would secure our self-interest. But because of prisoner's dilemma and tragedy of the commons type situations, we do best to forgo the goal of self-interest when it would be ruled out by principles that would be agreed to by other self-interested agents. A discourse that successfully leads us each to pursue self-interest only under these conditions would, on balance, function to solve a variety of bargaining problems.

But what would the shape of such a discourse be? The Hobbesian answer is that it is shaped by whatever rational, self-interested contractors would agree to on the condition that others do so as well. We can formulate a bridge principle on this basis, namely SCT:

> **SCT:** An action is wrong (or right) if and only if rational, self-interested contractors would agree, on the condition that others do so as well, to rule it out (or allow it) in deliberation about what to do.

Why call this a bridge principle? Because it stands between internal engagement in moral discourse and external considerations about the nature of that discourse. On the one hand, it gives shape to the moral domain. On the other, it is explained by appeal to the function of that domain.

Notice that one question this kind of explanation is meant to answer is "why this bridge principle, and not some other? ". This anticipates one part of my answer to Heathwood's Bridging Problem. For what is questionable about brute moral facts is that they can be controversial but are inexplicable in principle. Bridge principles, of course, are even more likely to be controversial, and asserting them as brute, inexplicable facts would be utterly unsatisfying. But this sort of functional explanation provides a principled way for choosing among possibilities. Such a theoretical choice can be justified by appeal to an explanation external to the target discourse.

But I have not yet presented the resources for denying the Heathwood line that such bridge principles still necessarily involve appeal to brute *moral* facts, regardless of the nonmoral explanation for choosing them. In the next section, I will consider my toy Hobbesian model in more detail in order to show that such an account can provide a (possible) explanation of moral truth without depending on brute moral facts.

3 A Constructivist Model of Moral Discourse

According to my brief sketch of a Hobbesian constructivist theory, its elements were as follows:

Function: (Bargaining) Solve collective bargaining problems.

Success Conditions: (Mutual Regulation) Regulate our behavior in such a way as to secure self-interest, but within limitations acceptable to any self-interested agent, in order to make possible optimal collective responses to prisoner's dilemmas, tragedy of the commons scenarios, etc.

Bridge Principle: (SCT) An action is wrong (or right) if and only if rational, self-interested contractors would agree, on the condition that others do so as well, to rule it out (or allow it) in deliberation about what to do.

My purpose is not to defend this model of moral discourse, but instead to explore whether it falls prey to the problems raised by Heathwood. If it does not, then it would turn out they are not problems necessarily facing constructivist theories. The dialectical advantage claimed by constructivists *vis a vis* brute moral facts could be regained.

To review, Heathwood's point was that constructivists are unable to formulate metaethical theories that are free of appeal to brute, inexplicable *moral* facts. In particular, because bridge principles will always at least partially involve a Rossian Structure, they will directly involve brute moral claims (albeit in conditional form). Can the functionalist social contract theory I've described avoid these problems?

Let's begin by considering the purported *autonomy* of moral discourse, a feature emphasized by Dworkin, Blackburn, Strawson, McDowell, Scanlon, Darwall, and others. Although these thinkers understand this feature in different ways, there are some useful points of similarity. First, we can only address particular moral problems from the committed moral point of view, as reasoners engaged in the moral project. Ethics is a distinctively philosophical subject matter; it cannot be the direct object of natural scientific study. It is only internally that we can grapple with ethical questions.

Second, our moral conclusions can only be justified by appeal to further moral facts or principles. This intuition helps explain objections to "naturalistic fallacies". Say it turns out that moral attitudes and practices developed in large part because they improved the genetic fitness of our ancestors. This fact does not yet justify any particular moral conclusions (for example, that it is wrong to do things that diminish the genetic fitness of our species, or that we are morally required to promote it).

Based on these considerations, we can define an autonomous domain of discourse as follows:

Autonomous Domain: A domain of discourse D is autonomous if and only if particular questions in D can only be addressed from within D, and particular D-claims can only be warranted by at least partial appeal to further D-claims.

The autonomy of morality centrally concerns what I have been calling the domain's internal view (the set of all warranted moral beliefs). This set is fixed from within according to internal standards of

warrant.

In light of these considerations, one way to reframe Heathwood's view is to say that since bridge principles have implications for which moral beliefs are justified, they must themselves be internal claims, given the autonomy of the domain. But notice that my explanatory framework (a) draws a distinction between the shape and the internal view of a domain of discourse, and (b) posits bridge principles in order to explain the shape alone. A constructivist about truth is centrally interested in the nature of *truth*, not warrant. Where these come apart, we must form our beliefs according to our best standards of warrant. A bridge principle might tell us what *makes* beliefs in the domain true, but this does not mean that it will be a useful tool for forming those beliefs in a reliable way.

My claim is this: a bridge principle figuring in a functional explanation need not be an internal principle at all. This will depend on the nature of the bridge principle and the details of the functional explanation in question. If intuitionism is the correct account of mathematical truth, for example, then it seems plausible that "only believe those propositions for which a suitable proof has been constructed" would stand up to scrutiny from within the mathematical point of view. But notice that in this case the shape and internal view of the domain are equal. So here it happens to be the case that the principle that shapes the discourse is also the principle governing which mathematical beliefs are warranted. Must the same be true in the case of the social contract theory?

Consider the following question: would rational contractors agree to SCT as the principle governing moral reasoning? It is possible that they would, but there are reasons to be hesitant here. For it may not be realistic to think that we can determine with reasonable certainty just what rational contractors would in fact agree to. That's because we are far from "rational" in the strong sense needed for the theory. This is one common objection to social contract theory in normative ethics: it doesn't render a very useful decision principle.

But useful or not, the SCT principle might still shape moral discourse in the technical sense I've been discussing. The metaethical theory I've sketched was not built according to the methods familiar from normative ethics. In normative ethics, we often begin from our core moral intuitions and attempt to derive principles that make the best sense of them. If a candidate principle yields deeply counterintuitive conclusions, this counts against the theory. We then evaluate how reasonable it is to take those principles as guides for moral living.

The debate between objective and subjective versions of utilitarianism (and associated concepts of actual and expected utility) provides one example of this kind of theorizing. What I'm suggesting is that the principles derived in this way are best seen as principles governing or explaining the *internal view* of morality. Understood in this way, we can see the particularist debate, for example, as a debate about how moral conclusions are justified. The particularist thinks derived principles are at best generalizations; her opponent might think that there are *a priori* principles governing the domain. If these debates are interpreted as *internal*, then they are consistent with a wide array of metaethical theories of moral truth.

In contrast, the metaethical theory I sketched was built by looking carefully at the nature of moral discourse as a whole, viewed from without. Why did it come to exist in the first place? What are agents doing when they engage in it? What distinctive function or functions does it play? These questions can help us pin down which properties play the truth-role in the moral domain, while remaining neutral on internal questions about the justification of moral beliefs and actions. Whether or not the bridge principle derived from this method is one that would be ratified from within moral discourse is a *further* question. If it would be, and morality comprises an autonomous domain, then it is not *because* of the correct metaethical theory that it would be so ratified. So we need to make a familiar distinction between metaethical and normative principles. And my model provides us with reasons for doing so.

4 Conclusion: Must Constructivists Posit Brute Moral Facts?

So where does this leave us regarding Heathwood's Bridging Problem? Doesn't SCT, in part, bear a Rossian Structure? To review, the questionable part of the principle is as follows:

> **SCT*:** If rational contractors would agree, on the condition that others do so as well, to rule out φ-ing in deliberation about what to do, then, in virtue of that, φ-ing is wrong.

The important question is whether or not this is a brute, inexplicable moral fact. A tempting strategy is to suggest that it is a moral fact, but it is not brute, since its status as a moral fact is itself a function of the fact that rational contractors would agree to take account of it in deliberation about what to do. But Heathwood argues that this kind of move results in a regress (Heathwood, 2012, 6-7). For we now have a higher level principle of the form:

> **SCT**:** If rational contractors would agree on SCT, then, in virtue of that, ignoring SCT is wrong (or something along these lines).

And since this also bears a Rossian Structure, we find ourselves in the same situation, *ad nauseum*.

We can avoid this problem by drawing a distinction between internal moral claims and claims *about* moral discourse along the lines laid out by my theory. A moral claim has as its subject matter particular moral truths (facts). Claims of this kind constitute the basic moves within the domain of moral discourse. A claim *about* moral discourse, on the other hand, looks at the domain from the outside as a practice susceptible to different kinds of analysis (psychological, sociological, historical, and, most importantly here, metaethical). My metaethical constructivist proposes that domains of discourse emerge to serve some function, and that we can understand the nature of the truth-predicate ranging over that domain in terms of the success conditions for serving that function. This gives us insight into the shape of the domain. The bridge principle discovered through this analysis is what gives the domain its distinctive shape.

In the case of SCT, the proposal is that moral truth is constructed from whatever rational contractors would agree to under certain conditions. But this is not a moral claim, despite its partial Rossian Structure. First, it is not justified, nor could it be justified, by direct appeal to moral facts. Rather, it is justified by metaethical explanatory considerations. Second, it does not have obvious *normative force*. In any particular circumstance, whether or not the principle underwrites a moral reason to perform or avoid some action must be worked out from within the moral point of view.

If the principle would be ratified from within, then it will have normative force, but only as a result of being ratified in this way. If it would not be ratified from within, then it turns out to provide an unhelpful decision procedure. The principle is not itself a reliable method for forming beliefs that are consistent with that very principle. If moral discourse were governed internally by the principle, it would fail to play its distinctive function successfully. So we have an explanation of the scenario where it is not ratified from within despite giving shape to the domain.

It should be pointed out that my theory would almost certainly disappoint those who are attracted to constructivism for the reasons Korsgaard emphasizes in *The Sources of Normativity*. Her explicit aim is to find a metaethical principle that is ratifiable from within the moral point of view. It is meant not only to account for the objectivity of moral truth, but also to serve as a reliable decision procedure. And furthermore, it is meant to ground a universally necessary moral principle that has normative force for any rational agent (Korsgaard, 1996). The version of constructivism defended here does not necessarily bear this kind of unity. That will depend on the details of the account. And it will almost certainly fail to satisfy Korsgaard's Kantian hopes. But what I have suggested in this paper is that a theory of this kind can nevertheless make further progress on both the internal and external accommodation projects than its prominent realist and anti-realist rivals. And it can do this without making dubious appeals to constitutively necessary commitments or other ambitious rationalist claims that weaken the Korsgaard-style constructivist's prospects for external accommodation.

References

Blackburn, S. (1993). Essays in Quasi-Realism, Oxford: Oxford University.

Darwall, S. L. (2006). The Second-Person Standpoint: Morality, Respect, and Accountability, Harvard University Press.

Dworkin, R. (1996). Objectivity and truth: You'd better believe it, Philosophy and Public Affairs, 25(2): 87–139.

Heathwood, C. (2012). Could morality have a source?, Journal of Ethics and Social Philosophy 6(2): 1–19.

Korsgaard, C. (1996). The Sources of Normativity, Cambridge: Cambridge.
McDowell, J. H. (1998). Mind, Value, and Reality, Harvard University Press.

McGrath, S. (2012). Relax? don't do it! why moral realism won't come cheap. Unpublished manuscript.

Functional Explanations for Constructivists

Scanlon, T. (1998). What We Owe to Each Other, Belknap Press of Harvard University Press.

Strawson, P. F. (1962). Freedom and resentment, Proceedings of the British Academy 48: 1–25.

Timmons, M. (1999). Morality Without Foundations: A Defense of Ethical Contextualism, Oxford University Press.

Carving a Niche for Immoderate Moral Realism[1]

Abstract

I outline a problem from disagreement for moral philosophy, specifically one that arises from the disagreement prevalent amongst moral philosophers themselves. I look at one "moderate" response to this problem given by Ralph Wedgwood, show why the moral realist shouldn't be satisfied with such a moderate response, and suggest another, "immoderate" response to the problem from disagreement.

1. Introduction

My goal here is (a) to briefly introduce a problem from disagreement for the practice of moral philosophy, (b) to analyze the dialectic between the non-realist who pushes the problem and a moderate realist response, and (c) to argue that at least some realists shouldn't be satisfied with a moderate response. My hope is to map and expand the dialectic surrounding this particular problem for moral philosophy; some details must be left fuzzy, but hopefully the ball will be moved forward. And in the end I will have provided what I hope are some compelling hints as to the direction the moral realist ought to take.

2. A Problem for Moral Philosophy

I'll define moral realism here as "the view that moral beliefs have *non-relativistic truth-values*," the denial of which is moral non-realism. (Wedgwood forthcoming, 1) There's a distinct argument from disagreement for moral non-realism that can be found implicitly in Nietzsche. This argument "calls attention not to 'ordinary' or 'folk' moral disagreement, but rather to what should be the single most important and embarrassing fact about the history of moral theorizing by philosophers over the last two millennia: namely, that no rational consensus has been secured on any substantive, foundational proposition about morality." (Leiter 2010)

[1] Thanks go to Danny Simpson, Patrick Kain, Jarod Sickler, David Horner, and Nathaniel Warne for helpful conversation and feedback on this topic, as well as the folks at the 2013 Talbot Philosophical Society Graduate Conference.

The argument has a couple important features. First, it targets *moral philosophers in particular*. Moral disagreement among moral philosophers, those who've spent lives devoted to discovering the truth about moral questions, is especially conspicuous.[2] Second, the disagreement being highlighted here is *disagreement about moral theories*, that which the Nietzschian calls the "substantive, foundational proposition[s] about morality." Typically, moral realists think of the collection of moral beliefs in a loosely three-tiered fashion. On the bottom tier are the moral judgments we make about specific situations. Call these specific moral judgments. At the next tier up, the middle tier, are the "general moral judgments" under which we subsume our specific moral judgments. The top tier is where moral theories reside. Our moral theories, minimally, are meant to explain our low and middle tier moral beliefs. The problem is that deeply entrenched disagreement is very wide-spread at the theoretical tier, much moreso than at the lower tiers. Most philosophers agree that one ought not murder, or that one should tell the truth. But there is no such near-universal agreement over the ideological foundations for these agreed upon lower-tier truths.

This fact about disagreement demands an explanation. The Nietzschian explanation of these facts, provided by Brian Leiter, is that the various moral theories "answer to the psychological needs of philosophers. And the reason it is possible to construct 'apparent' dialectical justification for differing moral propositions is because, given the diversity of psychological needs of persons (including philosophers), it is always possible to find people for whom the premises of these dialectical justifications are acceptable." (Leiter 2010) The moral non-realist needn't say this exactly. He need only provide some non-realist explanation of the

[2] And it's not as if conversions from Kantianism to consequentialism or from consequentialism to virtue ethics, for example, occur with any sort of regularity; the disagreements are deeply entrenched. This unyielding disagreement amongst moral philosophers, one might think, demands an explanation with more force than everyday folk moral disagreement, just as disagreement about historical facts among historians demands explanation with more force than everyday folk historical disagreement.

widespread disagreement among moral philosophers and provide reasons to think his explanation is simpler than a realist explanation, that it accounts better for the empirical data, and that it has whatever else constitutes better-making features of explanations. In doing so, he attempts to expose the practice of moral philosophy as illegitimate.

3. A Moderate Response

Ralph Wedgwood provides a response to this sort of argument that advances the dialectic in favor of moral realism by articulating a "non-skeptical moral realist" alternative to the Nietzschian explanation of disagreement. We can agree with Wedgwood in saying that "[a] version of moral realism counts as 'non-sceptical' if and only if it does not make it implausible to claim that a reasonably large number of ordinary thinkers know a reasonably large number of moral truths." (Wedgwood forthcoming, 1) Wedgwood's preferred version of non-skeptical moral realism is a *moderate* version. This is to say, Wedgwood thinks that we can plausibly claim that a reasonably large number of ordinary thinkers know a reasonably large number of moral truths, but that there are a number of moral truths that are much harder (if not impossible) to know, among which he includes theoretical moral truths.

First, Wedgwood points out the importance of the fact that there is a great amount of *agreement* amongst moral philosophers about *non-theoretical moral truths*. He puts it this way: "Certain central moral truths are equally widely agreed. Almost everyone agrees that we should normally keep our promises, refrain from killing and stealing, be grateful to those who have been kind to us, and so on." (Wedgwood forthcoming, 18) The sorts of moral beliefs Wedgwood focuses in on here are general moral judgments. By taking these general moral judgments to pick out the "central moral truths," he begins to turn the Nietzschian description of things on its head, since the problem the Nietzschian is trying to drive home is that the moral propositions upon

which moral philosophers agree are the trivial sort, not the substantive or central sort. Theoretical moral beliefs may very well be *ideologically* or *explanatorily* substantive, since they purport to tell us what makes the lower-tier truths true. If Wedgwood is right to think of our theories as inferences to the best explanation of our lower-tier moral beliefs, however, our general moral beliefs are *epistemologically* or *justificatorily* substantive.[3] They form the foundation upon which we base our belief in any given theory.

Next, Wedgwood assumes (to fix ideas) that our moral intuitions are given to us by our emotional dispositions and that our moral intuitions are the basis for our specific and general moral judgments. According to Wedgwood's moral psychology, we have certain emotional reactions to actual events in which we find ourselves and hypothetical events we simulate in our imagination. It's these emotional responses to both real and imagined stimuli that act as the basis for our moral intuitions. Our moral intuitions then serve as grounds for our specific and general moral judgments, which are explained by our moral theories.

As we've already seen, Wedgwood thinks that there is rather wide-scale agreement about these specific and general moral judgments. Obviously, there is *some* disagreement at the lower tiers though. Wedgwood acknowledges this. He claims that "[t]ypically, the areas where philosophical disputes arise concern… cases that are relatively peripheral to most people's moral sensibility (such as the ethical status of human foetuses and non-human animals)…." Here, Wedgwood's view begins to allow for a limited skepticism, for those non-theoretical moral judgments on which we agree are all explained by any number of normative theories. The

[3] Wedgwood is certainly not alone in thinking this way about the origin of our theories. Those who endorse reflective equilibrium as the method of settling on a moral theory are in agreement with Wedgwood on this score. For a good representative picture of reflective equilibrium, see DePaul (2006). Wedgwood explicitly endorses reflective equilibrium at Wedgwood (forthcoming, 10). Those who speak of evaluative doxastic practices by which we form moral beliefs also seem to view moral theories as explanations of our lower tier moral beliefs. See, for example, Alston (1989) and Adams (1999).

explanation of the non-theoretical moral judgments on which we agree is *largely overdetermined*. Thus, in order to settle on a theory, one must appeal to one's moral intuitions about cases that are contentious and exceptional: trolley cases, complex cases dealing with disputed issues, cases where rules seem to conflict, etc.

Wedgwood thinks that we ought to be less trustful of our moral intuitions in the abnormal and exceptional cases, however. This is because of his view about the origins of our moral intuitions. Our moral intuitions, according to Wedgwood, are grounded in our emotional responses to cases. It's hard for us to form clear emotional reactions (and thus to have clear moral intuitions) about abnormal and exceptional cases, though. It's not important for our purposes whether Wedgwood's account of the origins of our moral intuitions is correct. What's important is to see how his view leads to a kind of theoretical skepticism. If the only way to determine the correct moral theory is to appeal to our moral intuitions about abnormal cases, and our moral intuitions about abnormal cases are likely to be unreliable, then we shouldn't be very confident in our ability to discern which is the correct moral theory. Wedgwood reaches this exact conclusion, and it's this conclusion that makes his view a “moderate” non-skeptical moral realism. This is not the only problem, on Wedgwood's view, for the attempt to settle on a true ethical theory. Not only is it difficult to come by reliable moral intuitions by means of which one can arbitrate between theories, the task of arbitration itself is rather difficult. Settling on a *best explanation* of the myriad of moral beliefs one has is hard. (Wedgwood forthcoming, 22)

Wedgwood remarks, rightfully I think, that this doesn't at all imply that there's no truth of the matter about what moral theory best explains the true general and specific moral judgments folks have; that is, our difficulty in coming to settle on a moral theory doesn't entail that there's no true moral theory. However, it does yield a sort of moderate moral skepticism. It takes

endorsing the view that "if moral philosophers were perfectly rational, they would not have complete confidence in any particular ethical theory at all. They would have a mere partial degree of belief instead." (Wedgwood forthcoming, 23) This reservation is a result of the undercutting defeaters Wedgwood thinks we have for the moral intuitions that help us settle on a theory. In concluding so, Wedgwood appears to save moral realism from the problem from disagreement, but only at the cost of severe justificatory limits on our theories.[4]

4. An Immoderate Alternative

Now, I'd like to motivate the notion that Wedgwood's explanation of the problem from moral disagreement proves unsatisfactory for the moral realist, since the justificatory limits on moral theories once again call the practice of moral philosophy into question. I'll begin by asking the following question: *what role(s) should our moral theories play?* Wedgwood's conception of the method of moral theory, as was stated earlier, is that of an inference to the best explanation (Wedgwood forthcoming, 21) It seems fairly clear, though, that Wedgwood has in mind no other role for a moral theory than that of explanation. Just as a theory in a science is meant to explain specific data and general scientific principles, a moral theory is meant to explain specific and general moral judgments. But this elicits a further question: *should moral philosophy look just like science and metaphysics?*

There are some who would answer this second question with an emphatic, "No!," but would then go on to recommend that we engage in no moral theorizing at all.[5] I'm inclined to

[4] Wedgwood gives an artificial example that provides some indication of just how reserved he thinks one ought to be about one's belief in one's chosen theory. According to his example, one can proportion one's belief so as to consider one theory to be 40\% likely and two competitors each to be 30\% likely. In this case, one could believe the first theory is the correct one, since it's the most likely of the three, but it's still more likely to be incorrect, all things considered. It's not clear \emph{exactly} how reserved one ought to be about one's beliefs concerning moral theories on Wedgwood's view. But it's quite unlikely that one will be in a position to hold much confidence in one particular theory.

[5] See, for example, the work of Annette Baier, as well as Setiya (2010), Clarke (1987 & 1989), and Millgram (2002).

agree about the "no" answer concerning whether moral theorizing ought to look just like scientific theorizing, but not with the call to abolish moral theorizing altogether. There's good reason to think there's some true moral theory that explains the lower-tier moral truths. If there weren't a true moral theory there would be no explanation of why murder and lying share the property of moral wrongness, and this would be quite surprising.[6] But it's quite plausible, given moral realism, that our moral theories ought to do more than *merely* explain. After all, the realm of the moral has something the scientific, metaphysical, historical, and some others don't: it's normative. More specifically, morality is about *practice*, i.e. about how we ought to live. Thus, it's quite plausible to think that our moral theories ought to have practical significance for living; they ought to be directed towards making people good.[7]

The point of suggesting this is to say that if moral realism is true, then moral theorizing, unlike theorizing in metaphysics or chemistry, but very *like* theorizing in, say, medicine, is and ought to be directed towards a practical goal. Moral realists, believing in objective goodness, shouldn't settle for moral theorizing that has nothing to do with promoting the good. But Wedgwood's moderate realism doesn't allow for theorizing to be directed at anything beyond either psychological fulfillment or the satisfaction of intellectual curiosity.[8] So it looks like there's room in the dialectic for an *im*moderate non-skeptical moral realist, one who expects her theories to do more work than the moderate realist who thinks moral theories should be held

[6] I obviously brush aside concerns about particularism here, but only because that's entirely too much to introduce here. The particularist, however, will be dissatisfied with Wedgwood's theory as well, just for different reasons.

[7] This suggestion may seem, at first glance, to amount to a commitment to a specific moral theory: virtue ethics. But this isn't so. For goodness of a person can be defined (if one is so inclined) in terms of goodness of their actions, and can thus accord with non-aretaic moral theories.

[8] It's another question altogether whether our moral theories are satisfactorily fulfilling the role of explanation if we are unable to hold them with much confidence. One might think that such an explanation, even if successful, is trivial since if one doesn't believe it with much confidence it doesn't look that it contributes to one's *understanding* of morality.

loosely and are merely explanatory in nature. The following argument clarifies the immoderate realist's dissatisfaction with the moderate realist:

(1) Morality is about being and doing good.

(2) Moral philosophers, as experts in morality, ought to be able to increase understanding about and proficiency in morality.[9]

(3) Thus, the work of moral philosophers ought to be aimed at increasing understanding about and proficiency in being and doing good. (made plausible by (1) and (2))

(4) Moral philosophers are unjustified in putting much confidence in any specific moral theory. (assumption: the conclusion of Wedgwood's view)

(5) But, if moral philosophers can't justifiably be very confident in their theories, then they can't justifiably recommend their theories as helpful in increasing understanding about and proficiency in being and doing good, (and thus their work can't be helpful in this sense).

(6) Thus, the work of moral philosophers can't increase understanding about and proficiency in being and doing good.

(7) Thus, the work of moral philosophers can't do what it ought to.

(1), I think, is a conceptual truth and shouldn't provoke much argument. (2) is implicit in the suggestion above that our theories ought to contribute to making us good. It certainly is quite

[9] For a representative example of one influential moral philosopher who suggests this sort of view about moral philosophy, see Robert Audi, who gives an account of how virtues can be understood as a "second-order understanding\dots of how to deal with conflicting moral considerations." (Audi 2001, 632) In the context of Audi's paper, a first-order understanding involves having correct general moral judgments; so Audi here is advocating a view according to which a moral theory can provide a higher-order understanding of the general principles that correspond to those judgments. He is advocating moral theory as a an integral part of ordering one's moral beliefs and resolving conflicts between them for the sake of directing action. This isn't abnormal. A number of other philosophers working in normative ethics and metaethics are (plausibly) folks who intend their work to be geared towards enabling a better understanding of ethics *for the sake of promoting goodness*. See, for example, the moral philosophy of Robert Adams, Robert Roberts, and Peter Singer.

plausible, however, as an instance of a general principle about the roles of experts. Namely, the following:

> (2′) Experts ought to increase understanding about and proficiency in their respective fields.

(2′) is, I think, even more plausible. What could experts be for if not for those roles? Of course, how much emphasis is placed on increasing understanding versus increasing proficiency will vary, depending on the field. It's not clear, for example, that the expert metaphysician can fulfill the proficiency requirement in any meaningful sense. Even the expert physicist, however, can play this role, increasing our proficiency in inventing and engineering artifacts by which we increase our power to master our surroundings. If our best theory of physics didn't make us more proficient in this sort of manipulation of and control over nature, we wouldn't hold it in such high esteem. How much more should the expert of a normative field like morality aim to increase our proficiency in that field? Thus, unless one can provide reason for doubting (2′), or for thinking that morality is somehow different from other disciplines such that those whose expertise is in morality don't fall under this general principle, then (2) ought to be accepted.[10] (3) is plausibly inferable from (1) and (2). (4) is the assumption that Wedgwood's moderate non-skeptical moral realism is true; thus, whatever conclusion the argument yields is a result of accepting Wedgwood's view. (5) is the only premise that needs further sustained defense, since (6) and (7) follow from what comes before them. Thus, if (5) is right, then Wedgwood's view leads to the view that moral philosophers can't accomplish what they ought to.

---

[10] Obviously the details as they concern moral experts are going to be complicated. One can't rightly make others good by lecturing at them about theory. Moreover, in any given field, some experts might justifiably focus more on the goal of increasing understanding while others focus more on the goal of increasing proficiency. But (2) acts as a plausible constraint on a realist view of what moral theories are for. If they can't contribute to the goals of increasing understanding and proficiency, then they aren't doing what they should.

Before attempting a defense of (5), I want to address a second concern one might have with (2). Being a moral philosopher, one might say, does not make one a moral expert, at least not in the sense of being good. So why should one expect moral philosophy to contribute to being good? In response, I can offer two thoughts. One is to point back to my reasoning above: those who study a field are the supposed experts of that field, and experts ought to be able to promote understanding and proficiency in their given field. If a given moral philosopher can't do so, he's not very good at his work. Second, I can point to the precedent one finds for my view of moral philosophers throughout the history of philosophy dating back to Plato and Aristotle. In 463-466 and 500 of the *Gorgias*, for example, Socrates assigns philosophy a task which is analogous to medicine for the soul. The just person, the one who *practices philosophy*, is able to ensure the health of her own soul and of others' souls. In Aristotle, one finds the role of moral philosopher and medical expert compared once again. At 1094a8 of the *Nicomachean Ethics*, he introduces the analogy, and medical analogies remain throughout the work, demonstrating the fact that he takes the work of ethics to involve the health of the soul (which we may take, loosely, to involve moral goodness), and the work of the moral expert to be that of promoting health in the soul.[11] One can see this immoderate response to Wedgwood, then, to be carving a niche for those who buy into this picture of moral philosophy.

Why think that if moral philosophers can't be justifiably confident in their theories, then they can't justifiably recommend their theories as helpful in increasing proficiency in being and doing good? That is to say, what reason is there to think (5) is true? In order to formulate an answer to this question, it's necessary that we return our attention to the reason that Wedgwood thinks we ought to be reserved in our beliefs about theory. Wedgwood thinks that the explanation of the normal moral beliefs on which there's general agreement is overdetermined by the various

[11] For a fuller picture of Plato and Aristotle's medical analogies, see Jaeger (1957), and Lloyd (1968).

moral theories. The only way one can settle on an individual theory, then, is to appeal to moral judgments about abnormal cases, judgments that Wedgwood's moral psychology tells us are less likely to be reliable. But the unreliability of our intuitions about these abnormal cases ensures that anything derived from them is also unreliable. Thus, we can't justifiably place much confidence in our moral theories, since they are derived from unreliable sources; all moral theories come with undercutting defeaters.

If moral theories are meant to increase understanding and proficiency in morality, however, then they'll need to be able to resolve issues of conflict and disagreement *concerning the very abnormal cases that provide the grounds for distinguishing between various moral theories*. The point is this. Moral philosophers aren't needed to teach people that murder is wrong or that feeding the poor is good. The moral judgments that moral philosophers all share are shared by everyone else as well. What moral philosophers need to be able to do, if they're to fulfill the role of increasing understanding about and proficiency in being and doing good, is to resolve the difficult, abnormal, and exceptional cases on which normal folks disagree. Moral philosophers, and the theories they form, must be able to guide people past the easy cases on which there's general agreement. Some moral philosophers already take their theories to do just this. Thus, one meets Kantians who have resolved never to lie and consequentialists who are willing to happily lie through their teeth to produce good results. Such philosophers make internal appeal to their theories to resolve difficult moral cases.

The problem with this is that, if the moderate moral realist is right, then our moral theories are unfit to act as guides in such treacherous epistemological waters, since no one can justifiably have a high-level confidence that any particular theory is true. No moral philosopher can be justifiably confident enough in his theory to be able to use his theory to resolve difficult

moral questions about abnormal cases or cases of apparent conflict between general moral judgments. Stealing a Platonic analogy, say moral theories are like maps, with each map meant to show one the road to Larissa. Then every map comes with a disclaimer by the cartographer that he's not very confident the map is a reliable guide to Larissa. If moral philosophers can't be justifiably confident in their theories, then it doesn't look as if their theories can be of much practical use. If they are, it will have been purely accidental; someone will have gotten a lucky true belief, but he won't *know* that his belief is true, and regardless it would be irrational for him to act on his belief, given the undercutting defeater he has for it.[12]

One way the moral philosopher might attempt to save face is to claim that, even if we can't get theoretical moral knowledge, we can at least *rule out* certain moral theories.[13] Call this attempting to get at the correct moral theory by a *via negativa*. How useful this sort of moral theorizing will be is not immediately clear. It depends on just how many theories can be ruled out, and on what grounds. Even if the moral philosopher is able to narrow down the field to just two viable theories, the theories will conflict on certain important points and leave the philosopher unable to know which is correct (given the reasons already rehearsed). Moreover, it's clear that at least some moral philosophers don't take themselves to be doing moral philosophy in this purely negative fashion. Thus, if the moderate moral realist is right, then at least some moral philosophers must amend their current practices to fit this fact. Specifically, they must restrict their confidence in their preferred theories and withhold any bold judgments as to which theory is right.

---

[12] One possible pushback here, pointed out to me by Trevor Nyman, is to point out that one might need to get to Larissa for a very important reason, and taking a map that might be unreliable could be better than having no map whatever. This seems right to me. However, any moral philosopher who recommends his theory for help in being and doing good takes the risk of misleading the one to whom he recommends the theory. There is quite a risk involved, then, and it's not clear to me in what situations the risk will be worth taking.

[13] Thanks to Danny Simpson for this suggestion.

If moral philosophers can't be justifiably confident in their theories, then they can't justifiably recommend their theories as helpful in increasing proficiency in being and doing good, (and thus their work can't be helpful in this sense). That is to say, (5) is right. It looks as if on a moderate non-skeptical moral realism the work of moral philosophers (insofar as they articulate moral theories) is largely a waste of time. It can't provide anyone with an increase in practical direction. This is, in some ways, a better situation than the Nietzschian picture. It is a limited skepticism rather than a full-blown non-realism about morality. But it is not at all ideal for the moral realist, since he's likely to be inclined to think that moral philosophy is a worthwhile endeavor. On a moderate realism, moral philosophy might satisfy some intellectual curiosity, but it has no chance of accomplishing the roles the moral realist should set for it.

5. Conclusion

My goal has been to map out the dialectic between the Nietzchian and the moderate realist and to provide some (hopefully) compelling reasons to think that a moderate response to the problem from disagreement isn't enough for a satisfactory moral realism. If Wedgwood's moderate realism is right, then moral non-realism is no longer a threat, but moral philosophy appears to be radically unhelpful and unable to accomplish the goals appropriate to it. Anyone convinced by my argument can take it one of two ways. Either it suggests that the moral realist is unable to give a satisfactory response to this problem from disagreement, or it suggests that an immoderate moral realist explanation of said disagreement needs to be developed. It is an important question what form the immoderate non-skeptical moral realist explanation of moral disagreement amongst philosophers might take. This, I think, is an important project, one which would require a deal of work in and of itself. It would likely require the replacement or adaptation of Wedgwood's moral psychology and some fine-grained work on a theory of

reflective equilibrium. My hope, however, is that I've shown that a moderate realist response to this Nietzschian problem from moral disagreement is only an anemic defense of the legitimacy of the practice of moral philosophy. If moral philosophy is to be defended from such an attack, a bit of immoderation is required.

Aristotle on choosing virtuous action for its own sake

Yannig Luthra

1 Introduction

Aristotle claims that ethically virtuous action is to be chosen for its own sake.[1] But Aristotle also acknowledges that at least some ethically virtuous action is to be chosen as a means to further ends. He claims that political activity, which manifests practical virtue, aims at securing eudaimonia for oneself and for others.[2] And he claims that just actions aim at producing an equal distribution of goods,[3] and generous actions aim at benefitting others.[4] Indeed, the arc of the Nicomachean Ethics suggests that, for Aristotle, contemplation is choiceworthy for its own sake and not for the sake of further ends, whereas practically virtuous action is choiceworthy both for its own sake and for the sake of further ends.

Much ethically virtuous action really does seem to aim at further ends. Moreover, the choiceworthiness of many such actions seems to depend essentially on their reasonably perceived potential to serve further ends—like medical care or nourishment for recipients of charitable giving. It would be a mistake to give charity if it were not reasonable to think it had a chance of being useful.

It is natural to assume that an action is to be chosen for its own sake only when it is to be chosen independently of whether it serves further ends. Aristotle might seem to express a view along these lines in characterizing things choiceworthy for their own sakes as things that we pursue apart from further ends.[5] As a result, it seems that an action can be chosen as a means to further ends and for its own sake only if its choiceworthiness is overdetermined, like a meal that is delicious and nutritious. Given that the choiceworthiness of an action depends essentially on its serving further ends, it seems that, as a matter of definition, the action is not choiceworthy for its own sake. It is hard to see, then, how virtuous actions could be choiceworthy both for their own sakes, and as means to further ends.[6]

The interest of this difficulty goes beyond Aristotle interpretation. There

is something attractive-yet-elusive about the idea that many virtuous actions essentially serve further ends, but are also choiceworthy for their own sakes.

Usually this problem is approached by trying to show that virtuous action in fact does not depend for its choiceworthiness on its serving further valued ends. This approach is understandable, since it seems to be a matter of definition that actions choiceworthy for their own sakes are not to be chosen for the sake of further ends. In what follows, I criticize two such accounts of what Aristotle means when he says that virtuous actions are to be chosen for their own sakes, due to John Ackrill and Jennifer Whiting, respectively. Then, as an alternative, I propose an interpretation of what it means for an action to be choiceworthy for its own sake, according to which such an action can also depend for its choiceworthiness on it serving further valued ends.

I suggest that choosing an action for its own sake should be understood in a way that contrasts primarily with choosing an action as a necessary evil. A necessary evil, like undergoing painful treatment for an injury, is a misfortune for the agent. Some rightly chosen actions constitute misfortunes for the agent, whereas others are actions the agent is glad to have the chance to do. Such actions are appropriate sources of fulfillment, gratification, and pleasure for the agent. Such actions are rightly valued by the agent in a way that necessary evils are not. My proposal is that for an action to be choiceworthy for its own sake is for the action to be rightly valued in this way. I argue that, following Plato, Aristotle has this conception of an action's being choiceworthy for its own sake. This interpretation provides a way of understanding how virtuous actions can be choiceworthy both for their own sakes, and as means to further ends. An action can be an appropriate source of fulfillment or gratification for the agent, even if it is to be undertaken as a means to further ends.

2 Two attempts to resolve the difficulty

2.1 Ackrill's proposal

I want to begin by considering Akrill's remarks about choosing virtuous action for its own sake.[7] Ackrill's discussion of choosing action for its own sake is part of a discussion of the difference between *praxis* and *poiesis*. According to Aristotle, *praxis* is action that is chosen for its own sake, whereas *poiesis* (production, roughly) is not. It would seem that instances of *praxis* —like virtuous actions—

often are productive acts. Ackrill offers the example of a person who acts justly in mending a neighbor's fence. Ackrill's aims to explain how Aristotle's distinction applies to examples like that. Is mending the fence a case of *praxis* or *poiesis* ?[8]

Ackrill suggests that the action is a means to further ends under the description "mending the fence." And the action is done for its own sake under the description "acting justly." Ackrill says, "when it is asked whether the doer chose to do it for itself the question is of course whether he chose to do it because it was just, not whether he chose to do it because it was mending a neighbour's fence." According to Ackrill, the agent acts justly for its own sake in that she performs her just action because the action is just. More generally, to carry out a virtuous action for its own sake is to carry out that action because it is virtuous.[9]

Ackrill's proposal seems to be that an agent φs for its own sake when she would justify her φing by pointing out that her action is an instance of φing. Simlarly, an action is choiceworthy for its own sake when it is right to justify the action in this way. This proposal is initially plausible. If an action is to be chosen for the sake of further ends, it would seem that one would have to adduce those further ends as reasons for φing, instead of just pointing out that one's action is an instance of φing.

When an action is described as virtuous, one does not need to adduce further ends served by the action to establish that the action is choiceworthy. If we are given that an action is an instance of acting justly, courageously, or generously, that suffices to establish that the action is choiceworthy.[10]

However, this characteristic of virtuous actions does not mean that virtuous actions are choiceworthy for their own sakes. Every choiceworthy action falls under a description such that, given that it falls under that description, no further ends need to be adduced to justify the action.[11] The fact that an action falls under such a description does not reveal it to be choiceworthy for its own sake.

If an action is choiceworthy as a means to further ends, one can describe the action in a way that builds in its instrumental profile. For example, injecting insulin to help with diabetes can be described as maintaining one's health. Described that way, further ends do not have to be adduced to explain why the action is choiceworthy. But that is because the action's serving the valued ends

is already captured in the description of the action. A shortcut would be to describe an action undertaken as a means to further ends as a means to valued further ends. If an individual taking her medicine says "I am acting prudently" or "I am taking appropriate means to valued further ends," it would not make sense to ask her "what use is that?" But that does not mean the action is choiceworthy for its own sake. The usefulness of the action is already built into the description of the action. To tell whether an action of a certain type is choiceworthy for its own sake it is not enough just to consider whether describing the action as being of that type suffices to establish that it is choiceworthy. One must seek a full explanation of why actions of that type are choiceworthy.

It is true that if one describes an action as generous, or beneficent, or helpful, one does not need to adduce further ends to establish that the action is choiceworthy. But to tell whether such an action is choiceworthy for its own sake, one must seek a full explanation of why actions of that sort are choiceworthy. It seems that the choiceworthiness of such actions depends essentially on their serving further ends, like nutrition and medicine for people in need. The fact that describing an action as generous makes it unnecessary to mention further ends to justify the action provides no reason to doubt that. Indeed, describing an action as generous seems to build in that the action is worth doing as a means to further ends, in roughly the way that describing an action as prudent builds in that the action is worth doing as a means to further ends. If an action is not undertaken as a means to providing something of value to others, that would undermine its claim to counting as a generous action.

3 Whiting's proposal

In outline, Whiting's account is that choosing virtuous actions for their own sakes is choosing the actions "simply for being actions of a certain sort and insofar as each is just the action it is."[12] The core claim of her account is that, "Aristotle's notion of virtuous action is expansive, and so takes in, as it were, the external results at which it aims." For example, the effects of generous action, like benefits to others, are "included within my virtuous action itself."[13]

The ends of virtuous action are included within the action in that the realization of those ends "completes" the actions.[14] According to Aristotle, the activity of teaching is completed by learning in the student.[15] Whiting suggests

that, in a similar way, the improved well-being of a beneficiary of charitable giving completes that virtuous act. She suggests that, in this way, the improved well-being is part of the virtuous agent's activity of giving.

So, in aiming at ends like the well-being of others, virtuous agents aim at an aspect of the virtuous action itself. Virtuous agents value these sorts of ends for their own sakes, and not for the sake of something further. So their actions are not chosen for the sake of any end external to the action itself. Virtuous actions are chosen for the sake of ends that are contained within the actions themselves, and not for the sake of further ends beyond the action itself.[16]

3.1 Concerns about whiting's proposal

I want to quickly sketch three concerns about Whiting's proposal. First, it seems that an action can be choosen as a means to a further end, even if that end completes the action. For example, consider a reluctant teacher, who values her students' learning for its own sake, but wishes it was not her who had to teach them. Even if it is true that the students' learning completes her action of teaching, still she chooses to teach as a means to that further end, and not for its own sake. If there were some way for them to learn other than by her teaching, she would much prefer that they learn that way. If there is a sense in which the end of teaching is contained within the action, one must claim that choosing an action for the sake of an end contained within the action does not always amount to choosing the action for its own sake. The proposal that the "further" ends of virtuous actions complete those actions does not show that they are chosen for their own sakes.[17]

The second worry is that Aristotle specifically claims that the ends of some virtuous actions are different from the actions themselves. In NE 10.7, Aristotle argues for the superiority of contemplation over virtuous political activity partly on the grounds that political activity aims at further ends—namely eudaimonia for oneself and for fellow citizens. Aristotle specifies that these ends are different from political activity itself, and that we seek them as being different. This passage suggests that the ends of virtuous political activity are not included within the activity itself. Still, political activity is choiceworthy for its own sake.

The third worry is that Whiting's proposal cannot succeed in explaining how virtuous actions are to be chosen both as means to further ends and for their

own sakes. Whiting's proposal is that the ends of virtuous actions are included within the actions themselves, so choosing the virtuous actions for the sake of those ends is choosing the actions for their own sakes. But if that is right, then there are no ends beyond the actions themselves for the sake of which virtuous actions are to be chosen. So virtous actions are not to be chosen for the sake of further ends.[18]

4 An alternative proposal

It is natural think an action is choiceworthy for its own sake only if it is worth doing independently of its reasonably perceived potential to serve further ends. If one accepts this view, then either virtuous actions are worth doing independently of their potential to serve further ends, or they are not to be chosen for their own sakes. I want to suggest that Aristotle has a conception of an action's being choiceworthy for its own sake which does not require that such actions be choiceworthy independently of its potential to serve further ends.

My proposal is that, for Aristotle, an action is choiceworthy for its own sake if the action is rightly valued by the agent in a certain way. An activity's being valued in this way contrasts with an activity's being regarded as a necessary evil, like undergoing an amputation (or a necessary waste of time, like brushing one's teeth). There is a family of ideas which contrast with being a mere necessity. These include, for example, being fulfilling, gratifying, rewarding, enjoyable, satisfying, meaningful, and so on. An action is choiceworthy for its own sake when there is something good about the action which makes ideas in this family apply to the action.[19]

Before arguing that this view is Aristotle's, let me explain how it helps with the puzzle under discussion. An action can be rightly valued in this way even if it is to be chosen as a means of securing further ends. Serving at a soup kitchen can be gratifying, and a wonderful thing to have the chance to do. There is no conflict between its being gratifying, and its being choiceworthy because it is a means to helping people.

Consider Middlemarch's Dorothea, who, feeling her idle life empty and unfulfilling, wants the chance to spend her time doing useful things. Her lack of usefulness to others is not only a misfortune to the people who might benefit from her help, but a misfortune and legitimate source of unhappiness to

Dorothea herself. When she embraces opportunities to help others, her useful helping actions are a source of fulfilmment and meaning in her life. Helping others in the community has a value in Dorothea's life that mere necessities lack. Helping to found a hospital for the poor is rightly valued as meaningful and fulfilling, whereas undergoing an amputation is not.

One may well ask why it is that some means to ends are rightly valued as fulfilling, gratifying, and so on, while others are mere necessities. It may be relevant that some means to further ends involve an exercise of one's talents, or appreciation from others for one's efforts, or an atmosphere of fellowship and common purpose with others. But I have no general account of what makes some means to ends appropriately valued in these ways. The important point for present purposes is that they can be valued in these ways. An action that is to be chosen as a means to further ends can also be an appropriate source of fulfillment, enjoyment, meaning, and so on for the agent. I want to suggest that, for Aristotle, such actions are choiceworthy for their own sakes. If so, a virtuous action that is to be chosen as a means to further ends can also be choiceworthy for its own sake.

Some support for this interpretation is provided by considering what the con- trast is supposed to be between actions that are choiceworthy for their own sakes and actions that are not choiceworthy for their own sakes. Some indica- tion of how Aristotle might think of the contrast can be found in the way Plato characterizes things that are good merely as means, in contrast to things that are good for their own sakes, or things that are good both as means and for their own sakes. When Plato discusses these three kinds of goods in the Re- public,[20] he gives examples of activities that are good merely as means. They include physical training, medical treatment, and ways of making money. What these activities have in common is that they are "toilsome but beneficial." The characteristic of actions that contrasts with their being good for their own sakes is being toilsome. Roughly, the idea seems to be that these actions are not good for their own sakes because there is a kind of misfortune in having to do them in order to secure the benefits they promise.

Aristotle seems to think in a similar way about the contrast between activ- ities that are choiceworthy for their own sakes and activities that are not. This comes through in his discussion of "mixed actions," which are called for in bad

circumstances.[21] In NE 3.1 Aristotle discusses actions which are done because the agent fears a worse alternative, like throwing cargo overboard to save the lives of those aboard a ship. Aristotle claims that such actions are chosen, but that no one would choose such actions for their own sakes. These actions are recognized by the agent as the right thing to do in the circumstances, but they are nonetheless not chosen for their own sakes. It is natural to suppose that such actions are not to be chosen for their own sakes because it is a misfortune for the agent to have to carry out those actions.[22]

Aristotle's view that the productive activity of craftsmen is ignoble, and to be avoided, also illustrates his contrast between actions choiceworthy for their own sakes, and actions that are a kind of misfortune for the agent to have to do. Aristotle describes the life of craftsmen and merchants as "ignoble," and contrary to virtue."[23] The activity of the productive classes is a mere necessity, and to be avoided if possible. Productive craft activities are not to be chosen for their own sake. Those productive activities are not mere necessities just in virtue of their being means to further ends. They are mere necessity, and not choiceworthy for their own sakes, because (according to Aristotle) those activities are servile and undignified. It seems likely that for Aristotle these productive activities are not to be chosen for their own sakes because having to do those activities constitutes a kind of misfortune for the agent.[24]

In seeking what Aristotle thinks it is for an action to be choiceworthy for its own sake, we should look for a characteristic that contrasts with an action's being a misfortune. An action's being a legitimate source of fulfillment or gratification, contrasts in a natural way with an action's being a misfortune. Aristotle's discussion of the relation between virtuous action and pleasure suggests that an action's being choiceworthy for its own sake contrasts with being a misfortune in this sort of way.

Aristotle claims that virtuous activity is "objectively" pleasant. This characteristic of virtuous action contrasts with the regrettable character of mixed actions. Aristotle claims that virtuous people rightly find virtuous actions pleasant. Virtuous agents experience virtuous actions as pleasant, and there is something about the actions themselves—wherein the actions are objectively pleasant—that makes it correct to experience them as pleasant.[25]

Many virtuous actions do in fact seem to be objectively pleasant in the way

Aristotle suggests. Helping a friend, giving charity, or serving one's community as a politician, seem to be legitimate sources of a certain kind of pleasure. These are activities the agent could rightly be glad to do. Concepts like fulfillment and gratification are useful for characterizing the way in which actions like these are legitimate sources of a serious kind of pleasure. Virtuous actions like these seem to be proper objects of pleasure in that they are proper objects of satisfaction, gratification, fulfillment, or other forms of appreciation along these lines.

Aristotle's claim that virtuous actions are pleasant suggests that virtuous action has a dimension of goodness which mixed actions lack, even though they are rightly chosen as appropriate means to worthwhile ends. Whereas it is a misfortune to the agent to have to perform a "mixed" action like throwing goods overboard, virtuous actions are rightly found pleasant by their agents. So, for Aristotle, there seems to be something good about virtuous actions that goes beyond their being appropriate means to worthwhile further ends.[26] Inasmuch as virtuous actions have a kind of goodness that goes beyond their being appropriate means to further ends, there seems to be a sense in which they are good for their own sakes. This kind of goodness is marked by the correctness of finding virtuous activity pleasant, meaningful, or fulfilling.

Now, one might doubt Aristotle's view that it is always correct to take any kind of pleasure in virtuous actions. It makes sense to find pleasure or gratification in helping a friend, because that's a wonderful thing to have the chance to do. In contrast, it does not make sense to find pleasure in turning in one's child to the police, because that is a horrible thing to have to do, even if it is the right thing to do. Turning in one's child might be a virtuous, just action even though it is a necessary evil, and not something one should take pleasure in doing.

But it is indeed Aristotle's view is that virtuous actions are rightly found pleasant. To address this worry, he could claim that turning in one's own child is not in fact a case of virtuous action. A life full of actions of this kind would be like Priam's. While possessing a certain dignity, it would not be a eudaimon life. If so, such a life would not be a life of practical virtue, since such a life is eudaimon, at least to a secondary degree. Alternatively, Aristotle could claim that it is correct to take a certain kind of pleasure even in tragic virtuous actions—perhaps a kind of pride in doing the right thing in difficult circumstances, akin to the gratification that an ideal warrior could find

in sacrificing her life for a good cause.

If indeed virtuous actions are rightly found pleasant, meaningful, or fulfilling—as Aristotle claims they are—then they have a dimension of goodness beyond their being instruments for achieving further ends. Virtuous actions are choice-worthy for their own sakes in that they are rightly valued as wonderful, fulfilling, gratifying, and so on. Aristotle's claim that virtuous actions are choiceworthy for their own sakes may be interpreted to mean that virtuous actions are rightly valued in this way. When an action is valued in this way, it is valued as having a dimension of goodness that goes beyond its usefulness as a means to further ends.

An action can be rightly valued as fulfilling or meaningful even if it is to be chosen as a means to further ends. One can rightly find fulfillment in giving charity even if the charitable act is aimed at benefits for others. Indeed, it makes sense to find fulfillment in giving charity largely because doing so serves worthwhile further ends. Thus, virtuous actions can be choiceworthy both for their own sakes and as means to further ends.

Notes

[1] NE 1105a35, 1140b7, 1144b20

[2] Pol 1103b5, NE 1094b5, 1099b38, 1177b3

[3] NE 1132b31-1133a2

[4] NE 1120a21

[5] NE 1096b8

[6] There is another puzzle about Aristotle's claim that we choose virtuous action for its own sake, besides the one at issue in this paper. Aristotle claims that virtuous action is chosen both for its own sake, and for the sake of eudaimonia (NE 1097b2-4). Kraut plausibly suggests that these claims are not incompatible because virtuous action may be chosen as constitutive of eudaimonia, rather than being chosen as an instrumental means to eudaimonia. Kraut's suggestion leaves unaddressed the present issue: how can virtuous action be choiceworthy for its own sake if its choiceworthiness depends on its being an instrumental means to further ends, like health and nutrition for people in need?

[7] The remarks are from "Aristotle on Action," Mind 87 (1978) pp. 595-601.

[8] The proposal in this paper about how to understand what it is to choose virtuous action for its own sake can be adapted to shed light on the difference between praxis and poiesis. In my view, praxis is action that is valued in the way sketched above, and poiesis is productive action that is not valued in that way. I do not develop this suggestion here.

[9] David Charles's view is similar. See "Aristotle: Ontology and Moral Reasoning," Oxford Studies in Ancient Philosophy 4 (1986) pp. 119-144. To act generously for its own sake means

that one carries out the generous action because it is generous. Charles claims that whatever it is specifically that one does so as to act generously—making a charitable donation, say—is done as a means to further ends in that it is done as a means to acting generously.

[10] This claim may need qualification. Perhaps an action that is generous could fail to be choiceworthy because it is unjust. On the other hand, maybe a strong version of the thesis of the unity of virtue is true. These complexities do not affect the main line of argument here.

[11] There is a sense in which any choiceworthy action is good to do apart from whether it brings about further ends. Whether one acts well in giving charity does not depend on what effects the act actually turns out to have. I assume that the goodness of virtuous actions is not consequentialist in this sense. But the resilience of the choiceworthiness of virtuous action to vicissitudes of cause-and-effect does not show that virtuous actions are to be chosen according to themselves. An action rightly undertaken just as a means to further ends remains well-chosen and good to have done, even if its expected effects do not materialize.

[12] Whiting, Op. Cit. p. 276

[13] Ibid pp. 290-291

[14] Ibid p. 288

[15] Physics 202b5-22

[16] Note that this proposal tries to explain how virtuous action is choiceworthy for its own sake by making claims about the metaphysics of virtuous action—where an action ends and its results begin. The proposal I suggest below lets the metaphysics of virtuous actions turn out as they may.

[17] One might also worry that the ends of virtuous actions do not complete those actions in the way that learning completes an act of teaching. One may have carried out an act of charitable giving without having improved things for anyone, if, say, the funds were misused. In contrast, according to Aristotle, one cannot have carried out an act of teaching unless the student has learned. So one might doubt that the realization of the ends of virtuous action are included within the actions themselves. (Note that this point is compatible with the thought that it is part of the nature of charitable actions, so described, that they be undertaken for the sake of certain further ends.)

[18] A possible reply here is to bite the bullet, and to insist that Aristotle's real view is that ethically virtuous action is not chosen for the sake of further ends. (Compare NE 1176b6.)

[19] This proposal bears some similarity to Korsgaard's "Two distinctions in Goodness" (in The Philosophical Review 92 (1983) pp. 169-195). Korsgaard distinguishes between intrinsic goods and final goods. Final goods are good ends that are not to be pursued as means to further ends, and intrinsic goods are good ends that do not get their goodness from an external source. Korsgaard argues that an end may be a final good without being intrinsically good. (In particular, she argues that an end can be a final good partly because of the interest someone takes in that end.) The proposal here is a "distinction in goodness" between means to ends that are merely instrumentally good, and means to ends that are valued in their own right.

[20] Republic 2.357c

[21] It is also worth noting Aristotle contrasts pleasures that are to be chosen for their own sakes with necessary pleasures (NE 7.4 and 7.7). This contrast suggests that, to understand how virtues actions can be choiceworthy for their own sakes we should try to

understand how an action can be choiceworthy as a means to further ends without being a mere necessity.

[22] Aristotle's view that the productive activity of craftsmen is ignoble, and to be avoided, also illustrates Aristotle's contrast between actions choiceworthy for their own sake, and actions that are a kind of misfortune for the agent to have to do. (See Pol. 1328b39. Also compare NE 1329a1-5.) I do not try to develop this point here.

[23] Pol. 1328b39-1329a2. Also compare NE 1329a1-5.

[24] Unless, as Aristotle chillingly suggests, the defects in the agent's nature match the defects in the activity.

[25] See 1.8.1099a3-15. Also compare 3.1 on the relation of pleasure to virtue, and discussion of encratic agents in book 7. 2.1.3

[26] Aristotle's view that virtuous actions are kalon provides another way into seeing how the value of virtuous actions goes beyond their being means to further ends. Viewing a virtuous action is fine, or noble, is a way of valuing the action as something more than a mere necessity, which is compatible with choosing the action as a means to further ends. I do not

The Salience of Moral Character

1. Introduction: Moral Rules

Rules are prominent in moral cognition. The Ten Commandments and other prescriptions drawn from canonized religious texts are treated by many as paradigms of moral content, and guiding oneself by this sort of prescription is treated by many as a paradigm of moral judgment. Many people also think of themselves as following a partly self-authored code of conduct, either in addition to or in place of religious and other social rules. These personal codes are sometimes less explicit than social prescriptions, but they also often take the form of rules, as in "do not take supplies home from the office" or "hold the door open for someone walking in behind".

Since rules are commonplace in ordinary moral thinking, a moral theory must illuminate the proper place of rules in moral thought. On one approach, commonly known as "deontology", morality fundamentally concerns the formulation and observance of rules. Deontology is often presented as a representative theoretical position in introductory courses in moral theory and in the stage-setting material of scholarly articles.

There are at least two distinct approaches to moral theory commonly grouped under this heading. One is known as "intuitionist deontology", and is best exemplified in the work of W. D. Ross.[1] On this view there are several moral rules – Ross calls them "prima facie duties" – which generate obligations in context. A prima facie duty of fidelity can generate an obligation to show up on time for a meeting, for example, and a prima facie duty of beneficence can generate an obligation to help an elderly person who has fallen on the sidewalk. These duties are not always dispositive in moral judgment, for they can be defeated, as when a prima facie duty to aid an elderly person in distress defeats a prima facie duty to show up on time for a meeting. Although they can fail to generate all things considered obligations in context, these prima facie duties are nonetheless well understood as duties, since in normal circumstances they manifest as all things considered obligations.[2]

Ross's is perhaps the paradigm form of deontology, and it is intuitionistic in at least two respects. The first is that in his view our access to the content of our duties is quasi-perceptual, the exercise of a putative cognitive faculty of moral intuition.[3] The second is that these intuitively grasped elements of our moral understanding resist more systematic explanation in terms of other values, as is attempted in the theories of Immanuel Kant and the utilitarians. This latter feature of Ross's deontology exposes a limitation, in his view, of what moral theory can accomplish. Since there is no more general value or more systematic moral understanding which gives rise to the prima facie duties, we cannot inform judgment about cases of apparently conflicting prima facie duties by reference to any such value or system. The relative strength of our prima facie duties is

---

[1] David Ross, *The Right and the Good*, ed. Philip Stratton-Lake (Oxford, 1930).

[2] I make no distinction between duties and obligations. Through much of the twentieth century obligations were understood as a subcategory of duties, namely those which arise from social roles or relationships; see especially John Rawls, *A Theory of Justice* (Harvard, 1971), 108-117.

[3] Ross *The Right and the Good*, 29-31.

discerned through contextualized judgment, and there is not much a theorist can say about how to best exercise this judgment.

The other approach to moral theory often labeled deontological is an interpretation of Kant's writings that was popular throughout the twentieth century. On this interpretation Kant's Categorical Imperative generates more specific rules of conduct that people should use to regulate action. This interpretation leans heavily on Kant's illustrations of the Categorical Imperative in the *Groundwork of the Metaphysics of Morals*.[4] Kant there uses the Categorical Imperative in an effort to explain the wrongfulness of making false promises and of cheating customers, and in so doing he does not appear to trade on contextually local features. Thus if these explanations are successful, the upshot appears to be not only that the actions under consideration are wrong, but more generally that any other action of the relevant type is also wrong. Hence these arguments are taken to purport to show that rules like "do not make a promise you do not intend to fulfill" and "do not give a customer incorrect change for the sake of higher profits" have the force of all things considered moral obligation. The task of fleshing out the implications of the Categorical Imperative, on this interpretation, is to produce more rules of this kind until we have a set of rules rich enough to navigate all domains of human life.

2. Rules of Moral Salience

This understanding of Kant's moral theory came under criticism in the 1980s and 90s.[5] Two significant difficulties with it are whether all the rules Kant purportedly derives from the Categorical Imperative genuinely have the force of all things considered moral obligation and what further rules, beyond those which emerge immediately from Kant's own illustrations, could plausibly be derived from the Categorical Imperative in the appropriate way. These difficulties operate in tandem. The more the normative force of a rule is relaxed by allowing for defeaters or exceptions, the more plausibly that rule might be derived from the Categorical Imperative; but to the extent that a theorist pursues this strategy to more completely cover the domain of morality, the more that theorist forfeits the advantage of a more systematic theory.

No one was more central to the effort to re-understand this dimension of Kantian moral theory than Barbara Herman. She rejects an understanding of Kant's Categorical Imperative as generating moral rules with the normative force of obligation, proposing instead a conception of moral rules as *rules of moral salience*.[6] These rules are not directly action-guiding. Their role in moral judgment is instead to occasion deliberation, to prompt explicit thought about an action's permissibility. On her view the bulk of our practical life is routinized, not the product of explicit deliberation. We were trained as children into patterns of sensitivity to certain sorts of reasons, such as reasons not to invade other people's bodies and reasons not to use for our own purposes objects that belong to other people. To varying degrees we train ourselves as adults into further sensitivities, as we become involved in a more variegated social environment and we learn more

---

[4] Immanuel Kant, *Groundwork of the Metaphysics of Morals*, ed. Mary Gregor (Cambridge, 1997). See also Kant's *Metaphysics of Morals*, ed. Mary Gregor (Cambridge, 1996), and his *Lectures on Ethics* (Harper, 1963).

[5] Barbara Herman's collection *The Practice of Moral Judgment* (Harvard, 1993) warrants mention in particular; other important texts in this connection include Marcia Baron's *Kantian Ethics Almost Without Apology* (Cornell, 1995) and Nancy Sherman's *Making a Necessity of Virtue* (Cambridge, 1997).

[6] Herman, *The Practice of Moral Judgment*, 73-93.

about the peculiar moral dangers of our historical circumstances. Someone raised in a small town may move to a city, for example, and so come to acquire sensitivities necessary to interact with people with different background assumptions about conduct. Or someone may learn of implicit sexist or racist biases prevalent in her historical moment, and so better develop her ability to detect these biases in herself and better develop sensitivities needed to interact with those who sometimes exhibit them unknowingly.

Most of the time these trained sensitivities run on a kind of autopilot. We instantiate routines, like refraining from cutting in line at the grocery store and expressing gratitude to the clerk who scans our groceries, without thinking about them much. But sometimes circumstances are atypical, and we find ourselves in a context where routine action may be inappropriate. The function of rules of moral salience is to alert us to these contexts, and normally also to occasion explicit moral deliberation. Their purpose is to prompt us to switch off the autopilot, that is, and to assume the controls.

An illustration will perhaps help clarify this idea. Normally while driving we routinely respond to certain classes of reasons by keeping adequate distance from the car in front, signaling changes of lane, refraining from passing suddenly, and so forth. But if a passenger in the car is suffering a heart attack, we recognize this as morally salient, which prompts explicit judgment about whether and how to adjust our routine reasons-responsiveness. The driver's rule of moral salience picks up on the heart attack symptoms as cause for explicit deliberation, and the ensuing deliberation may result in practical conclusions at variance with routine, such as driving faster, following closer, and passing more suddenly than usual.

Note that the role of the rule of moral salience in practical judgment is not to sort actions right from wrong. The rule alerts the driver to an unusual possibility, namely justifiably driving in a mode less safe than normal. But the rule does not itself deliver the judgment that violating the routines is justified; that depends on other features of the environment. The rule itself simply calls attention to a morally significant fact, one which could potentially justify non-routine action. As Herman deploys the idea, after a rule of moral salience triggers explicit judgment, the formulas of the Categorical Imperative may then be introduced to help guide action.[7]

3. Maxims of Action

To complete the sketch of the role of rules of moral salience in Herman's view, we must consider the deliberation prompted by morally salient considerations. Herman follows Kant in claiming this deliberation concerns actions under rationalized descriptions, Kant's "maxims of action".[8] On a helpful rough-and-ready account, maxims have the form:[9]

I will perform act *A* in circumstances *C* for reasons *R*.

An example of a maxim of action is thus:

---

[7] Herman *The Practice of Moral Judgment*, 147-158.

[8] Kant, *Groundwork of the Metaphysics of Morals*, 13f. For Herman's account of maxims of action, see *The Practice of Moral Judgment*, 132-183.

[9] For a classic account of Kantian maxims, see Onora O'Neill's *Acting on Principle* (Columbia, 1975); see also her *Constructions of Reason* (Cambridge, 1989).

> I will pass suddenly when there is no immediate danger in doing so for the reason that my passenger urgently needs medical attention.

As Herman observes, Kant directs attention to this sort of action-description because it is the form appropriate for moral assessment. This distinguishes his view from a flatfooted version of Ten Commandments morality where moral rules pertain to acts only, and not to circumstances or to justifying reasons. On that naïve position, surely not the best interpretation of the commandments as part of any actual social practice, all we need for moral assessment is a description of what is done, not any description of how and why. That a passenger in the car is suffering a heart attack, on this view, is neither here nor there with respect to the permissibility of passing suddenly.

This naïve view appears obviously mistaken to anyone with moral understanding. The heart attack is clearly relevant to the permissibility of the act in question, and we must describe acts in a way that captures all their morally relevant features. We can thus introduce a stipulated distinction between "acts" understood narrowly – such as killing, stealing, passing suddenly, and so forth – and "actions", which include in their description the circumstances in which the act is performed and the reasons for which it is performed.[10] We are then in a position to formulate the claim that only actions, not acts as such, are of the proper form for moral assessment.

In ordinary language we sometimes predicate permissibility or impermissibility of acts in the narrow sense. But this does not undermine Kant's insight that maxims are the locus of moral assessment, for when we assess acts morally we implicitly fill in typical circumstances of action and typical reasons for which the act is performed. Thus someone who claims passing suddenly is wrong has in mind something like: passing suddenly in normal traffic to get to one's destination more quickly is wrong. Absent implicit appeals to these further features, there is no fact of the matter about whether passing suddenly is wrong; it is an act, not an action, and so (as such) is not permissibility-apt.

To summarize: Herman's analysis of moral judgment has three main components: routine judgment, rules of moral salience to occasion explicit judgment, and the Categorical Imperative for informing non-routine judgment. When Kant's moral theory is understood this way, it is more distant from deontology than is standardly believed. It is utterly different from naïve deontology in insisting on maxims rather than acts as the locus of assessment. It is also importantly different from Ross's deontology in its account of moral rules. Moral rules do not articulate prima facie duties whose force we intuitively apprehend but whose force can be defeated by other prima facie duties whose force we intuitively apprehend. Rather, most moral judgment proceeds by means of sensitivities trained into the routines of ordinary life, without reference to duties or rules at all; the rules needed to supplement routine judgment are calls to explicit deliberation and judgment, not apprehension of considerations with the force of obligation unless opposed by considerations of comparable force. When explicit moral judgment is called for, moreover, it can be informed by an overarching value like humanity or by a systematic understanding of morality like the Categorical Imperative.

---

[10] This distinction achieved currency among Kantian theorists decades ago; for an account of it in print see Christine M. Korsgaard, *Self-Constitution* (Oxford, 2009).

The first two of these differences, at least, move our understanding of Kantian theory not only away from deontology but toward virtue theory.[11] Virtue theories also standardly understand most practical judgment as deployment of trained routines of reasons-sensitivity; and while the virtue tradition has not explicitly formulated the idea of rules of moral salience, the idea resonates with its themes. Virtue consists not only in stable patterns of reasons-responsiveness but also in sensitivity to how atypical circumstances give rise to atypical appropriate actions. These affinities between Herman's Kantian moral theory and virtue theory explain another claim these views share: that only someone with appropriate training, who makes routine judgments well and tends to be sensitive to morally relevant features of the environment, is expected to exercise judgment well in context. Most obviously this is true when routine judgment itself generates non-virtuous action, either directly or through the omission of necessary deliberation. But Herman also claims that when a poorly trained person appropriately deploys a rule of moral salience to prompt moral deliberation, the ensuing deliberation is apt to be conducted poorly. Trained routine action does the lion's share of the work of moral judgment, and it is not expected that a person lacking these dispositions often compensates well by guiding action with the Categorical Imperative in explicit deliberation. If these claims are roughly correct, then Kantian moral theory has more in common with virtue theories than with paradigmatically deontological ones. Accordingly any taxonomy that groups Kantian theory with deontology on one side, and virtue theory with consequentialism on another, is misleading; and in developing a Kantian moral theory we should, in Herman's apt phrase, leave deontology behind.[12]

4. Moral Worth

Thus far I have reviewed the case for moving away from an understanding of Kantian moral theory that shares structure with intuitionistic deontology and in the direction of one that shares structure with virtue theory. I turn now to suggest that this trajectory of interpretation be continued further, toward a more thoroughly virtue-oriented understanding of morality. This can be accomplished within a theory which captures vital motivations behind Kant's approach even as it may differ from Kant's own view.

To articulate these suggestions I turn to another important idea in Kant's theory, namely that of *moral worth*. Moral worth is a property actions have, in Kant's view, if they are done from duty or (equivalently) with a good will. This is a more stringent condition than permissibility; it demands that an act be performed from a specifically moral motive. To use a familiar illustration, consider a shopkeeper who acts on the maxim:

> I will give customers correct change when they make purchases in my shop since that conduces to the long-term success of my business.

This maxim is permissible, but it fails as such to have moral worth, for it exhibits no commitment to moral self-regulation. If circumstances change – some competitors go out of business, say, or a new form of transaction makes it more difficult for customers to discern when they have received

[11] In the *Groundwork* Kant does not develop a theory of virtue, but he does so in his *Metaphysics of Morals* and *Lectures on Ethics*.
[12] Herman, *The Practice of Moral Judgment*, 208-240.

correct change – this maxim may become inert, since giving correct change may not conduce to business success. If this is the only maxim pertaining to the giving of change that the shopkeeper observes, then under those differing economic conditions the shopkeeper will not give correct change. By contrast if the shopkeeper's maxim in giving correct change is:

> I will give customers correct change when they make purchases in my shop out of respect for their status as fellow persons.

This maxim of action is plausibly understood to have moral worth. And if so, on a Kantian theory this means that an action genuinely performed under this description is done with a will which is good without qualification.[13]

I want to suggest that consideration of puzzle cases for a Kantian account of moral worth points in the direction of a Kantian theory even more distant from intuitionistic deontology and even more in line with virtue theory. The first puzzle case, that of Huckleberry Finn, has become a stock example in recent literature on moral psychology.[14] The protagonist of Mark Twain's *Adventures of Huckleberry Finn* is prominent in these discussions because his self-understanding and his patterns of reasons-responsiveness are at odds with each other. Huck Finn undertakes considerable risks to help his friend Jim, an escaped slave in antebellum Missouri, head north toward relative safety. All the while Huck believes he is doing wrong, but feels compelled to help Jim despite his pangs of conscience. Jim's escape attempt fails, but in the end he gains freedom through the will of his recently deceased owner, Miss Watson.

Huck Finn is a puzzle case for an account of moral worth because of his skewed moral understanding. His racist views entail that his routine heuristics of judgment are not morally adequate. Nevertheless his rules of moral salience in the case in question get him onto the most morally relevant features of his decision: concern for Jim's liberty and well-being, on one hand, and concern for Miss Watson's property, on another. But when he explicitly deliberates about how to respond to those features in action, he comes down clearly on the side of the latter, even though in action he is moved by the former. He does the right thing (helping Jim in his effort to escape) for the right reason (out of concern for Jim), but not under the guise of the right (since his conscience has it that what he does is neither the right thing nor done for the right reasons). Huck himself appears to lack moral worth, if we understand this property of *persons* to entail full moral virtue, for his moral understanding is skewed. But this does not settle the question of the moral worth of his *action* in this case.

I will approach this difficult question by noting there are moral categories intermediate between acting with full moral virtue and acting permissibly. Perhaps the most familiar of these intermediate categories is that of praiseworthy action. Strongly overlapping with this category is that of actions which help train people in the direction of virtue. Let us call an action "laudable" if it falls into such an intermediate category. This enables us to ask: does the laudability of an action

---

[13] There is a complication here, since Kant insists that the motive of duty as such involves respect for the moral law rather than respect for the moral status of other individuals. See Kant (1785), 13-14. This does not undermine the claims in the text, however, and in any case I would recommend abandoning this feature of Kant's view.

[14] The use of Huckleberry Finn as an example was inaugurated by Jonathan Bennett's "The Conscience of Huckleberry Finn", *Philosophy* 49. For another excellent discussion see Nomy Arpaly, *Unprincipled Virtue* (Oxford, 2003).

suffice for its moral worth? In the context of a Kantian theory this becomes: does the laudability of an action suffice to entail it is done with an unqualifiedly good will?

Framing the question in this way does not immediately settle whether Huck's action has moral worth, in part because the notion of unqualified value is not familiar from ordinary life. But it seems to me that this framing pushes against the claim that his act has moral worth. For though he does the right thing for the right reasons, both his belief that he acts wrongly and his reluctance to act as he does appear to be qualifications of the value of the will he exhibits in his action.

Nomy Arpaly articulates a case we can use to push the point further.[15] She distinguishes a "diehard" philanthropist, a "fair-weather" philanthropist, and a "capricious" philanthropist. Each of these individuals commits significant resources to help others in need, and each acts in the belief that what she does is obligatory. The difference is that the first would do this even if it were highly burdensome to do so, perhaps because of a personal crisis demanding resources or because of a descent into depression sapping her compassion for those whom she helps. The second would not help in the event of a major life crisis like these, but otherwise has a stable disposition to help. The third treats doing the obligatory thing as a lark, and has no stable commitment to fulfilling her obligations in general.

The capricious philanthropist responds to need appropriately in this case, but the transient nature of her disposition to do so calls into question whether it is really the need she responds to, as opposed to something which happens to coincide with the need in this case. It is thus unclear whether the capricious philanthropist even does the right thing for the right reasons. In the case of the fair-weather philanthropist, by contrast, there is no need to doubt she does the right thing for the right reasons. Fair-weather philanthropy is laudable: it is an action people are appropriately praised for performing, and it is an action that an uncharitable person – a person incapable, in the short term, of diehard philanthropy – might perform to train herself into better responsiveness to need. But the fair-weather philanthropist does not act with moral worth in the Kantian sense of acting with an unqualifiedly good will. If the act were genuinely motivated by the specifically moral motives of concern and respect for those in need, it would be more stable. Laudability thus should be distinguished from Kantian moral worth as a property of actions.

Once we mark this distinction, it becomes even more dubious to attribute moral worth to Huck Finn's action. The point is not that he is a fair-weather friend; on the contrary, he is in some respects a diehard friend, one who helps appropriately even at considerable risk to himself. The point is rather that our characterization of his act as praiseworthy or as putting him on the road toward virtue does not suffice for its moral worth. While it is unclear whether Huck's false beliefs about morality are compatible with morally worthy action, I hope it is clear that his compromised routine judgments are incompatible with the stability of motive across time and circumstances necessary for his action to have moral worth.[16]

If this is the correct way to think about Huck's case, it calls into question whether moral worth is a property that it even makes sense to attribute to actions as such. It could instead be a property actions have only derivatively, when they are the actions of a morally worthy – a fully

---

[15] Arpaly, *Unprincipled Virtue*, 87-93.

[16] If we stipulate the necessary stability of motive, we must be careful to note the manifold adjustments to Huck's dispositions this entails. It may be possible to stipulate him into full moral virtue without thereby eliminating all his false beliefs about morality, but it is difficult to stipulate him into full moral virtue while leaving constant his thoroughly racist moral beliefs.

morally virtuous – person. For unlike the space between full moral virtue and mere permissibility, which needs to filled with a category like laudability, it is unclear there is a significant category intermediate between laudability and moral virtue.

5. Moral Character

Suppose the observations from the preceding section are correct, and moral worth is a property actions have only derivatively, insofar as they are the actions of a fully morally virtuous person. On a Kantian theory, the moral worth of an action tracks its being done with a good will; the resultant position is thus that having a good will tracks being fully morally virtuous. On this position, the presence or absence of a good will is not a temporally local feature of actions, but emerges only from stable patterns of action in a range of circumstances over time.[17] These, then, are Kantian analogs to the Aristotelian claims that virtue consists in stable character traits and that virtuous action consists in the sort of action a virtuous person performs. Laudable action, on this understanding of Kantian theory, corresponds to the Aristotelian category of as-if-virtuous action; this is doing the right thing for the right reasons for the purpose of acquiring the stable character in which moral virtue consists.[18]

Note that the claim here is not that the best formulation of Kantian moral theory is the same as Aristotle's theory. The conception of moral judgment I am sketching is not eudaimonist; it does not assert that human flourishing consists mainly in, or necessarily coincides with, moral virtue. Nor is it all-encompassing of practical judgment. It purports to account for the specifically moral phenomena of obligation and moral worth, not to encompass all value-responsiveness. Nor does it assert that a virtuous person must take pleasure in doing the right thing or that a virtuous person must do the right thing with ease.

Notwithstanding these caveats, this sketch of a Kantian virtue theory is not Herman's view.[19] Despite her virtue-friendly emphases on routine judgment and rules of moral salience, her account leaves open the possibility that in those cases where explicit judgment occurs, a good will can manifest regardless of a person's broader patterns of reasons-responsiveness. Against this I want to suggest that the more thoroughly virtue-emphasizing understanding of Kantian theory is to be preferred, and hence that we should continue further in pursuing the program that Herman and others instigated.

The principal concern motivating this extension of the project is the empirical adequacy of a view that permits temporally local attributions of good willing in light of the large and ever-growing psychological literature about the non-self-transparency of motives.[20] I do not suggest Herman or other Kantians are unaware of this literature, or that their view that a good will can be a temporally local phenomenon is refuted by it. But insofar as their view characterizes people as always having considerable control, in a temporally local way, over their maxims of action, this

---

[17] Related views are defended as interpretations of Kant in Samuel Kerstein, *Kant's Search for the Supreme Principle of Morality* (Cambridge, 2002), and in Richard Dean, *The Value of Humanity in Kant's Moral Theory* (Oxford, 2006).

[18] Aristotle, *Nicomachean Ethics*, ed. Roger Crisp (Cambridge, 2000), 23-36.

[19] The affinity Herman sees between Kant's theory and virtue theory increases over time; see her collection *Moral Literacy* (Harvard, 2007). But the view in the text goes further than she would countenance.

[20] Many of these empirical results are summarized in Daniel Kahneman's *Thinking, Fast and Slow* (Farrar, Straus, and Giroux, 2011).

psychological literature calls it into question. This is necessary, if the Categorical Imperative or another systematic account of morality is to significantly direct moral judgment, even if explicit deliberation is infrequent and is needed only when prompted by rules of moral salience.

Against this, the efficacy of human self-regulation through explicit deliberation may be constrained to cases where we can clearly identify morally problematic maxims and refrain from acting on them for the reason that they are problematic. We will often be unable in the moment to detect our morally problematic maxims, and we will have in general an extremely limited ability to consciously choose a maxim on which to act. This typically makes it impossible to achieve moral worth in the moment, through occurrent thought about the morally relevant features of a deliberative context. But this is not a problem if we contend that moral worth emerges only from stable dispositions of reasons-responsiveness; for on that position it is expected that we normally cannot achieve moral worth in a temporally local way.[21]

6. Conflicting Rules

It is appropriate at this juncture to revisit the apparatus of rules of moral salience in light of these observations about moral worth and moral character. Recall that Herman doubts explicit judgment guided by the Categorical Imperative is the centerpiece of the best moral psychology; against this she conceives of moral judgment as consisting mainly in routine judgment and rules of moral salience which prompt explicit deliberation, and only atypically as involving explicit judgment guided by the Categorical Imperative. As I indicated I see this as movement in the right direction, and I have begun to sketch a position yet further in the same direction. On the view I propose, explicit judgment guided by the Categorical Imperative or some other moral system is even less significant. Such explicit judgment may be possible, sometimes even required. But we should not expect it to be very effective, and even perfect instantiations of it fail as such to suffice for morally worthy action, since moral worth is a property which emerges only from diachronic features of agency.

This position accordingly places even greater emphasis on routine judgment and rules of moral salience. But these features alone cannot constitute a complete account of moral judgment, for no humanly realizable routine heuristics are adequate to the entire range of contexts of human judgment. This fact is less important than it seems, however, for even in practice sorting actions right from wrong is a less important part of moral life than is standardly believed. Lest this claim be misunderstood, I want to emphasize the importance of reliably and routinely sorting right from wrong and the importance of making a lifelong project of improving one's heuristics for routine judgment. I want to emphasize also the importance of developing moral sensitivities in the form of rules of moral salience signaling a need to suspend routine and the importance of attempting in good faith to do the right thing in atypical circumstances or when morally salient considerations appear to conflict. Notwithstanding these appropriate emphases on permissibility, conscientious

[21] It may seem odd to invoke affinity with virtue theory as a response to a charge of empirical adequacy, in view of the prominence in the moral psychology literature of the so-called "situationist" objection to virtue ethics. Although I cannot give the issue adequate attention here, my own view is that this objection is very weak and rests on a fairly gross mischaracterization of what virtue theorists standardly assert. For examples of the charge that virtue theory is empirical inadequate see Gilbert Harman, "Moral Philosophy Meets Social Psychology", *Proceedings of the Aristotelian Society* (1998-1999), and especially John Doris, *Lack of Character* (Cambridge, 2002).

sensitivity to morally salient considerations is often more important than permissibility, not only for assessing a moral theory but also for the practice of moral judgment.[22]

One implication of this claim is that often the appropriate focus of moral interest is a person's seriousness and sensitivity over time rather than the permissibility of what he has done. Consider G. E. M. Anscombe's famous criticism of Oxford University for granting an honorary degree to the man who authorized use of nuclear weapons on Japanese cities.[23] Let us suppose, as I believe, that Harry Truman was a morally serious person, and that he cannot credibly be charged with indifference to the salience of destroying innocent human life, including lives of Japanese civilians during the war. Thus he had and deployed a rule of moral salience alerting him to the moral significance of taking innocent human life, which no doubt prompted explicit deliberation in the extraordinary case. (No one can doubt the context of this decision was extraordinary; its gravity is scarcely possible to appreciate fully.) Suppose next what is at least not entirely obvious, that Truman acted wrongly in authorizing the use of the nuclear weapons without offering the Japanese further opportunities to negotiate terms of surrender. Even granting this stipulation, the fault which is justifiably laid at Truman's door is a mistake in explicit judgment, not a mistake in routine reasons-responsiveness or in picking up on morally salient features of context. The fact that so much was at stake, so far from entailing that Truman is unworthy of an honorary degree for his error, even supposing it to be such, actually points to the consideration in virtue of which we should not judge him too harshly. It is not beyond the pale to ask if Anscombe's eagerness to assume the higher horse more clearly constitutes a moral fault.[24]

Consider next how the view I am developing treats more mundane cases of conflicting obligations. It is wrong not to fulfill a genuine moral obligation, but a genuine moral obligation must be something a person *can* fulfill. Hence there is a puzzle, if I promise Ted that I will meet him at three o'clock in the seminar room and I promise Max that I will meet him at three o'clock at the university center. Whatever I do, I fail to fulfill an obligation, in apparent violation of the ought-implies-can constraint just articulated. It is at this point, I believe, that a Kantian theorist's head is supposed to explode.

It has long been recognized that progress toward explaining at least some such cases can be made by appeal to the person's past moral mistakes. In making at least one of the promises in question I act wrongly. *That* is the genuinely wrongful action in this case, it might be claimed, not my failure to fulfill one of the promises when the time comes.[25]

There is something correct about this line of thought, but I find it does not entirely dispel the sense that failing to fulfill the promise is itself wrong. I want to suggest that the puzzle here

---

[22] Recall also that rules of moral salience are partly relative to circumstance. A lifeguard or an emergency medical technician may have reason to routinize judgment about triage in a way that is not to be expected of everyone.

[23] G. E. M. Anscombe, *Mr. Truman's Degree* (Oxford, 1958).

[24] It is perhaps worth pausing to compare Truman's decision to use nuclear weapons on Japanese cities to George W. Bush's decision to invade Iraq in 2003. Even if the latter decision might have been permissible if done in the right way for the right reasons by a person of moral sensitivity, the course of action in fact undertaken was a moral travesty. There were many dimensions to this travesty, but perhaps none is more significant than the manifest lack of moral seriousness and sensitivity of the person mainly responsible for initiating it.

[25] For an exemplary treatment, see Ruth Barcan Marcus, "Moral Dilemmas and Consistency", *Journal of Philosophy* 77, 1980.

dissolves once we adopt the orientation, already motivated by independent considerations above, of focusing moral assessment in the first instance on a person as temporally extended. The wrong in this case is what common sense indicates it is, namely the failure to fulfill a promise. Where common sense may mislead is in locating that wrong too narrowly in time, at the moment when the decision is made to meet Ted rather than Max. I suggest instead that the wrong is located in the whole process of making the promise but failing to fulfill it; this explains both what is correct about locating the moral mistake in the making of the promise and about saying that the mistake consists in failing to fulfill that promise.

Focus on the temporally extended character of agency also helps explain the nature of the wrong in negligence, a problem that bedevils motivation-emphasizing accounts of wrongfulness. Unlike recklessness, indifference, and malice, negligence need not involve awareness of morally salient possibilities. Rather, it involves a failure to think about and prepare for these possibilities. The puzzle is to explain what is wrong with the motivations of a negligent person, when all the temporally local positive content of her motivations is unobjectionable. On this point there is no biting the bullet; backing a car out of a driveway without looking behind is wrong, if anything is. But negligence is less puzzling when we *begin* moral assessment with character, with assessment of a person as temporally extended. The moral problem is the conjunction of inadequate thought or preparation with the present circumstance, and there need be nothing odd or exceptional about invoking the usual range of moral assessments to evaluate these temporally extended features. Our maxims extend over these broad temporal ranges, hence so too do our actions insofar as they are morally evaluable. Our moral character, as we might put the point, is something we do.

7. Conclusion: Social Rules

Thus far I have focused on obligation and first-personal moral judgment, what we might call the supply side of morality. In this concluding section I use the resources already articulated to make a conjecture about morality's demand side, rights and entitlements. More specifically, I suggest we understand moral rights as corresponding to rules of moral salience. Since some rules of moral salience do not depend for their appropriateness on any actual social structures, this vindicates the possibility of natural rights. Natural rights are those that emerge simply from the conjunction of features of individuals in virtue of which they have moral standing (humanity, in Kant's view, though I suspect consciousness is a better candidate) and the general circumstances of their lives. Thus there are natural rights to life, bodily integrity, and the use of objects, which correspond to rules of moral salience that occasion explicit judgment when a morally competent person in a state of nature considers killing someone, invading his person, or taking his things. But as with rules of moral salience, most of the necessary action-guidance is accomplished not by natural rights, but by the more specific rights present in a particular social context. Thus a person has a right against sexual harassment, on this proposal, just in case in her social context potential harassment is experienced by any morally competent person as morally salient.

Social rules often make explicit the connection between considerations to which people ought to be sensitive and forms of treatment to which people are entitled. In some cases actual social rules make obligations and entitlements specific enough to be action-guiding, when they otherwise would not be. Consider a social rule requiring people to pay taxes in part to forestall

generation of morally salient needs among the elderly, for example, or a rule requiring people to pay for garbage collection to forestall morally salient wastefulness.[26]

Although it is less apparent, another purpose of social rules is to enable discernment of who has adequate rules of moral salience which they sincerely deploy. In some cases, such as the examples from the preceding paragraph, the social rule removes from serious consideration the possibility that another person acts poorly because she is unaware of certain salient features of her environment. The rules of taxation and garbage collection make it practically impossible to be sensitive to a salient consideration without performing certain actions in a public way. Violation of these rules without special justification is not only evidence a person acts wrongly, it is also evidence a person lacks a rule of treating elder need and waste as morally salient. This is a red flag, alerting others of the need for caution when associating with such a person: among the most important morally salient features of our environment are other people's rules of moral salience.

This in turn may explain another puzzling feature of rights, namely their residues and remainders. Even when a rights-violation is justified, it is often the case that rights-violators have a responsibility to apologize or otherwise indicate mindfulness of the right in question. In cases where rights-violation is justified, this may seem odd; if a person is justified in doing what she did, how could she have a responsibility to apologize for doing it? I would close by suggesting that this phenomenon is explained by the importance in human social life of a meta-rule of moral salience, of the need to be sensitive to other people's routines and rules of morally salience. By apologizing for using your telephone without permission to call an ambulance for a stranger who collapsed at a restaurant, I signal that I treat use of other people's property as a morally salient consideration, which in turn marks me as someone with whom it is safe to associate. That the need to apologize for justified action tracks the existence of a defeated moral right against the action is thus neatly explained by the claim that moral rights correspond to normal rules of moral salience. This claim demands a much fuller development on another occasion, of course, but it strikes me as promising to pursue.

---

[26] Cases of this sort are given illuminating discussion in Chapter 5 of Onora O'Neill, *Towards Justice and Virtue* (Cambridge, 1996).

Autonomy: Incoherent or Unimportant?

Mike Valdman

We seem capable of self-government, of controlling and shaping our lives. This also seems to be one of our more important attributes; it seems to be something that *matters*, especially with regard to how we should be treated. Theories of personal autonomy attempt to explain this, but no such theory, I'll argue, is likely to deliver a coherent account of what self-government involves without undermining the case for its mattering. Autonomy theorists, I'll argue, no matter the details of their view, face a potentially intractable dilemma.

Very briefly, the dilemma stems from a choice that theorists confront when considering an agent's role in the process that confers autonomy upon her desires. An agent's being autonomous either requires her active involvement in this process or it does not. If it does, then being autonomous will require that agents control or govern the very desires that must control or govern them, which, I'll argue, is incoherent. If it doesn't, then we'll be left with a normatively uninteresting conception of autonomy – one that, among other shortcomings, won't be able to ground a presumption in favor of letting people pursue their interests without coercion, manipulation, or interference. Or so I shall argue.

1. Background

My target in this paper is *personal autonomy* – the idea of being self-governing, self-directing, or the author of one's life. Conditions for self-government (or self-direction, etc.) are hotly disputed, but most will agree that being self-governing is largely a matter of being properly *motivated* – of having and acting on the right desires.[1] Following common practice, I'll refer to these as *autonomous desires*. The standard view is that not every desire may be autonomous, and that a person is autonomous when, and perhaps to the extent that, she maintains and acts on the ones that are.

[1] Throughout I will use "desire" in a broad sense to include any and all motivating elements of a person's psychology. A desire, on my view, is anything that would be a member of what Bernard Williams once referred to as one's subjective motivational set.

There are three kinds of theories as to what makes desires autonomous. According to *structural* theories, a desire D is autonomous (roughly) if it is related in the right way to (a certain subset of) its bearer's other desires.[2] According to *historical* theories, D is autonomous (roughly) if it was formed in the right way.[3] According to *rationalistic* theories, D is autonomous (roughly) if its bearer maintains, endorses, or acts on D for good reasons.[4] Many theorists defend a version of one of these theories. Some defend hybrid views that incorporate elements of all three.[5]

A fundamental question facing defenders of these theories, and, indeed, of any theory of autonomy, concerns an agent's role in *making* her desires autonomous – her role in the process that confers autonomy upon her desires (hereafter the *autonomy-conferring process*). Notice that defenders of historical, structural, and rationalistic theories needn't require an agent's active involvement in this process. A historicist, for instance, could claim that a desire D is autonomous if it was formed in the absence of coercion or manipulation, whether or not its bearer approved of its formation, shaped its content, or engaged with it in any meaningful way. A structuralist could claim that D is autonomous if D coheres with its bearer's other desires, whether or not its bearer endorses D or the desires that D coheres with. A "rationalist" could claim that D is autonomous if there are reasons to endorse or to act on D, whether or not its bearer endorses or acts on D *for* those reasons. On these views, agents needn't be actively involved in the autonomy-conferring process; *they needn't do anything to make their desires autonomous*. I will call such views *mere authenticity* views.

---

[2] Structural theories come in hierarchical and non-hierarchical varieties, but that distinction won't matter for our purposes. For a hierarchical theory see Harry Frankfurt (1971), "Freedom of the Will"; Gerald Dworkin, *The Theory and Practice of Autonomy* (Cambridge: Cambridge University Press, 1988). For a non-hierarchical theory see Laura Ekstrom (1993), "A Coherence Theory of Autonomy," *Philosophy and Phenomenological Research* 53 (1993): 599-616.

[3] John Christman is among those who have defended a historical theory of autonomy. See his "Autonomy and Personal History," *Canadian Journal of Philosophy* 21 (1991): 1-24.

[4] See George Sher, "Liberal Neutrality and the Value of Autonomy", *Social Philosophy and Policy* (1995): 136-59; Sigurdur Kristinsson, "The Limits of Neutrality: Toward a Weakly Substantive Account of Autonomy," *Canadian Journal of Philosophy* (2000), 30 (2): 257-86.

[5] I mention several hybrid theories in section 3.

Alternatively, one could claim that, for a desire to be autonomous, its bearer must have actively engaged with it, perhaps through some process of critical reflection and evaluation, so as to have conferred upon it its special status. Thus a historicist could claim that a desire D is autonomous only if its bearer *guided* its development or *crafted* its content. A structuralist could claim that D is autonomous only if it coheres with desires that its bearer *endorses*. A "rationalist" could claim that D is autonomous only if its bearer *recognizes* reasons for maintaining D and only if she maintains D *for* those reasons. On these views, autonomous agents must make their desires autonomous by actively engaging with them in the right way; they must be *autonomy-conferrers*. I will call such views *agent-government* views.

Whichever theory of autonomy one accepts, then, whether historical, structural, rationalistic, or a hybrid, one must choose between an agent-government view and a mere authenticity view – between a view that requires agents to be autonomy-conferrers and one that does not.[6] And here lies the dilemma, for agent-government views render autonomy incoherent while mere authenticity views render it unimportant. Requiring agents to be actively involved in the autonomy-conferring process would require them to govern or control the very desires that must govern or control them, which, I'll argue, is incoherent (section 2). But not requiring such involvement makes it hard to see what meaningful role autonomy could play in normative discourse (section 3). Mere authenticity views might not strip autonomy of all normative importance, and I can't definitively rule out the possibility that, on some version of this view, autonomy might matter for some purpose or other. But I'm convinced that such views can't vindicate autonomy's vaunted status in contemporary moral and political argument.

2. Agent-Government

[6] More precisely, one must decide whether to incorporate an agent-government *condition* into one's theory of autonomy. I don't mean to rule out the possibility of a hybrid view according to which autonomy is sometimes conferred actively, in accordance with agent-government, and sometimes passively, in accordance with mere-authenticity.

The case for agent-government's incoherence is fairly straightforward, so I begin there. First there is the much discussed threat of regress. To see the worry, recall that, on an agent-government view, agents must *do* something to make their desires autonomous. But must they also be autonomous with respect to these doings? It seems that we should answer in the affirmative since, if these doings are not themselves autonomous, it's hard to see how they could function as autonomy-conferrers. But an affirmative answer seems to generate a regress since, in keeping with an agent-government model, we'd then need to posit further agential doings to make the doings in question autonomous, then further doings to make those doings autonomous, and so on, *ad infinitum*.

Consider next a related problem, and one that reveals more clearly why agent-government is incoherent. Defenders of agent-government see autonomous persons as genuine shapers of their lives – as persons who, in a robust sense, govern or author their desires and actions. They see autonomous persons as having *authority* over their desires – as capable of exerting a kind of *managerial control* over them, with the ability to stand back from their desires, asses them at a distance, and decide which to act on, to ignore, and to shed. Theorists disagree about what having such control involves, but they all seem to think that autonomous agents must have some such control, and that its exercise is the means by which agents put their stamp of approval on their desires, so to speak, thereby making them their own.

But there is a deep problem with this view no matter how the idea of managerial control is unpacked. For consider. Having such control over one's desires surely requires that one engage in a deliberative process whose purpose is to determine whether some desire is worth having. But this deliberative process, it seems, must be guided by some psychological entity or other, whether it's a desire, a value, or something else; one can't, presumably, deliberate from nothing or according to nothing. But now consider the status of these guiding entities. Must they too be under the agent's control? Must they too bear his stamp of approval? If not, then it seems as if these entities, and not the agent who bears them, would have ultimate governing power. On

this model, autonomy would have to consist in there being the right relations between these guiding entities and an agent's other desires, actions, and choices (this would then be a type of mere authenticity view). But if these entities are under the agent's control, then the picture would be of agents having deliberative control over the very entities that guide their deliberations, governing the very processes that determine how they govern. That seems untenable. Agents, surely, can't control, via deliberation, the very entities that guide their deliberations. These entities can't be both an agent's servant and his master.

In this paper's longer version I consider three replies. The first appeals to an analogy with democratic self-government. The second identifies the aforementioned guiding entities with the agent himself, claiming that they can constitute his identity. The third attempts to show that agents can have managerial control over their desires through a kind of pure deliberation, untainted by the motivating elements of their psychology. I lack the space to discuss these replies here, but I don't think they succeed.

3. Mere Authenticity

Mere authenticity views may seem unpromising. It's odd, after all, to think that autonomous persons needn't be autonomy-conferrers – that they needn't do anything to make their desires autonomous. Indeed, such views seem to relegate "autonomous" agents to *spectators* or *bystanders* vis-à-vis their desires. And while most mere authenticity theorists will insist on what can perhaps be described as an engaged form of spectatorship (e.g. that agents not have certain negative attitudes towards their desires, that those desires not be inconsistent with their core convictions, and/or that agents be satisfied with their desires in the passive sense that they lack an interest in changing them), seeing autonomous agents as spectators vis-à-vis their desires – even as engaged spectators – makes it hard to see what important role autonomy could play in normative discourse. I'll offer two arguments for this claim. In sections 3.1 and 3.2 I'll argue that, on a mere authenticity view, autonomy can't ground a presumption in favor of letting people live their lives without coercion, manipulation, or interference. In section 3.3 I'll argue

that the price of escaping the aforementioned regress by way of a mere authenticity view is inheriting a set of normatively irrelevant distinctions. All this won't show definitively that mere authenticity views lack normative importance, but, together, they put the burden squarely on my opponent to justify his or her conviction to the contrary.

3.1 Mere Authenticity and Personal Sovereignty

Consider the role envisioned for autonomy in moral and political argument. Some believe that its role is profound – that it grounds our moral status, our most basic rights, and the state's duty to take its citizens' interests seriously.[7] Arguments for these views, however, tend to rely not on autonomy itself but on the capacity for it, which is more widespread.[8] Actual self-government's primary role, it seems, is to ground a constraint against certain kinds of manipulation, coercion, and interference – to justify a strong presumption in favor of letting people live their lives and pursue their interests even if they're likely to make sub-optimal choices.[9] Its role, in short, is to ground a presumption against interferences that undermine a person's ability to govern himself or that thwart his will. Call this the *presumption of personal sovereignty,* or PPS, for short. I'll argue that mere authenticity views can't ground it.

Begin with a worry about grounding PPS in any conception of personal autonomy, mere authenticity or otherwise. Such autonomy, notice, isn't had just by being a person. Whether autonomy is understood historically, structurally, or rationalistically, persons will be autonomous to varying degrees, with some potentially lacking it entirely. Yet PPS, it seems, is meant to protect *all* persons, regardless of the quality, origins, or structural coherence of their desires. As long as a person isn't harming anyone (himself included), it seems that we should let him act on his desires without interference even if his desires are silly, mutually inconsistent, or even if they

[7] See, for instance, David Richards, "Rights and Autonomy," *Ethics* 92 (October 1981), 3-20; Robert Nozick, *Anarchy, State, and Utopia*, (New York: Basic Books 1974) p. 48-51.
[8] Of course, if I am right that autonomy is either incoherent or unimportant, then it isn't clear that the *capacity* for it will be able to play an important role in moral and political argument either.
[9] Steven Wall may be the most explicit on this point, but it's widely accepted. See his (1998), *Liberalism, Perfectionism and Restraint* (New York: Cambridge University Press): 140-146.

were implanted in him by a wizard. And if he were harming others we'd plainly have reason to violate his personal sovereignty *regardless* of his autonomy; his autonomy would then offer him no protection at all and may even be an aggravating factor.[10] And so it appears that one can act non-autonomously yet be protected by PPS and one can act autonomously yet not be protected by PPS. Autonomy, then, does not seem to be PPS's ground.

Of course, even if autonomy doesn't ground PPS, it could still contribute to its strength. Perhaps there is *more* reason not to violate the personal sovereignty of an autonomous person, all else being equal, than that of a non-autonomous person.[11] Or, to frame the issue in terms of desires, one might think that while most desires should be respected, autonomous desires should get more respect (perhaps much more) than their non-autonomous counterparts.

Such views seem plausible, but how could one defend them? The most natural way, I think, is to link PPS with a duty to respect persons. It's natural to think that PPS is ultimately grounded in this duty, especially if PPS covers *all* persons. And it's natural to think that a duty to respect persons includes a duty to respect their desires. But which ones? A tempting answer is those that have a deep connection to their bearer such that, by granting those desires special deference, we'd be showing respect for the person who bears them. And that suggests a natural connection between respecting persons and respecting desires that satisfy agent-government criteria since, on that view, autonomous desires owe their special status to their bearer's active approval or engagement. A desire's autonomy on the mere authenticity view, however, requires no such thing. Mere authenticity views, recall, treat agents as spectators vis-à-vis their desires, and that makes it hard to see why granting those desires special deference should count as showing respect for the person who bears them. Let me explain why.

---

[10] It may seem that, in the case of autonomous wrongdoing, we have more reason to violate the perpetrator's sovereignty than we would if he were acting non-autonomously. See Raz, J. (1986), *The Morality of Freedom*, p. 380.

[11] Except, perhaps, in the case of autonomous wrongdoing. The thought, then, could be that autonomy strengthens PPS in the case of morally appropriate behavior but weakens it in the case of immoral behavior.

Start with an analogy. Suppose that you're invited to the home of a famous artist (Pierre). His home is littered with artwork, some of which he acquired, some of which he made, some of which were gifts, and some of which were left behind by others. If you wish to show respect for Pierre as an artist, which of his pieces should you single out for *special* praise? The answer, presumably, is those that he made and perhaps those that he had a hand in acquiring. It would be odd, though, to heap praise on the pieces that satisfy *only* the aesthetic analog of the mere authenticity view: those that cohere with their surroundings (or with Pierre's aesthetic preferences), are worth having, and weren't acquired, say, by theft or fraud. And that, presumably, is best explained by noting the tenuous connection between those pieces and Pierre *qua* artist – to the fact that, with respect to those pieces, he is more aptly described as a spectator than as a creator or acquirer.

Consider next the case of unconscious desires. Such desires, notice, could satisfy all plausible mere authenticity conditions; they could have been uncoercively formed (they could be innate), they could cohere with their bearer's other desires, and they could be supported by reasons. Yet it's hard to believe that such desires should receive *special* deference – that, all else being equal, the presumption against interfering with actions that flow from them should be much stronger than the presumption against interfering with actions that flow from their inauthentic counterparts. And it's especially hard to believe that part of what it is to respect someone as a *person* is to grant such desires special deference.

Of course, one could always work into one's mere authenticity account a condition that excludes unconscious desires from contention. A historicist, for instance, could claim that, in order for a desire to be autonomous, its bearer must have approved of its formation in the sense that she was aware of it and didn't resist it.[12] But such approaches are bound to disappoint so long as the awareness and non-resistance condition is understood *passively*, as requiring only that

[12] John Christman includes a passive awareness and non-resistance condition in his historical model of autonomy. See his "Autonomy and Personal History": 10. Harry Frankfurt's satisfaction requirement can be considered a non-historical version of this condition.

an agent not have resisted the desire in question (or that he not now resist it). If it seems otherwise, that may be because a desire's satisfying that condition could provide *evidence* that its bearer has actively engaged with it in a way that satisfies the conditions for agent-government (whatever that may involve). But if we consider cases in which that evidential link is severed, it will be clear that D's satisfying that condition has no bearing on its respect-worthiness.

To see that, return to the case of the artist and suppose that a sculpture S that he received as a gift satisfies not only the aesthetic analog of the mere authenticity view but also a passive awareness and non-resistance condition – Pierre was aware that S was being offered to him and he did not resist its inclusion in his collection. If you wish to show respect for Pierre as an artist, should you then single out S for special praise? Well, notice that there are many reasons why Pierre might not have resisted S's inclusion in his collection, not all of which provide evidence of active approval (whatever that may involve). His non-resistance could have been due to indifference, for instance, or to a desire to not embarrass S's giver. And in that case Pierre can't be said to "own" S in the relevant sense, undermining the thought that his passive non-resistance to S makes showing respect for S akin to showing respect for Pierre as an artist. A desire's respect-worthiness, it seems, is not enhanced merely by satisfying a passive awareness and non-resistance condition.

3.2 Justified Interference

Things get much worse for the mere authenticity view when we consider another crucial aspect of PPS. PPS, clearly, is meant to protect us from unwanted interference. It's meant to protect us from being forced to live according to someone else's conception of a good life. But it isn't meant to protect us from all interferences or even to raise the bar against them. Odysseus's mates, for instance, don't violate his personal sovereignty by tightening his bonds as they sail past the Sirens.[13] They interfere with him, to be sure, preventing him from acting on what is then his strongest desire. But this accords with his deeper wishes, so it's not the sort of interference that

[13] Odysseus orders his men to tie him to the mast and to ignore any later orders he might give to be set free.

PPS is meant to discourage. Crucially, this isn't a case in which PPS is simply outweighed.[14] Rather, in light of Odysseus's preferences, this is a case in which PPS doesn't apply. Or if it applies, it does so in the opposite direction, licensing, and perhaps even demanding, the interference in question.

Grounding PPS in autonomy might explain all this. Being autonomous, after all, is largely about having and acting on one's own desires. And so, when others impose goods on us that we reject – when they impose on us *their* desires – they fail to respect our autonomy. But there may be no such failure when they "impose" goods on us that we welcome – when they act in accordance with our deepest desires, imposing on us that which we consider to be good by our own lights. In such cases they may interfere with our actions but not with our autonomy, and we should expect a PPS grounded in autonomy to acknowledge the difference – to set a high bar against interferences that seek to impose goods on us that we reject but not against interferences that seek, and that can be reasonably expected, to further our own desires, interests, and goals.

It would be a mistake, however, to count among our own desires, interests, and goals in the preceding formula those that satisfy only mere authenticity conditions. The case of unconscious desires is once again illuminating. Surely PPS shouldn't be inoperative when people interfere with us with the reasonable expectation that that will help us act on and satisfy some of our unconscious desires, even if those desires are supported by reasons, are structurally coherent, and were uncoercively formed. Our assessment of the Odysseus case would change drastically if his desire to hear the Sirens were one he didn't even know he had. It would clearly infringe PPS to tie someone to a mast and to keep him there despite his protestations on the grounds that that is the only way for him to obtain some benefit that he unconsciously desires or to avoid some harm that he unconsciously wishes to avoid (even if his desire to receive the benefit in question is stronger than his desire not to be restrained). And adding a passive awareness or non-resistance

[14] It may be outweighed because of the risks to Odysseus and his crew. But many, I think, believe that securing Odysseus to the mast would be justified regardless of the potential harms.

condition wouldn't help. Even if Odysseus were aware that he had a desire to hear the Sirens and even if he hadn't (passively) resisted its acquisition, it still seems that we would be a great distance from justifying the interference in question.

In general, it seems that in order for a desire D to be an agent's *own* in the sense that promoting its satisfaction could be seen as promoting the *agent's* interests, the agent should have played an active role in conferring upon D its privileged status (a condition of agent-government). At a minimum, while interferences that help people satisfy desires they've actively approved of or endorsed may not violate their sovereignty, the same cannot be said of interferences that compel people to act on desires that satisfy only mere authenticity conditions and passive awareness and non-resistance conditions. There are those desires that we claim as our own and there are those that came about in the absence of coercion and manipulation, that cohere with our other desires, and that are supported by reasons. These sets will often overlap, but they won't always do so. And when they don't, we expect PPS to throw its weight behind desires in the former category – to protect us from those who would interfere with how we take ourselves to have chosen to live. And that means that a mere authenticity view cannot be PPS's ground.

3.3 Normatively Irrelevant Distinctions

To broaden my attack on mere authenticity's normative importance, consider Robert Noggle's view.[15] Noggle seeks a way to make sense of autonomous action – or, more precisely, of authentic desires, where these, on his view, are necessary for autonomous action – in light of two problems that threaten authenticity's coherence. The first is the regress problem, which we've already encountered. The second is the *ab-initio* problem, which consists in specifying how an inauthentic element of a person's psychology can be authenticity-imparting. It's odd, after all, to think that some desire of yours could be authentic if it's the product of desires and processes that aren't themselves authentic (the brainwashing cases that pervade the autonomy

[15] Robert Noggle, "Autonomy and the Paradox of Self-Creation: Infinite Regresses, Finite Selves, and the Limits of Authenticity," in James Stacey Taylor (ed.), *Personal Autonomy: New Essays on Personal Autonomy and its Role in Contemporary Moral Philosophy*, Cambridge 2005: 87-108.

literature often exploit this idea). But if only the authentic can give rise to the authentic, then a regress ensues, raising worries about authenticity's coherence (and if authenticity is necessary for autonomy, then for autonomy's coherence as well).[16]

Noggle's solution is to reject the ab-initio requirement. A theory of authenticity, he writes, must "leave open the possibility of another means by which authenticity can arise besides having it be conferred by some other element that is already authentic."[17] That seems like the right move, and, indeed, the only move that preserves authenticity's coherence. But what are the other means to which Noggle refers? How can an authentic self emerge from inauthentic sources? Where does the authentic self come from?

Noggle's answer is that an authentic self can, and often does, emerge gradually in childhood through ordinary developmental processes like operant, aversive, and classical conditioning, imitation, blind obedience, and the internalizing of norms. "Out of a seemingly unpromising beginning – a sort of chaotic 'psychological soup'", he writes, "the child's self gradually emerges as her cognitive and motivational systems develop the kind of structure and stability and the rational and reflective capacities necessary for the existence of a coherent and stable self that can be the source of authenticity."[18]

But what should we say when these developmental processes are manipulated? Consider Noggle's *Oppressed Olivia*, who:

> …has been raised (using standard child-rearing techniques) to abide by and adopt the sexist attitudes of the patriarchal society in which she lives. Consequently, she shapes her ideals, aspirations, and activities in ways that reflect these attitudes. As Olivia reaches adulthood, her convictions include a belief in the naturalness of women's subservient role, and her deepest aspiration is to be a housewife.[19]

[16] The ab-initio requirement poses a problem for autonomy even if authenticity isn't necessary for it. After all, it seems just as odd that a desire could be autonomous if it's the product of desires and processes that aren't themselves autonomous.
[17] Ibid., 99
[18] Ibid., 101
[19] Ibid., 102.

Should we regard Olivia's subservient convictions as authentically hers? Noggle thinks we should. Authenticity, he notes, is a two-place relation. "Before the self initially arises, there is no other self for the initial self to bear any authenticity-grounding relation to."[20] "[I]t is meaningless," he continues, "to ask whether the initial self that arises in [Olivia] is authentic. When that initial self forms, it is the only self that there is."[21]

Noggle is quick to add, though, that "it makes a great deal of difference whether such processes [e.g. conditioning, imitation] are being used to build an *initial* self, or whether they are being used to implant psychological elements into an *existing* self."[22] Consider *Brainwashed Ben*, who was raised Catholic but was then brainwashed by a cult, which, with the aid of drugs, and using the aforementioned socializing techniques, got him to adopt their religious views. Ben's newly acquired religious convictions, thinks Noggle, are likely inauthentic. "There is a big difference," he writes, "between the application of brainwashing and related techniques to a person with a fully formed self and the application of very similar techniques during the early stages of child rearing."[23]

Let me stress the plausibility of Noggle's view. The ab-initio requirement, though compelling, raises a deep conceptual problem for authenticity (and for autonomy), giving us reason to reject this requirement. If we reject it, we must say that an authentic self can arise from inauthentic sources. And if we say that, and we're convinced both that autonomy is possible for creatures like us (i.e. creatures that emerged from "psychological soup") and that at least some forms of manipulation, like the kind in *Brainwashed Ben*, undermine autonomy, we'll be drawn to a view like Noggle's that distinguishes between manipulation involved in a self's creation and manipulation that alters an existing self. In broad outline it's hard to see an alternate path.[24] But

[20] Ibid., 103
[21] Ibid., 103
[22] Ibid., 104
[23] Ibid., 105
[24] One alternative would be to locate the authenticity-imparting sources outside the agent. But I agree with Noggle that such a model would not qualify as a model of *self*-government.

where does that path lead? Is there really a *big* difference between manipulation involved in a self's creation and manipulation that seeks to alter an existing self, as Noggle claims? Is it a difference that matters?

I don't see how. Suppose that Olivia could take a pill that would make her less subservient, leaving in place all her other desires and commitments. Does the *mere* fact that subservience is part of her initial self give her any reason not to take it? Suppose that Bill could take a pill that would restore his Catholicism. Does the *mere* fact that Catholicism is part of his initial self give him any reason to take it? The answer to both questions, I believe, is No. Olivia and Bill might have many reasons to take their respective pills (and many reasons not to), but considerations of *authenticity* don't seem to be their source. To think otherwise – *to think that mere authenticity matters* – is to think that there is something special about the self, or, more precisely, the bundle of desires, dispositions, and commitments, *that just happened to get there first*. But why should anything of normative importance turn on that? If we inquire into how we should live our lives or what we owe to each other, can the answer really depend, even in part, on something as arbitrary as the order of desire acquisition? An independent Olivia and a cultish Ben might be inauthentic, but it's hard to see how that judgment has moral weight if it means only that they aren't acting in accordance with desires that were first to emerge from what Noggle aptly describes as "psychological soup."

Noggle's version of the mere authenticity view, of course, is just one among many. Still, it highlights a problem confronting all such versions. Embracing an ab-initio requirement on either autonomy or authenticity leads to a regress. Rejecting it leads to a view like Noggle's that distinguishes between desires that came about as a result of conditioning processes prior to one's having a developed self and those that came about through similar processes after the self's development. That distinction is fine for purposes of (mere) classification. But if one were to ask how one should live one's life, it would be bizarre to say that one should live it (or even that one always has at least some reason to live it) according to the very first bundle of desires that one

was conditioned to have (or according to the desires that emerged, via proper means, from that bundle). There is nothing special about the desires that got there first. There is nothing special about authenticity *per se*.

4. Conclusion

We have reached a startling conclusion. Theories of autonomy must be either mere authenticity theories or agent-government theories. The latter requires agents to govern that which governs them, which is incoherent. The former turns agents into spectators vis-à-vis their desires, making it hard to see why we should care about the desires that pass the mere authenticity test. And so it seems that autonomy, if it's to be coherent, may also be normatively unimportant, or at least much less important than is typically believed.

Personal Ideals, Rational Agency, and Moral Requirements
Sarah Buss

What is it that we disagree about when we disagree about whether someone is a good mother? or a good friend? or a good person? What is wrong with someone whose every choice reflects her belief about what is permitted or required by some general rule? How can anyone aim at being a good person when she has only the vaguest notion of what it would take to succeed? What is the relationship between being *morally* good and being a good mother or friend – or artist or philosopher or citizen? What is the relationship between the reasons to which a morally admirable person responds and the reasons to which a prudent person responds? Can there be genuine moral dilemmas?

We cannot go very far down the trail that leads from any one of these questions without finding ourselves heading off in the direction of the others. I propose to illustrate this claim by exploring the role that ideals play in our lives. More specifically, I propose to explore the thick and thin elements of personal ideals, and the implication of the fact that every ideal is both in one respect thick and in another respect thin. The distinction I have in mind is not the distinction between the descriptive and evaluative elements of ideals, though it is certainly closely related. It is, rather, the distinction between our conception of what is required in order to realize a given ideal and our appreciation of the fact that this conception is a provisional approximation. If we can better understand these two aspects of our ideals and their interdependence, and if, additionally, we can understand the role each aspect plays in the relationships among our ideals, then we will have the resources we need to tackle the questions with which this paper began.

This, at any rate, is my motivating assumption. Perhaps it would be more accurate to call it a "hunch." Operating on this hunch, I will begin by considering an ideal that has played an important role in debates over the necessary conditions of moral motivation. I will suggest that this ideal calls our attention to the basic conditions we must satisfy in order to be motivated by *any* ideal. I will then appeal to these conditions to explain what is right about the widespread assumption that whenever we do things for a reason, we act "under the guise of the good." This discussion will lead, in turn, to a brief review of how we revise and refine our conceptions of the ways we want to be good. Reflecting on this process will naturally raise questions about the relations among our ideals. And the answers to these questions will shed light on the possibility of moral dilemmas. Moral dilemmas involve irresolvable conflicts between the requirements we must satisfy in order to be morally good. Having considered how such internecine conflicts can arise, I will then turn to the conflict between (i) the constraints we must respect in order to be morally good and (ii) the constraints that spell out what it is to be good in nonmoral ways.

In reaching this point, I will, I hope, not merely have addressed some interesting questions. I will also have shown how the answers to these questions are elements

of a single story. This is the story of how we manage to treat a rather motley collection of ideals as guides to action, even as we are aware of how imperfectly we understand what is involved in realizing these ideals, and how little we can say to justify our assumption that they are worthy guides -- and even as we are thus aware that nothing can rule out the possibility that we are mistaken to rely on these guides.

Chapter 1.

It is widely agreed that even if someone does what she has good reason to do because she believes she has good reason to do it, there is something wrong with her if she is not able to "see" any of the features of her circumstances *as reasons* for acting as she does. Such a person is, we are told, deficient as a rational agent. But what, exactly, does this deficiency amount to? There is, it is said, something "fetishistic" about her motivational structure: rather than being directly ("nonderivatively") moved by the features or facts of her circumstances -- rather than, for example, being directly moved by the plight of the person who fell overboard, or the fact that this person is her wife -- she is moved by the fact that certain actions pass the test she believes she must apply in order to determine what she has reason to do.

But what is wrong with doing something because one endorses some principle according to which acting this way is the reasonable, or right, thing to do? Can there really be anything wrong with being so motivated? To answer this question, we need to understand the role that action-guiding principles play in seeing a fact *as a reason*. Whatever reasons are, it seems that *seeing* a concrete feature of one's circumstances *as a reason* to perform a particular action differs from merely *believing that* there is some formal principle according to which we are justified in performing the action under these circumstances. But what does this phenomenological distinction amount to?

Let us set to one side any reasonable concerns we may have about whether any principle or principles could possibly generate determinate recommendations for every situation. Let us assume that our fetishist is confident that the action-guiding principles on which she relies yield adequate verdicts for the choice situations in which she finds herself. Every time she acts, the fetishist does so only because she believes that her action passes the test spelled out by the principles she consults. Our question is, then: what is wrong with relating to one's action in this way? What is the ideal of agency in relation to which the fetishist falls short?

Is the problem simply that she is more self-conscious about the basis of her responses? Is it that, like the moral novices Aristotle discusses, her responses do not "come naturally"? But what is problematic about this -- if, as we can suppose, she is able to respond as swiftly and confidently as the rest of us? The idea seems to be that the fetishist's apprehension of the normative significance of her circumstances is flawed in something like the way her apprehension of sizes and shapes would be flawed if she could not see two lines as having two different

lengths, but had to measure them both in order to reach this conclusion -- with great rapidity, perhaps, and even by exercising some special mental faculty that has no impact on her perceptions.

Note that, contrary to what some who bring the charge of "fetishism" suggest, someone would be just as handicapped if "the fact that this person is in need" or "the fact that this person is my wife" were what moved her, and yet she did not perceive these facts *as reasons*. If an agent simply *believes that* someone's need is a reason to help her -- perhaps because this is what she has heard people say -- and if, despite this belief, the needs of others "leave her cold," she is as alienated from the normative force of her actions as is someone who does what she does simply because she believes that otherwise she would violate a formal principle of rationality.

What the fetishist *believes* is just what the nonfetshist *sees*. How, then, can we cash out the phenomenological difference between their two different ways of relating to their reasons for action? This question prompts a search for analogies. Consider, for example, the difference between a well-trained autistic person's *belief that* a given facial expression is an expression of puzzlement and someone else's *experience* of the expression *as* an expression of puzzlement. Consider, too, what distinguishes the ordinary experience as of interacting with one's beloved from the experience of those who are afflicted with Capgras syndrome. If no one ever had the extremely distressing experience as of her beloved's being a stranger, it would be natural for us to assume that our sense of a person's familiarity is secured by our perception of her other properties. But people really do sometimes experience their spouses *as imposters*. So we are forced to concede that there is something more to the experience of familiarity than can be explained by our other perceptions -- something more, too, than can be explained by our well-developed habits.

If someone cannot *just see* the normative significance of her circumstances, then she must derive this significance from a principle or rule. If she cannot see the premises of her reasoning under the guise of a normative principle, then she must treat a normative principle as one of the premises in her reasoning. Is this extra "thought" her way of compensating for her perceptual handicap? Or is it simply the form that this handicap takes? In any case, whereas the nonfetishist is able to reason from the fact that her wife is drowning to the conclusion that she has sufficient reason to jump in the water to save her, the fetishist must appeal to a principle according to which if one's wife is drowning, one has reason to jump in to save her. This principle must be a premise in her reasoning -- it is an extra thought she must think -- because she does not "see" the fact that her wife is drowning as having the significance spelled out by the principle. What's more, though she appreciates that applying the principle is the thing she has reason to do, the content of this belief is so thin that there is almost nothing she understands in believing it.

It is important to distinguish this case of "one thought too many" from a case that involves a far less profound alienation from the normative significance of one's

circumstances and actions. The extra thought in the second sort of case is the thought that treating F as a reason to do A is an instance of doing something that can be described in more general, more abstract, terms. Someone who needs to think this sort of thought in order to believe that F is a reason to do A *sees* F as a reason in the sense that she sees treating it as a reason *as justified*. Her alienation from the normative significance of F is thus very different from that of the paradigm fetishist. Her failing is that she cannot see F as a reason without seeing a reason to see things this way.

To acknowledge that someone is handicapped if she is alienated from her reasons in either of the ways just sketched is not to reject the value of self-reflection. Critics of normative fetishism can readily concede that a responsible rational agent is prepared to call her normative and evaluative assumptions into question, and to review the considerations for and against these assumptions. Their point is that though there are many circumstances in which a well-functioning rational agent *could* reflect on the rationale for attributing a certain normative significance to certain facts, and though there are even some circumstances under which she *would* do so, she rarely *needs* to do this in order to discover what she has reason to do. In calling our attention to this fact, they remind us that in order for someone to do anything for a reason, there must be some fact whose normative significance she appreciates without engaging in any reasoning.

This does not mean that the nonfetishist's relations to reasons is more reliable than the fetishist's. Any direct perception as of some fact's being a reason to do something could well be a *mis*perception. And when a person's normative assumptions are behind the scenes, it is often especially difficult for her to call them into question. As Nietzsche points out, those who internalize Christian ideals overlook the contingency of their evaluative scheme; their perception of "good" and "evil" prevents them from appreciating the possibility of employing a very different scheme instead. (This is why, according to Nietzsche, the most admirable human beings are those whose "nature is designed entirely for brief habits": such rational agents "always believe" that they see things aright and are "to be envied" for having discovered the truth; but after some time passes, their "faith" weakens, and they replace their old ideals with new ones.)

The fetishist's handicap is that she cannot relate to her circumstances under the guise of thick, substantive assumptions regarding what is important and good. Even, however, as there is an important thick element in the ideal rational agent's experience of reasons, so too, there is an important thin element: no one can experience certain features of her circumstances as having a certain value and importance without experiencing this very experience *as justified*; it is not possible to see certain facts as reasons without regarding oneself as having *good reason* to draw the inferences one does, *whatever this good reason may be.* Even if we assume that in a wide range of circumstances, a person who does what she has sufficient reason to do is a person who tries to save those who are drowning nearby, we also

know that this assumption is an extremely partial, and possibly flawed, way of filling out what it is to be a good rational agent, conceived -- ever-so-thinly -- as such.

It seems, then, that the "seeing as" which characterizes the ideal experience of reasons is a rather special sort. In the familiar case exemplified by metaphors, to see one thing as a second thing requires having a determinate conception of the second thing. In contrast, seeing a fact as a reason involves treating a relatively determinate -- relatively thick -- guide to action (e.g., Try to prevent someone from drowning, especially if this someone is your wife) as a stand-in for something very indeterminate, or thin (e.g., Do only what you have sufficient reason to do). It is only because the nonfetishist can see the relatively thick as a stand in for the vanishingly thin that she does not need to treat the ideal of rational agency as a fetish. To put the point somewhat paradoxically, if she did not share the extreme fetishist's commitment to doing what she has sufficient reason to do, *under this description*, she would be forced to turn this commitment into a fetish.

What is true of the ideal of rational agency is true, too, of every other personal ideal. Ideals are ways of being good – including ways of being good enough - which an agent takes to be good ways of being. When "internalized," ideals determine the normative significance the agent attributes to the nonnormative features of her circumstances. (This is just what it is for them to be "internalized.") More specifically, they determine how those who have internalized them evaluate various actual and possible states of affairs -- which actions strike them as "worth performing," and which they ought to perform, which people appear to be worth emulating, which states of affairs appear to be worth bringing about, which things appear to be worth trying to obtain. But however thick one's conception of one's ideals may be -- and however many facts or features they thus enable one to see as reasons for doing some things and not others, there is no way one can possibly describe everything that is involved in being a good mother, a good teacher, a good citizen, a good colleague, a good person. In short, every conception of every personal ideal is necessarily indeterminate and provisional, and this is because every conception of every personal ideal takes the form of a description, or descriptions, of what it is to realize this ideal (or a set of requirements one must follow in order to realize it), where these thick elements are necessarily approximations, or stand-ins, for something thin: whatever it really takes to be good in this way.

Of course, in describing the kind whose way of being good is at issue in any given ideal, one need not be describing anything that approximates, or stands in for, something unknown. Thus, for example, to say that an ideal mother is a woman with a child is simply to say that she is a mother. Insofar, however, as the descriptive elements in our conception of this ideal represent what it is to be a *good* mother, something thin enters the picture. An ideal mother, we say, is a woman who feeds and clothes her child, talks to her regularly, reproves her when she throws rocks at car windows, etc. But these descriptions spell out an adequate guide to action only insofar as they are adequate descriptions of what a good mother

would do. So a mother's assumption that this is the way she ought to conduct herself is inseparable from the assumption that this is what she must do in order to meet a standard of which she has such a thin conception that she cannot rule out the possibility that she is mistaken.

Chapter 2.

It might seem that even if there is a thin element in every experience as of responding appropriately to our circumstances, it is, in principle, possible to flesh this element out. It might seem, moreover, that ideals had *better* be determinate standards: how else could they function as standards? I want to explain why it is not possible to expunge the thin element from our ideals -- not even in principle. This element is not, I want to argue, a contingent feature of our experience of doing things for reasons. Nor is it simply a necessary feature of reasoning. The thin aspect of our ideals is inseparable from what we are aiming at when we aim to realize these ideals. In particular, the experience and structure of reasoning reflects the fact that there could not possibly be a complete description of what someone must do in order to realize an ideal. Our contingent inability to get by without direct normative perceptions is inseparable from the fact that in doing things for reasons we aspire to achieve something we cannot possibly know whether we have achieved -- under this very description. The failure to appreciate this fundamental feature of nonfetishistic action for reasons underlies the resistance to the assumption that acting for a reason is acting "under the guise of the good." So, at any rate, I hope to show.

Whether the ideal we seek to "live up to" is the ideal of motherhood, friendship, or rational agency, as long as it plays a role in the reasons we see, we can intelligibly wonder whether in acting for these reasons, we have really done what a good mother, friend, rational agent would do. This is not simply because we are fallible. It is also because, no matter how much someone knows, there is an infinity of additional conceptually possible circumstances she could consider (if only because there is no conceptually necessary end to the causal chain of events). This means that there is an infinite number of circumstances in which any given ideal might be realized -- even if there are also an infinite number of circumstances (e.g., those in which there are no living things) in which it cannot be realized. And this means that none of us can possibly know just what it is to be good in a certain way in every possible circumstance.

This conceptual limitation on our knowledge of what we are aiming at in aiming to be good in some way has nothing essential to do with the fact that our ideals govern our actions by determining what reasons we see. It simply reflects the fact that most ways, or kinds, of being are not such that we can fully characterize what it is to be good in this way. But there is another reason why the endless supply of conceptually possible circumstances ensures that our conceptions of our ideals will always be incomplete; and this reason is a function of the normative significance we attribute to our ideals. Because we take it for granted that our ideals are constraints on what we have reason to do, our beliefs about what we have reason to do are

relevant to what we can and must do to realize our ideals. And this means that we can never be in a position to declare that we fully understand what we can and must do to realize them.

No description of what is involved in realizing a given ideal is beyond question because every conceptually possible circumstance, and every feature of these circumstances, has some normative implication for how it is good to be -- if only because it does not count against any evaluative judgment. Not only, then, is it impossible for anyone to know everything there is to know about what it is to be a good mother or sister or spouse. It is also impossible for anyone to rule out the possibility that she is mistaken about what a good mother or sister or spouse would do in a given circumstance. For all she can know, there may be a consideration that counts against some aspect of her conception of ideal motherhood. She may have overlooked a decisive reason against her assumption that a good mother would *never do that*.

This imperfect access to our own ideals is something most of us take for granted -- at least when we are not doing philosophy. We understand that in aiming to be a good mother, daughter, sister, wife, friend, neighbor, citizen, philosopher, person, we aim to achieve something an essential aspect of which is that our own conception of it is necessarily indeterminate, imperfect, provisional, incomplete. If someone fails to appreciate that whatever substantive aim she pursues is a provisional and imperfect stand-in for some form of the good, then she is ethically blind, even if she never does anything wrong. This is a second way in which someone can fall short of the ideal experience of acting for reasons.

If in aiming to realize various ideals, our target is necessarily pretty fuzzy -- if it may not even be in the place we are seeking it -- and if in determining what we have reason to do on any occasion, we rely on a plethora of internalized ideals, then acting for reasons essentially involves being motivated by aspirations we only dimly comprehend. This is not only a *common* feature of rational agency. It is a *necessary* feature. To be sure, we often forget the fact that we see certain substantive considerations as reasons only because we have no choice but to rely on approximations of the ideal standards that elude us. But as long as we have the experience as of its making sense to be thus motivated, we implicitly acknowledge the thin, aspirational element at the heart of our actions.

But this means that we must reject an argument David Velleman offers against the conception of acting for reasons as acting "under the guise of the good." According to Velleman, "the good" could not be the aim of action for the simple reason that in aiming at "the good," we would be aiming at nothing at all. "The good", Velleman says, is an "empty" concept. Aiming to do what is good is thus like aiming to capture "the quarry": unless one has some determinate conception of what the quarry is, one has nothing to hunt.

If the preceding observations are correct, then Velleman's comparison is inapt. To be sure, we must form particular, determinate conceptions of the good in order to have something to aim at in action. But satisfying this requirement is compatible with regarding these conceptions as imperfect and provisional -- as approximations of something, we know not what. Indeed, if what I have said is correct, to regard such stand-in ideals as something to-be-realized just is to regard them as open to question.

Again, most of us take this for granted. Thus, for example, when someone's conception of what it is to be a good mother seems to imply that she ought to allow her daughter to wear her pajamas to school, she may find it quite natural to ask herself*:* "But is this really what a good mother would do?" All questions of this sort are *open* questions. Indeed, it seems to me -- though this is an argument for another paper -- that it is precisely because such questions are always *normatively* open that we must reject all reductive accounts of what reasons are. Any conception of what it is to be good can be challenged as long as we take it as a substantive normative proposal.

No similar challenge makes sense if we have decided to hunt for mushrooms. Precisely because what we hunt is a matter of stipulation, if we know what that stipulation was, it makes no sense to ask: "But is this *really* the quarry? Might it not really be foxes instead?" This question makes no sense because the only aspiration involved in hunting for mushrooms is the aspiration of finding *mushrooms*.

If our aspiration were to find the best mushrooms in the forest (the ones it would make most sense to seek), then something "empty" (something thin) would have crept into our goal. We know enough to know that we have failed to achieve this goal if we come home with a basket full of nettles. (This is the unprovisional, domain-fixing aspect of our goal.) But when we are seeking the best mushrooms in the forest, no basket of mushrooms is such that we cannot intelligibly wonder: is that really what we were trying to find? We can, of course, stipulate, that we are looking for the biggest mushrooms, or the ones with the "strongest" taste, or the ones that taste most like truffles. The point is that insofar as this is all we are looking for, we are not concerned about whether this is a good thing to be hunting. Insofar as we are aiming to find the biggest mushrooms *as a way of aiming to be the best mushroom hunters we can be*, the question opens up again: But does gathering these mushrooms really qualify as realizing this aim? To insist that *this* question does not make sense is to be ethically blind. The question makes sense, even though no determinate conception of what would count as a satisfactory answer can possibly close the question. It makes sense *precisely because* such closure is not a conceptual possibility.

Again, this is not to deny that we must have ends that are determinate enough to indicate which steps we can take to achieve them; we must have ends that can serve as a guide to our choice of means. The point, however, is that insofar as we take ourselves to have reason to achieve these ends, our aim in acting is necessarily also

indeterminate enough to represent the distinction between any ideals we might *actually aim* to realize and whatever ideals are really *worth* realizing. It is, in other words, indeterminate enough to represent the distinction between (i) whatever determinate conception we could possibly have of what it would take to realize these ideals and (ii) what it would really take to realize them. In acting for reasons, we are guided by determinate-enough ends, conceived as provisional stand-ins for whatever is really and truly the worthy object of aspiration we seek.

Note that "the quarry" is not an end of this sort. It is not determinate enough to indicate how anyone should go about looking for it. (This is Velleman's point.) Nor is it related to anything of this sort in such a way as to leave open the possibility that the members of a given hunting party are mistaken about whether they are really looking for the right thing, in the right place, in the right way. We cannot aim at finding "the quarry" if this is to aim at finding whatever-it-is-we-*are*-in-fact-aiming-at, under this description. I hope it is obvious that this is not an appropriate model for what goes on when we aim at "the good."

Chapter 3.
To say that our ideals constrain our choices is to say that they determine which facts, or features, we *see as* reasons to do certain things and not others. I have a rough-and-ready sense of what it is to be a good friend. With this rather indeterminate, provisional conception of my ideal as a guide, I assume that in most circumstances I would betray this ideal if I were to make no effort to help a drowning friend. To the extent that I have internalized this ideal, I do not have to draw any inferences in order to *see* my friend's plight *as* a reason to help her. To internalize this ideal just is to see her plight in this way.

As our conception of our ideals becomes thicker, the range of substantive nonnormative facts we can see as reasons increases. But how do our conceptions of our ideals become thicker? There are at least three, closely related, contributing factors.

First, if some ideal *I* really is one of our ideals, then we take there to be some considerations that count in favor of realizing it -- considerations that account for its status as a worthy guide to action. This means that any facts we already see as reasons for or against various responses to our circumstances are constraints on what can count as realizing ideal *I*. If, for example, it seems to me that telling a lie would be a way of realizing *I* under these circumstances, but I am also convinced that there are good reasons *not* to tell a lie under these circumstances, then it seems to me that I have reason to rethink m*y* conception of what is involved in realizing ideal *I.*

This basis for revising and refining our ideals is closely related to another: because we have many different ideals, the aims associated with each one impose limits not only on what we can do to realize the others, but also, more specifically, on what we must do to realize each one. If, for example, I have the ideal of being a good mother,

then if I also have the ideal of being a kind person, this second ideal will play a role in my understanding of what a good mother would do when her son shows her his art project.

A third factor that interacts with the others is a factor to which I have already alluded: our circumstances are constantly changing. My indeterminate conception of what it is to be a good mother may not spell out what a good mother would do when her daughter insists on wearing her pajamas to school; or it may seem to imply that I ought to do something which, given my other ideals, I am pretty sure I ought *not* to do. In either case, in confronting this novel circumstance, I am forced to reconceive what it is to be a good mother.

Given that our ideals start out pretty thin and that we are open to revising them in light of each other, there is some reason for us to hope that we can realize the meta-ideal of realizing all our ideals. But given that not just anything could count as being a good mother, philosopher, ballet dancer, or lion tamer, there is also reason to be skeptical. This is not simply because we are finite, mortal beings who cannot play arpeggios while juggling five balls on a high wire. The more interesting, and deeper, reason why so many of our ideals force us to give up the ideal of realizing them all is because the thick ingredients of each ideal are not arbitrary. To be sure, we must be prepared to change our minds about what counts as realizing a given ideal. But there must also be limits on what can count. There must be substantive constraints internal to each ideal -- impossible though it may be for us to say what they are. If we discover overriding reason to ignore these constraints, we will have discovered that we do not hold the ideals after all.

I find it quite puzzling how our ideals can constrain us, even as we are open to reconceiving what they require. It seems to me, however, that at least some of the normative pressures against *abandoning* an ideal are also pressures against *reconceiving* it. In particular, as long as we can discover no reason to repudiate the goal we associate with being good in a particular way, and as long as the features we identify with being good in this way appear to be conditions someone must satisfy in order to achieve this goal, nothing will count for us as an adequate thickening of our conception of this ideal if it does not include these features. If, for example, we are convinced that there are overriding considerations in favor of the assumption that no good mother would be indifferent to whether she cares for her child, and if we also believe that neglecting to give someone the food she needs to survive is no way to care for her, then we cannot endorse any alleged conception of a good mother if it implies that a good mother is someone who withholds food from her child.

It is quite likely that constraints of this sort are built into most, if not all, of an agent's ideals. If this is the case, and if some of the relevant ideals seem to ground conflicting requirements, then the agent might well take herself to have insufficient reason to revise either one in such a way as to eliminate this conflict. Indeed, she might reasonably doubt that such revision is possible. What then? It might seem

that she could give up one of these ideals -– or replace it with something less demanding in the relevant respect. But this might not be possible: after all, these *are* her ideals; she really does want to be good in these ways. If at least one of these ideals -- say the ideal of being brave, or being a good painter-- has not yet been incorporated into her identity -- if she has only just begun trying to see things the way a brave person, or a good painter, would, and to respond accordingly -- then replacing these ideals with something else may not be so difficult: she now wants to be brave under circumstances where the personal costs are not too high, or a lover of visual art who dabbles a bit just for fun. But what if she is already a person who, according to her own standards, is brave -- and a pretty good painter too? Can she just decide not to see her reasons for action in light of these ideals? More importantly, is this something she necessarily has reason to do? It seems to me that the answer is "no." It is sometimes intelligible -- and even reasonable -- for someone to endorse two ideals, even though she realizes that she may not always be able to satisfy the requirements grounded in each. In short, given the nature of ideals, a person need not be confused if it seems to her as if she faces a moral dilemma**.**

Chapter 4.

The possibility of moral dilemmas is the possibility that some apparently irresolvable conflicts among moral requirements really cannot be resolved -- that there is no way to eliminate these conflicts by introducing further exceptions, qualifications, and other refinements. If moral dilemmas are possible, then it is possible for an ideal moral agent to wittingly and deliberately do something wrong. In other words, it is possible for a morally good person to rightly believe that she is doing something wrong, even as she rightly believes that she has no more justifiable alternative.

I want to challenge the resistance to this possibility by focusing on the thick, descriptive aspect of our ideals. It is not just that a description of a well-working train engine cannot possibly be a description of a good mother; a description of someone who does an excellent job of whipping her child also falls far short of the mark. Though we can never acquire a complete understanding of what someone must do in order to be a good mother, we have at least a dim grasp of what we wish to understand better. We thus face constraints on which alterations in our conceptions reflect greater insight into the specific sort of goods they aim to describe.

Since these constraints come from the ideals themselves, there is no guarantee that concerns external to any given ideal can provide sufficient justification for revising our understanding of what is required to realize it. This means, in particular, that such a revision might not be justified, even if there is no other way to avoid facing irreconcilable requirements. And this means that nothing rules out the possibility that some of the ideals we have good reason to endorse are sufficiently thick to ground requirements that render us vulnerable to moral dilemmas.

To see the important role that thick elements play in determining the distinction between right and wrong, it may help to consider a case in which (i) an agent has *overriding reason* to violate a given requirement, and (ii) this requirement can plausibly be interpreted as spelling out what she must do, under even these circumstances, if she is to be good *in a certain way*. Imagine, for example, that the relevant way of being good is being kind. Plausibly, this is not compatible with telling a small child how disgusting and stupid and boring and ugly and useless she is, calling her names, taking her prized possessions, locking her in a small dark room, and ignoring her every appeal for help and affection. Yet, surely, one is morally permitted -- even obligated -- to behave this way if doing so is necessary to save the child's life.

Those who reject the possibility of moral dilemmas are committed to one of two interpretations of this situation: they must say that, under these special circumstances, this child-tormentor is not really behaving unkindly; or they must say that, though she is behaving unkindly, it does not follow that there is anything morally problematic about what she does. On both interpretations, though the person's behavior certainly has features that normally, generally, typically are morally problematic, this is the full extent of their moral significance in this particular case. Since, moreover, her behavior has no further moral significance, she has no reason to regard herself as falling morally short. That she takes herself to have violated a moral requirement may be a commendable indication of how deeply committed she is to avoiding wrongdoing. If so, it is a commendable mistake.

Is this really the best way to interpret the situation? What, exactly, is the cost of insisting that the person who behaves as described would be right to think that she has treated the child unkindly, and that in behaving unkindly she has not only harmed the child, but wronged her? Even if, as I earlier noted, a novel circumstance can enable an agent to gain insight into her ideals, nothing follows regarding whether someone is justified in revising her ideals in light of what she learns on a given occasion. In particular, from the mere fact that a novel circumstance reveals a conflict among the requirements that are grounded in someone's ideals, it does not follow that the ideally rational response is to revise at least one of these requirements. This does not follow because revising at least one of these requirements in the way necessary to prevent the conflict might not do justice to the ideal the requirement spells out, and because it might be quite reasonable for the agent to have internalized this ideal. Thus, for example, from the mere fact that someone is unfortunate enough to find herself in a circumstance in which it is not only permissible, but obligatory, to treat a child in the way just described, it does not follow that, under these circumstances, she can treat a child in this way without treating her cruelly, and thereby violating a moral requirement. So, too, if this person has *two* children, and if she must torture one of them in order to save the other, the mere fact that she has sufficient reason to do this need not itself be a sufficient reason for her to conclude that she is not really violating any duties of motherhood -- that under these special circumstances, she can be a good mother to

her torture victim because, under these circumstances, she has no morally better alternative.

One needs reasons to revise one's conception of what is required in order to be kind or modest or morally good; and not just any reasons will do. In particular, and trivially, if neither of the ideals that underlie two conflicting requirements provides a justification for altering either requirement in a way that eliminates the conflict, then this fact is not itself a reason to alter one of the requirements. To say that one cannot avoid doing something wrong under these circumstances is to say that one cannot avoid violating at least one of the requirements grounded in at least one of one's ideals. In contrast, to say that one is nonetheless justified in acting this way is to say that under the circumstances, one lacks sufficient reason to do otherwise – even if moral reasons are overriding.

This second judgment reflects a conceptual constraint on rational agency: an ideal rational agent never does anything she cannot possibly do because, trivially, doing something requires the capacity to do it. My point is that respecting the implications of this conceptual necessity is compatible with being guided by the more substantive standards associated with our personal ideals. This is what we acknowledge when we say that, under special circumstances, it is morally permissible to behave cruelly, or to do something else morally wrong. We acknowledge that what we are required to do in order to avoid violating any moral standards is sometimes what we are not required to do, all things considered.

According to those who reject this possibility, a person is always justified in revising a worthy ideal if she must do so in order to prevent the associated requirements from conflicting with those of another worthy ideal. What could this justification be, if it must override the reasons that support each of the conflicting ideals, as she conceives them? In this sort of case, the only justification available is that she cannot comply with both requirements in all circumstances. If, however, a person's only reason for revising her understanding of what she is morally required to do is the very thin reason that otherwise an ideally moral agent could knowingly act wrongly, then to recommend that she revise her understanding in this way is to recommend that she adopt the stance of a moral fetishist. It is to suggest that whenever it seems to her as if she faces a moral dilemma, she ought to believe that what she is really required to do is to comply with only one of the two apparent requirements. She ought to believe this, even though none of her moral ideals provides a rationale for this disjunctive requirement, and even though she is thus unable to see this requirement as spelling out a condition sufficient for treating all human beings with concern and respect.

Here, then, is what follows if the ideal moral agent is someone who is invulnerable to moral dilemmas. If the ideal moral agent has ideals -- and how could she not? – and if she has acquired these ideals as her life unfolded -- and what is the alternative? -- then if she is invulnerable to moral dilemmas, she may sometimes find it necessary to include among the premises in her reasoning the proposition

that anyone who cannot avoid doing A or B is morally required to do only A or B. There will be occasions on which she is forced to add this requirement to her premises because there will be occasions on which though she appreciates that she has sufficient reason not to do A or not to do B, she is powerless to see either omission as morally unproblematic. Precisely because she cannot see either omission as beyond moral reproach, she has no choice but to think an extra thought from which she can derive this conclusion.

On this conception of the ideal moral agent, there are circumstances under which this agent will fall short of the experiential ideal of rational agency. Worse still, she will experience her action as that of a morally bad person, even as she is confident that there is nothing morally problematic about what she is doing. From what I can tell, this is a coherent ideal. But do we really want to endorse it? Surely, whatever attraction it has can be traced to an ideal of rationality: ideally, the order of reasons would not be such as to yield requirements that cannot be reconciled with each other. I endorse this regulative ideal. I believe, in particular, that we should do what we can to revise our conception of what it is to be a morally good person so that it implies that this ideal can be realized. If I am right, however, our ideals themselves can impose limits on what we can do to realize this meta-ideal.

If this is, indeed, possible, then we must not only acknowledge the possibility of moral dilemmas; we must also rethink our conception of the ideal rational agent. If the thick aspect of our ideals is an impediment to unifying them all, then in fleshing out our ideal of rational agency, we seem to have two options: (1) we can concede that there are occasions on which an ideal rational agent will wittingly violate a requirement, even while acknowledging its status as a requirement, or (2) we can concede that an ideal rational agent would avoid acquiring ideals that render her vulnerable to being pulled in more than one direction at once. The latter ideal seems as unattractive as it is unattainable. Can it really, then, be *our* ideal of rational agency?

Chapter 5.
Having explored the possibility of moral requirements whose conflicts we cannot resolve, I want now to turn, in this final chapter, to conflicts between moral and nonmoral requirements. Many philosophers and nonphilosophers are attracted to the view that if it is not *morally permissible* to do something, then we do not have *sufficient reason* to do it. Others argue that the reasons allied with certain nonmoral ideals -- reasons of prudence, or whatever reasons are tied to living the most "fulfilling" or "interesting" or "meaningful" life one can live -- are no less fundamental than reasons of morality, and that when these reasons support conduct that is forbidden by moral considerations, there is no rational basis for resolving the conflict. I want to suggest that our investigation into the nature of ideals helps us to see that both sides in this debate are right about something. My secondary aims are, first, to explain why it is natural to regard moral imperatives as the imperatives of "pure reason," and second, to consider what the possibility of conflict between our

moral and nonmoral ideals implies about the relation between the ideal rational agent and the perfectly coherent agent.

The ideal of being a morally good person is, very roughly, the ideal of treating human beings with concern and respect. Call this "the moral ideal." It is natural to think of the moral ideal as the meta-ideal of realizing two distinct ideals. At the same time, however, anyone who aims to treat others with "concern and respect" is under pressure to regard each ideal as an aspect of a unified whole. We cannot understand what we must do in order to treat others with respect without understanding what we must do in order to treat them with concern, and -- conversely -- thickening our conception of someone who gives the interests of others their due requires that we thicken our conception of someone who respects the capacity of others to set and pursue their own ends. This is how we acquire a more determinate idea of what is involved in assigning other people a "proper" role in our lives. So this is the way we acquire a more determinate idea of what it is to be a "morally good person."

To aim at being a morally good person involves committing oneself to giving others a proper place in one's life. But what role it is proper for others to play in our lives depends on what counts as treating others with concern and respect, *given that it is a good thing for us to promote and protect our own long-term interests, raise families, maintain friendships, cultivate our talents, and realize any number of other ideals*. In short, the indeterminate moral ideal of "treating others with concern and respect" gains determinacy only insofar as it accommodates other ideals. Of course, not any accommodation will do. The goal is to accommodate our other ideals in such a way as to do justice to the conditions we must satisfy -- whatever these may be -- if we are to treat other people with concern and respect.

The moral ideal is an ideal of accommodating other ideals because to be a morally good person is to treat others "properly," "appropriately," to give them their "due," and because this means that the task of gaining a more determinate conception of the moral ideal is the task of gaining a more determinate conception of which accommodations we *ought* to make to others, where this requires determining the relative significance of other action-governing norms. In short, the moral ideal becomes more determinate by becoming responsive to a wider range of considerations -- a wider range of reasons. And this means that -- though we cannot forget the possibility of moral dilemmas -- it becomes more determinate by approximating what we have all-things-considered reason to do. This is why it makes sense to think of moral demands as the demands of reason: they are the demands that are grounded in an ideal that has already taken into account the reasons that are tied to our other ideals.

In contrast, the imperatives grounded in the ideal of living an interesting, or otherwise aesthetically ideal, life do not reflect any accommodation with other ideals. To be sure, my life might be more interesting if I were to pursue a wide variety of goods; and if this were the case, then whether a given action would

contribute most to my living an interesting life might well depend on what is required in order to realize various other ideals. But this would not be because what it is to be good in one of these other ways imposes a constraint on what it is to be interesting or beautiful or sublime. The point is simply that variety might be an essential component of an aesthetically good life.

More generally, even if we cannot understand what is required in order to realize a given ideal without understanding what is required in order to realize a second ideal, the first ideal need not be "accommodationist" in the sense I have just introduced; understanding the first ideal need not require understanding how it accommodates the independent requirements identified with the second. Consider, for example, instrumental relations. It is difficult to realize the ideal of being a good athlete if one smokes two packs of cigarettes every day. But this does not mean that what it is to be a good athlete is to be an athlete in a way that properly accommodates the ideals of health, or prudence. Rather, the point is simply that in order to be a good athlete, one must be guided by these other ideals too.

The same point applies to the relation of inclusion, or constitutive means. In order to realize the ideal of being a good dentist, one needs to realize the ideal of not being a clumsy person. But this relationship simply follows from what it is to be a good dentist: it is because of what it is to be a good dentist that one cannot realize this ideal if one handles one's instruments in a clumsy manner. To aim at being a good dentist is not to aim at being a good dentist *in such a way that one accommodates the independent ideal of not being clumsy*. The ideal of not being clumsy does not impose a constraint on what it is to be a good dentist; rather, *having the ideal* of being a good dentist ensures that one also *has the ideal* of not being clumsy.

[It is perhaps worth stressing that an "accommodationist" ideal need not, and could not, qualify as such in relation to every other ideal. If, for example, someone believes that it is not possible to be morally good without being brave, this is not a matter of her conception of the first ideal accommodating her conception of the second. In believing that she cannot realize the first ideal without realizing the second, her point is *not* that a morally good person is someone who allocates a proper place in her life for an ideal that threatens to pull her in a different direction. (She does not believe that to be a morally good person is to treat others in the way they ought to be treated, given the independent value of finding opportunities for being brave.) Though in order for a morally good person's conception of what it is to be just and benevolent to gain determinacy, she must accommodate this conception to her commitment to being prudent, and to living an interesting life, the unity constituted by her virtues of justice and benevolence and her virtue of courage is not itself the product of accommodation.]

As far as I can tell, the ideal of prudence resembles the ideal of living an aesthetically admirable life: at least as it figures in philosophical discussions**,** prudence is not an accommodationist ideal. It is, of course, prudent to do the morally right thing in a wide range of circumstances. But this is simply because doing the right thing is

often the best means of promoting one's interests. If a proper conception of what it is to thrive implies that we do not promote our own interests unless we treat others with concern and respect, this is not because what it is to be prudent is a function of what counts as pursuing one's self-interest, *given that one has reason to treat others with concern and respect*. The "thin" aspect of the ideal of prudence is the "whatever is good for me" aspect. What one must do in order to be prudent does not depend on what is required for one's good-for-me conduct to be good in some other ways too.

[Note: The fact that aesthetic ideals resemble the ideal of prudence in this way explains the sense in which the selfless dedication to art is nonetheless more "selfish" than the self-interested commitment to being the best doctor one can possibly be. Dedicating oneself to art is indifferent to moral constraints in a way in which dedicating oneself to medicine (whatever one's motives) is not; and this is because only the latter involves dedicating oneself to an ideal that gains determinacy by accommodating such other ideals as kindness, honesty, fairness, etc.]

Because the permissions and requirements the ideal of prudence underwrites are independent of any extra-prudential considerations that count for or against them, these permissions and requirements do not have a claim to be the permissions and requirements of reason. Accordingly, when the demands of prudence clash with the demands of morality, anyone who has internalized both ideals is committed to regarding the moral demands as overriding. The moral demands are overriding in this case not because they are the demands of (pure) reason, but because insofar as they are determinate enough to serve as guides to action, they reflect a reasoned accommodation to the demands of other ideals, including most importantly, the ideal of prudence. Since the moral ideal gains determinacy by accommodating other ideals, anyone who has internalized both ideals necessarily gives priority to the demands of morality. In determining what is required in order to treat others with concern and respect, a person has already given prudential considerations their due.

But what if someone has *not* internalized both ideals? It is important to stress that if such a person is wondering whether she has reason to comply with a moral requirement, she cannot settle the matter by appealing to the story I have just told about why moral demands are overriding. Nothing I have said suggests that the moral ideal trumps the ideal of prudence. Nothing I have said implies that moral reasons are reasons that apply to every rational agent. Of course, there could well be compelling nonmoral reasons for each of us to care about being a good person. The point is that, for all I have said, these reasons (including prudential reasons) are no less basic than whatever moral reasons there may be for behaving prudently. In the final pages of this paper, I want to pursue this issue a bit further, in the context of taking a closer look at the relationship between the ideal rational agent and a perfectly coherent rational agent.

As long as at least some of a person's ideals are not of the accommodationist variety, her psyche lacks perfect unity. So, in particular, as long as she has internalized both the ideal of prudence and the moral ideal, there will be a lack of unity in her motivational structure: there will be occasions on which she cannot live up to *at least* one of her ideals. Insofar as someone in this condition puts a high premium on unity, she will do what she can to come closer to realizing this meta-ideal. She might, for example, revise her conception of what it is to be prudent -- or, if this is different, replace her initial ideal with an accommodationist version of prudence. From what I can tell, this is a pretty popular move. Few people really regard the maximal promotion of self-interest as, in itself, an ideal. Most of us do not call someone "imprudent" whenever she sacrifices her self-interest in order to help someone in need. Someone's behavior is "prudent," we say, as long as it reflects a *reasonable* degree of self-interestedness. By thus incorporating an appeal to nonprudential reasons into our ideal of prudence, we take a significant step toward unifying the virtues. (So, too, a dentist endorses an accomodationist conception of the ideal dentist insofar as she assumes that to be good of her kind is to be good in a way that does not prevent her from treating others with concern and respect.)

It is interesting to contrast this typical approach to the ideal of prudence with our typical approach to aesthetic ideals. Rather than replacing these ideals with an accommodationist version, people typically content themselves with making the accommodations *themselves*. If on some occasion a poet or painter cannot realize the ideals of her art without violating a moral requirement, then, assuming that she wants to be a morally good person, she will probably decide to violate the aesthetic requirements instead. More generally, most artists (though not all) accept moral limits on their artistic endeavors. Rather than regarding the moral ideal as relevant to what counts as aesthetically good, they regard it as a constraint on whether, and if so, by what means, they have reason to realize this good. To this extent, their aspirations lack unity. The same is true of the rest of us who aim to live interesting, and even exciting, sublime, and beautiful lives.

I have noted that when someone's ideals lack unity in this way, she falls short of a meta-ideal. In the context of the preceding remarks, it is worth considering the suggestion that this meta-ideal is itself an aesthetic ideal. Whatever we make of this suggestion, it prompts us to consider what place the ideal of motivational unity ought to have in our lives. Is there an alternative meta-ideal according to which being torn between aesthetic and moral ideals is a better way of being human? Must we answer these questions in order to understand what it is to be ideally rational?

To forego an interesting or sublime experience because it would require one to do something wrong is not only to sacrifice something of value; it is to be to some extent torn. In evaluating the costs and benefits of being thus disunified, it is important to note that -- leaving to one side whatever emotional distress this condition may directly provoke -- the costs do not include being vulnerable to living a worse life than one otherwise would under the circumstances in which one finds oneself. For consider. If the moral ideal is one's own ideal, then in treating it as a

guide to action, one does not take oneself to be sacrificing a better life when one decides to act rightly, rather than to do what is interesting or self-promoting. If someone endorses the moral ideal, then she necessarily takes herself to have overriding reason to do what she can to realize it. For she necessarily regards this ideal as the best possible expression of the normative relations among her ideals. But if doing what one can to realize the moral ideal is doing what one can to do justice to the normative relations among one's ideals, then doing what one can to realize the moral ideal is doing what one can to live what -- by one's own lights -- is the best life one can live, given the cards one has been dealt. The life of a morally good person may not be the most pleasure-filled life, or the most intellectually challenging, or the longest and healthiest. The point is simply that, given the heterogeneity of a morally good person's ideals, the best life she can possibly live (the best life *for her*) is likely to be a life in which she must sacrifice many things she values for their own sake.

Of course, someone who does not embrace the moral ideal will see things differently. According to the amoralist, if a person refrains from realizing her nonmoral ideals for no reason other than that doing so lacks moral support, then she is not living the best life she could live. Is it possible to adjudicate between the moral and the amoral agent on this point? What does the answer to this question suggest about the ideal of rational agency?

To answer these questions, we need to consider whether we could possibly render the ideal of rational agency just thick enough to serve as a standard for evaluating the two competing conceptions of living well, but not so thick that it incorporates these conceptions themselves. The prospects for this sort of strategy do not look very good. One might think, for example, that in order to evaluate the competing claims of the moral and amoral agents, an extremely thin conception of rational agency will do. One might, for example, think we can appeal to a conception of the ideal rational agent according to which someone realizes this ideal if and only if she acts only for reasons that others can rationally accept. But can the ideal, so thinly conceived, really do the job? In order for it to serve as a standard for adjudicating between the moralist and the amoralist, we will, it seems, need to add a bit of thickening into our conception of what others can "rationally accept." And isn't it an open normative question whether we have reason to prefer the thickening provided by the moral ideal over that provided by prudence alone?

I mean this question to sound rhetorical. But I am not sure whether that is the right way to hear it. This is in part because it seems to me that it would be extremely difficult for someone to live among other rational agents without acquiring some accommodationist ideals. To take it for granted that one shares the world with others is to be caught up in ways of being that thrust on one the opportunity of being more or less good in these ways. One discovers, for example, that one is someone's daughter. But this means that one is in a position to be a more or less *good* daughter. Could someone really be indifferent to the possibilities this reality suggests if she cares about whether her conduct can be justified? Isn't such a person

committed to conducting herself as a daughter has reason to conduct herself? But to do this, mustn't she form an action-guiding conception of what it is to be a "good daughter?" And can she do this without determining what is required in order to treat her parents properly? Can she flesh out this ideal without considering how it is reasonable to respond to their interests and ends? And can she do *this* without considering how it is proper to respond, *given that she has her own interests and ends*?

This train of thought yields an apprehension of reasons that are invisible to anyone who has no accommodationist ideals. So it yields a conception of the ideal rational agent according to which the pure egoist and pure aesthete fall short. Can this conception become more determinate without incorporating a conception of the good moral agent? Despite what so many moral philosophers tell us, we know that it can: whether someone sees all other sentient rational agents as proper objects of concern and respect depends, in large part, on various contingent features of her social world.

We are all familiar with the philosophers' favorite response to worries about reason's capacity to free itself from the constraints imposed by contingent features of a reasoner's psychology, history, or "way of life": they wheel in the ideal of coherence. We should, however, be wary of this response. Where coherence is not simply a necessary condition of agency, it is a formal norm, indifferent to what coheres with what. It thus provides no basis for endorsing the moral ideal. Perhaps more importantly (and to return to a question that has been with us since the end of the previous chapter), it is far from obvious that maximal coherence is compatible with living a full life. Faced with a choice between the ambivalences prompted by competing ideals and a psychic unity whose perfection consists in the absence of any such competition, which option would an ideally rational agent pick? If we, ourselves, cannot help seeing reasons that correspond to competing ideals, then we ourselves believe that an ideally rational agent would share our vision. It seems, then, that, as far as we can tell, an ideally rational agent would not wholeheartedly endorse her actions in a wide range of circumstances. Of course, we could be mistaken. If the thick aspect of our ideals makes it difficult for us to take this possibility seriously, the thin, aspirational, aspect leaves the question wide open.

On The Justness of Defensive Wars

It is widely believed that some wars are just, and some unjust, and that the justice of a war depends on the justice of the cause. The defense of sovereignty, understood as the rights of political independence and territorial integrity, is commonly accepted as the paradigm case of a just cause. And so while the UN Charter generally forbids "the threat or use of force against the territorial integrity or political independence of any state,"[1] it provides the notable exception that "[n]othing in the present Charter shall impair the inherent right of individual or collective self-defence if an armed attack occurs against a Member of the United Nations."[2, 3] A state may fight an aggressing state in self-defense because its sovereignty is being threatened.[4]

What makes a defensive war a just war? It cannot simply be that such wars save lives, for a state might save the lives of its citizens by surrendering. There is radical uncertainty, before the fact of the war, about whether more lives might be saved than lost by prosecuting a defensive war. Nor can it simply be that sovereignty is so excellent a goal that it somehow overcomes the presumption against killing, for we recognize many

---

[1] http://www.un.org/en/documents/charter/chapter1.shtml

[2] http://www.un.org/en/documents/charter/chapter7.shtml. In this article, "individual" refers to an individual state, "collective" refers to more than one state. *See, e.g.,* Dinstein, War, Aggression and Self-Defence, Fourth Edition (Cambridge Univ. Press, 2005) at p. 252. I'll use "individual" to refer to individual people, and talk about "individual self-defense." I'll use "defensive war" to refer to killings authorized by a state in its defense.

[3] For the purposes of this paper, I'll consider only the limited case of traditional wars covered by Arts. 2(4) and 51, that is, wars fought by armies under the authority of states. There is much debate about how the Charter bears on civil wars, aggression by guerrillas, and humanitarian intervention or peace-keeping by troops under transnational or international authorities, as well as cases of genocide, enslavement or crimes against humanity, and I hope to consider some of these problems in future work.

[4] *See, e.g.,* Dinstein, at p. 177, explaining that "[t]he provision of Article 51 has to be read in conjunction with Article 2(4) of the Charter."

excellent goals that cannot overcome the presumption. A state may not declare war on another state in order to redistribute their food to feed their own starving population.

In order to understand why a defensive war is a just war, we will have to try to understand sovereignty in a different way, as something that you can fight for in a way that's explained by the same moral considerations that underlie the permission for individual self-defense. In the just war tradition, we find a strong conceptual link between the permissibility of individual self-defense and defensive war. A common strategy used to justify defensive war is to infer its permissibility from the individual case. So, for example, Walzer argues that "territorial integrity and political sovereignty can be defended in exactly the same way as individual life and liberty."[5] The state protects the community that the individuals have made together, and this is "why we assume the justice of [its] defensive war[]."[6] States, like individuals, are rights-bearers,[7] and if individuals are permitted to kill in defense of (some of) their rights, then so may states.

In considering the question of what, if anything, makes a defensive war just, I'll adopt a similar strategy – I'll begin by developing what I find to be the most compelling account of the permissibility of individual self-defensive killing, and then show how that very same permission might justify defensive war. I'll then argue for the permissibility of pacifism in the individual case, and show how pacifism could be a permissible alternative to fighting a defensive war. Walzer argues that citizens who are aggressed against by another state are forced "to risk their lives for the sake of their rights," and "in

---

[5] Just and Unjust Wars, Fourth Edition, Basic Books (2006) at p. 54.
[6] *Id.*
[7] "The Moral Standing of States: A Response to Four Critics," *Philosophy and Public Affairs*, vol. 9, no. 3 (Spring, 1980), pp. 209-229 at p. 212.

most cases, given that harsh choice, fighting is the morally preferred response."[8] Although I am skeptical of the claim that any war could be a just war, I will not try to argue here that a defensive war could never be justified. What I do hope to show is that pacifism is a real moral alternative, and if that is true, then I think it becomes less clear that fighting is the morally preferred response.

I. The Problem of Self-Defensive Killing

Why is killing in self-defense permissible? It cannot be enough that it's your life or mine. This is intuitive enough – if you and I are adrift in a life raft with only enough supplies for one person, I cannot permissibly throw you overboard that I might have all the food and water for myself and live.

To help us get started thinking about the problem, consider the following scenario presented by Thomson.

> *Case Innocent Threat*: You are lying in the sun on your deck. Up in the cliff-top park above your house, a [] man is sitting on a bench…. A villain now pushes the [] man off the cliff down toward you. If you do nothing, the [] man will fall on you, and be safe. But … if he falls on you, he will squash you flat and thereby kill you. If [you shift the position of your awning] the [] man will be deflected away from you… down onto the road below.

You may shift the awning, and this is because unless you kill him, the falling man will violate your right not to be killed. Neither fault nor agency is relevant to the question of whether your right is about to be violated, and so neither is relevant to the question of whether you may kill aggressors and threats.[9]

---

[8] Just and Unjust Wars, p. 51.
[9] "Self-Defense," *Philosophy and Public Affairs*, vol. 20, no. 4 (Autumn, 1991), pp. 283-310, at p. 301-2.

Perhaps it's true that the falling man will violate your right not to be killed by him, perhaps not.[10] But if we're uneasy about the conclusion that it's permissible to deflect the falling man to his death, I think it would be worthwhile to consider the possibility that the killing intention of the aggressor does matter. In the next section, I'd like to turn to Barbara Herman's view, according to which the killing intention of the aggressor matters, and it matters because it gives my own violence a character it wouldn't otherwise have had – namely of resistance or self-respect.[11]

II. The Agential Solution

According to Herman, what's wrong with aggressive killing cannot rest solely on the fact that the victim dies, for dying is part of what it is to be human. What's wrong with aggressive killing has to do with the maxim under which the aggressor acts. When the aggressor decides to kill the victim, whether it's because she wants the victim's wallet or because the victim stands between the aggressor and some other goal she has, the aggressor treats the victim as something to be used and destroyed for the purpose of securing the aggressor's private end. She treats the victim as a mere means.

Acting on such a maxim is incompatible with recognizing the victim as a rational agent, and seeing her as an end in herself. It is not possible to act on such a maxim and at the same time recognize those features that characterize the limits of our powers as human agents – that we are physically vulnerable, mortal, and need the help of others.

---

[10] *See, e.g.,* Otsuka, "Killing the Innocent in Self-Defense," *Philosophy and Public Affairs*, vol. 23, no. 1 (Winter, 1994), pp. 74-94 at 79-84 (rejecting the claim that the innocent threat will violate your right not to be killed by him by killing you).

[11] "Murder and Mayhem" from The Practice of Moral Judgment, Harvard University Press (1993).

As human agents, our lives are a necessary condition for the continued exercise of our agency. We must take the fact of a life as a reason not to destroy it.

The aggressor who treats the victim as a mere means fails to recognize the victim as a rational agent, and so fails to correctly value the victim's life. Because of this mistaken valuation, the aggressor dismisses the victim's life as a reason not to destroy it, and tries to use the victim and her death for her own private purposes. What is it to fail to respect agency, and so fail to take life as a reason not to destroy it? The aggressor is deciding for the victim what should be done with her life only in terms of the aggressor's own life.

For the victim to fail to resist the aggressor who acts on the impermissible maxim would be for her to go along with the aggressor's plan to use her as a mere means. And she cannot allow herself to be used in this way. The victim must resist because she must not be complicit in her own subjugation.

As in the case of the aggressor, the moral character of the victim's action can be determined by the maxim under which she acts. The victim must respond to the aggressor by acting under a maxim of resistance. By acting under a maxim of resistance, the victim is "asserting [her] status as a rational agent.[12] In some cases, resistance might involve fighting back and possibly killing the aggressor. In other cases, though, it might not be possible for the victim to fight back because, for example, she might be physically restrained or because if she attempts a physical defense, she might kill innocent bystanders. But even in such circumstances, the victim can still act under a maxim of

[12] *Id.* at p. 130.

resistance, in part by recognizing that the aggressor is impermissibly discounting her agency and condemning the aggression.

Because the aggressor is a rational agent, I still owe him respect as an agent, and it is by "limiting my action where possible [that] I demonstrate the moral regard he is still owed."[13] The requirement of proportionality of response requires that the victim limit her counter-violence to what is necessary to defuse the threat. She cannot use more counter-violence than she thinks is necessary to defend her agency, since any excess violence cannot be justified as a necessary defense of her agency.

So the maxim of resistance is not a blanket permission to kill. Consider the following case.

> *Case Innocent Bystander:* An aggressor is trying to kill her victim. The victim can neither deflect the threat nor retreat. The only way she can stop the threat is by killing the aggressor. But to kill the aggressor, the victim will also have to kill an innocent bystander ("Innocent").

Thomson and Herman agree that the victim cannot kill Innocent in order to save her own life, because to do so would be to use Innocent as a mere means. But it is not immediately clear in what way the victim would be using Innocent as a mere means. After all, it's not like the victim is pushing Innocent into the path of a bullet intended for the victim, or throwing her onto the tracks to stop a trolley from crushing five people to death. To kill Innocent in order to deflect the lethal threat would be to use her as mere means because the victim would be treating Innocent's life as merely part of the causal story that will save (or promote) her own life; she is not reasoning about what to do while recognizing Innocent as an end in herself. And this is exactly what makes aggressive killing wrong.

---

[13] *Id.* at p. 130.

Acting on a maxim of resistance, the victim must only use as much violence as is minimally necessary to defuse the threat, *and* must restrict her actions as required by other regulative maxims and concerns.[14] In Case Innocent Bystander, the victim's killing of Innocent would be opposed by other moral reasons. What the victim owes the innocent bystander in that case is serious enough to make it the case that she should not fight back. When the victim deliberates about what to do, she's not weighing the value of her life against the value of the aggressor's and Innocent's lives. The value of human lives is not merely additive, such that two lives are more valuable than one.[15]

Refraining from doing what will kill Innocent constitutes good resistance. The victim's failure to land a lethal blow against the aggressor in that case does not make her complicit in her own subjugation. What qualifies as good resistance will depend on the exigencies of the particular case. In general, where a victim finds herself in a situation like Case Innocent Bystander, she will count as resisting even if she does not do what will kill the aggressor where she (a) recognizes the aggression as impermissible, (b) condemns the aggression either silently or out loud, and (c) decides to limit her violence against the aggressor for the sake of Innocent.

---

[14] *Id.* at p. 130.

[15] *See, e.g.,* Taurek, "Should the Numbers Count?" *Philosophy and Public Affairs*, vol. 6, no. 4 (Summer, 1977), pp. 293-316. Considering the question of whether you may kill one to save five, Taurek writes, "It seems to me that those who … would have me count the relative number of people involved as something itself of significance, would have me attach importance to human beings and what happens to them in merely the way I would to objects which I valued." But "it is the loss *to the person* that I focus on … It is the loss to the individual that matters to me, not the loss of the individual." (p. 307.) *But see* Gregory S. Kavka, "The Numbers Should Count," *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition*, vol. 36, no. 3 (October, 1979), pp. 285-294, arguing Taurek fails to show that numbers *shouldn't* count; *and* John T. Sanders, "Why the Numbers Should Sometimes Count," *Philosophy and Public Affairs*, Vol. 17, No. 1 (Winter, 1988), pp. 3-14, criticizing Taurek for failing to see the significance of the distinction between a loss *to* a person and the loss *of* a person.

III. Political Community and an Agential Case for Defensive War

Now that we see what a more promising account of the permissibility of self-defensive killing might look like, I'd like to turn to the state case and see how such an account might bear on the permissibility of defensive war. Consider the following scenario: State Alpha announces that it is taking over the entirety of State Omega; Omega, as a state, is over. Alpha has no unconditional plans to kill anyone; as long as Omega surrenders to the annexation, no lives will be lost.

We have reason to doubt that the justification to prosecute a defensive war comes from the fact that fighting back will save lives. For in the scenario we are considering, it will be *by* fighting a defensive war that Omegan lives will be lost. We saw in the individual case that the justification for killing an aggressor in self-defense comes from the fact that the individual must not be complicit in her own subjugation by the aggressor, and fighting back against the aggressor is how she asserts her agency. We need to find a similar value on the state side. Walzer suggests that "[w]hen states are attacked, it is their members who are challenged, not only in their lives, but also in the sum of the things they value most, including the political association they have made. We recognize and explain this challenge by referring to their rights. … How these rights are themselves founded I cannot try to explain here. It is enough to say that they are somehow entailed by our sense of what it means to be a human being."[16]

So what kind of political association have the Omegans made? The Omegans have decided that they will make important decisions with each other about how they will

[16] Just and Unjust Wars, pp. 53-54.

live together, including how they will educate their children, protect against rights violations and peacefully resolve disputes, create public spaces, and satisfy the basic needs required to live a decent human life. These social and political institutions are an expression of the community's values, and connect the present Omegans to the past and future Omegans. The right to political independence protects this kind of self-determination.

The right to territorial integrity is necessary because, as Walzer argues, it is a necessary condition for political independence. Just as an individual cannot be secure in her life or liberty unless there's some space within which she is safe from intrusion, the political community "requires the existence of 'relatively self-enclosed arenas of political development.'"[17] To cast it in a slightly different light, territorial integrity is important because the citizens of Omega are doing various projects that require their presence in and use of the territory. Alpha's assumption of sovereign power over the territory could amount to shutting down those projects and deciding for the Omegans how they will associate.

So the decision Omega must make when faced with Alpha's coercive threat isn't whether sovereignty is such a worthy goal that it's worth killing or being killed for. If the Omegans decide to fight back, their defensive war will be justified as a refusal to be complicit in their subjugation by Alpha.

Fighting for sovereignty, then, might be understood as an instance of following a maxim of resistance. When a state fights permissibly in a defensive war, the permission doesn't come from the fact that the goods of territorial integrity and political

[17] "The Moral Standing of States," at p. 228.

independence somehow outweigh or overcome the prohibition against killing. The state fighting a defensive war is not fighting in the pursuit of some goods. Unlike wars of aggression, defensive war is not aimed at private gain. The state fights so that the community can continue to be self-determined. Fighting just is resisting, the very thing characterized in the individual case, and so we might expect it to carry the same kinds of permissions.

IV. A Pacific Interpretation of the Maxim of Resistance

Having developed an account of how the justification of individual self-defense might justify defensive war, I'd like to revisit our initial account of the permissibility of individual self-defensive killing, and try to show that non-lethal resistance, even where there is no innocent bystander, can also count as good resistance. To develop a pacific interpretation of the maxim of resistance, it seems the question for us is – what is it in virtue of that a person counts as resisting even where she doesn't kill, or try to kill, the aggressor? What counts as good resistance will depend on the situation. But we saw from Case Innocent Bystander that non-lethal resistance will count as good resistance because the victim (a) recognizes the aggression as impermissible, (b) condemns the aggression either silently or out loud, and (c) decides to limit her violence against the aggressor for the sake of Innocent.

But this is not the only form that resistance can take. In some cases, the permissibility of an action cannot be judged in momentary isolation. In those cases, it's necessary to step back and consider the moral character of an entire course of action. To borrow a phrase from Barbara Herman, we shouldn't "shrink the moral moment." The

moral moment can last beyond the aggressive act. Once the aggressive act is over, the victim can still go on acting on her own reasons – she can condemn the violent act, report it to the police, join a neighborhood watch. She can act against the violent act that has already happened, and continue to act on the maxim of resistance.

It might seem that in the case of lethal violence, unlike coercion or beating, once the violent act is done, the moral moment is truly over. But I think we can take the idea of not shrinking the moral moment even further. Even for the person who faces a murderous aggressor, I'd like to suggest that the victim's non-lethal resistance can still count as good resistance, and that the moral moment can go on. Imagine that the victim lived her life taking others' lives as reasons not to kill them, and she treated people with respect, maybe she even tried to convince the murderous aggressor what she was doing was wrong. The victim, by living her life according to the good maxim and taking all lives as reasons not to end them out of respect for the life-bearer's humanity, and by living with others according to the good maxim, will qualify as resisting her own subjugation through that activity (both before her death and continuing on after), even though she refrained from landing a lethal blow to destroy her aggressor. We should not characterize the victim's restraint in this case as a failure to respect herself.

This view, that non-lethal resistance can count as good resistance, becomes more plausible when we notice that moral moments are interpersonal. They are not just about the victim. They are about the victim, and the aggressor, and bystanders. And because the moral moment lasts, it might also include the police, and the victim's neighbors and others to whom the victim might tell her story, all of whom share the victim's activity of respecting the rational nature of others. If moral moments are not just about the victim,

then it seems the moral moment *could* continue on even after the victim is killed. The moral moment could be filled out by the victim's friends, and her family, and the police, *etc.* And maybe this is part of the reason why we embed ourselves in moral communities. So it looks like not shrinking the moral moment can characterize the victim's restraint against the murderous aggressor in the same way it can characterize the victim's restraint in cases of non-murderous aggressors, namely, as permissible.

If we think the pacific interpretation is a good one, I think we now bear the burden of showing why self-defensive killing, as opposed to non-lethal self-defensive violence, is justified. Having accepted that what's at stake in Case Innocent is important enough such that we have to hold fire, I think that to make a really compelling case for the permissibility of self-defensive killing, we need to make sure that that same value isn't also present when the victim is confronted by a murderous aggressor alone. (Or, if that same value is also present, we have to be able to account for why it should factor in our reasoning about what to do differently than it does in Case Innocent.)

V. The Agential Theory of Defensive War Reconsidered

In light of the doubts about the permissibility of individual self-defensive killing, the question for us now is whether defensive war can still be justified. Because of some important differences between the individual and the political community, I think the answer is not immediately clear.

At the international level, there are two ways the moral moment could be filled out. The first way is along an interpersonal dimension: given that there is no global legal system and no international moral community of pacific resisters, Omega, when faced

with the threat of annexation, can't count on others to carry on the moral moment by continuing to resist. Here is the first disanalogy between the individual and the political community. The individual's act of resistance can be carried on by her friends and family, and by the police and justice system. But if there is no such analogue at the international level, then for Omega to choose not to fight might be for it to go along with its own subjugation. So on this dimension the alternatives to violence look worse than in the individual case. And this suggests the possibility that collective self-defense might be permissible even while individual self-defense isn't.

Whether there is a strong and important disanalogy along the interpersonal dimension will depend on whether we have an international community of resisters. There are organizations that have been created to try to create an international community and an international system of law, most obviously the UN, and also the International Criminal Court and the International Court of Justice. (Even if one is skeptical of whether these systems are robust enough to serve the purpose for which they were created, I think it's worthwhile to suggest that states do not exist in a state of nature.[18] With respect to war, even before these institutions, states didn't interact with each other as in a state of nature. The war activity is rule governed, and states that participate in it are governed by law and custom.) Perhaps seeking recourse through these institutions is neither efficient nor timely, but wars are not without their own terrible costs. And seeking recourse through the UN has the added benefit that it's not morally

[18] For a skeptical view of whether states could exist in a state of nature, *see* "Anarchy is What States Make of It."

impermissible.[19] We have (at least nascent) global institutions of peace, through which international democratic action is possible.[20]

The second way the moral moment could be filled out is along a time dimension: unless the invader is going to kill me, I can resist, later, by acting, myself, for the sake of restoring our original association. Here is the second disanalogy. Compared to the individual case, there are even more opportunities for resistance to take a non-violent form in the collective case. Citizens of the invaded territory can make it very difficult for invaders to rule them. Possibilities include civil disobedience, protests, mass strikes, destruction of infrastructure, exclusion of invaders from civil society.[21] The invaded can make it very costly for the invaders to try and stay in the newly annexed territory. Those citizens of the invaded territory who resist the take-over might be killed, and their deaths will be terrible, but their deaths will be part of the greater resistance that will be carried on by their compatriots. The individual self-defender has no equivalent to this option. And so this opens the possibility that defensive war will be condemned by a maxim that permits individual self-defense.

VI. Conclusion

I hope I've shown that pacifism can be a real moral alternative to killing, both for the individual facing an aggressor and for the state facing an aggressing state. If this is true, then fighting a defensive war is not morally obligatory. I have not shown here that

[19] I will not consider here the permissibility of UN intervention by peacekeepers, as the peacekeepers are not a traditional army. This is an issue, along with genocide and crimes against humanity, that I hope to consider in future work.
[20] That the initial framework exists, and might be the only way out of international violence, might obligate us to develop these global systems.
[21] *See, e.g.,* Sharpe, The Methods of Nonviolent Action, Porter Sargent Publishers (1980)

fighting a defensive war is impermissible. But if it is a permissible choice, it is now just one of two. And given our general presumption against the permissibility of killing, I think this is enough to shift the burden back on those who would justify killing in a defensive war to show why that choice is better.

Bibliography

Anscombe, Elizabeth, "War and Murder," from Richard A. Wasserstrom (ed.), War and Morality, Wadsworth (1970)

Buchanan, Allen, "Institutionalizing the Just War," *Philosophy & Public Affairs*, vol. 34, no. 1 (Winter, 2006), pp. 2-38

Dinstein, Yoram, War, Aggression and Self-Defence, Fourth Edition, Cambridge University Press (2005)

Fotion, Nicholas and Gerard Elfstrom, Military Ethics: Guidelines for Peace and War, Routledge and Kegan Paul (1986)

Grossman, Dave (Lt. Col.), On Killing: The Psychological Cost of Learning to Kill in War and Society, Little, Brown and Company (1995)

Hauerwas, Stanley, "Should War Be Eliminated?" from Against the Nations: War and Survival in a Liberal Society, University of Notre Dame Press (1992)

Herman, "Murder and Mayhem" from The Practice of Moral Judgment, Harvard University Press (1993)

Holmes, Robert L., "Can War Be Morally Justified? The Just War Theory" from On War and Morality, Princeton University Press (1989)

James, William, "The Moral Equivalent of War," from Richard A. Wasserstrom (ed.), War and Morality, Wadsworth (1970)

Jenkins, I., "The Conditions of Peace," *The Monist*, 57 (1973), 507-26

Kavka, Gregory S., "The Numbers Should Count," *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition*, vol. 36, no. 3 (Oct., 1979), pp. 285-293

King, Martin Luther Jr., "Loving Your Enemies," http://www.thekingcenter.org/archive/document/loving-your-enemies-1

--- "Why I Am Opposed To The War In Vietnam," http://www.thekingcenter.org/archive/document/mlk-sermon-why-i-am-opposed-war-vietnam

Luban, "Just War and Human Rights," *Philosophy and Public Affairs*, vol. 9, no. 2 (Winter, 1980), pp. 160-181

Martin, Michael "On an Argument Against Pacifism," *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition*, vol. 26, no. 5/6 (December, 1974), pp. 437-442

Mavrodes, George I., "Conventions and the Morality of War," *Philosophy and Public Affairs*, vol. 4, no. 2 (Winter, 1975), pp. 117-131

McMahan, Jeff, "Just Cause for War," *Ethics and International Affairs*, vol. 19, issue 3 (December, 2005), pp. 1-21

--- "The Basis of Morality Liability to Defensive Killing," *Philosophical Issues,* vol. 15, no. 1 (October, 2005), pp. 386-405

--- "War as Self-Defense," *Ethics and International Affairs*, vol. 18, no. 1 (2004), pp. 75-80

--- "The Ethics of Killing in War," *Ethics*, vol. 114, no. 4 (July, 2004), pp. 693-733

--- "Innocence, Self-Defense and Killing in War," *The Journal of Political Philosophy*, vol. 2, no. 3 (1994), pp. 193-221

Nagel, Thomas, "War and Massacre," *Philosophy and Public Affairs*, vol. 1, no. 2 (Winter, 1972), pp. 123-144

Narveson, Jan, "Pacifism: A Philosophical Analysis," *Ethics*, vol. 75, no. 4 (July 1965), pp. 259-271
Norman, Richard, Ethics, Killing and War, Cambridge University Press (1995)
Otsuka, Michael, "Killing the Innocent in Self-Defense," *Philosophy and Public Affairs*, vol. 23, no. 1 (Winter, 1994), pp. 74-94
Paskins, Barrie and Michael Dockrill, The Ethics of War, Duckworth (1979)
Pick, Daniel, War Machine: The Rationalization of Slaughter in the Modern Age, Yale University Press (1996)
Rawls, John, The Law of Peoples, Harvard University Press (2002)
Rodin, David, "Terrorism Without Intention," *Ethics*, vol. 114, no. 4 (July, 2004), pp. 752-771
--- War and Self-Defense, Oxford University Press (2002)
Russell, Bertrand, "The Ethics of War," *International Journal of Ethics*, vol. 25, no. 2 (January, 1915), pp. 127-142
Ryan, Cheyney C., "Self-Defense, Pacifism, and the Possibility of Killing," *Ethics*, vol. 93, no. 3 (April, 1983), pp. 508-524
Sanders, John T., "Why the Numbers Should Sometimes Count," *Philosophy and Public Affairs*, vol. 17, no. 1 (Winter, 1988), pp. 3-14
Sharp, Gene, There Are Realistic Alternatives, The Albert Einstein Institution (2003)
--- The Methods of Nonviolent Action, Porter Sargent Publishers (1980)
--- "The Meanings of Non-Violence: A Typology (Revised)," *The Journal of Conflict Resolution*, Vol. 3, No. 1, Studies from the Institute for Social Research, Oslo, Norway (March, 1959), pp. 41-66
Taurek, John M., "Should the Numbers Count?" *Philosophy and Public Affairs*, vol. 6, no. 4 (Summer, 1997), pp. 293-316
Thomson, Judith Jarvis, "Self-Defense," *Philosophy and Public Affairs*, vol. 20, no. 4 (Autumn, 19910, pp. 283-310
Timmerman, Jens, "The Individualist Lottery: How People Count, but Not Their Numbers," *Analysis*, vol. 64, no. 2 (April, 2004), pp. 106-112
Walzer, Michael, Just and Unjust Wars, Fourth Edition, Basic Books (2006)
--- "The Moral Standing of States: A Response to Four Critics," *Philosophy and Public Affairs*, Vol. 9, No. 3 (Spring, 1980), pp. 209-229
Wasserman, David, "Justifying Self-Defense," *Philosophy and Public Affairs*, vol. 16, no. 4 (Autumn, 1987), pp. 356-378
Wasserstrom, Richard A., "War, Nuclear War, and Nuclear Deterrence: Some Conceptual and Moral Issues," *Ethics*, vol. 95, no. 3 (Apr., 1985), pp. 424-444
--- "On the Morality of War: A Preliminary Inquiry," *Stanford Law Review*, vol. 21, no. 6 (June, 1969), pp. 1627-1656
--- "Three Arguments Concerning the Morality of War," *The Journal of Philosophy*, vol. 65, no. 19 (October, 1968), pp. 578-590
Wendt, Alexander, "Anarchy is What States Make of it: The Social Construction of Power Politics," *International Organization*, Vol. 46, No. 2 (Spring, 1992), pp. 391-425
Yoder, John Howard, "How Many Ways Are There To Think Morally About War?" *Journal of Law and Religion*, vol. 11, no. 1 (1994-1995), pp. 83-107

What Good Are The Humanities? An Impassioned and Impolitic Defense

Or: Reflections on What I Hope Is Not Your Future

Tal Brewer

You may wonder why I feel compelled to speak up in defense of the humanities today. I can begin to answer by sharing a vignette from the dramatic events that unfolded at UVA last summer. There we were, a couple thousand faculty members, students and staff, gathered on the main lawn of the University of Virginia, eyes fixed expectantly on the door just behind the pillars of Thomas Jefferson's Rotunda. The door opens and once and future UVA President Theresa Sullivan emerges. We burst into cheers and some begin to chant her name. Sullivan bathes for a moment in the enthusiasm. Then she takes the podium to give her restoration speech.

Two weeks earlier, Sullivan had been stripped of her office in an abrupt coup orchestrated by Helen Dragas, Rector of the University's Board of Visitors. When news of Sullivan's removal began to spread, and when it became clear that it had been done without consulting anyone actually working at the university (with the possible exception of the chief financial officer), the faculty went into full rebellion. The Faculty Senate – a body that ordinarily trundles along without a clear and discernible sense of its own mission – promptly passed a resolution of no confidence in the Board of Visitors. The Senate was joined by a number of other faculty groups in calling for the reinstatement of President Sullivan and the resignation of Dragas and her co-conspirator, Vice Rector Mark Kington. The university grounds, ordinarily sleepy in the summer, were soon in full uproar. Reporters began digging into the motives for the coup.

Dragas and Kington had traced their actions to "philosophical" differences with Sullivan, yet declined to say which of the great questions of existence had divided them so irrevocably as to require her dismissal. Enterprising reporters from the student newspaper made use of the Freedom of Information Act to shed some light on this mysterious philosophical disagreement. It turned out that Dragas and Kington had come to believe that the rise of on-line learning would soon pose an existential threat to the university, and that UVA had to join this movement quickly if it was to avoid being buried by it. Sullivan had been reluctant to move in this direction with the boldness they thought necessary. Sullivan was threatened with imminent dismissal, and agreed under duress to step down.

At first blush this does not sound like a philosophical disagreement. It sounds like an ordinary empirical disagreement about whether, and under what conditions, UVA will be able to attract enough qualified undergraduates to sustain itself. One party to the conflict, President Sullivan, was less impressed than her adversaries by Clayton Christensen's widely discussed prediction that on-line instruction would prove to be a "disruptive innovation" – one that poses an existential threat to traditional suppliers not by providing a better product but by providing an inferior substitute that is either vastly cheaper or more convenient. Christensen had insisted that traditional providers of higher education

could survive this disruption only by "changing their DNA" – that is, fundamentally changing their mode of instruction, partly by using on-line instruction to lower costs and reach more students. The leaders of UVA's board of visitors had apparently concluded that Christensen was right, and had instructed Sullivan to move quickly towards a radical institutional makeover. Sullivan seems to have thought that this was alarmist and that no program of on-line instruction would soon convince parents to forego the rite of passage into adulthood that we call "going to college".

(SKIP: For the record, my guess is that Sullivan is probably right about this empirical dispute. Going to college involves a great deal more than attending and completing a certain number of classes. It is a comprehensive experience that helps youths to navigate all aspects of the difficult transition from adolescence to young adulthood, and relatively few elements of this experience can be replicated on-line. Furthermore, it is hard to suddenly disrupt the market for a product whose dollar value is supported by largely unreflective patterns of public prejudice. A new kitchen appliance that is a tenth as expensive and almost as good as its competitors will quickly win adherents, because its purchaser immediately enjoys the full 90% cost reduction associated with it. If on-line education were every bit as good as the best on-campus education and cost a tenth as much, its consumers would not enjoy a full proportion of the associated 90% cost savings unless the public fully internalized this fact and began to extend the same esteem, and the same opportunities and pay levels, to the possessor of an on-line degree as to the possessor of a Harvard degree. Until this shift in public opinion occurred, the possessor of the on-line diploma might well end up paying more for his degree than the possessor of the Harvard diploma. Disrupting the value of an elite college diploma is the meritocratic (or perhaps faux-meritocratic) equivalent of disrupting the value of an aristocratic accent: it could happen, but it would require a decades-long shift in deep-seated public prejudices.)

On the surface, then, the conflict between President Sullivan and the leaders of the Board of Visitors seems to have been a difference in speculative market prognostications and not a difference that could be termed philosophical, even in the loose and popular deployment of that term. Yet I believe there are important philosophical disagreements in the background. Discussions among members of the Board of Visitors touched not only on the importance of taking bold steps to deliver instruction via the internet but also on the importance of taking bold steps to trim away departments with relatively few majors. German had been mentioned; so had Classics. The guiding idea of the would-be reformers was that the university should be continuously reshaped to meet changes in student demand. Consumer sovereignty should be extended from the problem of determining which products shall be displayed on the shelves of which stores, to the problem of determining what the proper education of a young adult should consist in.

This extension of consumer sovereignty is objectionable, in the first instance, because college-age youths cannot be presumed already to know what it would behoove them to know, but are in need of education concerning what sort of education they need. But there is another objection to this invocation of consumer sovereignty, one rooted in the very identity of liberal arts colleges and universities. Perhaps I can explain this second

objection by telling you about a recent phone conversation with a friend who teaches Spanish literature at Mary Baldwin College, a small women's liberal arts college in the town of Staunton, Virginia, about half an hour from UVA. Maybe it's a case of local contagion, or maybe there is a nationwide trend here that is flying under the radar screen of the national press, but there is a stand-off underway between Mary Baldwin's Board of Trustees and its faculty that is strikingly reminiscent of last summer's stand-off at UVA. The Board is imposing major changes upon the college, against the resistance of the faculty, and the faculty is contemplating a vote of no confidence in the university's president and board. The changes being dictated by the trustees include the closing or consolidation of several humanities departments and the shifting of resources to a new set of undergraduate majors in career-specific fields such as criminal justice, marketing, human resource management, social work, public health, special education, and (this one's my favorite) coaching and exercise management.

The trustees seem to believe that these changes are needed to ensure the college's survival. This makes me wonder how exactly they conceive of the college entrusted to them, such that this alteration could count as its survival. It obviously would not count as the survival of, say, Plato's Academy if, at some point in the waning of Athens' golden cultural era, it had taken up the training of military leaders, or merchants, or rhetoricians. Nor would it count as the survival of a Franciscan monastery if it responded to a decline in religious enthusiasm by filling its vacant rooms with, say, vacationing nudists. What is a liberal arts college, such that it can rededicate its facilities to career training for marketers, police officers and coaches without losing its soul? How could such a change be counted as securing the college's survival rather than managing its demise? One gets the impression that Mary Baldwin's trustees take themselves to have been charged with the mission of efficiently using a certain number of physical resources – dormitories, classrooms and the like – for a course of post-secondary pedagogy sufficient to justify the conferral of a bachelor's degree . . . in something or another. Commitment to the liberal arts is not deemed essential to the institutional mission.

I think I know the provenance of this picture of proper institutional governance. It has its home in the business corporation, whose one essential purpose is to find a profitable market niche, and whose products and services are to be changed when necessary to fulfill this fixed purpose with optimal efficiency. It would be fiduciary irresponsibility for the board of a publicly held company to stick doggedly with the production and marketing of a particular product even in the face of clear indications that the market for said product was declining and that some other product could be made and sold at a greater profit. Consumer sovereignty converges here with responsible board governance: one cannot responsibly insist upon some vision of what the world needs that is out of fashion with most consumers.

If one imports this picture of responsible governance into the trusteeship of colleges and universities, one might think that this would at least increase the chances of institutional survival. What I'm suggesting is that this is a mistake. Like Plato's academy, so too liberal arts colleges and research universities are vulnerable to finding themselves in situations where they must choose between filling their corridors with instructional

activities that pay the bills yet secure the mere semblance of survival, or struggling to sustain the liberal arts mission even in the face of serious uncertainty about its financial viability. It is the sort of institution that can remain financially solvent in a faithless or faithful way, and whose path to financial insolvency can be noble or ignoble. The business corporation knows no such distinction: if it remains profitable, it remains true to and successful at its identity-conferring mission.

It matters to the little drama we've lived through at UVA that the fate of our university had been entrusted to a board composed almost entirely of business executives and without a single career educator. Dragas herself is the CEO of a large real estate development firm founded by her father. The two board members who worked with her most closely in plotting Sullivan's ouster were real estate developer Hunter Craig and Vice Rector Mark Kington, co-founder of Columbia Capital and managing director of X-10 Management Corporation. These three were apparently deemed qualified to run UVA because of their success in the world of business. They plotted Sullivan's dismissal in cooperation with a small group of influential UVA alumni who have made fortunes on Wall Street, including Jeffrey C. Walker, the former managing partner of J. P. Morgan Partners and Chairman of CCMP Capital Advisors LLC; Jeffrey Nuechterlein, founder and managing partner of Isis Capital LLC; and former Goldman Sachs partner Peter Kiernan.

One might expect this group to have some qualms about its capacity to plot the future of an institution whose history, structure and purpose differ so greatly from a Wall Street investment house. Yet their email communications suggest an unshakable confidence in their competency and a smug condescension towards the judgments of career academics like Sullivan. Nuechterlein was moved to support the firing partly because he had asked Sullivan about her plans to counter the threat posed by on-line courses and "was not impressed" with her "rather pedestrian answer". Kiernan faulted Sullivan's approach to funding challenges and on-line education for its lack of "strategic dynamism" – a buzzword in the business world for leadership that emphasizes bold risk-taking and nimble responsiveness to changing circumstances. Author of a book called *Becoming China's Bitch: And Nine More Catastrophes We Must Avoid Right Now*, Kiernan is not shy about presenting himself as a seer and re-arranger of the future, one who has mastered the skill Machiavelli thought crucial to a leader – the skill of mastering Lady Fortuna with bold action. He proved somewhat less skilled at foreseeing the perils of abruptly removing UVA's first female president from office without consulting a single member of the faculty or, by all appearances, anyone with significant experience on the academic side of any institution of higher education. His active participation in Sullivan's ouster, which seems to have involved coordination with the governor's office, led to his resignation under duress from the Board of Trustees of UVA's business school foundation just four days after he had written an email telling his fellow board members "Trust me, Helen [Dragas] has things well in hand."

She didn't. To our great surprise we won last summer's battle, but we certainly have not won the war. It continues to unfold, now under the guise of struggle over the content of a forthcoming strategic plan, one that may well initiate the changes that Sullivan was

resisting and that Dragas and her allies wanted. I've told the story of the battle to explain the basis for my current understanding of the nature and stakes of the conflict, and to provide you with what I'm afraid may be a preview of similar events coming soon to campuses near you. If I may try my hand at Kiernan-style prognostication, I think that what we are witnessing is the sharpening of a long-standing conflict between academia and an ascendant Zeitgeist that affirms material productivity and economic competitiveness over all other human ends, and that celebrates the all-purpose managerial acumen of the corporate leader. In my lifetime, we've gone from dismissing corporate mucky-mucks as conformists in grey flannel suits to counting business success as the reliable sign of an all-purpose capacity to lead any enterprise at all, right on up to the Free World (here I have in mind Romney's continuous invocation of his experience in corporate restructuring as proof of his fitness for the presidency).

(SKIP: It is not just in moments of dramatic conflict, such as those we lived through at UVA last summer, that we feel the encroachment of this Zeitgeist in the academy. We feel it daily, in the slow-motion transfer of power from the faculty to administrators who usually lack a higher degree in an academic field. These managers can exercise power only if the activity they manage can be conceived as something whose attainment they are in a position to ascertain and measure first-hand, without relying upon the testimony of the trained professionals whose activities they are trying to rationalize. Hence their rise tends to be accompanied by a change in institutional aims and ambitions – a change promoted under the abstract name of accountability that in fact alters what we professors are accountable for, in a metric that can be appreciated from a vantage point external to our fields. This is the source of the increase of paperwork and the shrinking of prerogative that the professoriate has experienced in recent decades. Here is the point of entry for the accursed paperwork requirements that at UVA go under the name of learning outcome assessments, and which have rightly been called "a disciplinary instrument masquerading as a pedagogic one."[i])

(SKIP: We've seen very similar patterns of institutional change in other professions, including social work, elementary and secondary school teaching, and medical care. All of these fields have seen a rise in managerial oversight of trained professionals, with dramatic expansions of reporting requirements and equally dramatic restrictions in the room for case-specific professional discretion. Yet unlike these other professions, colleges and universities are not mere innocent bystanders to the rising power of the all-purpose manager. They have greatly abetted it by profiting from new masters-level programs in public and business management. Colleges have always derived a large part of their market value from their capacity to provide students with a large boost in the contest for positions of prestige and high remuneration. One plausible explanation of the rise of MA programs in management is that it is a strategic response to the overproduction of BAs relative to the supply of truly advantageous career starting points. The value of the BA has been bolstered, and a new revenue stream created, by the introduction of management degrees that require the BA yet are kept in much shorter supply. There is, then, a kind of poetic justice in the remaking of academia by the sort of hyper-active and disciplinary business-style management that it has helped to unleash on other professions. Not that this makes me like it one bit more.)

The words ‘scholar’ and ‘school’ come to us from the Greek ‘*scholé*’, which can be translated at least roughly as leisure, in the sense of time clear and free from the need to labor. Aristotle regarded such freedom as the precondition of the most valuable sorts of thought and action, and in particular for the sort of thought that is its own end. On this point he was not striking out on his own; he was giving voice to a prevailing Greek opinion that manifests itself in the etymological path connecting ‘*scholé’* to ‘school’ and ‘scholar’. On this characteristically Greek view, leisure opens the possibility of genuinely free thought – thought borne of wonder and free to unfold in accordance with its own internal demands, rather than thought borne of consciousness of lack or need. To be sure, *scholé* also opens the possibility of other autotelic activities – that is, activities that are their own end and are free to unfold in accordance with the demands of this internal *telos* rather than activities whose point lies in some conceptually separate state of affairs that they are calculated to produce. Yet the term evolved in such a way as to suggest a specifically Greek taste for free thought itself, coming first to mean ‘discussion’ and later to designate the philosophical schools that regarded themselves as custodians of the best and highest human discussions.

This broadly Greek understanding of the opposition between leisure and leisurelessness is taken up into Latin thought by such figures as Cicero and Seneca with the terms ‘*otium*’ and its contrary ‘*negotium*’. This latter Latin term itself traversed a long etymological path, one of whose endpoints is the contemporary Spanish word ‘*negocio*,’ which means business. A somewhat less direct affinity between commercial life and the absence of leisure is embedded in the English-language term ‘business’ – though ordinarily the term spills out of the mouth too quickly for its etymology to register in the ear. I wonder if business school might not lose some of its appeal if its name were enunciated less hastily. Wouldn’t people think twice before signing up for a school of busy-ness?

Maybe they wouldn’t. But I hope at least that you can now hear the lingering oxymoron in that phrase. A school of business. A *scholé* of the negation of *scholé*. An institution devoted to discussion and thought unfolding under its own internal demands, yet offering training for the sort of life that has no place for such thought – the sort that places thought in service of need. Indeed, the contrast is rather more stark these days, since business busies itself not merely with the navigation of need but with the creation and intensification of felt need, hence with continuous amplification of the realm of human life in which thought takes direction from something alien to it. The leaders of last summer’s coup at UVA are practitioners of this mode of thought. Indeed three of the four key organizers of the ouster are closely associated with UVA’s own *scholé* of *ascholia*, the Darden Business School: Dragas and Kington are Darden graduates and Kiernan chaired Darden’s foundation board until he resigned under duress last summer.

The turbulence at UVA last summer, then, can be understood as a clash between scholarship in the Ancient sense – which is to say thought unfolding in freedom, thought that does not take direction from anything alien to itself – and the contrary forms of thought that are appropriate when basic needs deprive human beings of the opportunity for more valuable uses of their defining mental capacities. This echoes the tension that

we find in the ongoing contest over the soul of Mary Baldwin College. Historically Mary Baldwin has identified itself as a liberal arts college, and a large portion of its faculty remains attached to that identity. The liberal arts have traditionally taken their identity from a contrast with the servile arts, which is to say, the arts needed for efficiently navigating the condition of captivity to need, the condition that the Greeks called *ascholia*. The liberal arts are those it makes sense to develop and exercise when one is lucky enough to be free from the compulsion of genuine material need. The purpose of the servile arts is to remain alive and healthy. The purpose of the liberal arts is to engage in activities that are worthwhile in themselves, activities that can give point to remaining alive and healthy. When a college retreats from the liberal to the servile arts, it announces to its students that times are too lean to permit four years of indulgence in something capable of giving point to remaining alive and healthy, and that they must concentrate on the task of enhancing our material means. And when it includes marketing among the servile arts that it recommends to those who come to its doors seeking education, it departs even more decisively from its prior mission: it announces that times call for the creation and intensification of new needs, needs that will redirect the minds of one's fellow human beings from the liberal arts even under conditions that might otherwise conduce to the development and exercise of those arts.

Like many of my fellow professors, I gave serious thought last summer to the possibility of publishing a defense of what we do at UVA. I was particularly eager to defend the humanities, since the limited information that had been published concerning the intentions of the Board of Visitors suggested that they had keener doubts about the justifiability of humanities departments than about the justifiability of programs in the currently favored STEM programs – Science, Technology, Engineering and Math – or in such areas of social science as politics and economics, or in the new arrival on campus, so-called leadership studies. I produced rough versions of possible op-eds but did not try to publish anything.

What paralyzed me was that my attempted defenses of the humanities seemed to fall into two categories: those that might conceivably help our cause but that were not heartfelt, and those that were downright impolitic. These latter efforts were in fact doubly impolitic: they traced the value of the humanities to a vision of the human good that ran against the political currents of our place and time, and they pointed towards a serious indictment of the form actually taken by the humanities in my own university and in colleges and universities around the country. They were, in short, politically ineffective defenses of an ideal that, if taken seriously, would provide fresh grounds for attacking us. I thought that in the long term it might be salutary to consider this ideal, and the novel objections to which it gives rise, but I did not want to publish my thoughts about these matters while the battle was in the public eye. I was even less eager to write something politically savvy that did not faithfully express my passion for the humanities. It seemed wrong to seek to preserve social space for philosophy with the tools of philosophy's traditional nemesis, which is to say, with sophistry. So I decided to wait until a later, less fraught and less public occasion to attempt to articulate the grounds of my devotion to the humanities in general, and to philosophical reflection in particular. And here I am.

When I look at what others have written in defense of the humanities, the arguments seem to fall into three categories. The first strategy is to bring out the pecuniary benefits of the humanities, whether in terms of individual career advantage or of national economic competitiveness. Perhaps it will not come as a surprise that I do not favor this approach. I don't mean to say that there is nothing to it. I recently traveled to Peking University to trade ideas about higher education with Chinese educators, and it seems that the country's leaders have concluded that they should invest in the humanities precisely because of the creativity that such an education incites, and the contribution of this sort of creativity to market innovation. The American Association of Colleges and Universities has been trying to promote liberal arts education on precisely this ground: "In an economy fueled by innovation, the capabilities developed through a liberal education have become America's most valuable economic asset."[ii] Similar observations have prompted Carol T. Christ, president of Smith College, to argue that "it would be a mistake to find irrelevant a system that has proven so disproportionally successful that its methods are being adopted by some of America's strongest economic competitors."[iii] Yet whatever the truth about the aggregate economic benefits of widespread training in the humanities, the employment and earnings prospects of humanities majors are dim when compared to other majors. Recent history majors have an unemployment rate more than one third higher than engineering or business majors, and humanities majors who do manage to get jobs earn an average salary of only $37,000, compared to about $62,000 for engineers. While I've sometimes heard it said that in the long run humanities majors end up earning at least as much as engineers, the data do not bear out this rumor. In fact engineers continue to out-earn humanities majors by 30% in the mid-career years. So pecuniary considerations do not seem to count in favor of majoring in the humanities, and there are indications that students may be turning away from the humanities for precisely this reason. Eighty eight percent of first-year college and university students now cite "getting a better job" as the top reason for pursuing their studies, up from 71 percent prior to the economic downturn, and only about one in 12 now chooses to major in the humanities – less than half the portion that made this choice in 1967.[iv]

These economic considerations have created a political backlash against public subsidies for humanities departments at state universities. Florida is currently considering higher tuitions for humanities majors than for "strategic" degrees because their studies contribute less to the state's economic health. In the words of Florida Governor Rick Scott, "If I'm going to take money from a citizen to put into education, then I'm going to take that money to create jobs. Is it a vital interest of the state to have more anthropologists? I don't think so." Wisconsin Governor Scott Walker and North Carolina Governor Patrick McCrory seem to be thinking along similar lines. "If you want to take gender studies that's fine. Go to a private school and take it," McCrory said during a recent interview. "But I don't want to subsidize that if that's not going to get someone a job."[v] Those who attempt to defend the humanities in pecuniary terms are not only missing the point of the activity they seek to defend; they are lending credence to the threatening views of these governors by conceding one of their most contestable premises – namely, the premise that subsidizing the education of fellow citizens is justifiable only if it conduces to career success or economic growth. This is precisely the pattern one sees in President Obama's public pronouncements on education. He has tied the

justifiability of education at all levels to economic competitiveness, and he has praised the educational systems of Singapore and other Far Eastern nations for emphasizing science and technology, thereby "spending less time teaching things that don't matter, and more time teaching things that do."[vi]

One reason I think these governors are mistaken to suggest that economic growth is the only public good that could justify investments in education is that I do not believe continued economic growth, in any thing like its recent historical form, to be good at all. Here we come to what is arguably the crux of the Zeitgeist that surrounds us, and that makes it so difficult to put forward an impassioned defense of the humanities that is not impolitic. I can do no more than to take a glancing look at this topic today. A good place to begin is with a little essay published in 1930 by John Maynard Keynes under the title "Economic Possibilities for our Grandchildren". Keynes projected that increased worker productivity would soon make it possible to sustain a decent standard of living for all while reducing the average work week to 15 hours. This would bring human beings face-to-face with what Keynes regarded as their real and permanent problem: what to do with genuine freedom – that is, with *scholé*. A proper answer would require the arts of *scholé* – that is, the liberal arts, the art of identifying, refining and pursing life activities that are genuinely valuable in themselves. Keynes was deeply pessimistic about the human capacity to meet this challenge. Looking around him, he observed that the wives of the wealthy were already faced with it and were failing badly. These women "cannot find it sufficiently amusing, when deprived of the spur of economic necessity, to cook and clean and mend, yet are quite unable to find anything more amusing."[vii]

Recent history suggests that Keynes's pessimism was well-founded. We seem inclined to evade our "permanent problem" even in conditions of remarkable prosperity. As Juliet Schor observes, the material consumption rates achieved in 1948 could have been sustained in the 1990s even if every single worker took every other year off.[viii] Yet the average American family contributed 16 *more* weeks of full-time work to the formal economy in 1988 than in 1967.[ix] This is not just a reflection of the entry of more family workers into the workforce. The trend holds good at the individual level as well as at the family level. The average U.S. worker put in 148 more hours per year in 1996 than in 1973.[x] Nor does this seem to be a function of economic need. The percentage of U.S. workers putting in more than 49 hours per week grew from 13% in 1976 to 19% in 1998, while the percentage of managers working that many hours grew from 40% to 45%.[xi] This is the flip side of the vast increase in consumption experienced in recent decades in the U.S. and Western Europe. We are working more and we are spending more – not now in the name of mere survival or creature comforts but as our first attempt to answer what Keynes called the permanent question of humankind. We have effectively refused the offer of increased free time in exchange for increases in paid work and consumption. Indeed, in the United States our consumption patterns have become so lavish that they could be sustainably enjoyed by all of our fellow human beings only if we can somehow get our hands on five more planets that are roughly as well endowed with natural resources as the one we've got. Given this, our *de facto* answer to the permanent question of humanity is not available on a sustainable basis for humankind as a whole.

This seems to me to ground both a practical and a moral demand to rethink our answer in a vastly less production- and consumption-oriented direction. I envision the humanities as central among the liberal arts, in the sense that they are themselves fecund sources of intrinsically valuable activities, and they deepen virtually all of the activities of those who permit their psyches to be reshaped by sustained engagement in them. They deepen friendships, they deepen neighborly social relations, they deepen loves and marriages and parent-child relations, walks in the woods, idle musings, creative and expressive activity, and contemplation of the creative and expressive products of others. The moment has come when we can afford to democratize this life-enhancing form of education. If we opt instead to remake ourselves as a kind of commercial Sparta, whose educational systems are geared primarily to the enhancement of economic productivity, we will leave future generations with a pillaged natural environment and a badly degraded cultural environment.

Yet if we do wish to cultivate a deeper public appreciation of the humanities, we will face some impressive obstacles. We will have to counteract the effects of a pervasive form of socialization by commercial enterprises – one that represents the largest and best-funded program of proselytism ever mustered in the history of humankind. The telling novelty of this form of proselytism is that it is automatic: it can go forward without a single true believer in the wisdom of the consumerist vision of the good on which its many and varied communications overlap. This tuition-free, corporate-sponsored schooling begins long before the first day of kindergarten and does not adjourn or go out of session until we die. The average six-year-old child in the United States sees 40,000 commercial messages per year and can name 200 brands. Nor does this open-air school have to limp along on bake sales. Global expenditures on advertising totaled something in the neighborhood of $650 billion in 2003, making advertising the world's second biggest industry (after weapons).[xii]

The advertising industry can be conceived as a continuous high-stakes bidding war for the precious commodity of human attention. Its messages must surely bend prevailing evaluative sensibilities in an anti-contemplative direction. But they have other effects as well. First, they produce an environment of attentional overload, one that inures us to all but the loudest and most ingratiating focal objects. Second, they flood us in messages whose content is expressly devised to be manipulative, and this cannot help but leave its mark on the prevailing understanding of the normal and proper use of language.

(SKIP: I was thinking about this while flying out to San Francisco for this year's APA. I was hungry. On a flight like this you used to get a meal. Now United Airlines has instituted a Choice menu: you have a choice of several different meals, or no meal at all. (One of these choices -- the last one -- is still free.) The item I end up choosing is called Tapas. It turns out to be a box of seven or eight carefully packaged samples of appetizers. What is most notable about the enclosed offerings is not what's in the packages but the packages themselves. If you lifted all the text from these packages and wrote it on a legal pad, you'd have several pages of text. Interspersed with lists of ingredients whose nutritional value you would need a food chemist to assess, you'd find a recurring tale. It begins with some particular person – someone with a common, ordinary

name like Tom or Suzie or Stacy. This person had a great passion for making whatever is in the package. But she never had a thought in the world about going into business: she made her treats from love, for a small circle of family and neighbors. Then word got around about these love-drenched wonders, and friends and family members convinced her that she ought to share the love by creating a tiny little company. The purpose was not in the least pecuniary. You yourself, here in the plane, flying over the United States looking down at the farms below, you are practically being invited right into the living room to share the love of this good soul, whose name you now know. These pita chips are made with love and care because "That's the Stacy's Way". And so on. You begin to realize that what you've purchased in the Tapas box is really eight advertisements, and you wonder if in fact United is paid by these loving souls, or the companies they've created, to include their little ad-wrapped samples in the Tapas box. For now the real purpose of "Tapas" seems to be to sell larger packages of the kind found in the box – packages that are available in the grocery stores 30,000 feet below. Then you take your napkin out to wipe your mouth, and you see that it too is an advertisement, in this case for the United App for iPhones. It's not an ordinary nap; it's an app ad nap.)

We have devised a world in which mercenary words and images press upon us. (SKIP: I had seen the same phenomenon a couple of hours earlier, when I lifted my shoes and laptop from the obligatory security inspection and found myself staring at a bank ad.) Wherever our eyes are likely to alight, someone is willing to pay for images and text that they can alight upon. They elbow us out of the quiet repose, the *scholé*, that contemplation requires. We adapt, and teach our children to adapt, to a contested and interested domain of image and language where the interesting is continuously revealed to be a mere effect, produced from someone's interest in interesting us. And when amid this clamor of manipulative messages there suddenly appears something quite different, something called the humanities, it is not easy to adjust our form of attention to open ourselves to this newcomer. The attentional environment has not encouraged the traits required for proper engagement with philosophy and other humanities disciplines: the habit of sustained attention and of patience and generosity in interpretation; the openness to finding camaraderie and illumination from others in the more treacherous passages of human life; the expressive conscience that cannot rest until it lights upon exactly the right words for one's own incipient thoughts, words that have nothing to do with manipulation. It is, then, an historical monstrosity to suggest that the humanities must be justified or condemned on the basis of their contribution to the vitality of the commercial sphere. The vitality of the commercial sphere, as currently constituted, poses an existential threat to the repose of mind, the *scholé*, that the humanities require if they are to flourish.

If the humanities are not to be justified in terms of economic productivity, perhaps they can be justified by their contribution to the life of the polity. This is the second of the three strategies one finds in literature on the topic, and it has some extremely able proponents, including Martha Nussbaum. Nussbaum is drawn to this justificatory strategy partly because she thinks that a more direct insistence on the value of public funding for the humanities, or of providing an education in the humanities to all youths, would offend against a properly liberal commitment to pluralism concerning the human good. As Nussbaum sees it, modern democracies "are societies in which the meaning

and ultimate goals of human life are topics of reasonable disagreement among citizens who hold many different religious and secular views, and these citizens will naturally differ about how far various types of humanistic education serve their own particular goals. What we can agree about is that young people all over the world, in any nation lucky enough to be democratic, need to grow up to be participants in a form of government in which people inform themselves about crucial issues they will address as voters and, sometimes, as elected or appointed officials."[xiii] She argues that the humanities cultivate the analytical capacities, historical and intercultural understanding, and mutual respect and concern needed for proper participation in the difficult activity of democratic self-rule.

Like the economic argument discussed above, so too this argument appeals to a politically potent value. However, unlike the economic argument, it ties the humanities to a value they really might reasonably be thought to promote, at least to some degree. Still, I think there are two serious difficulties with the argument. The first difficult is that few democracies actually provide room for the sort of active citizenship on which Nussbaum's argument depends. The actual practice of democracy in most western democratic nations is unfortunately well-captured by Joseph Schumpeter's uninspiring definition of democracy as that institutional arrangement for political decision-making in which the power to make decisions is acquired by means of a competitive struggle, among political elites, for the peoples' vote.[xiv] The role of most citizens in this institutional arrangement is sharply limited: to vote at regular intervals for the elites they prefer. If we ask whether a particular citizen ought to invest in a long and expensive course of education for the sole reason that this will make her votes better informed, more comprehending, and more respectful and empathetic, the answer seems clearly to be 'no'. The educative effort is virtually guaranteed to make no difference in political outcomes. We all recall the counting of the hanging chads in Florida 13 years ago. This episode is sometimes mentioned as part of an argument that every vote counts, even though it's not even clear that every vote was counted. But what it really seems to show is that it doesn't matter one bit how you vote or whether you vote at all. Even in the closest presidential election in the nation's history, in the single state whose electoral votes ended up determining the winner, no one would have changed a thing by staying home or by voting for the losing candidate rather than the winner. Bush would have won the state by 536 or 535 votes rather than 537.

It may of course be true that it behooves even a Schumpeterian polity to promote thoughtful and competent voting as part of the solution to a collective choice problem that makes political ignorance individually rational. Yet it seems to me that the problem at hand is deeper and more systematic than a mere decision-theoretic paradox. There are forms of democratic self-rule that really would provide a proper forum for most citizens to exercise a rich historical and cultural understanding and a well-developed mutual concern. But these forms seem to be workable only in very small scale. The Madisonian corrective to the dangers of popular rule is to make the so-called republic vast, with a citizenry too numerous, dispersed and varied for coherent deliberation or coordinated action. If this corrective works at all, it works by leaving most citizens with no effective means of civic engagement. Contemporary polities could of course be radically

downsized, in the image of the Greek City-State, to permit a more active and engaged form of citizenship. Perhaps they should be. But unless and until this restructuring is carried out, to cultivate a capacity and taste for active political engagement is to prepare the citizenry for unactionable discontent.

A second problem with Nussbaum's argument is that there is a very tenuous connection between immersing oneself in the humanities and becoming a good citizen. If one wants to encourage youths to engage in politics in a respectful and mutually comprehending spirit, it might help to have them read Locke's *Second Treatise* and study *The Federalist Papers*, but is it really necessary, or even greatly helpful, to have them read *Madame Bovary*, *To the Lighthouse*, *The Brothers Karamavoz,* or *Remembrance of Things Past*? Would it be greatly helpful to study aboriginal sculpture, Renaissance painting or baroque music? To read Kant's *First Critique*, Aristotle's *Metaphysics*, or Kierkegaard's *Either/Or*? I doubt it. Even if one were to accept the premise that post-secondary education should be sculpted so as optimally to promote the virtues of engaged and compassionate citizenship, these works would hardly suggest themselves as the surest way to attain this effect.

Here's what I think is driving Nussbaum's argument. The fixed point is that the humanities are valuable and must be defended against impending threats. The obvious line of defense is to articulate the value of the humanities as one experiences them, since it is one's lively sense of this value that sends one looking for a defense in the first place. But this direct argument is deemed inadmissible on liberal neutralitarian grounds – that is, because it turns on claims about the human good about which citizens can reasonably disagree. The neutralitarian liberal is permitted to take an official concern with the character and the evaluative outlook of the citizenry only to the extent that this is necessary for fundamentally important political purposes such as the stabilization of liberal rights and the proper functioning of democratic institutions. Yet the real value of the humanities cannot shine through under these constraints. The main opportunities for exercising the special capacities honed by close reading of Proust and Joyce seem to lie not in the political forum but at the café, or over the dinner table, or perhaps in the bedroom (where they greatly multiply the menu of available pathologies). Despite Nussbaum's notable gifts as a thinker and writer, and her obvious love of literature and philosophy, the resulting argument wobbles under scrutiny.

I do not think that liberalism calls for strict neutralitarian limits on state concern with the acculturation of successive generations of citizens. Indeed, it is precisely this stance of principled abstention that has effectively ceded the task of shaping future citizens to the highest bidders – namely, corporations in their capacity as advertisers – and that has therefore helped to make the background culture so hostile to the humanities. Against the backdrop of today's communications technology, and especially in the context of what is less a democracy than a corporatocracy, neutralitarianism has non-neutral and baleful effects. So I think that Nussbaum's argument turns on a picture of the proprieties of public debate whose real-world effects are bad in general, and especially bad for the humanities.

I think, too, that this picture of proper public debate threatens to undermine Nussbaum's argument in a more direct way. The liberal neutralitarian way of accommodating pluralism is to require laws and basic political institutions to be justifiable to bearers of any reasonable conception of the good, whatever its content. Normative talk is pared down to a *lingua franca* that may include such things as basic rights and liberties, career opportunities, income and wealth, self-respect and basic human functionings. The point of honoring rights, affording opportunities and equalizing wealth is supposed to be visible from a standpoint that abstracts from the content of these plural and conflicting conceptions of the human good – e.g. from Rawls' Original Position.[xv] If we recognized that the task of deliberating together in the public form requires careful thought about goods beyond the reach of the aforementioned *lingua franca*, then we would arguably need to accept a more demanding conception of the virtues of citizenship, one that would give greater plausibility to Nussbaum's argument. But if citizens renounce the task of deliberating together about the human good, then it will be implausible to suppose that they need a sophisticated understanding of the humanities in their capacity as citizens. The vocation for which the humanities prepare us – the vocation of understanding and enacting the meaningful possibilities for living and acting that are opened up by our cultural inheritance and our place in history – will be extra-political.

None of this means that Nussbaum's argument won't work, in the practical and political sense of 'work'. No competent politician would confess to a Schumpeterian vision of politics. Telling the people to their face that they are not in charge would be a serious misstep in the Schumpeterian contest for the popular vote. Actual democracies can be relied upon to pretend that they are governments "of the people, for the people and by the people" even when the people have relatively little say. Given the political power of this pretense, and given public uncertainty about just what it means to be a good citizen, Nussbaum's strategy for defending the humanities is perhaps as wise as any. In any case, I have no more promising strategy to offer. I've been up front about my intention to make an impolitic case for the humanities. I turn to that case now.

If I ask myself why I recoil from the arguments canvassed above, it's because they so thoroughly miss the appeal of the form of thought and life that I seek to share with my students. For me, its appeal has nothing to do with adequating my students for any pre-given social role in the market or the forum. I feel I have something especially valuable to offer those who recoil from whatever satisfactions may be available in the cubicle of an accounting firm or in politics as currently practiced. That these are not enough to live a fully human life, to participate fully in the most meaningful currents of the human project on earth, that these are at best means and at worst impediments – this much seems to me to be an essential starting point for recognizing of the value of the humanities. To tie the worth of what I do to my role in preparing young minds for these pursuits would be to betray the impulse that drew me down the path to this profession, and to turn my back on the special bonds of friendship I think I've established with those few students I believe I've really "reached".

I am not alone in refusing to locate the value of the humanities in the useful preparation they provide for excelling in future roles or professions. Stanley Fish, for instance,

insists that the humanities are their own reward, and that they can be justified only in terms of the special pleasure they afford their initiates, and that humanities professors should be pleased to admit their uselessness since this is tantamount to insisting upon the autonomous and intrinsic value of their pursuits.[xvi] This sort of stance is obvious fodder for the reformist agenda of the likes of Dragas, who recently scolded President Sullivan and her faculty supporters for failing to appreciate that UVA is "not an academic playground."[xvii] So Fish's view certainly fits the bill of an impolitic one. And there is a good deal right about it as well. The humanities really can be pleasurable, they really are intrinsically rewarding, and it would be a serious mistake to turn to them because of their usefulness. Yet in my view, Fish is not careful enough to distinguish the truth that the humanities are not to be used from the falsehood – I want to say, the slander – that they are useless. To say they are useless is to say that they bring nothing of value to our lives beyond the transient pleasure of engaging in them. But this is surely wrong. They are, I think, a gateway to and instigator of a lifelong activity of free self-cultivation – self-cultivation not directed by need. The change they provoke is not always for the happier, or the more remunerative, or the more civically engaged, but when things go passably well it is for the deeper, the more reflective and the more thoughtful. And it is connects our lives with a human vocation that is different in kind from, and potentially more meaningful than, commerce or politics (though in the end the lines between these spheres can break down, and commercial and political activities can themselves be infused with, and made more meaningful by, the extra measure of understanding we might hope to cultivate by our engagement with history, literature, the visual arts, philosophy and the like).

I want to flesh out these suggestions by focusing on the case of philosophy, and by looking in particular at the moment in which western philosophy comes to self-consciousness as an alternative to the Athenian practice of providing youths with training in rhetoric. As you know, the opposition between these two visions of pedagogy is dramatized in Plato's *Gorgias*, which portrays a dialogue between Plato's teacher, Socrates, and a group of teachers of rhetoric. Like any fundamental inquiry into pedagogy, Socrates' exploration turns on a conception of the good human life and the good human being. Socrates grounds his conception in a distinctive and valuable capacity that marks us off from the other animals. The lives of other animals are fixed to a very great degree by instinct. We humans have a capacity that the Greeks called '*logos*' – a capacity for speech and thought; a capacity to take hold of words and exchange them with our fellow human beings in an attempt not merely to understand our world but also to give a distinctive, non-instinctual shape to ourselves and our communities.

The rhetorician (at least as portrayed by Plato) thinks that *logos* is best used as a tool for persuading people to believe whatever one happens to want them to believe. Its use, then, is guided not by truth but by the exigencies of power. Since one can't completely ignore truth and still be persuasive, it is guided in many cases by the ring of truth, by plausibility. The sort of philosophical thought practiced by Socrates is fundamentally different. Socrates' first rule of dialogue was: say exactly what you think. Not what you think will impress your fellow students, or fellow citizens, or what you think your teacher

or your employer or anyone else wants to hear. Speak your mind. Make your thoughts answerable to the phenomena under discussion, and find words that faithfully capture your vision of them. When you speak your mind in sustained and careful conversation about important topics, while refusing the impulse to find rationalizations when serious objections are raised, you put yourself into play. You draw yourself out, make your stance in the world more self-conscious, more determinate and more refined. For Socrates, this is what the defining human capacity of *logos* is for, this is its *telos*: to develop ourselves in freedom from want, articulating and refining our views – both individual and communal – about the world we live in and about what sort of life we are to pursue in it. For this, he thought that we need sincere, persistent, thoughtful and compassionate conversational partners. And we need time. Lots of time. Not, say, a four year stay at college, but a lifetime.

Plato has Socrates open the *Gorgias* by requesting that a particular question be put to the rhetorician. What Socrates says is this: Ask him who he is. This is initially taken up in a rather superficial form. The first viable answer that emerges is: a teacher who can instruct students in the artifice of persuasion. But I think Plato wants to show us something more illuminating about who the rhetorician is – that is, about how his psyche has come to be constituted, due in part to the cumulative influence of his chosen life pursuit. What he wants to show us is that the rhetorician is ill-constituted for serious inquiry into the value of his own chosen way of life. Under pressure of Socratic questioning, the rhetorician acts defensively, insulating his own beliefs about the human good from searching critical reflection. The reason is that his *logos* has become subservient to his *thumos* – that is, his taste for public honor and acclaim. As a result, he views refutation of any element of his view as a kind of rebuff, and reacts not by changing his mind but by redoubling his conviction and searching for rhetorical means of fortification. In other words, the rhetorician becomes a pliable audience for his own rhetorical gifts. His *logos* is turned against its own autonomous standards, hence in a sense against the rhetorician himself, and serves alien elements of his psyche. His own convictions about important matters are held in place not by *logos* but by other psychological forces: vanity, or greed, or consoling fantasies of importance. Consequently his thought lacks lucidity even when shielded from public view.

(SKIP: One way to put the Platonic view at issue here is this: there is no form of the sophist. There is a form of the human being, specifying a particular relation among the parts of the soul, and it is in light of this form that philosophers and sophists alike are identifying as the kinds of things that they are. Thus the sophist has a peculiar relation to the form in light of which he is known. The sophist counts as a unity only because and to the extent that he approximates this form, yet his mode of thought continually betrays this form. Sophistry is at heart the chosen and active pursuit of a kind of falling away from being. It is a refusal to pull oneself together. The sophist does not wish to be in error: thought cannot proceed under the forthright attempt to realize that wish. Yet he also does not wish to be corrected. He views refutation as a harm and not as a gift. Plato dramatizes this state of inner conflict by showing the sophist as alternating between a prematurely triumphant eagerness to provide a genuine account of his chosen life and its

value, and recourse to self-stabilizing rhetoric as soon as the possibility of refutation looms.)

We are now in a position to broach the question why Plato chose to write dialogues rather than treatises. For Plato, conveying philosophy to those who have not yet felt its appeal is not a matter of serving up a list of ethical or metaphysical claims along with arguments for them – though Plato is sometimes taught as if this were his primary intention. It is instead a matter of bringing out the appeal of a certain kind of constitution that a thinker can have, and of the form of thought and way of life that bodies forth from that constitution. Plato's most important argument is in a sense *ad hominem*. What he offers in defense of the philosophical form of life is the person of Socrates. What he offers by way of critique of the sophistical form of life is a series of portraits of sophists, including his portrait of Thrasymachus in the *Republic* and his portrait of Callicles in the *Gorgias*. When Socrates asks who the rhetorician is, he is not asking what sort of instruction he offers but what sort of psyche one exhibits and reinforces if one masters and enacts that instruction. His answer is that one exhibits and reinforces a badly disordered psyche, one incapable of a forthright and sustained effort to uncover the truth about important matters, hence one incapable of the sort of life that gives full expression to our best and highest capacities.

In the concluding pages of the *Phaedrus*, Socrates argues that the written word cannot itself capture and deliver over what needs to be understood, but can at best incite readers to turn towards the phenomena themselves and secure understanding through a more immediate apprehension of them. Further, the *written* word is not ideally suited to play even this indirect role, since its author is not there to respond to successive attempts, on the part of the reader, to attain a first-hand discernment of the phenomena that inspire it. What Platonic philosophy hopes to deliver up to students is no more amenable to summary statement in a treatise or textbook than what Freudian psychotherapy hopes to deliver up to patients. The quest for understanding is irreducibly idiosyncratic, because the sources of blindness and delusion are irreducibly idiosyncratic. If the reader cannot speak to the author, the possibility of useful communication is greatly reduced. If this is right, then the spoken word taken in itself – delivered, say, in the form of a lecture rather than in the course of a conversation – is no better a vehicle for philosophical enlightenment than the written word. Philosophy lives in conversation. The student must be called upon to speak, and to do so sincerely rather than strategically – e.g. with an eye to a grade. This is what puts the student himself or herself into play. If this does not happen, then philosophy does not happen. Thus philosophy does not happen in the passive uptake of lectures – whether they are delivered in a large lecture hall or in a Massive Open Online Course (a MOOC, as they are quickly coming to be known). If university-style philosophy is in danger of being replaced by the "disruptive innovation" of on-line education, perhaps this is because university philosophy classes have assumed a deficient form, one suited for fields whose findings can be mastered in passive uptake.

The task of the philosophy professor is to enact philosophical thinking in conversation with students. Since one can never know for sure what students will say, there is no sure recipe for making things come off well, and no saying in advance where exactly the

conversation will lead. When this activity is remade so as efficiently to produce some pre-envisioned outcome whose attainment can be certified by those who are not party to the conversation, it is not thereby improved; it is annulled. This is why learning outcome assessments are so desperately out of place in the humanities classroom. They obviate the improvisatory conversational exploration that the humanities require in order not to be replaced by something else.[xviii]

The ideal teacher of philosophy is not someone whose opinions are to be accepted but someone whose form of thought is worth emulating. The Socrates we know is a dramatic persona that Plato puts forward as worthy of emulation. What would that emulation consist in? I think it would consist in serious-minded lifelong engagement (engagement "unto death," as Plato wishes to make clear in the *Phaedo* and *Crito*) in the activity of self-articulation, which is to say, the activity of bringing oneself into a more determinate form by bringing oneself into words. Here the word 'articulation' is meant in both of its common senses: we give a more fine-toothed and determinate shape to our views about important matters (i.e. give them greater articulation) by bring them into the space of words (i.e. by articulating them).[xix] This activity requires faithfulness to our actual outlook, but it also alters that outlook by finding words for it that we are prepared to live by, hence it sets the stage for another, more adequate articulation. If this is philosophy, then philosophy is continuous with the sort of self-formative activity that all human beings go through again and again in the course of their lives, provided that they live with even a modicum of deliberateness.

But this is as it should be. Philosophy is not a recherché topic through which certain bookish human beings cultivate an optional or adventitious expertise. It is the intensification and refinement of an inescapable human task – the task of "being-underway towards what is to be uncovered" (to borrow Heidegger's phrase for the Greek notion of the fundamental posture of the human being).[xx] One way in which this posture can be intensified is when certain words appeal to us, they seem like the right thing to say, yet we are not entirely certain what they mean. They are at the moving horizon of our understanding. They call us to an effort to understand them, to uncover what we find appealing about them. We relate to such words as the instruments of our own becoming.

The intensification of this form of living needn't end at graduation, and it needn't be restricted to book groups or evenings at the theater. It is a daily possibility that can infuse the daily activities, including the daily work, of almost any way of life. In a sense, then, the humanities can be said to be useful for any career whatsoever, but the use lies not in increasing one's capacity to secure the characteristic ends of those careers but in deepening one's experience of the characteristic activities of those careers.

You may have noticed that this talk has drifted from a wide discussion of the humanities to a narrower discussion of Platonic philosophy. Have I changed the subject, offering an impolitic defense only of the latter and leaving the rest of the humanities to find some other way of shooting themselves in the foot? I don't think so. It seems to me that the self-formative and culture-formative form of thought whose value I've defended is present in nearly all fields of the humanities. I should add, though, that the category of

the humanities has haphazard boundaries. It is not clear to me, for instance, why logic is in and mathematics out, or why history is in and cultural anthropology out. Perhaps there are fields within the humanities that do not encourage the self-cultivating and culture-shaping use of thought to which I've called attention, and fields beyond the humanities that do. If so, this might be a ground for revisiting the traditional understanding of the demarcation, but this would be an argument for another day. What matters for now is that I believe myself to have rested my argument on a form of thought that is found not only in philosophy but also in classics, in history, and in the circle of creative expression and critical response that characterizes the fine arts and the academic study of drama, poetry, literature and the visual arts.

It is important here that what I claim to be defending is the humanities as they might be and sometimes are, and not all that goes under the name of the humanities in actual colleges and universities. The humanities as practiced have internalized certain aims and aspirations that are alien to them, so they are threatened not only from without but also from within. They have for instance compromised their transformative potential by adopting a degree of specialization that would make sense only if they were in the business of delivering up reliable findings from which others could benefit without themselves traversing the paths of thought that led to them. This hyper-specialization permits practitioners to advance their careers by mastering the extant literature on some arcane debate and "making a contribution" to it. It has become an unspoken requirement that journal articles must put forward something unprecedented under the name 'my view' and show how it is different from and superior to an array of surrounding views that have appeared in recent journal articles. This self-promotional sort of originality may or many not coincide with what is original in the Socratic sense that I would favor – that is, in the sense of originating in one's struggle to find words for one's own deepest preoccupations. Socrates insists that the rhetorician lacks this latter sort of originality, however clever and unprecedented his arguments may be, because the origin of his thought lies outside of him, in the opinions of the *demos*. The alternative, Socratic sort of originality is fully compatible with repeating the words of another, provided that these words genuinely serve to bring own inchoate thoughts into greater clarity. I take it that in the *Symposium*, when Socrates repeats Diotima's views of love, he is speaking with this sort of originality.

Indoctrination into academic specialization, and into the careerist search for a distinctive and unprecedented niche, can easily cause us to lose track of what initially prompted us to throw ourselves into philosophy (or, I imagine, into literature, drama, music, etc.). We can get talked out of our own questions, and come to pursue replacement questions that have a recognized place in the field. There is a fine line between those cases in which we've found a clearer expression of what was bugging us all along, and those cases in which we've been deflected (to use my colleague Cora Diamond's word for it) from our own puzzlements into something else that is more tractable, or more widely discussed in the field. In my own view, though, the latter sort of deflection is very common in academic philosophical training, and perhaps in other academic disciplines as well.
If so, and if my defense of the humanities is on target, we ought to regard this as a serious failing, as it severs the proper connection between the academy and the examined life.

CONCLUDING REMARKS

It's said that college is not the real world, and in a sense I'm happy to affirm that. But I do not see it as mere preparation for the things of real substance and value – that's not the mode of its remove from reality. I see it instead as a kind of polis apart, with a few permanent members and a revolving temporary citizenry of youths. What happens in this polis, when it's in good working order, is a kind of intensification of a form of reflective self-cultivation that can and ought to be a continuous life activity. It is the stuff of a good life and not some mere instrumental means, and it can be intertwined with, and can deepen, almost any subsequent life activity. This parallel polis provides an important counterweight to the culture-shaping effects that arise from the melding of corporate capitalism and contemporary communications technology. It would be a devastating loss if we remade this parallel polis in accordance with the guiding values of the corporation. Because the academy encourages an open-ended form of self-cultivation, and because it provides an important counterweight to an outlook on value that threatens to render us a monoculture, it can be defended in the name of liberal pluralism, and the liberal should not adopt standards of public argument that prevent us from bringing its value into view.

---

[i] Frank Furedi, "Learning Outcomes Are Corrosive," *Times Higher Education* (November 29, 2012), available at http://www.cautbulletin.ca/en_article.asp?articleid=3575.

[ii] http://www.csmonitor.com/Commentary/Opinion/2011/0725/Liberate-liberal-arts-from-the-myth-of-irrelevance

[iii] http://www.acenet.edu/the-presidency/columns-and-features/Pages/Myth-A-Liberal-Arts-Education-Is-Becoming-Irrelevant.aspx

[iv] The data in this paragraph come from *The Hechinger Report*, a non-profit education news outlet based at Columbia University's Teachers College. See: http://nation.time.com/2013/03/07/who-needs-philosophy-colleges-defend-the-humanities-despite-high-costs-dim-job-prospects/#ixzz2QYRxU7FL

[v] *Ibid.*

[vi] See Martha Nussbaum, *Not for Profit: Why Democracy Needs the Humanities* (Princeton: Princeton University Press, 2010), 136-8.

[vii] John Maynard Keynes, "Economic Possibilities for Our Grandchildren" (1930), 4. Available at: http://www.econ.yale.edu/smith/econ116a/keynes1.pdf.

[viii] Juliet B. Schor, *The Overworked American: The Unexpected Decline of Leisure* (New York: Basic Books, 1993), 2.

[ix] Barry Bluestone and Stephen Rose, "Overworked and Underemployed: Unraveling the Economic Enigmas," in *The American Prospect*, March 1997.

[x] Franco Berardi, *The Soul at Work: From Alienation to Autonomy* (Los Angeles: Semiotext(e), 2009), 78.

[xi] These U.S. Bureau of Labor Statistics figures are cited by Berardi in *The Soul at Work*, 78.

[xii] Serge Latouche, *Farewell to Growth* (Cambridge: Polity Press, 2009), 17.

[xiii] Nussbaum, *Not for Profit*, 9.

[xiv] Joseph Schumpeter, *Capitalism, Socialism and Democracy* (London: Routledge & Kegan Paul, 2003), 269.

[xv] Of course, there will be cases in which more detailed knowledge is needed, so as to assess the weight of a case for some special demand (think here of the Amish and mandatory schooling requirements, or Native American worshippers and laws against peyote use). But it's not clear that someone who takes little interest in these cases, trusting that the courts will get them right, has thereby failed to be a good citizen.

[xvi] Stanley Fish, "Will the Humanities Save Us?" *New York Times* (January 6, 2008). See http://opinionator.blogs.nytimes.com/2008/01/06/will-the-humanities-save-us/

[xvii] http://articles.washingtonpost.com/2013-03-01/local/37374353_1_president-teresa-sullivan-board-leader-u-va

[xviii] See Furedi, *Op. Cit.*

[xix] I owe this idea of self-articulation to Charles Taylor, who sets it out in "Responsibility for Self," in Amelie Rorty, ed., *The Identities of Persons* (Berkeley and Los Angeles: University of California Press, 1976), 281-300.

[xx] Martin Heidegger, *Plato's* Sophist, tr. Richard Rojcewicz and André Schuwer (Bloomington: Indiana University Press, 1997), 255.