

*This is a penultimate draft. For a final draft see the draft in Metaphilosophy 31 (2010) pp. 189-213. The article is also reprinted in Virtue and Vice: Epistemic and Moral, ed. Heather Battaly (Blackwell, 2010).*

## Epistemic Malevolence

Jason Baehr

Loyola Marymount University

There is considerable structural symmetry between moral and intellectual character virtues. This is evident in the fact that many moral virtues have straightforward counterparts among the intellectual virtues. We speak, for instance, of moral but also of *intellectual* courage, honesty, integrity, fairness, autonomy, and humility. The same goes for moral and intellectual *vices*. A person can be a moral or intellectual coward, morally or intellectually dishonest, morally or intellectually unfair, and the like.<sup>1</sup>

In light of this symmetry, it is surprising that what is perhaps the paradigm moral vice has no obvious intellectual counterpart. The moral vice in question is malevolence. While there may be such a thing as “intellectual” or “epistemic malevolence,” this is hardly a familiar notion: it does not appear on any standard lists of intellectual vices; nor does it occupy anything like the central role in our thinking about intellectual vice that malevolence proper or “moral malevolence” occupies in our thinking about moral vice. This, then, suggests a notable structural *asymmetry* between moral and intellectual vice.

The aim of this paper is to explore this asymmetry in some depth. The discussion will be guided by three main questions: (1) *Is there a counterpart to malevolence proper or “moral malevolence” among the intellectual virtues?* (2) *If so, what does it amount to?* (3) *And why does it fail to occupy the central role in our thinking about intellectual vice that malevolence proper occupies in our thinking about moral vice?* To begin to answer these questions, we first

must attempt to get clear on the basic character of malevolence proper. This, it turns out, is no easy task. Accordingly, the first half of the paper is spent addressing a range of questions about malevolence proper. I then go on, in light of this discussion, to identify and illustrate a conception of *epistemic* malevolence. In the final part of the paper, I address the forementioned question concerning comparative role of epistemic malevolence in standard ways of thinking about intellectual vice as a whole.<sup>2</sup>

## 1. Malevolence Proper

I begin by with a very concise and general account of malevolence proper or “moral malevolence.” As I am thinking of it, malevolence is essentially or paradigmatically a matter of *opposition to the good as such*.<sup>3</sup> Its spirit is captured in the infamous manifesto of Milton’s Satan: “Evil be thou my good,” which, to better fit the definition just noted, may be recast as: *Good be thou my enemy*. While an acceptable first approximation, this conception bears considerable scrutiny and elaboration.

### 1.1 Opposition to the good

First, what *kind* of opposition is essential to malevolence? Very briefly, malevolence involves “opposition to the good” that is *robustly volitional*, *active*, and “*personally deep*.” To say that this opposition is “robustly volitional” is to say that it centrally and fundamentally involves the *will*. Malevolence it is not a mere *conviction* that the good should be opposed, nor a mere *preference* for such. Rather it involves a kind of hostility or contempt the good. To say that the opposition characteristic of malevolence is “active” is to say that it tends to issue in actual *attempts* to stop, diminish, undermine, destroy, speak out, or turn others against the good. It is not a passive orientation. Finally, the opposition in question is “personally deep” in the sense that it reflects the malevolent person’s fundamental cares and concerns. A malevolent person cannot

simply or easily give up or repudiate her malevolence: her opposition to the good is not a commitment or orientation that she can simply take or leave. Rather, it is central to her very identity or self-conception.

### 1.2 Opposition to the good *as such*

A second, trickier aspect of malevolence concerns the idea that it involves opposition to the good *as such*.<sup>4</sup> What exactly is it to be against the good—or anything, for that matter—“as such”?

While I cannot fully defend the view here, I think this notion is plausibly understood in terms of making something or someone an *enemy*. Just as we can choose to make someone a friend—to *befriend* another person—so too we can choose to make or “take up” another person or thing *as* an enemy. To coin a phrase, we can *enemize*. My suggestion is that to be opposed to X “as such” is to take or regard X as an enemy; it is to *enemize* X. This fits well with the idea that the kind of opposition characteristic of malevolence is “personally deep,” since our friends and enemies are among the things most personal to us. It also fits the gloss of Satan’s manifesto noted above: namely, “Good be thou my *enemy*.”

In a recent discussion of the vice of *malice*, Robert Adams (2007) suggests an alternative conception of what it is to be opposed to something “as such.” He defines malice as “opposition to the good for its own sake” and maintains that a person is opposed to X “for its own sake” just in case she is opposed to X “non-instrumentally,” that is, just in case she is opposed to X not *merely* for the sake of achieving some other end or goal (42-43). I shall assume that Adams and I are concerned with essentially the same trait and that his “for its own sake” is interchangeable with my “as such.”

Noninstrumental opposition and “enmity” do typically go hand in hand. However, the latter arguably provides a better way of understanding the claim that malevolence is opposition to the good “as such,” for malevolence would appear to be consistent with *instrumental* opposition to a good—a possibility ruled out by Adams’s account. Note, first, that in the same way that “the enemy of my enemy is my friend,” we can say that “the friend of my enemy is my enemy.” Accordingly, suppose Jones is my enemy and Smith is the good friend and close ally of Jones. Apart from his relation to Jones, I have nothing against Smith. Nevertheless, in an effort to harm or spite Jones, I might make or declare Smith my enemy.<sup>5</sup> This in turn might bring about an opposition to Smith that is robustly volitional, active, and personally deep. I see no reason why this orientation might not also be *malevolent*, notwithstanding the fact my only *reason* for opposing Smith is to bring down Jones. Alternatively, imagine a soldier who, out of a love for his country, becomes deeply, actively, and personally opposed to his enemies in combat. These are people whom, if not for the war, the soldier would have nothing against. Here again it seems the soldier’s orientation toward his enemies might count as malevolent while nevertheless being instrumental in nature. Thus while malevolence may typically involve noninstrumental opposition, this would not appear to be a *requirement*. If so, it is a mistake to think of the relevant “as such” qualification in terms of such opposition. A better account of this qualification is one that appeals to the notion of making someone or something an *enemy*.

### 1.3 Impersonal vs. personal malevolence

A third important point concerns the *object* of malevolence itself. If we take Milton’s Satan as a paradigm, it looks as though the immediate object of malevolence is rather *abstract* or *impersonal*, for Satan is apparently opposed to *the* good or to *goodness* itself. While I do think malevolence can take something like this form, surely this is not a requirement. A person can be

malevolent simply on account of his orientation toward other *persons*, for example, by opposing another person's well-being or *share* in the good. Such an orientation is independent of the malevolent person's orientation toward goodness in general. This, then, suggests a distinction between "impersonal" and "personal" malevolence. In the case of impersonal malevolence, the object is impersonal in the sense just noted; in the case of personal malevolence, the object is a person's or a group of people's well-being or share in the good.

It might be thought, however, that the very notion of impersonal malevolence is problematic: that, for instance, there is something odd or implausible in the idea of opposition to something abstract or impersonal like goodness or the good in general. While I share this reservation to some extent, I think there remain good reasons to take seriously the idea of impersonal malevolence. First, we sometimes think or speak of, say, a person's love of *justice* or of *beauty*. Here the object in question presumably is justice or beauty *itself*, not merely, say, the set or any subset of just states of affairs or beautiful objects. But if we can be *for* an impersonal end like justice or beauty, why doubt that we can be *opposed* to the good or to goodness in general? Second, the notion of impersonal malevolence does have at least some intuitive traction. Again, Milton's Satan would not appear to be opposed merely or even primarily to the well-being of any *one* person or *group* of people. Nor does it seem quite right to think that he is opposed, say, to the well-being of *every* person besides himself. Instead, his wickedness seems to run deeper and to be more formal in character. He is, it seems, opposed to goodness *itself* or to the good *in general*.

Moreover, we need not limit our attention in this context to the Judeo-Christian character of Satan. Super-villains like the Joker, for instance, while no doubt opposed to the well-being of many individuals, also seems fundamentally opposed to goodness itself or to some such general

or abstract end.<sup>6</sup> The latter, rather than the well-being of any individual or group of individuals, is an intuitively more plausible characterization of the object of the Joker's malevolence. Indeed, this would appear to be what makes him, and his diabolical counterpart, so frightening and wicked.

Note, finally, that the notion of impersonal malevolence is consistent with the possibility that the object of such malevolence is a kind of abstraction or idealization. It may be, for instance, that Milton's Satan shifts from being against, say, the well-being of God or of God and God's followers, to being against the good or goodness in general, where the latter is, in Satan's mind at least, an abstraction of the former.<sup>7</sup> It does not follow from this that Satan's malevolence is "personal" in the relevant sense; nor that it is *itself* an abstraction or idealization. Thus while there may be something *prima facie* odd about impersonal malevolence, it is a coherent concept, and I shall continue to take it seriously.<sup>8</sup>

#### 1.4 The scope of malevolence

A fourth and related aspect of malevolence that merits attention is its *scope*. Particularly when thinking of malevolence on the impersonal model just identified, it is difficult to escape the impression that the scope of malevolence is maximally *broad*. However, there is something at least *prima facie* problematic about this suggestion. For instance, if I am really opposed to goodness itself, then it seems I shall be opposed to all that *is* good, and thus to anyone's *share* in the good, including my *own*. But of course we do not tend to think of malevolent persons as opposed to their own well-being. Similarly, suppose that human well-being consists partly in an ability to freely choose what to be for and what to be against. If so, then a person opposed to the good as a whole or to goodness itself apparently must be opposed to her own *opposition*, since this opposition represents an exercise of the relevant, valuable ability. But again, this conflicts

with our intuitive understanding of malevolence, for malevolent persons certainly need not—and perhaps *cannot*—be against their own opposition to the good. This suggests that there is, after all, something problematic about the very idea of impersonal malevolence.

There are two problems with this objection. First, it is not difficult to imagine that someone like Satan or the Joker might be opposed to goodness in general in the sense that he *takes* goodness or the good in general to be something worth opposing, and even “enemizes” the good as he conceives of it, but nonetheless fails to appreciate or take seriously the fact this opposition “commits” him to being opposed even to his *own* well-being and opposition to the good. Impersonal malevolence might, in other words, be accompanied by a kind of logical inconsistency and self-deception. The malevolent person might be committed, in principle at least, to opposing his own good, but might fail, through intellectual carelessness or dishonesty, to see that this is so, and thus to actually or actively “follow through” with his commitment.<sup>9</sup> Again, while this might involve ascribing to the person in question a certain kind or degree of *irrationality*, such irrationality seems entirely consistent with malevolence.

Second, the scope of impersonal malevolence need not be as broad as the discussion thus far has suggested. For an impersonally malevolent agent, instead of being opposed to the good as a *whole*, might be opposed to some limited *part* or *dimension* of the good. Such a person might, for instance, actively and vehemently oppose *generosity* or *neighborliness*; and she might oppose it *universally*—opposing even the show of generosity or neighborliness toward *herself*. Here the object of malevolence is still abstract or impersonal in the relevant sense, and yet it is not so broad as to raise any of the self-referential worries just noted. For this reason as well, the notion of personal malevolence remains worth taking seriously.

## 1.5 The psychological coherence of malevolence

A fifth and closely related feature of malevolence also concerns an apparent tension within its psychology. We have said that to be malevolent is to be opposed to the good as such (or to some *dimension* of the good or to one or more person's *share* in the good). It is plausible to think, however, that to be *opposed* to something—at least in any sense relevant to malevolence—is to regard it “under the aspect” of the *bad* or to regard it *as* bad. If so, it can look as if to be malevolent is to regard as *bad* that which one regards as *good*: that it is ascribe *contrary* qualities to the object of malevolence. We have already seen that it would be a mistake to think of malevolence as an especially rational state. I take it, however, that intuitively it is not *this* irrational or irrational in this *way*. The malevolent person is not confused or of two minds about that which she opposes. In the two sections that follow, I consider how we might think about the psychology of malevolence so as to accommodate these facts. Doing so will require addressing some even “deeper” and more complicated aspects of the psychology of malevolence than those considered thus far. The payoff, however, is that we shall finally be in a position to identify an epistemic counterpart of malevolence.

### 1.5.1 Subjective vs. objective conceptions of the *object* of malevolence

We may begin by noting that the worry in question arises—or arises in full force, at any rate—only if malevolence is understood as opposition to that which one *regards* as good. If malevolence is better understood as, say, opposition to that which *in fact* is good, then the forementioned point entails merely that malevolence is a matter of *regarding* as bad that which *in fact* is good; and this, on the surface at least, is considerably less objectionable than the claim that malevolence is a matter of regarding as bad that which one *regards* as good. We would do well, then, to distinguish between what I shall call a “subjective conception” of the object of

malevolence and an “objective conception,” and to try to determine which of these conceptions is the more plausible. Again, according to the subjective conception, malevolence is a matter of being opposed to something that one *regards* as good; and according to the objective conception, it is a matter of opposition to something which *in fact* is good.<sup>10</sup>

Which of these provides a more plausible characterization of malevolence? There are, I think, good reasons for preferring the subjective conception. Specifically, X’s being objectively or *in fact* good would appear to be neither necessary nor sufficient for a person’s opposition to X to count as malevolent. That it is not *sufficient* is evident in connection with cases of what I shall call “benign opposition” to a genuine good. Suppose that a person S has grown up in a community the identity of which is rooted in its opposition to a genuine good G. S has long been taught of the problems, limits, even evils, associated with G. S even has good testimonial grounds for thinking that G is bad; and nothing about his own experience or knowledge relative to G threatens to defeat or undercut these grounds. As S matures, he develops a staunch *personal* opposition to G, one that manifests itself in S’s actively and vehemently opposing G. While S is opposed in the relevant sense to a genuine good, I take it that we would *not* regard S as *malevolent*. For again, S is opposing what he has good reason to think is a genuine *evil* (which, in general, is actually a *good* or *appropriate* stance to take). This suggests that X’s being objectively good is not sufficient for opposition to X’s being malevolent.

Neither is it necessary. For we can imagine cases in which someone is personally and actively opposed to what he has good reasons—reasons of the sort just noted, say—to *think* is a genuine good, but which in fact is not. Again, if the person is opposed in the relevant, robustly volitional and “personal” way to what he *thinks* or *perceives*—and thinks or perceives on reasonable *grounds*—is a genuine *good*, I take it that we *would* consider him malevolent, even if

the object of his opposition were in fact evil. If so, we may conclude that X's being objectively good is neither necessary nor sufficient for opposition to X's being malevolent, and thus that the subjective conception of the object of malevolence is preferable to the objective conception.<sup>11</sup>

The subjective conception also fits well with certain intuitive ways of thinking about malevolence. Malevolence is an especially pernicious or insidious vice; it is a form of *wickedness*. And part of what *makes* it wicked, it seems, is that it involves something like a "knowing" opposition to a "genuine" good. Intuitively, a malevolent person opposes something or someone which she is *aware* of as having a certain value or worth; this is part of what makes her opposition *malevolent* rather than, say, simply misguided or mistaken.<sup>12</sup>

### 1.5.2 Resolving the tension

We are now in a position to return to the challenge noted above concerning how to make coherent sense of the psychology of malevolence. Recall that if we accept a subjective account of the object of malevolence, together with the further claim that one's "opposing" something in the relevant sense requires that one "regard" that thing as bad, it turns out that malevolence necessarily involves regarding as bad that which one also regards as good. But again, this can suggest that malevolence involves a kind of psychological duality or incoherence that we do not tend to ascribe to it. In what coherent and intuitively plausible sense, then, might malevolence involve the relevant contrasting attitudes? It is important to note that there need not be just a single right answer to this question. There may be multiple ways in which the psychological tension can be resolved: multiple ways, that is, in which malevolence can plausibly be thought of as involving both a positive and a negative assessment of its object. In what follows, I shall identify four such ways.

First, in certain cases, there may exist a semantic *gap* between the relevant judgments, such that the malevolent person does not, strictly speaking, ascribe *contrary* qualities to the object of his malevolence, and thus may not exhibit an obviously problematic level or kind of irrationality. For instance, someone might acknowledge the object of his opposition as a “good,” on the one hand, but nevertheless regard it as “worthy of opposition” or as “meriting destruction or diminishment,” on the other. Such an orientation would not involve an acceptance of two explicitly contrary propositions. Would it be consistent with malevolence? It seems to me that indeed it might. That said, if the person were to regard the relevant object as “good” and as “worthy of opposition” *at the same cognitive “level,”* so to speak, and to do so while invoking a *univocal* standard of goodness or worth, then perhaps he might be of two minds or ambivalent in a way that malevolent persons intuitively are not. For this reason, while I think the present possibility goes some way toward resolving the tension in question, it is most plausible when combined with some of the other possibilities noted below.

A second way, just alluded to, in which a malevolent person might plausibly regard the object of her malevolence as both good and bad is if she employs alternate concepts of good and bad in her assessment of this object. Someone might, for instance, oppose the well-being of a particular group of people, which she regards as *morally* significant or valuable, on the grounds that doing otherwise would have *politically* adverse results. Thus she might regard as *politically* bad something which she nevertheless regards as *morally* good. Assuming that her opposition is sufficiently personal, volitional, and so forth, and that she “enemizes” the relevant people or their well-being, there is little reason to deny that her opposition might be malevolent.

A third way of alleviating the relevant tension, also alluded to above, requires distinguishing between two different ways of “regarding” something as good or bad. It is not

difficult to imagine a malevolent person who, say, in an *immediate* or *personal* way, regards as bad that which, in a rather *distant*, *abstract*, or *impersonal* manner, she also acknowledges or regards as good.<sup>13</sup> This person might *experience* the object of her malevolence *as* bad, while still acknowledging, at some level or in some way, that this object is good or has positive worth or value. While actually making or holding to the relevant judgments in this case might involve a certain amount of self-deception or irrationality, it represents a genuine psychological possibility, and one that falls within the boundaries of intuitive ways of thinking about malevolence. It does not make malevolence look implausibly irrational or cognitively dissonant.

Fourth, and finally, a malevolent person might regard the object of her malevolence as both bad and good by regarding it as bad *in one respect* or *under one description* and as good in a *different* respect or under a *different* description. Suppose I acknowledge that the professional success of Jones would add to his happiness and in this respect regard it as a good thing. But suppose I am also filled with contempt for Jones and everything he stands for, such that, relative to my *own* happiness, I regard Jones's success as bad. This characterization does not employ different standards of "good" and "bad": *happiness* is the sole normative criterion. Further, the co-instantiation of the relevant judgments, while perhaps not fully rational, is not irrational or incoherent in a way that conflicts with our ordinary ways of thinking about malevolence. Again, while I regard Jones's success as both "good" and "bad" in a single sense, I regard it as good under one description (as a contributor to *his* happiness) and bad under a different description (as a potential threat to my *own* happiness). Once again it is plausible to think that my orientation toward Jones might be malevolent.

We have examined four different ways in which a malevolent person might, in some *sense*, at some *level*, or in some *respect*, regard the object of his malevolence as both good and

bad without manifesting a problematic kind or level of irrationality. It is important to note that the ways in question are not mutually exclusive. To the contrary, it is plausible to think that malevolence can, and often does, involve some combination of these ways. For instance, an individual might, in an immediate and experiential way, regard the object of his malevolence as “worthy of opposition,” while, at another, more distant and perhaps semi-conscious level, regarding it as “good.” This scenario combines the “semantic gap” and “different ways of regarding” possibilities identified above. Alternatively, a malevolent person might regard the object of his malevolence, at one level and in one respect, as bad, while, at another level and in a different respect, regarding it as good, thereby illustrating the “different ways of regarding” and “different respects” possibilities. We may conclude that the dual judgments central to malevolence do not ultimately generate a problematic psychological conflict within the psychology of malevolence as we are conceiving of it.

## 2. An Epistemic Counterpart of Malevolence

We are now in a position to turn our attention to the idea of *epistemic* malevolence, and specifically, to the question of whether there *is* such a thing as epistemic malevolence, and if so, just what it amounts to. In fact, the groundwork for an answer to these questions has already been laid. In the previous section, we saw that malevolence proper is reasonably understood in terms of “opposition to the good as such,” and that it admits of both impersonal and personal varieties. It is plausible, I submit, to think of *epistemic* malevolence as opposition to the *epistemic* good as such, and to maintain that it too admits of both impersonal and personal varieties. Let us, for the moment, identify the epistemic good with *knowledge*. Accordingly, we can think of “impersonal” epistemic malevolence as (roughly) opposition to *knowledge* as such; and “personal” epistemic malevolence as opposition to another person’s *share* in knowledge, or

to her epistemic *well-being*, as such. I turn now to consider five candidate cases of epistemic malevolence, all of which involve some kind of opposition to knowledge. Doing so will add some flesh and bring some clarity to the conception of epistemic malevolence just noted. While several of the cases are rather extreme or hypothetical, it is not difficult to identify elements of these cases in ordinary experience.

(1) Consider, first, the philosophical *skeptic*. This person is, in some sense at least, “opposed to knowledge as such,” for she denies the very *possibility* of knowledge, or at least of some significant *variety* of it (moral, philosophical, religious, scientific, etc.). However, given the model just sketched, she does not really qualify as epistemically malevolent. Among other things, the skeptic is not opposed to knowledge in a robustly *volitional* or *active* way; rather, her opposition consists merely in a denial of the possibility of knowledge. Similarly, she does not regard knowledge as an *enemy*. She might even wish that knowledge were possible and that skepticism were false: she might value knowledge as something that would be well worth pursuing if only it could be achieved. Finally, a commitment to skepticism seems actually to *rule out* the possibility of epistemic malevolence. For presumably, malevolence of any sort requires, at a minimum, belief in the *possibility* of its object. But this is precisely what the skeptic lacks.

(2) Next consider the roughly Foucault-inspired “suspicionist.” This person regards knowledge as a dangerous idea or concept. She thinks knowledge has an inherently corrupting effect on human beings. On her view, persons who pursue and acquire knowledge do so as a way of trying to control and dominate others. Thus she sees all (or nearly all) knowledge claims as *power plays*. Suppose, further, that the suspicionist is adamantly *for* social justice. The result is that she is personally, vehemently, and actively opposed to knowledge: she regularly speaks out against this alleged “good,” discourages its pursuit among her friends and colleagues, and so on.

Is the suspicionist an example of epistemic malevolence? This depends on the precise character of her opposition. We stipulated that she is *for* social justice and that this goes some way toward explaining her opposition to knowledge. This underscores the possibility that she is not really opposed to knowledge *as such*: that she may just be strongly in *favor* of social justice, while not really regarding knowledge as an *enemy*. But of course it is also possible that she *does* regard knowledge as enemy: that her opposition to knowledge is sufficiently or relevantly entrenched in her psychology such that she is not merely for social justice and only “incidentally” against knowledge. To the extent that the latter description is right, the suspicionist does provide a good illustration of epistemic malevolence (and of *impersonal* epistemic malevolence in particular).

It might be objected that the suspicionist is not really epistemically malevolent on the grounds that there is, after all, something to her belief that knowledge corrupts and fosters social injustice in the relevant way.<sup>14</sup> It might be thought that her opposition to knowledge is rational or justified in a way that malevolence is not. I will not stop here to assess the credentials of the suspicionist’s belief; however, in keeping with the discussion in section 1.5.1 above, I do think that if her belief about the disvalue of knowledge *were* in fact rational or well-supported, then she could not really be considered malevolent. For in that case, she would be opposed something she has genuine reason to think is *bad*, which again would seem to make her orientation something less than malevolent. Accordingly, we can draw the additional conclusion that a malevolent agent necessarily lacks good reasons for thinking that the object of her malevolence is bad. While questions can be raised about how exactly to understand the notions of “rational” or “good reasons,” I take it that the referent of these terms is sufficiently familiar and intuitive.<sup>15</sup>

The case of the suspicionist also generates an important question concerning the *scope* of epistemic malevolence. It is tempting to think of the suspicionist as one who is opposed to knowledge in *general* or to *all* knowledge. But this cannot be right. First, the suspicionist may very well take herself to *know* that knowledge is dangerous in the relevant sense, but without regarding *this* knowledge as dangerous. Second, her campaign against knowledge is itself likely to involve knowledge—knowledge which, again, she is unlikely to consider objectionable. Convincing her friends and acquaintances of the disvalue of knowledge is likely to involve, at a minimum, a considerable amount of practical or circumstantial knowledge (e.g. knowledge about their concern for knowledge, their likely reaction to her diatribes against knowledge, better and worse ways of convincing them of the disvalue of knowledge, and so on). Thus, as with the impersonal variety of malevolence proper discussed in section 1.4 above, the scope of impersonal *epistemic* malevolence is not maximally broad.<sup>16</sup>

(3) A third candidate case of epistemic malevolence is found in the character of O'Brien in George Orwell's *1984*. O'Brien is a master of mind control. He aims to acquire absolute control over the thinking and reasoning capacities of Winston Smith and his other subjects. He attempts to condition in them various submissive, shallow, and contradictory ways of thinking and to prohibit any kind of intellectual autonomy or reflective questioning. He explains that his subjects must undergo "an elaborate mental training ... [which] makes [them] unwilling and unable to think too deeply on any subject whatsoever." His goal is that they develop "the power of not grasping analogies, of failing to perceive logical errors, of misunderstanding the simplest arguments," as well as "the power of holding two contradictory beliefs in one's mind simultaneously, and accepting both of them" (1992, 174, 177).<sup>17</sup>

While ultimately a good illustration of epistemic malevolence, this case does not fit perfectly with the model sketched above. For it appears that the *object* of O'Brien's opposition is not principally or primarily his subjects' acquisition of *knowledge*, but rather their capacity to *think* and *reason* in a free and rational way. We might, then, draw the conclusion that O'Brien is not really epistemically malevolent (since again, we are thinking of epistemic malevolence as opposition to the epistemic good as such and have identified the epistemic good as knowledge). A more plausible response, however, would be to *expand* our conception of the epistemic good or epistemic well-being such that it *includes* an ability to think and reason in a free and rational way. For, a person with a great deal of knowledge, but without the ability to think and reason in the ways forbidden by O'Brien, surely is not very epistemically well off. If we broaden our conception of the epistemic good in this way, and if we assume that O'Brien regards his subjects' intellectual autonomy as an "enemy," then we may conclude that O'Brien is indeed epistemically malevolent (in the *personal* sense).

(4) A similar, though more realistic (and hence more tragic), example of personal epistemic malevolence is found in Frederick Douglass's famous autobiography, in which he recounts some of his early attempts at self-education. Douglass's mistress, Sophia Auld, initially strikes him as "a woman of kindest heart and finest feelings" who had "been in a good degree preserved from the blighting and dehumanizing effects of slavery." As Douglass explains, she "commenced to teach me the A, B, C. After I had learned this, she assisted me in learning to spell words of three or four letters." At this point, however, Sophia's husband Tom intervenes, insisting that Sophia cease all instruction,

... telling her, among other things, that it was unlawful, as well as unsafe, to teach a slave to read. To use his own words, further, he said, 'If you give a nigger an inch, he will take

an ell. A nigger should know nothing but to obey his master—to do as he is told.

Learning would spoil the best nigger in the world. Now,’ said he, ‘if you teach that nigger (speaking of myself) how to read, there would be no keeping him ... He would at once become unmanageable, and of no value to his master. As to himself, it could do him no good, but a great deal of harm. It would make him discontented and unhappy.’

Before long, the corrupting influence of slavery lays hold of Sophia as well, divesting her of all of her “heavenly qualities”:

Under its influence, the tender heart became stone, and the lamblike disposition gave way to one of tiger-like fierceness. The first step in her downward course was in her ceasing to instruct me. She now commenced to practice her husband’s precepts ... Nothing seemed to make her more angry than to see me with a newspaper. She seemed to think that here lay the danger. I have had her rush at me with a face made up all of fury, and snatch from me a newspaper, in a manner that fully revealed her apprehension. (1999, 42-43)

Mr. and Mrs. Auld seem clearly to “enemize” Douglass’s epistemic well-being. They staunchly oppose and take active measures to thwart his acquisition of knowledge and attempts to think critically and autonomously.

This case also illustrates the claim made above that a malevolent person’s belief concerning the disvalue of the object of his malevolence is necessarily irrational. Mr. Auld suggests that part of his opposition to Douglass’s education is that it would only add to Douglass’s discontent and that therefore it is in his best interest prevent it. But surely Auld’s belief here is *not* well-supported or rational; and presumably this is part of what makes his attitude toward Douglass malevolent.<sup>18</sup>

(5) A final illustration is the notorious Cartesian demon—a *systematic deceiver*. The demon, let us suppose, delights in filling his subjects' minds with lies, in creating an ever deepening and insurmountable chasm between appearance and reality. Unlike O'Brien and Mr. Auld, he is not principally concerned with controlling or manipulating his subjects' thought processes. Rather, his goal is that his subjects *end up* with as many false beliefs and as few true beliefs as possible, even if this involves their *thinking* or *reasoning* in a reasonably autonomous and rational way.<sup>19</sup> On the surface, this seems to be a rather pure and straightforward case of personal epistemic malevolence (and, of course, Cartesian demons are sometimes referred to as "malevolent").

My only reservation concerns the possible motivation for the demon's deception. Our characterization—as with many other characterizations of Cartesian demons—at least leaves open the possibility that the demon's motivation is mere *amusement*: that he simply gets a kick out of systematically misleading his subjects. But if this is right, then while the deceiver may be epistemically twisted or perverse, it is not clear that he is malevolent, for his opposition to his subjects' epistemic well-being may not be sufficiently *personal* or *negative*. He may not regard their flourishing as bad or as a genuine *enemy*, and thus his orientation may not be adversarial enough to count as malevolent. That said, if we were to stipulate that the demon is personally opposed to their epistemic well-being, that he is bent on deceiving them, and that his opposition is fueled by something like spite or anger or hatred, then it would be right to view him as a paradigm case of personal epistemic malevolence.

One of the central tasks of the present paper is to identify what an epistemic counterpart of malevolence proper or "moral malevolence" might look like. Between the discussion of malevolence proper in the previous section, and the analysis of several cases in the present

section, the basic nature and structure of epistemic malevolence should now be fairly clear. Again, epistemic malevolence is plausibly understood as opposition to the *epistemic* good as such, where the good in question includes, but is not limited to, knowledge. It is also plausible to think of epistemic malevolence as admitting of both “personal” and “impersonal” varieties in the senses outlined above.

### 3. Epistemic Malevolence and Intellectual Vice

We may now turn now to address the third of our initial three questions: namely, why don't we think of epistemic malevolence as epitomizing intellectual vice in the way that we think of malevolence proper as epitomizing moral vice? There are, in fact, two separate questions here, and they are best dealt with in turn. First, *does* epistemic malevolence epitomize intellectual vice? Is it a clear and informative paradigm element of it? And second, in the event that it is, why don't we think of it as such? Why doesn't it top, or at least make an appearance on, any standard lists of intellectual vices? The first is a question about the actual place of epistemic malevolence within the normative *structure* of intellectual vice. The second is a question about its place in our *thinking* about intellectual vice.

In response to the first question, I think that indeed epistemic malevolence conceived in the present way does epitomize intellectual vice. This is evident from certain standard ways of thinking about intellectual virtues and vices. First, several virtue epistemologists have thought of something like a desire *for* truth or knowledge as the underlying and unifying feature of intellectual *virtues*: that intellectual virtues, by definition, are traits that arise or flow from a love of knowledge.<sup>20</sup> Second, consider what apparently underlies the traits that *do* tend to appear on standard lists of intellectual vices. These include traits like intellectual carelessness, superficial thinking, laziness in inquiry, dogmatism, narrow-mindedness, ignoring of evidence, and so forth.

Arguably, what these traits have in common, and what (at least in part) explains their status as intellectual vices, is either a straightforward *lack* of a desire for knowledge or an insufficient concern with knowledge *relative to* other concerns (e.g. a concern with power or status, or a fear of being mistaken). Given, then, that intellectual *virtue* is fundamentally a matter of loving or being *for* epistemic goods, and that many intellectual *vices* have in common a lack of or otherwise inadequate concern for such goods, it stands to reason that an outright *opposition* to these goods would epitomize intellectual vice. Indeed, such opposition, which of course is the very essence of epistemic malevolence, seems as much or more than any other intellectual vice to represent a kind of epistemic *wickedness*.

Now for the second question: why don't we tend to *think* of epistemic malevolence as epitomizing intellectual vice? I shall assume that this is because epistemic malevolence is less *familiar* to us than malevolence proper, and that this *in turn* is because epistemic malevolence is less *common*, that it occurs or is manifested less frequently than malevolence proper. But *why* is epistemic malevolence less common?

There are, in fact, two problems with this question that need to be addressed before we can attempt to answer it. First, I have thus far left unclear how I am conceiving of the relation between malevolence proper and epistemic malevolence. At times I have suggested that the latter is an "application" of the former, in which case there is a sense in which epistemic malevolence is *part* of and can be subsumed *under* malevolence proper, while elsewhere I have described epistemic malevolence as a "counterpart" of malevolence proper, which at least leaves open the possibility that it is a separate or distinct trait. Fortunately, we need not settle this issue before answering the question at hand. For instead of comparing the relative incidence of epistemic malevolence and malevolence proper, we can compare that of epistemic malevolence and what I

shall now refer to as “moral malevolence,” which is opposition to the *moral* good or to a person’s *moral* well-being, and where the latter consists in a person’s share in certain recognizably *moral* values like life, health, the experience of higher and lower pleasures, freedom of movement, freedom from physical and mental suffering, and so forth. I take it that moral malevolence at least approximates what we have thus far been discussing under the label of “malevolence proper.” But whether this is so, or whether the latter is a considerably broader concept, need not worry us here, for epistemic malevolence remains much less common than moral malevolence thus defined. And again, our question is *why* this should be.

The question also stands in need of an additional refinement. As stated, it suggests that both personal and impersonal epistemic malevolence are less common than their moral counterparts. But I am not sure that this is right. Specifically, it is questionable whether *impersonal* epistemic malevolence is really less common than impersonal moral malevolence. Neither, it seems, is really very common at all. The former is, perhaps, more *familiar*, but I suspect this is due more to certain memorable and compelling portrayals of characters like Satan and the Joker than it is to an actual higher incidence of impersonal moral malevolence. Assuming this is right, we would do best to limit our attention to the question of why *personal* epistemic malevolence is less common than personal moral malevolence.

How, then, might we go about explaining this discrepancy? Here again I think that there need not be, and in fact *is* not, just a single right answer. That is, I suspect the higher incidence of personal moral malevolence is attributable to a confluence of factors, four of which I shall attempt to identify in the remainder of the paper.<sup>21</sup>

First, the moral good or moral well-being arguably includes a greater *plurality* of goods or values than that of its epistemic counterpart. For reasons already indicated, I would not insist

that the epistemic good be *identified* with knowledge; nevertheless, as my earlier enumeration of the goods associated with moral well-being suggests, there does appear to more “packed in” to the notion of moral well-being compared with epistemic well-being. Consequently, it seems that, other things being equal, the greater breadth of moral well-being, or the greater number of moral goods, is likely to make for a greater number of occasions for personal moral malevolence. In short, within the moral realm, there is considerably more to be opposed to. This provides at least a partial explanation of the higher incidence of personal moral malevolence.

Second, I think the higher incidence of personal moral malevolence can be attributed in part to the comparative *value* of the goods to which this kind of malevolence is opposed. If person A is filled with contempt, hatred, spite, vengeance, and the like, for person B, and consequently wishes significantly to *harm* B, A is more likely, I take it, to strike at B’s share in certain fundamental *moral* goods—life, health, pleasure, freedom of movement, and the like—than he is to strike at B’s share in *knowledge*. And this, it seems, is due at least in part to the apparent greater value of the relevant moral goods. Thus, to the extent that malevolence is motivated, as it typically is, by hatred, revenge, spite, and the like, we have at least some reason to expect personal moral malevolence to be more common than personal epistemic malevolence.

Third, there is a sense in which individual *moral* well-being is more *vulnerable* vis-à-vis other agents compared with individual epistemic well-being. Alternatively, it is generally easier, in a world like ours, to undermine another person’s moral well-being than it is to undermine this person’s epistemic well-being.<sup>22</sup> It is, for instance, generally easier to undermine the pleasure, health, or practical autonomy of other people than it is to block their access to truth or knowledge or a good education. This underscores the more (though by no means exclusively) social and interpersonal dimension of morality and moral well-being compared with that of epistemic well-

being. The acquisition of knowledge, while by no means a strictly solitary enterprise, is generally *more* solitary (or capable of being so) than the acquisition of moral goods. This difference can go some way toward explaining the higher incidence of personal moral malevolence, for our greater *vulnerability* vis-à-vis moral goods makes it more likely, other things being equal, that others will be opposed to, and more likely to attempt to undermine, our share in these goods.

A fourth and final explanation is somewhat more complicated. It is rooted partly in the fact that malevolence often arises in connection with *competition* for certain goods. If you and I must compete for a certain mutually cherished good, this increases the probability, other things being equal, that enmity or malevolence will arise between us. My suggestion is that there is generally greater competition for moral goods than there is for epistemic goods and thus that we should expect a higher incidence of personal moral malevolence than personal epistemic malevolence.

In what sense or why is there greater competition for moral goods? I think the explanation here is manifold. First, as indicated above, we generally value moral goods more than we do epistemic goods, in which case we are more likely to compete and to compete more fiercely for these goods. Second, our world and the goods in question are such that it is generally easier to come by the relevant epistemic goods than it is the relevant moral goods. Knowledge, for instance, is easily transmitted via testimony; books can be purchased at reasonable prices; the internet and related media place a wealth of knowledge about a vast range of subjects at our fingertips. Knowledge is also relatively easy to come by in the sense that one person's acquiring knowledge about some subject matter X generally does not make it any less likely that someone else will be able to acquire knowledge about X. Knowledge is, in this respect, a "sharable resource."<sup>23</sup> The relevant moral goods, however, are often more expensive and less easily

accessible. Maintaining good health, for instance, requires resources that are finite and often in short supply: this includes anything from affordable nourishment to state of the art of medical treatments. Here, one person's laying hold of the relevant resources and resulting goods *is* more likely to pose a threat to another person's laying hold of the same goods. The result is that, other things being equal, there is likely to be more competition associated with the relevant moral goods than with the relevant epistemic goods. Third, epistemic well-being is generally easier to sustain compared with moral well being. Life, health, the experience of various pleasures, etc., require ongoing attention and investment. The good of knowledge, however, is relatively easy to sustain. This is due in no small part to the faculty of *memory*. If I satisfy a particular epistemic appetite—a desire to know about X—this appetite will remain satisfied as long as I retain and can recall the relevant information. If I satisfy a particular bodily appetite, however—a desire for food, for instance—this satisfaction will quickly dissipate. There is no counterpart to memory within the moral domain. Because moral well-being is more difficult to sustain than epistemic well-being, there is likely to be more competition, and thus more malevolence, associated with the former than with the latter.<sup>24</sup>

### Works Cited

- Baehr, 2006. "Character, Reliability, and Virtue Epistemology," *The Philosophical Quarterly* 56: 193-212.
- Baehr, 2008. "Four Varieties of Character-Based Virtue Epistemology," *Southern Journal of Philosophy*, forthcoming in December.
- Smith, Michael. 1994. *The Moral Problem* (Oxford: Blackwell).
- Hurka, Thomas. 2001. *Virtue, Vice, and Value* (Oxford: Oxford UP).
- Adams, Robert. 2006. *A Theory of Virtue* (Oxford: Oxford UP).
- Orwell, George. 1992. *1984* (Knopf Publishing, Everyman's Library Series).
- Zagzebski, Linda. 1996. *Virtues of the Mind* (Cambridge: Cambridge UP).
- Montmarquet, James. 1993. *Epistemic Virtue and Doxastic Responsibility* (Lanham, MD: Rowman & Littlefield).
- Douglass, Frederick. 1999. *Narrative of the Life of Frederick Douglass, An American Slave* (Oxford: Oxford UP).

---

<sup>1</sup> As this suggests, I am thinking of intellectual virtues as *character traits*, rather than as cognitive faculties or powers like memory, vision, introspection, reason, or the like. This difference corresponds to the difference between "responsibilist" or character-based and "reliabilist" or faculty-based approaches to virtue epistemology. For a discussion of this distinction, see Baehr 2006. For an up-to-date account of character-based virtue epistemology, see Baehr 2008.

<sup>2</sup> Throughout the discussion, my concern will be features of malevolence proper and epistemic malevolence that distinguish these traits from other moral and intellectual vices—*not* the full range of necessary and sufficient conditions for these traits. Thus, the satisfaction of the conditions I lay down for epistemic malevolence, for instance, may not be sufficient for the possession of a "full blown" intellectual vice; the latter may require a certain ill motivation. While I make some suggestions about what the motivation in question might amount to, this is not a central point or concern of my discussion. Thanks to Wayne Riggs for getting me to be clearer about this dimension of my paper.

<sup>3</sup> Some modifications to this definition will be called for; however, these can be viewed as "interpretations" or construals of the original definition, rather than as objections to it.

<sup>4</sup> Indeed, it might be wondered whether such an orientation is even *possible*: to be opposed to something, mustn't one regard it as *bad*? I address this worry below.

---

<sup>5</sup> Interestingly, Adams himself seems to want to allow for something like this possibility when he says that Satan might show malice toward God “simply to spite God” (40). This makes it sound as though Satan’s opposition to “important goods,” as Adams describes it, is indeed instrumental to a certain end (viz. that of frustrating or angering God).

<sup>6</sup> Thanks to Tom Hurka for this example. Another would be a Grinch- or Scrooge-like character who is—by appearances—opposed to something like happiness *in general*, not necessarily to the happiness of any particular person or group of persons. Thanks to Damon Evans for this example.

<sup>7</sup> As this suggests, one needn’t be a Platonist about properties to accept this conception of malevolence. Also, it is worth noting that the end in question may be more or *less* “abstract” or “impersonal.” A malevolent person might, for instance, be against a certain kind of good *activity* (versus, say, the property of goodness or some conceptualization of it). As long as the object here is the activity understood *generally*, it still makes sense to think of this person’s malevolence as “impersonal” in the relevant sense. Thanks to Anne Baril for raising this point.

<sup>8</sup> Another potential worry about the notion of impersonal malevolence, raised to me by Miranda Fricker, is that *benevolence*—the contrary of malevolence—would not appear to admit of an impersonal variety. I feel the force of this worry; but again, when weighed against the reasons for thinking that there is such a thing as impersonal malevolence, I think the more reasonable conclusion is that malevolence and benevolence are not structurally isomorphic in every respect.

<sup>9</sup> This would be similar to the person who declares with conviction that “there is no truth” or that “only empirically verifiable statements have any meaning.” There is a definite sense in which such a person is “opposed” to truth *in general* or to the *possibility* that a non-verifiable statement might have meaning, notwithstanding the problematic self-referential implications of these convictions.

<sup>10</sup> Alternatively, we might distinguish between a “*de re*” opposition to a good (objective) and a “*de dicto*” (subjective) opposition. It is also worth noting that the two conceptions are not mutually exclusive in the sense that one can be opposed to something which in fact is good and which one also regards as such. The worry arises in any case in which the subjective element is present. And if the subjective element is essential, the problem is general.

<sup>11</sup> See Hurka 2001 pp. 171-80 for a related discussion. Parts of Hurka’s discussion suggest that he might opt for an objective conception of the object of malevolence, and specifically, that he might think of malevolence as involving a kind of brute, non-cognitive, or, in his terms, “simple emotional” opposition to its object. Such a view would be problematic, however, for at least two reasons: first, as we have seen, it seems essential to the very wicked or pernicious quality of malevolence that the malevolent person actually be aware of the (apparent) positive worth of what she opposes; and second, if the alternative to the relevant cognitive requirement is, as Hurka’s discussion suggests, an affective or emotional one, such that a malevolent person is emotionally “for” the object of her malevolence (but does not cognitively regard it as valuable), this makes the corresponding view look even *less* plausible, since it is clear that a malevolent person’s feelings or emotions are overwhelmingly *opposed* to the relevant object (even if, cognitively, such a person were to be aware of the object’s having a certain positive value). A related question or worry is whether the subjective account is inconsistent with motivational internalism. My response, very briefly, is that it is consistent with any *plausible* version of motivational internalism. It is consistent, for instance, with the view that if S judges X to be valuable, then S will experience *some* (potentially very minimal) motivation to comply with this judgment; and it is consistent with even weaker versions of motivational internalism (e.g. Smith 1994) according to which the foregoing principle holds only for “good and strong-willed” agents (which, needless to say, a malevolent person is not). It is *inconsistent* with the view that if S judges X to be valuable, then S will experience an *overriding* motivation to comply with this judgment. But this view is much too strong.

<sup>12</sup> One additional kind of case requires a further, fairly minor amendment to the subjective conception. These cases in question are cases of what might be called “hardened malevolence.” Suppose, for instance, that I have spent so many years hating and opposing Jones that I no longer have any sense or awareness of his worth or well-being. It does seem possible that my orientation toward Jones might still

---

count as malevolent. Note that in such cases, it is malevolence *itself* that vitiates the malevolent person's awareness of the value of the relevant object (in this respect it differs from the cases considered above). Accordingly, we might say that malevolence is a matter of being opposed to that which one regards as good, *except* where the subject's malevolence itself has vitiated any awareness of the value of the end or object in question. Alternatively, we might say that malevolence is a matter of being opposed to that which one regards or *has* regarded as valuable, where the latter qualification accommodates cases of hardened malevolence. Either formulation provides a suitably modified, but still fundamentally subjective, account of the object of malevolence.

<sup>13</sup> This is, of course, suggestive of Aristotle's discussion of *akrasia* in Book VII of *Nicomachean Ethics*. The malevolent person here described lacks the kind of deep knowledge or awareness of the value of the object of his malevolence that the akratic person lacks of the good.

<sup>14</sup> Thanks to Aimee Koeplin for raising this worry.

<sup>15</sup> The case also underscores the fact, alluded to in note 2 above, that epistemic malevolence is consistent with good motives (in this case social justice), and thus that it is not always a vice (or at least not always a *moral* vice). I will not stop here to defend this position. Rather, I will simply note in passing that the same basic point applies in connection with several other cases and virtues: e.g. one is hard-pressed to deny that the terrorist (or daredevil) is *courageous*, even if this person's courage does not qualify as a genuine *virtue*.

<sup>16</sup> The considerations just noted should not lead us to doubt the very possibility of *impersonal* epistemic malevolence. First, while a completely general opposition to *truth* or *knowledge* may be problematic, such opposition to *other* epistemic ends or values (e.g. theoretical understanding) may not be. Second, the same combination of logical inconsistency and self-deception discussed in section 1.4 above may be operative in cases of impersonal epistemic malevolence. Thanks to Linda Zagzebski for forcing me to be clear about this.

<sup>17</sup> Thanks to Daniel Ambord for this example and for conversation related to this and other aspects of epistemic malevolence.

<sup>18</sup> Auld's irrationality lies not in his belief that learning will make Douglass discontent, but rather in his belief that he ought therefore to oppose Douglass's education. If Auld's belief is *not* irrational, then I take it that this is because his is a case of "hardened malevolence," described and dealt with in note 12 above.

<sup>19</sup> Thus he might focus his efforts on imparting false "basic beliefs," from which his subjects might reason impeccably, but to little epistemic avail.

<sup>20</sup> See e.g. Zagzebski 1996 and Montmarquet 1993.

<sup>21</sup> Clearly the answer to our question is largely an empirical one. However, as I hope momentarily to demonstrate, I think several helpful things can be said in response to it from the philosophical armchair.

<sup>22</sup> Thanks to Wayne Riggs for suggesting this point.

<sup>23</sup> There are, of course, exceptions to this and each of the considerations or suggestions I am putting forth in this section. My aim in this section is merely to identify the way things *tend* to be or how they *tend* to go relative to the goods in question (and indeed only how they tend to be or go *ceteris paribus*). This is all that is required for answering the central question of this section.

<sup>24</sup> I am grateful to many participants in the 2008 Fullerton International Philosophy Conference for helpful comments and feedback on this paper (see specific acknowledgments spread throughout the notes above). I owe a special thanks to Michael Pace for many hours of helpful conversation about epistemic malevolence and related issues.