

A new moral sentimentalism

Eric Vogelstein

Department of Philosophy and School of Nursing, Duquesne University, Pittsburgh, USA

ABSTRACT

This paper argues for a novel sentimentalist realist metaethical theory, according to which moral wrongness is analyzed in terms of the sentiments one has most reason to have. As opposed to standard sentimentalist views, the theory does not employ sentiments that are had in response to morally wrong action, but rather sentiments that antecedently dispose people to refrain from immoral behavior, specifically the sentiments of compassion and respect.

ARTICLE HISTORY Received 25 June 2015; Accepted 19 March 2016

KEYWORDS Sentimentalism; moral emotion; compassion; respect; moral reasons

1. Introduction

In this paper, I present and defend a novel *sentimentalist* metaethical theory. The mark of a sentimentalist metaethical theory is that it analyzes moral language, concepts, or properties in a way that essentially involves *sentiments* of a certain sort.¹ The theory I shall propose is about moral properties – in particular, the property of moral wrongness. I will argue that the nature of the property of moral wrongness is plausibly analyzed in terms of *reasons* for having certain sentiments.

There are two basic kinds of sentimentalist theories of moral properties: *dispositionalist* theories, according to which moral properties are analyzable in terms of the sentimental responses that some specified group of people is disposed to have under certain conditions, and *normative* sentimentalist theories, according to which moral properties are analyzable in terms of the sentiments one is *justified* in having, or *ought* to have, or has a *reason* to have, or in terms of some other properly normative feature of sentiments. Both dispositionalism and the *standard* kind of normative sentimentalist theory invoke *reactive* sentiments:

CONTACT Eric Vogelstein  VogelsteinE@duq.edu

© 2016 Canadian Journal of Philosophy

those sentiments that are had, or justified, *in response* to perceiving that the moral property in question is instantiated.

Although the theory I shall defend is a normative sentimental account, it does not follow the common sentimental model in that it does not invoke reactive sentiments. Rather, I shall argue that moral wrongness is plausibly analyzed in terms of sentiments that dispose people to refrain from morally wrong action – in particular, the sentiments of compassion and respect. This should be welcomed by those who have a generally sentimental bent, due to significant problems with models based on reactive sentiments. But, as I will argue, my theory should also be attractive to those who do not share any antecedent commitment to analyzing moral properties in sentimental terms. In particular, the plausibility of the theory I shall propose derives chiefly from its unique ability to account for two phenomena. First, the theory is specially poised to explain a highly plausible pre-theoretical view about the relationship between compassion, respect, and morally wrong action: that someone who does something wrong typically ought, in the commission of such an act, to have had *more* compassion or respect. Second, the theory is able to explain why morality is necessarily and universally normative in a way that avoids controversial commitments about practical reason. Explicating these two virtues of the theory will constitute my main argument in its favor. My intention is not to provide a conclusive argument for the theory, but rather to show that it is (a) an improvement over traditional realist versions of moral sentimentalism, and (b) overall, a plausible account of the nature of moral wrongness.

2. Sentimentalist analyses of moral properties

2.1. Property analysis

My aim in this paper will be to make some headway in elucidating the *nature of the property* of moral wrongness; that is, it shall be an investigation into what moral wrongness *fundamentally is*.²

But what is it to elucidate the nature of a property? Here is how I understand the task.³ To elucidate the nature of a property is to explain what it is to instantiate that property in terms of *other* properties that are, in an intuitive sense, *fundamental* to the property at issue. Elucidating the nature of a property is tantamount to providing an *analysis* or *reduction* of that property.⁴ For example, the nature of the property of being a brother is elucidated by explaining that to be a brother is to be a male sibling. In this case, the property of being a brother is analyzed in terms of other properties – the property of being male, and the property of being a sibling – that are fundamental to being a brother. Fundamentality here need not be identified with *necessity*; rather, it should be understood as something like *constitution*. Thus, an elucidation of the nature of a property can be thought of as a description of the other properties that

comprise that property's *constituent elements*, as well as the logical relations in which those properties stand that comprise the *structure* of the property at issue.⁵

2.2. Problems with leading sentimentalist analyses of moral properties

The two most prominent types of sentimentalist accounts of moral properties – *dispositionalism* and what we can call *standard normative sentimentalism* – have in common that they seek to analyze moral properties in terms of people's sentimental *responses* to objects with moral features. They differ primarily in that dispositionalism, unlike standard normative sentimentalism, does not invoke the sentimental responses that have some particular *normative* feature, but simply the sentimental responses that some specified group is *disposed* to have, under specified conditions.⁶

When invoked as accounts of the nature of moral wrongness, however, both theories face serious problems; indeed, dispositionalism does not even seem to be extensionally correct. The problem with dispositionalism emerges when we try to specify the group of people who are disposed to have the relevant reactions. Regardless of how we identify such people (e.g. via normalcy or statistical regularity) there is no guarantee that the dispositions of such people will conform to an intuitively acceptable view of morality – for any sentimental reaction the dispositionalist employs in her analysis, such people might be disposed to react in a way that is out of sync with clear moral standards (e.g. feeling guilty for having done something that is obviously morally permissible). For this reason, dispositionalism appears implausible (cf. Koons 2003; D'Arms and Jacobson 2006).

Standard normative sentimentalism avoids this worry. On such a view, what matters morally is not what the members of some group of people are disposed to feel, but rather what people in general *ought* to feel, what we are *justified* in feeling, what one has *reason* to feel, or some other similarly normative fact about sentimental responses. And it is highly plausible that the normative features of morally relevant reactive sentiments (unlike facts about our dispositions to have such sentiments) will necessarily sync up with clear moral standards (e.g. necessarily, one is justified in feeling guilty if and only if one has done something morally wrong) – thus, it is likely that at least some versions of standard normative sentimentalism will be extensionally correct. For that reason, of the extant, currently prominent realist sentimentalist views, standard normative sentimentalism appears to hold the best prospects for being true.

However, despite extensional correctness, there is a major worry for standard normative sentimentalist theories as elucidations of the nature of moral properties, based on a well-known circularity problem: the nature of any sentiment that would be invoked in such an analysis (such as guilt or indignation) will itself be plausibly elucidated in terms of the moral property in question (e.g. moral

wrongness), rendering the account circular (e.g. guilt and indignation are plausibly understood as sentiments that represent or are responsive to one's own morally wrong acts and the morally wrong acts of others, respectively) (Ross 1939, 276–9; D'Arms and Jacobson 2000a, 2000b; Rabinowicz and Rønnow-Rasmussen 2004). And elucidations of the nature of a property are supposed to be informative in a way that rules out circularity. Furthermore, and importantly, if we take seriously the notion that elucidations of a property's nature describe that property's *constituent elements*, i.e. its proper parts, then circular elucidations are *impossible*, since nothing can be a proper part of itself.

The upshot is that the sentimental project – if such a project is aimed at elucidating the nature of moral properties – has thus far been unsuccessful. That is not to say that the circularity problem for standard normative sentimentalism constitutes a decisive objection to the theory (although the extensionality problem for dispositionalism does seem to rise to that level); but the problem is sufficiently worrisome to warrant looking to alternative theories, and it is in that context that I present and defend my view here. I shall suggest that there exists a plausible, hitherto unexplored form of normative sentimentalism. Such a theory, I shall argue, not only avoids the circularity problem that plagues standard normative sentimentalism, but has important and unique theoretical virtues of its own. In what follows, I provide arguments in favor of such a view, explicate the view's details, and respond to objections.

3. A new normative sentimentalist theory

3.1. *Theoretical starting points: compassion and respect*

I shall begin by distinguishing between two kinds of immoral acts: those that are *callous* and those that are *non-callous yet immoral*. Callous acts are immoral acts that cause significant and unjustified harm to another. Callous acts can be *extremely* callous, e.g. torturing an innocent person for money, or less callous yet callous nonetheless, e.g. robbery. Non-callous yet immoral acts, on the other hand, are simply those immoral acts that are not callous. I shall have more to say about non-callous immoral acts shortly, but first let us consider how we are inclined to *react* to someone who commits an act that is callous. One way in which we are inclined to react is to wonder how someone could have so little concern for the suffering of another, so little compassion. This reaction does not reflect an empirical problem, but a normative one – we are at a loss in such cases not because we are unable to explain or understand the person's lack of compassion, but because the person's lack of compassion is *unwarranted*. That is, we think that this person *ought* to have had more compassion (just what sort of 'ought' is at issue will be discussed shortly).⁷

Now, consider how we are inclined to react to someone who commits an act that is *not* callous but is nevertheless morally wrong. There may be no particularly

helpful descriptive generalization of such acts, but we can list paradigm cases: stealing, cheating, lying, freeloading, promise-breaking, and unjustly distributing burdens and benefits.⁸ There is, however, a common element in how we are inclined to *react* to such behavior: we wonder how someone can have so little concern for the desires and interests of others, so little respect for people? This type of reaction to someone's lack of respect, like that described above regarding an immoral actor's lack of compassion, is based on a *normative* judgment: that those who commit such acts *ought* to have more respect for the people whose desires and interests are thwarted as a result of their behavior.⁹

If what I have said thus far is correct, then it appears, at least in general, that (1) someone who commits an immoral act that is callous ought to have had more compassion, and (2) someone who commits an immoral act that is not callous ought to have had more respect. But how strong a claim should we accept regarding the relationship between morally wrong action and lacking the amount or degree of compassion or respect that one ought to have? There are several possibilities – for example, the relationship could be *universal* (i.e. applying to *all* immoral action), or even *necessarily* universal (i.e. applying to all *possible* immoral action). But if the remarks I've made thus far are on point, then at the very least we should be willing to accept the following relatively weak claim:

S1 For *most* A, x: If it was morally wrong for A to do x, then when A did x, A did not have the degree of compassion or respect that A ought to have had.¹⁰

The phenomenon described by S1 – the consistent connection between performing a morally wrong act and failing to have the degree of compassion or respect that one ought to have – provides the starting point for the theory of the nature of moral wrongness I shall propose and defend in this paper. As I explain in the following section, a key virtue of that theory will be its ability to felicitously explain why S1 is true.

3.2. *The theory*

We can begin by noting that S1 describes an interesting connection between two otherwise disparate phenomena: morally wrong action, and lacking the degree of compassion or respect that one ought to have. Such a connection cries out for explanation. Thus, any theory that is able to provide an explanation of the truth of S1 would thereby gain plausibility; and the simpler and more powerful the explanation, the more plausibility that theory would enjoy. The theory I shall propose stands to explain, in an extremely simple and powerful way, why S1 is true. Here is my proposal:

Preventionist Normative Sentimentalism (PNS): The nature of moral wrongness is elucidated as follows: Necessarily, for all A, x: It is morally wrong for A to do x if and only if A's doing x would be ruled out¹¹ by A having the degree of compassion and respect that A ought to have.¹²

PNS provides a maximally *powerful* explanation of the truth of S1, since PNS *entails* S1.¹³ Furthermore, PNS is a decidedly *simple* explanation of the truth of S1, because the truth of PNS itself does *not* demand explanation. Explanation comes to an end with an elucidation of the nature of a property. There is no informative answer to the question why an elucidation of the nature of a property is what it is – the answer can only be (something like) ‘because that’s what it fundamentally *is* to be that property.’ Thus, we do not need to invoke any theory beyond PNS itself in order to explain the truth of S1. This kind of explanatory simplicity is an additional and important virtue of PNS; moreover, it is a virtue that is not shared by any previously proposed theory of the nature of moral wrongness. Indeed, PNS seems *uniquely* poised to explain the phenomenon described by S1, and thus it represents a distinct and plausible prospect for an elucidation of the nature of moral wrongness.

Furthermore, PNS avoids the kind of circularity problem that plagues standard normative sentimentalism. Recall the problem: sentiments that are justified *in response* to the instantiation of moral properties (e.g. guilt and indignation) are themselves plausibly analyzed in terms of the moral properties in question (e.g. moral wrongness). Standard normative sentimentalist theories analyze moral properties in terms of such sentiments, and thus run a high risk of circularity. PNS, however, does not invoke reactive sentiments – unlike guilt and indignation, for example, compassion and respect are not justified *in response* to moral wrongs, but rather (simply) dispose people to refrain from morally wrong behavior. And, there is no philosophical pressure to analyze compassion and respect *in terms* of the wrongness of the behavior they dispose against; rather, the acts ruled out by having a certain degree of those sentiments *just are* the morally wrong acts – or so I suggest.¹⁴

3.3. Ruling out action

PNS relies crucially on the notion that having certain sentiments to certain degrees of strength *rules out* the performance of certain acts. This requires explication. The relevant notion of ruling out is simply that of being prevented or precluded by the particular motivational elements involved in having the sentiments in question to the relevant degrees. The idea is that having the degree of compassion and respect that one ought to have entails having a certain motivational profile – a set of motivations of particular strengths to perform and to refrain from performing various acts (or act-types) – which is so composed that it precludes the performance of certain acts. That is, due to the particular makeup of the motivational profile in question (and the strength of the motivational states therein), one simply *cannot* have such a profile while being motivated *to the point of action* to perform certain acts.

But, one might question, are the motivational aspects of compassion and respect so tightly connected to the degree to which those sentiments are had?

In particular, might a person have the proper degree of compassion and respect while *not* being strongly motivated in the ways characteristic of those sentiments as long as *other* aspects of those sentiments, such as their purely affective dimensions (i.e. their 'raw feels'), manifest prominently? That seems to be a coherent view, and perhaps it is not implausible. However, there are reasons to favor a view more friendly to PNS. In particular, it seems that one simply *cannot* have a *high* degree of compassion or respect without having strong associated motivations, affective responses notwithstanding. For example, one cannot have a high degree of compassion for someone while lacking significant motivation to help that person avoid suffering, regardless of how strong one's affective response is to that person's suffering. Or, for instance, if one is not strongly motivated to share equally the fruits of equal labor with one's collaborator (all else being equal, e.g. desert and need), then one simply does not have a high degree of respect for that person. These thoughts lend significant credence to the PNS-friendly view: that in order to have a certain degree of compassion or respect, one *must* be motivated to a certain degree to perform the acts characteristic of having those sentiments. In any case, I submit, that view is at least sufficiently plausible to maintain the plausibility of PNS.

But what about the idea that having compassion and respect to certain degrees can actually *rule out* particular behavior? One skeptical of such a notion might argue as follows: even if having a certain degree of compassion and respect entails having a particular level of motivation not to perform immoral acts, one might still commit such an act while having *any* degree of compassion and respect if one is motivated to a strong enough degree *to* commit that act.

A plausible response to this worry holds that sentiments function to control motivation not simply by *providing* motivation, but also by *regulating* motivations whose source is external to those sentiments. On this view, having a certain level of compassion and respect not only involves having certain motivations of certain degrees of strength not to perform callous and disrespectful acts, but also involve a *suppression* of one's motivation *to* perform such acts (cf. Vogelstein 2011). Indeed, this view fits with our common-sense conception of what is involved in having those sentiments. For example, it stands to reason that one simply cannot be highly motivated (especially, motivated to the point of action) to perform heinously callous acts while having a very high level of compassion – that is simply not what someone filled with compassion would be strongly motivated to do, and certainly not what such a person would *do*. Thus, it should be plausible that part of the way in which compassion and respect function to influence one's overall motivational profile is by suppressing the strength of one's desires to commit certain sorts of acts; and that readily allows us to accept that having a great enough degree of compassion and respect can indeed rule out certain behavior.

3.4. *Sentiments, reasons, and oughts*

At this point, a key question about PNS remains: what is referred to by the relevant 'ought'? When we say that the callous person *ought* to have more compassion, and that the disrespectful person *ought* to have more respect, what are we saying? It might be thought that this is a purely *moral* 'ought' – that there is some moral failing on the part of one who fails to have compassion and respect to the relevant degrees. If that were the whole story, then, it seems, PNS would be rendered circular; but it is plausible that that is not the whole story. Even if it is the case that the callous and disrespectful fail in some moral respect by lacking compassion and respect to certain degrees, that such people *ought* to have had one or both of those sentiments to a greater degree seems to imply something more: that such people have failed to properly respond to their *reasons* for having those sentiments. One way of drawing out that implication is by recognizing that those who fail to have the degree of compassion or respect that they ought to have are justifiably *criticizable* on that basis, and, plausibly, a necessary condition of justifiable criticism is a person's failure to properly respond to her best reasons (Joyce 2001; Vogelstein 2013). If that is correct, then in saying that the callous or disrespectful person ought to have more compassion or respect, we are (at least in part) saying that such a person has *most reason* to have those sentiments to a greater degree.

In that case, the reasons at issue in PNS will be so-called *right-kind* reasons as opposed to *wrong-kind* reasons. Roughly, a wrong-kind reason to have attitude A is a reason that derives from the value of merely *having* A, regardless of what sort of attitude A happens to be (e.g. prudential reasons to believe).¹⁵ A right-kind reason to have A, on the other hand, is a reason that is tailored to the type of attitude that A is. Such reasons bear upon the fit, given by the nature of the kind of attitude in question, between that attitude and its object (e.g. evidential reasons to believe). Thus, in the same sense in which one ought *not* to be afraid in the absence of danger, or feel guilty when one has done nothing wrong, one ought to have compassion and respect for beings like us. And according to PNS, one ought to have those sentiments to a degree sufficient to rule out certain acts, and those very acts just are the acts that are morally wrong. In that way, the nature of moral wrongness is elucidated in terms of (right-kind) reasons for having compassion and respect.¹⁶

3.5. *Moral normativity*

In this section, I describe an additional and important advantage of PNS – one that derives from the notion that there is a special connection between morality and normative reasons. It is a commonplace that morality is genuinely and robustly normative, and a common way to understand that idea is via the thesis that morality involves a necessary and universal relation to normative reasons.

PNS gives us a novel account of that relation – an account that satisfies the dictum that necessarily, one who does wrong is not in conformity with her reasons, but an account that does *not* imply that those reasons are reasons against performing the immoral act. For all PNS says, the *only* way in which violating the dictates of morality implies that one has failed to conform to one's reasons is that performing immoral acts entails that one has failed to conform to one's right-kind reasons for having compassion and respect – there may be no necessary connection between immoral action and reasons *for action*.¹⁷ But this is a welcome result, due to the plausibility of *Humean* theories of reasons for action, according to which reasons for action require the presence of a desire that would be served by such action (Williams 1979; Hubin 1999, 2001; Schroeder 2007a, 2007b). Humeanism places considerable pressure on the traditional view of moral normativity (according to which moral reasons are reasons for action). After all, if reasons for action require the presence of desires that would be served by such action, then it is hard to see how any reasons for action could be necessary and universal, given that desires are not. Therefore, if moral reasons are reasons for action, Humeanism ought to be rejected, or else we ought to accept that necessarily, every moral agent has a desire that would be served by refraining from morally wrong action. Since the latter claim is highly controversial (if not simply implausible), and since Humeanism is a plausible and prominent view, the fact that the traditional view of moral normativity requires either accepting the latter claim or rejecting Humeanism is a significant cost of the traditional view.¹⁸

The account of moral normativity implied by PNS, on the other hand, allows us to preserve the necessary and universal connection between morality and normative reasons without incurring the controversial theoretical commitments at issue. Humeanism, as a thesis about reasons *for action*, does not apply to the reasons in PNS (and thus would not impugn the necessary and universal applicability of those reasons). Furthermore, there is no analogous, plausible 'Humeanism' about sentiments which the defender of PNS should be compelled to reject and whose rejection would be controversial – right-kind reasons for sentiments, and for compassion and respect in particular, do not plausibly depend on one's desires. Such reasons are reasons of *fit*, and whether a sentiment is a fitting response to the circumstances should not depend upon what one happens to want. For instance, the fact that you are suffering is a reason for me to have compassion towards you, and the fact that you are an autonomous agent is a reason for me to have respect for you, regardless of whether I have any desire that would be served by having such sentiments. The upshot is this: even if one's reasons for action are always contingent on one's desires, and thus even if we cannot necessarily say of someone who did wrong that she had most reason to do otherwise, or even *any* reason to do otherwise, given PNS we can, without incurring controversial commitments about practical reason, always say of such a person that she had most reason to have the compassion and

respect that would have precluded such action, and thus that in doing wrong she has failed to conform to her best reasons. This view of moral normativity represents a novel and important middle-ground, and is a significant and unique virtue of PNS.

Furthermore, the fact that PNS allows us to explain why morality is necessarily and universally normative *in general* is an advantage PNS has over any theory that lacks such a feature, including standard normative sentimentalism. According to standard normative sentimentalism, it remains unexplained why morality has any necessary and universal relation to normative reasons, such that when one does wrong one has failed to conform to one's reasons. To be sure, standard normative sentimentalism can explain the relationship between morally wrong action and reasons for having reactive sentiments such as guilt and indignation (although see Section 3.6 below for an explanation of why that is not a significant advantage of the theory); but that a person behaves in a way that justifies such reactions entails nothing about that person's failure to conform to her reasons – for all standard normative sentimentalism implies, a person might do wrong, feel guilty, and conform perfectly to her reasons (reasons for action as well as for having sentiments). Of course, we ordinarily have a reason to avoid feeling guilty, since guilt is usually experienced as unpleasant; but it seems possible that someone could lack such a reason. The upshot is that it does not follow from standard normative sentimentalism itself that doing wrong *necessitates* failing to conform to one's reasons. Thus, standard normative sentimentalism fails to explain the right sort of connection between morality and reasons. And it is unclear exactly how such a connection might be explained if not via the nature of the moral properties themselves – thus any theory of the nature of moral properties that can explain that connection, such as PNS, has a significant *prima facie* advantage over a theory that cannot, such as standard normative sentimentalism.

3.6. Other considerations for ethical theory

An important additional factor in the evaluation of any moral theory is the extent to which the theory coheres with intuitively strong first-order moral judgments, i.e. whether the theory appears to be extensionally correct. Here, both PNS and standard normative sentimentalism do well. Indeed, as was noted previously, standard normative sentimentalism's chief theoretical virtue is its apparent extensional correctness. PNS likewise holds good prospects for being extensionally correct. That is because (I submit) it will be difficult to imagine a case in which either (1) an act was wrong but we are intuitively inclined to think the agent had just the degree of compassion and respect that he ought to have had, or (2) an act was morally permissible yet we are intuitively inclined to think that if the agent had the compassion and respect he ought to have had, he would not have committed such an act. Thus, neither PNS nor standard normative

sentimentalism seems to run counter to our intuitive judgments about the morality of particular actions.

That said, it should be noted that standard normative sentimentalism does aim to explain a phenomenon that PNS does not aim to explain: the connection between the normative features of *reactive* sentiments and the instantiation of moral properties, e.g. that an act is morally wrong just in case guilt is a justified response to such an act. For example, a standard normative sentimentalist theory according to which to be a morally wrong act *just is* for guilt to be a justified response to that act explains *why* an act is morally wrong *if and only if* guilt is justified in response to it (which is a plausible thesis in its own right). PNS does not explain why that sort connection would hold. However, that explanatory ability of standard normative sentimentalism is not a *significant* point in its favor. To see *why*, recall from our discussion of the circularity worry for standard normative sentimentalism (Section 2.2) that there is a highly plausible explanation of the relevant connection between the normative features of reactive sentiments and morally wrong action that could obtain if standard normative sentimentalism were false: that such sentiments are themselves to be analyzed in terms of moral wrongness such that having those sentiments is justified just as a response to morally wrong behavior (e.g. by representing that an act was morally wrong). And here is why that is relevant: the fact that a theory is able to explain a particular phenomenon is a virtue of that theory only to the extent to which the phenomenon will be insufficiently explained if the theory is false – if the phenomenon has a perfectly good explanation even without that theory, then the theory's explanatory ability does not significantly raise the probability that the theory is true. Therefore, the ability of standard normative sentimentalism to explain why certain sentiments are justified if and only if certain moral properties are instantiated fails to be a *significant* virtue of that theory, since that fact has a highly plausible explanation in any case. That is not to say that standard normative sentimentalism's explanatory ability here is *no* advantage of the theory – after all, the other explanation (based on the natures of the sentiments themselves), even if highly plausible, might be false, in which case standard normative sentimentalism might indeed provide the correct explanation of the relevant phenomenon. But so long as the other explanation is highly plausible, standard normative sentimentalism's explanatory advantage here will be minor.

Furthermore, note that there appears to be no analogous highly plausible alternate explanation of S1 – that is an important part of why PNS's ability to simply and powerfully explain S1 (indeed, its unique ability to do so) is a strong point in its favor, and is a key reason why PNS stands in contrast with, and is superior to, standard normative sentimentalism *vis-à-vis* their respective explanatory virtues: S1 is in *need* of explanation, while the key *explanandum* for which standard normative sentimentalism is the *explanans* – the fact that an act is wrong if and only if relevant sentiment(s) are justified in response to such an act – is *not* in need of explanation.

3.7. *Moralized compassion and respect?*

In this section I respond to a final objection, based on the idea that compassion and respect are to be analyzed in moral terms after all. The objection is motivated by the initially plausible thought that one ought to have *less* compassion and respect for morally repugnant individuals than one ought to have for others, all else being equal. Indeed, if a person is morally bad enough, perhaps one ought to feel *no* compassion and respect towards that person; at the very least, it seems that one should mitigate one's compassion and respect in proportion to the moral badness of the individual – the worse a person is, morally speaking, the less right-kind reason there is to feel compassion or respect for him. But if that is correct, then a problem for PNS emerges: it seems, in that case, that the natures of compassion and respect must themselves involve moral properties. After all, the natural way to explain *why* one ought to tailor one's compassion and respect in response to certain moral features of a person is that those sentiments themselves are *moralized*, e.g. that they represent or are responsive to, at least in part, the moral properties instantiated by people (or their actions, motives, character, etc.). If that is correct, then PNS faces a circularity problem of its own.

Despite the initial appeal of the view, however, it is dubious whether one should indeed temper one's compassion and respect for the morally bad. Notably, the ideal according to which compassion and respect should *not* be mitigated in response to wrongdoing is a central tenet of certain cultural, philosophical, and religious traditions, e.g. many forms of Buddhism and Christianity (cf. Conway 2001; Cates 2003).¹⁹ Moreover, we can plausibly explain why it merely *seems* as if such mitigation is warranted: wrongdoing plausibly *does* warrant certain sentimental responses that themselves naturally diminish the level of compassion and respect one is inclined to have for the wrongdoer, such as anger (indignation and resentment), disrespect, and moral disgust. Plausibly, when we say of an evil person that one shouldn't feel compassion or respect for him, what we really mean is that one *should* feel anger, disrespect, or disgust towards him to a degree that would rule out feeling compassion or respect (or rule out having those sentiments to a significant degree). And that does not imply that one's right-kind reasons for having compassion and respect for the person are weaker than they otherwise would be – the strength of those reasons is not diminished by the existence of right-kind reasons (even conclusive reasons) for having *other* sentiments that naturally preclude or mitigate one's compassion and respect.²⁰ Thus, due to the ease with which we may be inclined to conflate the strength of our reasons for having reactive sentiments that diminish one's compassion and respect with a *lack* of reason to have compassion and respect themselves, we should be highly skeptical of the intuition that we have less reason to have compassion and respect for wrongdoers or for the morally bad.

Furthermore, there is a response open to the preventionist even supposing that compassion and respect *should* be had differentially according to the moral

features of agents. Such a phenomenon can be explained not by supposing that the sentiments themselves are *moralized*, but rather by invoking a *recursive* account of compassion and respect according to which they are representative of or responsive to the *degree of compassion and respect* that an agent has. On this view, compassion represents or is responsive to misfortune suffered by those who themselves have an appropriate degree of compassion and respect, while respect represents or is responsive to the interests and desires of those who likewise have an appropriate degree of those sentiments. Thus, if A lacks the degree of compassion or respect that A ought to have, the compassion or respect that we have towards A ought to be mitigated as well. In this way, we can explain why we should have lower levels of compassion and respect towards those who do wrong – because such people lack the degree of compassion and respect that would rule out such behavior, and the strength of our reasons for having compassion and respect depends upon the degree to which their target has those very sentiments – without having to invoke moral properties themselves in an elucidation of the nature of compassion or respect.²¹

To be sure, this is simply a sketch of a strategy that preventionists may employ in response to the kind of objection under consideration, and perhaps the details of the strategy's application will reveal it to be less plausible than I have suggested. For example, there is a legitimate question about whether this strategy results in a problematic regress, i.e. whether a recursive account may acceptably be invoked as an elucidation of a property's nature, given how I understand that task. But in concert with the first response (that it is plausible that compassion and respect are *not* justifiably mitigated in response to wrongdoing) this additional strategy comprises one component of a two-pronged reply to the objection at issue – the idea being that it is plausible that *one* of the two responses succeeds (i.e. that their disjunction is true) – a reply which, I suggest, when taken as a whole weakens the objection enough to maintain the plausibility of PNS.²²

4. Taking stock

At this point, we may tally the costs and benefits of PNS, both in relation to its chief sentimentalist competitor, standard normative sentimentalism, as well as with an eye towards assessing its overall plausibility.

Benefits of PNS that standard normative sentimentalism lacks

- (1) PNS simply and powerfully explains S1.
- (2) PNS (a) explains the commonplace that morality is necessarily and universally normative, i.e. why doing wrong entails that one has failed to conform to one's reasons, and (b) does so in a way that avoids

controversial commitments about practical reason (either to the rejection of Humeanism, or to the view that all moral agents necessarily have desires that would be satisfied by refraining from doing wrong).

Benefit of standard normative sentimentalism that PNS lacks

- (3) Standard normative sentimentalism explains why, for certain sentiments, an act is wrong just in case those sentiments are justified in response to such an act. However, this is a minor virtue of standard normative sentimentalism because that phenomenon has a highly plausible alternative explanation.

Cost of PNS that standard normative sentimentalism lacks

- (4) PNS relies on a potentially controversial (albeit plausible) view of the relationship between degrees of compassion and respect and their characteristic motivations: that having a particular degree of compassion or respect *entails* being motivated, in the way characteristic of that sentiment, to a particular degree of strength.

Benefits of both PNS and standard normative sentimentalism

- (5) Both theories accord with intuitively strong first-order moral judgments – they do not appear to suffer from any obvious counterexamples.

Cost of both PNS and standard normative sentimentalism

- (6) Both views are susceptible to circularity worries, although the problem is more acute for standard normative sentimentalism than for PNS. That is because it is not clear how to understand the natures of sentiments that are justified just as responses to the instantiation of moral properties if *not* via the thesis that the natures of those sentiments are themselves elucidated in terms of moral properties. The notion that compassion and respect are moralized, on the other hand, is highly controversial, and the intuition that drives that view (the sense that we should have less compassion and respect for wrongdoers) is explicable in ways that are consistent with PNS, either via a debunking explanation of the intuition, or via a recursive account of compassion and respect. Thus, standard normative sentimentalism, but not PNS, suffers from a *strong* circularity worry (even though there is *some* circularity worry for PNS).

The tally

PNS:

(1), (2), and (5) are major virtues

(4) is a minor cost

(6) is a minor to moderate cost

Standard normative sentimentalism:

(3) is a minor virtue

(5) is a major virtue

(6) is a major cost

Thus, PNS has three major benefits and two minor costs (or, at worst, one minor and one moderate cost), while standard normative sentimentalism has one major benefit, one minor benefit, and one major cost. Thus, the balance of costs and benefits stands in favor of PNS in comparison with standard normative sentimentalism. Moreover, (a) the benefits of PNS outweigh its costs, overall, and (b) two of PNS's major virtues – (1) and (2) – are enjoyed by PNS uniquely; thus PNS stands as a significant and plausible theory in its own right.

5. Further explication

In this final substantive section, I resolve some remaining issues about how we ought to understand PNS, in light of (a) the fact that compassion and respect might motivate towards different actions, (b) the possibility of having those sentiments to *too great* a degree, and (c) the distinction between subjective and objective moral wrongness. I discuss (a) and (b) (which are importantly related) in the following subsection; (c) is addressed in subsection 5.2.

5.1. *Conflicting and overabundant compassion and respect*

Paradigmatic immoral acts are criticizable in virtue of the fact that the immoral actor failed to have a *sufficient* degree of compassion and respect. As has been discussed, callous acts routinely involve a lack of sufficient compassion, and non-callous yet immoral acts involve a lack of sufficient respect. But it might also be that immoral acts involve a lack of the degree of compassion or respect that one ought to have not because one should have a *greater* degree of one of those sentiments, but because one has *too much* of one or the other.

Of course, in many cases having too much compassion or respect does *not* result in immoral action – after all, in many situations the actions that are ruled out by having the degree of compassion and respect that one ought to have will *also* be ruled out by having *even more* compassion or respect. But the fact that the motivational forces of compassion and respect can, in certain cases, pull in opposite directions opens the possibility that having too much of one or the other will result in a motivational profile that does *not* rule out the actions

that are ruled out by the correct degree of those sentiments. Let us explore that possibility.

Consider cases of punishment. Compassion might motivate towards mercy, while respect – in particular, respect for those who have been harmed or otherwise wronged by the immoral act – might motivate towards punishment. Or, consider a case in which honoring a person's autonomous choice entails causing or allowing significant harm to that very person. Compassion might motivate towards paternalism, i.e. contravening the person's choice in order to prevent the relevant harm, while respect for that person might motivate helping or letting him achieve what he has chosen despite it being a detriment to his welfare. In cases in which compassion and respect have contrary motivational valences, one of those sentiments will 'win out,' motivationally speaking. Note, in PNS the relevant 'preventer' of immoral action is the *combination* of the compassion one ought to have *and* the respect one ought to have. Thus, even if the sentiments involved in that combination pull in opposite directions in particular cases, the combination itself is able to rule out action based on the stronger and thus 'winning' motivational element.

But if one has *too much* compassion or respect, the winning motivation might *not* be in concert with what would result from the degree of compassion and respect that one ought to have. For example, suppose that punishment in a given case is morally required; an overabundance of compassion for the criminal might result in withholding that punishment. Or, suppose that punishment in a particular case is morally wrong; an overabundance of respect for the victims might result in providing that punishment. Similarly with paternalism: too much compassion might result in wrongfully paternalistic acts (i.e. the compassion one feels for a person, and the resulting motivation to prevent harm to that person, might overwhelm the respect one ought to have for her autonomous choice), while too much respect might make one refrain from paternalism that is morally called for, e.g. if a person's choice will result in serious and irreparable harm to themselves.

The upshot is that PNS does *not* imply that having the degree of compassion and respect that one ought to have *just is* having compassion and respect *above* a certain level. Rather, in order to avoid problems of overabundance, generated by the fact that the motivational forces of the two sentiments can sometimes pull in opposite directions, one must have the *precise* degree of compassion and respect that one has most right-kind reason to have – the degree that is called for by the situation and the relevant features of the parties involved.²³

5.2. Subjective and objective morality

A final complication with PNS arises from the fact that the actions that are ruled out by having compassion and respect to particular degrees depend upon one's

beliefs. No action is ruled out by having a certain degree of compassion and respect per se – rather, actions are ruled out by combinations of those sentiments and beliefs that are relevantly related to the motivational attitudes those sentiments involve. Suppose, for example, that I am exorbitantly compassionate and respectful, and that I have a choice of whether to push a button. Whether or not I push the button will depend on what I believe about the consequences of doing so. If I believe that pushing the button will cause 100 innocent people to needlessly suffer and die, I will not push the button; if I believe that pushing the button will *save* 100 people from needless suffering and death, I will push the button. Thus, we must specify PNS in terms of particular sets of beliefs that, in concert with the degree of compassion and respect that one has most reason to have, would indeed rule out certain action.

One natural thought is to specify PNS in terms of the beliefs one *actually* has. If, for example, one believes that pushing the button will save 100 people, even though it will in reality have the opposite effect, then there is a clear sense in which one does nothing morally wrong by pushing the button; indeed, it seems one does something morally wrong by *not* pushing the button. The sense in which it is wrong not to push the button in such a case is the *subjective* sense of moral wrongness. Roughly, the subjective wrongness of an act is determined by one's doxastic state (one's beliefs and/or credence levels). Objective wrongness, on the other hand, is determined by what is actually the case – thus, in this case, it would be objectively wrong to push the button due to the terrible consequences of doing so. So specifying PNS in terms of one's actual beliefs would seem to provide an account of subjectively morally wrong action, the idea being that subjectively wrong action just is action that is ruled out by the combination of (1) one's actual beliefs, and (2) the degree of compassion and respect that one has most reason to have.

But things are a bit more complicated, because the reasons that attach to sentiments also admit of a distinction between subjective and objective varieties, where objective reasons are based on what is actually true, while subjective reasons are based on one's doxastic state – roughly, subjective reasons can be thought of as the objective reasons one *would* have if one's beliefs were true, or as the objective reasons that are entailed by what one believes.²⁴ Thus we must specify which of those types of reasons – subjective or objective – is relevant to subjectively morally wrong action; and not surprisingly, the plausible answer is that subjective reasons are the relevant kind. We can see why objective reasons will not do by considering the following case:

I believe (1) that there are 1,000,000 people in chamber A, (2) that the only way to prevent those in chamber A from suffering and dying needlessly is to push the button, and (3) that there are three people in chamber B who would be killed by pushing the button. In reality, (1) and (2) are false: the beings in chamber A are in fact automatons and lack any form of sentience or rationality that would justifiably ground moral concern; (3), however, is true.

In this case, I have most *objective* reason to have compassion and respect only for those in chamber B, but most *subjective* reason to have compassion and respect for both those in chamber A and chamber B. Thus, given my actual beliefs, were I to have the compassion and respect that I have most objective reason to have I would not press the button (since I wouldn't be concerned with the beings in chamber A); but, it seems, I subjectively ought to press the button. If, however, I were to have the compassion and respect that I have most subjective reason to have, then I would indeed press the button (since in that case I would have compassion and respect for those in chamber A, and would, presumably, be more strongly motivated to prevent one million people from suffering and dying than to save the three people in chamber B). Therefore, the kind of reasons relevant to PNS, if PNS describes the nature of subjective moral wrongness, must be subjective reasons. Thus, we have:

Subjective preventionist normative sentimentalism (S-PNS): The nature of *subjective* moral wrongness is elucidated as follows: Necessarily, for all A, x: It is subjectively morally wrong for A to do x if and only if A's doing x would be ruled out by the combination of (1) A having the degree of compassion and respect that A has most subjective reason to have, and (2) A's actual beliefs and credences.

Subjective moral wrongness is supposed to be relative to one's doxastic state, and S-PNS captures that idea. Objective moral wrongness, on the other hand, is supposed to be relative to what is true; we can thus describe objective wrongness in terms of the compassion and respect one has most *objective* reason to have (reasons which are grounded on what is true, as opposed to one's doxastic state), and in terms of an agent having true beliefs and no false beliefs:

Objective preventionist normative sentimentalism (O-PNS): The nature of *objective* moral wrongness is elucidated as follows: Necessarily, for all A, x: It is objectively morally wrong for A to do x if and only if A's doing x would be ruled out by the combination of (1) A having the degree of compassion and respect that A has most objective reason to have, and (2) A believing all truths and no falsehoods.

We have already seen how S-PNS accounts for the fact that it is subjectively wrong for me *not* to push the button. Likewise, O-PNS accounts for the objective wrongness of *pushing* the button – if I were to believe, as is the case, that only those in chamber B are sentient and that pushing the button will cause them to die, and if I were to have the compassion and respect for those beings that I have most objective reason to have, I would not push the button. Thus, S-PNS grounds subjective wrongness in terms of the beliefs one has, while O-PNS grounds objective wrongness in terms of what is actually true; and that is just what we want out of an account of the distinction between subjective and objective wrongness.

6. Conclusion

Despite significant problems with standard realist sentimentalist accounts of the nature of moral properties, the nature of moral wrongness is plausibly

elucidated with reference to the normative features of sentiments via a *preventionist* theory that invokes the sentiments of compassion and respect. There is a strong basis for such a view due to the fact that morally wrong action routinely involves agents lacking the degree of compassion or respect that they ought to have. In addition, a preventionist theory allows us to account for the normativity of morality in a way that does not depend upon controversial views about practical reason, avoids the major circularity worry that plagues its chief sentimentalist competitor, and does not appear to suffer from any obvious counterexamples. Like any moral theory, preventionism is not problem-free; but its costs are moderate at worst, and are outweighed by its virtues. Thus, PNS and its specifications S-PNS and O-PNS not only represent an improved prospect for a sentimentalist analysis of the nature of moral wrongness, but should be quite plausible in their own right.

Notes

1. Here, 'sentiments' refers to a host of attitudes that bear a family resemblance, such as anger, joy, guilt, shame, love, sadness, jealousy, fear, pride, disgust, respect, and compassion. Sentiments are (at least) mostly, but need not be entirely, coextensive with the category of *emotions*.
2. I assume for the purposes of this paper that there is indeed a *property* of moral wrongness.
3. My thoughts about elucidating the nature of a property have been influenced by the work of Ralph Wedgwood (2007) and Mark Schroeder (2005, 2007a, 2007b), who themselves draw on the work of Kit Fine, and in Schroeder's case, Jeffery King (e.g. Fine [1994] and King [1998, 2002]). See these works for more detailed views on property analysis.
4. This is to be distinguished from *conceptual* or *semantic* analysis, in which the concept or meaning of the relevant term is analyzed.
5. This (rough but useful) understanding of property analysis will allow us to appreciate the full force of an important circularity problem for prominent sentimentalist views, which I discuss in the following section.
6. See e.g. Lewis (1989) and Brower (1993) for defenses of dispositionalism. McDowell (1985), Wiggins (1987), and Johnston (1989) defend versions of standard normative sentimentalism. Gibbard (1992) defends a similar view, although he treats normative language as noncognitive. Recent critique can be found in D'Arms and Jacobson (1994, 2000a, 2006).
7. In rough form, compassion can be understood as involving at least (1) some sort of emotional pain or distress directed at (what one believes to be, or perceives, or imagines as) the misfortune of another, and (2) an intrinsic desire to alleviate that misfortune. I shall not commit here to a full-fledged theory of compassion, nor do my arguments require any such commitment (for work on the nature of compassion, see e.g. Blum 1980; Nussbaum 2001; Cates 2003; Deigh 2004; Weber 2004; Cannon 2005; Crisp 2008).
8. Of course, acts of these kinds may very well be callous; but nothing implies that they must be. Theft, for example, may result in something less than significant harm to another, even if it does result in some harm. More generally, an act that

results in a decrease in well-being need not result in significant harm, but may result merely in annoyance, inconvenience, discomfort, or dissatisfaction.

9. The type of respect relevant to morality and thus to my analysis of moral wrongness is a form of *recognition* respect, as opposed to *appraisal* respect (cf. Darwall 1977). Appraisal respect involves thinking highly of someone; recognition respect, on the other hand, involves *due consideration* and the disposition to refrain from behavior one would otherwise engage in. The type of respect relevant to morality involves due consideration of other people's *desires* and *interests*, and the disposition to curtail *selfish* behavior. It is what goes under the famous label 'respect for persons.' It assigns to the desires and interests of others a baseline level of importance – when I have respect for someone, the desires and interests of that person *matter* to me, and do so because they matter (or should matter) to the person whose desires and interests they are.
10. Talk of *degrees* (or *levels*, *amounts*, etc.) of compassion and respect is simply a way of describing the fact that one can have *more* or *less* compassion or respect.
11. In Section 3.3, I explain how having a particular degree of compassion and respect can indeed rule out certain behavior.
12. Just what sort of 'ought' is at issue in PNS will be discussed in Section 3.4.
13. As I take it, the explanatory power of a theory *T* vis-à-vis an *explanandum* *S* is a matter of how probable *S* is given *T*. If *T* entails *S* then the probability of *S* given *T* is 1, and since that probability is maximally high the explanatory power of *T* vis-à-vis *S* will also be maximally high. But we should be sure not to conflate *explanatory power* and *overall explanatory value* – part of what makes an explanation a good one overall is its explanatory power, but other factors are relevant as well, e.g. the explanation's simplicity.
14. In Section 3.7, I respond to an objection according to which compassion and respect are to be analyzed in moral terms after all.
15. Although describing wrong-kind reasons in this way is common, it may be only roughly accurate; see e.g. Hieronymi (2005) and Schroeder (2012).
16. It might be thought that the attitude of respect is not a sentiment, properly construed, because it lacks an affective dimension, and thus that PNS is not properly or fully a sentimental theory. In response, I would suggest that respect indeed has an affective component – a distinctive 'feel' – even if it is subtle; furthermore, an affect-requirement for being a sentiment is controversial. In any case, whether respect is properly classified as a sentiment is not important for my purposes – what matters is that the attitude of respect answers to *reasons*. If that criterion is satisfied, then I am happy to accept that mine is a hybrid theory that analyzes moral properties partially in terms of the sentiment of compassion and partially in terms of the non-sentimental attitude of respect.
17. I argue for this sort of view of moral reasons in Vogelstein (2011).
18. It is worth noting that some philosophers have attempted to identify some desire or set of desires that everyone necessarily has, and which is necessarily satisfied by refraining from immoral action (e.g. Korsgaard 1996a, 1996b; Schroeder 2007a). Evaluating these proposals would take us too far afield for present purposes; suffice it to say that they have not enjoyed wide acceptance.
19. Conway (2001) argues that the notion that one should not have compassion towards those who do wrong (or that such compassion should be mitigated) is a modern Western cultural bias and thus fails to support the contention that the norms intrinsic to compassion require anything other than universal compassion for all beings that are capable of suffering misfortune.

20. Indeed, it may simply be impossible to have *all* the sentiments that one has most right-kind reason to have to the degree to which one has most reason to have them; e.g. perhaps one cannot have both the degree of indignation one has most reason to have towards a wrongdoer and the degree of compassion one has most reason to have for him.
21. It might be thought that this response is inconsistent with the idea that (recognition) respect assigns to people's desires and interests a *baseline* level of importance – a gloss which could suggest that the reasons for having respect must exist *unconditionally* and thus irrespective of whether the person themselves has a sufficient modicum of compassion or respect (I thank an anonymous reviewer for raising this issue). This thought, however, would be mistaken – although it represents one coherent view (and a view which, I should note, supports my previous response to the objection at issue), it is likewise coherent to suppose that the relevant baseline is not *fixed*, but might be determined by such things as the degree of compassion or respect that the person themselves has.
22. One might worry that a different circularity problem emerges from the plausible notion that reasons to have compassion and respect can be *moral* reasons. Specifically: given that we have reasons to refrain from doing what's morally wrong, we have moral reasons to have compassion and respect, because having those sentiments will help us avoid doing what's morally wrong; and if those are some of the right-kind reasons at issue in PNS, then, it could be argued, circularity looms. Note, however, that such reasons would not be right-kind reasons for having compassion and respect, i.e. reasons of *fit* – because they derive merely from the instrumental value of having those sentiments (albeit instrumentality vis-à-vis avoiding morally wrong behavior), as opposed to being reasons that speak to the fittingness of the sentiments themselves, they are wrong-kind reasons, and thus would not be included in PNS. I thank an anonymous editor for raising this issue.
23. One might object to the suggestion that there can be a *precise* degree to which one has compassion or respect. Of course, there is no unit of measure or precise method for measuring the degree or level of a person's compassion or respect, or of any sentiment (although there might be in principle). But if compassion and respect indeed come in degrees, i.e. if a person can have *more* or *less* compassion or respect, then the idea that such sentiments come in precise degrees should not be problematic, even if it is limited in its application; indeed, it is difficult to see what the plausible alternative would be – if degrees of compassion and respect are not precise, then they are vague or indeterminate, and the claim that compassion and respect can only be had to indeterminate degrees of strength seems far more problematic than the claim that such degrees are precise.
24. These are rough characterizations. For detailed accounts of subjective reasons, see Vogelstein (2012), Whiting (2014), and Sylvan (2015).

Acknowledgments

For helpful discussion and comments on earlier versions of this paper, I thank Mark Schroeder, Jonathan Dancy, Michael Slote, Justin D'Arms, and Daniel Jacobson; participants at the 2014 Pittsburgh Area Philosophy Colloquium, 2014 Georgetown University Philosophy Conference on Emotions and Emotionality, 2012 Pacific Division Meeting of the APA, and 2012 Fall Meeting of the Indiana Philosophical Association; and two anonymous reviewers and one anonymous editor for this journal.

Disclosure statement

No potential conflict of interest was reported by the author.

Notes on contributor

Eric Vogelstein is an assistant professor in the Department of Philosophy and the School of Nursing at Duquesne University in Pittsburgh, PA. He specializes in ethical theory and bioethics.

References

- Blum, Laurence. 1980. "Compassion." In *Explaining Emotions*, edited by Amélie Oksenberg Rorty, 507–518. Berkeley: University of California Press.
- Brower, Bruce W. 1993. "Dispositional Ethical Realism." *Ethics* 103: 221–249.
- Cannon, Laura. 2005. "Compassion: A Rebuttal of Nussbaum." In *Feminist Interventions in Ethics and Politics*, edited by Barbara S. Andrew, Jean Keller, and Lisa H. Schwartzman, 97–110. Lanham, MD: Rowman & Littlefield.
- Cates, Diana Fritz. 2003. "Conceiving Emotions: Martha Nussbaum's Upheavals of Thought." *Journal of Religious Ethics* 31: 325–341.
- Conway, Jeremiah. 2001. "A Buddhist Critique of Nussbaum's Account of Compassion." *Philosophy in the Contemporary World* 8: 7–12.
- Crisp, Roger. 2008. "Compassion and beyond." *Ethical Theory and Moral Practice* 11: 233–246.
- D'Arms, Justin, and Daniel Jacobson. 1994. "Expressivism, Morality, and the Emotions." *Ethics* 104: 739–763.
- D'Arms, Justin, and Daniel Jacobson. 2000a. "Sentiment and Value." *Ethics* 110: 722–748.
- D'Arms, Justin, and Daniel Jacobson. 2000b. "The Moralistic Fallacy: On the 'Appropriateness' of Emotions." *Philosophy and Phenomenological Research* 61: 65–90.
- D'Arms, Justin, and Daniel Jacobson. 2006. "Sensibility Theory and Projectivism." In *The Oxford Handbook of Ethical Theory*, edited by David Copp, 186–218. New York: Oxford University Press.
- Darwall, Stephen. 1977. "Two Kinds of Respect." *Ethics* 88: 36–49.
- Deigh, John. 2004. "Nussbaum's Account of Compassion." *Philosophy and Phenomenological Research* 68: 465–472.
- Fine, Kit. 1994. "Essence and Modality: The Second Philosophical Perspectives Lecture." *Philosophical Perspectives* 8: 1–16.
- Gibbard, Allan. 1992. *Wise Choices, Apt Feelings*. Cambridge: Harvard University Press.
- Hieronymi, Pamela. 2005. "The Wrong Kind of Reason." *Journal of Philosophy* 102: 437–457.
- Hubin, Donald. 1999. "What's Special about Humeanism." *Noûs* 33: 30–45.
- Hubin, Donald. 2001. "The Groundless Normativity of Instrumental Rationality." *The Journal of Philosophy* 98: 445–468.
- Johnston, Mark. 1989. "Dispositional Theories of Value." *Proceedings of the Aristotelian Society Supplementary Volume* 63: 139–174.
- Joyce, Richard. 2001. *The Myth of Morality*. New York: Cambridge University Press.
- King, Jeffery. 1998. "What is a Philosophical Analysis?" *Philosophical Studies* 90: 155–179.
- King, Jeffrey. 2002. "Two Sorts of Claims about Logical Form." In *Logical Form and Language*, edited by Gerhard Preyer and Georg Peter, 118–131. New York: Oxford University Press.

- Koons, Jeremy Randel. 2003. "Why Response-Dependence Theories of Morality Are False." *Ethical Theory and Moral Practice* 6: 275–294.
- Korsgaard, Christine. 1996a. *The Sources of Normativity*. New York: Cambridge University Press.
- Korsgaard, Christine. 1996b. *Creating the Kingdom of Ends*. New York: Cambridge University Press.
- Lewis, David. 1989. "Dispositional Theories of Value." *Proceedings of the Aristotelian Society Supplementary Volume* 63: 113–137.
- McDowell, John. 1985. "Values and Secondary Qualities." In *Morality and Objectivity: A Tribute to J.L. Mackie*, edited by Ted Honderich, 110–129. London: Routledge and Kegan Paul.
- Nussbaum, Martha. 2001. *Upheavals of Thought*. New York: Cambridge University Press.
- Rabinowicz, Wlodek, and Toni Rønnow-Rasmussen. 2004. "Strike of the Demon: On Fitting Pro-Attitudes and Value." *Ethics* 114: 391–423.
- Ross, W. D. 1939. *Foundations of Ethics*. Oxford: Clarendon Press.
- Schroeder, Mark. 2005. "Realism and Reduction: The Quest for Robustness." *The Philosophers' Imprint* 5. <http://hdl.handle.net/2027/spo.3521354.0005.001>.
- Schroeder, Mark. 2007a. *Slaves of the Passions*. New York: Oxford University Press.
- Schroeder, Mark. 2007b. "The Humean Theory of Reasons." In *Oxford Studies in Metaethics*. Vol. 2, edited by Russ Shafer-Landau, 195–219. New York: Oxford University Press.
- Schroeder, Mark. 2012. "The Ubiquity of State-Given Reasons." *Ethics* 122: 457–488.
- Sylvan, Kurt. 2015. "What Apparent Reasons Appear to Be." *Philosophical Studies* 172: 587–606.
- Vogelstein, Eric. 2011. "Morality, Reasons, and Sentiments." *Philosophical Studies* 155: 421–432.
- Vogelstein, Eric. 2012. "Subjective Reasons." *Ethical Theory and Moral Practice* 15: 239–257.
- Vogelstein, Eric. 2013. "Moral Normativity." *Philosophical Studies* 165: 1083–1095.
- Weber, M. 2004. "Compassion and Pity: An Evaluation of Nussbaum's Analysis and Defense." *Ethical Theory and Moral Practice* 7: 487–511.
- Wedgwood, Ralph. 2007. *The Nature of Normativity*. Oxford: Clarendon Press.
- Whiting, Daniel. 2014. "Keep Things in Perspective: Reasons, Rationality, and the A Priori." *Journal of Ethics and Social Philosophy* 8. <http://www.jesp.org>.
- Wiggins, David. 1987. "A Sensible Subjectivism." Chap. V in *Needs, Values, Truth*. Oxford: Blackwell.
- Williams, Bernard. 1979. "Internal and External Reasons." In *Rational Action: Studies in Philosophy and Social Science*, edited by Ross Harrison, 101–113. Cambridge: Cambridge University Press.