Received: 25 July 2021 | Revised: 30 October 2021 | Accepted: 16 November 2021

DOI: 10.1111/phib.12261

ORIGINAL ARTICLE

ANALYTIC PHILOSOPHY WILEY

When does self-interest distort moral belief?

Nicholas Smyth

Fordham University, New York, New York, USA

Correspondence
Nicholas Smyth, Fordham University, New York, NY, USA.
Email: nick.a.smyth@gmail.com

Funding information
Fordham University

Abstract

In this paper, I critically analyze the notion that self-interest distorts moral belief-formation. This belief is widely shared among modern moral epistemologists, and in this paper, I seek to undermine this near consensus. I then offer a principle which can help us to sort cases in which self-interest distorts moral belief from cases in which it does not. As it turns out, we cannot determine whether such distortion has occurred from the armchair; rather, we must inquire into mechanisms of social power and advantage before declaring that some moral position is distorted by self-interest.

1 | INTRODUCTION

In this paper, I argue against a dogma in contemporary moral epistemology: the idea that moral belief formation is distorted when it is influenced by self-interest.[1] Virtually, none of the writers who deploy this idea have tried to defend it, and as it turns out, the idea is very questionable on reflection. After offering some reasons to reject the idea, I will suggest that moral epistemology needs a more nuance and situational sensitivity than it often displays, and that ideas like "self-interest distorts moral belief" are far too simplistic to be of any serious use. Our real question should be: *under what conditions* are such forces as self-interest distorting? In this spirit, I will provide a principle which separates cases in which self-interest is distorting from cases in which it is not.

First, a few assumptions: the study of moral knowledge almost trivially implies that there can be such a thing as *true* and *justified moral belief*, and this already involves the assumption that extreme versions of non-cognitivism are false (Ayer, 1936). In addition, the sort of constructive inquiry I am engaging with tends to assume that moral error theory is false and that moral beliefs are not rendered automatically false or unjustified by some abstract metaphysical considerations

[1]For various deployments and articulations of the idea, see (Rawls, 1951, 179–180; Nagel, 1986, 148; Brink, 1989, 163; Sinnott-Armstrong, 1996, 27; Sturgeon, 1998, 229–230; Shafer-Landau, 2003, 219; Enoch, 2011, 192–193; Parfit, 2011, vol 2, 553; Fitzpatrick, 2014, 3; Wedgwood, 2014). Even those who the idea is deployed *against* seem to tacitly endorse it: see (Doris and Plakias 2008, 320).

or general facts about human evolution (Joyce, 2006; Mackie, 1977). Rather than engage directly with moral skepticism, the general assumption here is that at least some human moral beliefs are both true and justified, and that we are trying to locate this moral knowledge rather than argue over its mere possibility.[2] These assumptions are compatible with moral realism, quasi-realism, and constructivist anti-realism.[3]

Now, before criticizing the idea that self-interest necessarily distorts moral belief, I will first try to get clearer on just what it is and on the role it plays in contemporary moral epistemology.

2 | FILTERS IN EPISTEMOLOGY

First, let us note that there is an important distinction between examining self-interest from a normative-ethical perspective and examining it from a more distinctively moral-epistemological perspective. For example, Kant's normative theory says that our actions lack moral worth when self-interest is a motive for our actions. Nothing in this paper is meant to imply that this impartialist normative perspective is correct or incorrect. Rather, this paper is an essay about justified moral belief. As such, my target is the view that our beliefs about what we ought to do will *lack justification* if they are influenced by self-interest, and it is this view that appears to be nearly hegemonic in moral epistemology.

Here is one helpful way to see what the view is. In her important discussion of moral epistemology, Karen Jones outlines the idea of a "filter" on moral beliefs (Jones, 2005). Any constructive epistemology in any domain needs filters, which are just heuristics or methods for identifying unreliable or untrustworthy belief formation, or beliefs which cannot count as knowledge. One relatively uncontroversial filter is

> **(F1)** If a belief is based on reasoning which essentially involves a *false* belief, that resultant belief is untrustworthy.

For example, if I am reasoning about whether a person is virtuous or vicious, it is very unlikely that I will reason in a trustworthy manner if my eventual judgment is based on a false conception of their character. Any beliefs I form in this manner are unjustified. While this principle needs refinement, and while it may do less work for us than we think it does, it nonetheless seems like the right kind of filter for a positive epistemology to employ.[4]

Seen in this light, a first pass at the filter I am discussing in this paper looks like this:

> **(F2)** To the extent that a person's moral belief is explained by their desire to maintain or increase their own relative well-being, the belief is less justified.

[2]This distinction between constructive and anti-skeptical epistemology is helpfully outlined in Aaron Zimmerman's work. See (Zimmerman, 2010, 15).

[3]It is sometimes thought that anti-realists do not believe in moral knowledge or moral truth, but it should be borne in mind that Sharon Street and Christine Korsgaard both accept that moral truths are mind-dependent and that we can justifiably believe them. This variety of anti-realism is certainly consistent with the existence of justified true belief (Street, 2010). Moreover, while Simon Blackburn's quasi-realist theory adopts a non-realist metaphysics, it is reasonably clear that it must retain the concept of more or less justified belief, and of our *tracking* moral truths in some sense (Blackburn, 1991, though see; Golub, 2017).

[4]I criticize the idea that (F1) can explain away moral disagreement in Smyth 2021.

The idea concerns *relative* well-being because, I take it, and no one is concerned about an agent's desire to increase their own well-being as such. This would rule out the desire to increase everyone's well-being equally, which does not seem to be the object of moral-epistemological concern.

Something like this filter seems to do heavy lifting in two related debates. The first begins with the well-known fact that utilitarian theories have insisted that morality can require serious self-sacrifice from agents. Peter Singer and Peter Unger, for example, have separately insisted that the average person is seriously morally delinquent for this reason (Singer, 2010; Unger, 1996), but these admonitions have often met with strong resistance. The notion that we might be required to give up huge amounts of our time, energy, and bodily or financial resources to (for example) help the global poor is indeed hard for most of us to accept. Such intuitions have driven so-called "Demandingness" arguments, which allege that impartialist moral theories require too much of individual agents.

But even if my protests against this idea take the form of careful argumentation, some philosophers have suggested that my moral beliefs are not trustworthy or justified because they are *motivated* by my self-interested desire to keep my time, energy, and resources for myself. As David Enoch says (Enoch, 2011):

> There is no mystery about the almost universal belief that morality does not require all that Singer and Unger believe it does. Acknowledging that they are right would exert a high price: it would involve exposing "our illusion of innocence", leading us either to give up almost all of our belongings or to the horrible acknowledgment that we and our loved ones are morally horrendous persons. Refusing to see the (purported) truth of Singer's and Unger's claims thus has tremendous psychological payoffs
>
> (Enoch, 2011, 193).

And here is Brian Berkey, explicitly drawing the desired epistemological conclusion (Berkey, 2016):

> Those of us who are well off have self-interested reasons for favoring relatively undemanding views, and this fact might be thought to play a role in explaining our having the intuition that we cannot be obligated to make large sacrifices in order to, for example, aid the global poor. Since there seems to be at least some reason to worry that the intuitions that motivate the demandingness objection are generated by a disposition to accept views that are convenient for us, given our current place in the distribution of benefits and burdens, we have at least some reason to be suspicious of direct appeals to these intuitions.
>
> (Berkey, 2016, 3022).

The second (related) debate proceeds from the observation that moral disagreements are apparently quite widespread and intractable. Critics of moral objectivism have long insisted that this phenomenon makes trouble for the view. This is because objectivism predicts a certain amount of convergence in belief; to the extent that we do not observe this convergence among apparently rational agents, we should become less confident that moral truths are objective. The objectivist's move, at this point, is to deny those appearances by insisting that most agents are *not* rational, because the moral beliefs of many or most agents are strongly

influenced by self-interest (Brink, 1984; Enoch, 2009). Peter Seipel neatly summarizes the idea here (Seipel, 2019):

> On this method of explanation, intractable moral diversity is attributable to the distorting effects of our interests. People tend to disagree about moral matters because self-interest makes them prone to error, leading them to endorse self-serving moral beliefs.
>
> (Seipel, 2019)

In each case, a substantive philosophical conclusion is generated by that key, undefended assumption: that self-interest necessarily renders moral belief less justified, trustworthy or reliable.

Now, at this point, a clarification is in order. In order for these arguments to work, their proponents cannot really be drawing only on (F2) as stated above. This is because the kind of influence at work in these examples is rarely purely egoistic. Human beings in fact tend to display two marked tendencies, each of which is suggested by writers in this area. The first is indeed to form moral beliefs and values in a way that is disproportionately influenced by their own welfare, and the second is to form beliefs and values on the basis of the welfare of their loved ones, family, friends, communities, or fellow members of a social class.[5] Each of these is generally viewed as problematic by the philosophers I have been discussing, but for clarity's sake I will sometimes keep them separate. When I speak of *Egoistic Influence* I will be referring to the process whereby beliefs and values are adopted, entirely or in part, because they increase the believer's own welfare. On the contrary, when I speak of *Parochial Influence*, I will be referring to the process whereby beliefs and values are adopted, again entirely or in part, because they protect or increase the welfare of the believer's friends, family, and broader social compatriots. "Self-interest," as I will continue to use the term, refers to both types of influence.

Both types of influence are often cited. For example, Ralph Wedgwood, in discussing persistent moral disagreement among philosophers, writes that institutional incentives strongly encourage philosophers to "fall in love with the theory they are defending, and so come to believe the theory with greater confidence than they are really entitled to" (2014, 38). This is a paradigm case of purely egoistic influence, since the individual philosopher's well-being (as partly determined by institutional success) is said to be driving their strong acceptance of a theory. Wedgwood concludes that persistent disagreement is no skeptical threat, because most participants are not forming their beliefs in a reliable or trustworthy manner.

Parochial influence is also often cited. In discussing the meta-ethical argument against objectivism, for example, Richard Boyd insists that "class interests" have played a strong role in the maintenance of problematic moral beliefs throughout human history (Boyd, 1988). In addition, Catherine Wilson, discussing Susan Wolf and Bernard Williams' dismissal of "revisionary" moral philosophers like Singer, argues that this dismissal "constitutes a defense of a life of leisure and privilege" typical of the academic classes. She suggests that the fact that "that *our* ordinary way of life does appear immune from criticism... is due to the desire that it be made to appear so."(Wilson, 1984, 289) Thus, both Boyd and Wilson argue that bad or suspect moral intuitions are the product of parochial interest, of a person's desire to protect the welfare or interests of their social group.

[5]In evolutionary theory, this shows up as the distinction between individual selection and kin-selection.

The compound idea under discussion, then, is that both egoistic and parochial influences on moral belief are distorting and should therefore be filtered out by any reasonable moral epistemology (Hereafter: DSI). Before proceeding to criticize it, I should quickly address a possible interpretive concern.

DSI, as it stands, is a very strong principle. As will be seen, this means that it is vulnerable to a number of counterexamples. It might be thought that such a wide array of philosophers could not believe something which is so clearly vulnerable. This may be true, and it may be that philosophers in these debates only believe that self-interest distorts moral belief in *some* circumstances. However, one searches the literature in vain for any acknowledgment of this qualified position. When the principle is explicitly stated, it does not, to my knowledge, come with qualifications. One should, writes Michael Huemer, "distrust intuitions that differentially favor oneself, that is, that specially benefit or positively evaluate oneself, as opposed to others."(Huemer, 2008, 381–382) There is no indication here that self-interest may be a distorting influence in only some situations. Moreover, as I have just shown, philosophers often simply *cite* the influence of self-interest in order to attack some position or intuition, and they do not go on to tell us how this particular case is one that falls under the scope of a more narrow or restricted principle. If indeed meta-ethicists do not believe DSI, they should come out and say so, and this paper will be valuable as a way of provoking this recognition.

But why should not we believe that DSI is true? In what follows, I will offer a series of reasons to reject it.

3 | COUNTEREXAMPLES

If there are uncontroversial cases of moral knowledge acquisition which seem to centrally involve the operation of self-interest, then DSI is at least called into question. In fact, I think that there are such cases.

Consider, for example, a long-term victim of domestic violence who tragically accepts her fate because she has internalized sexist or misogynist values. That is, she acquiesces to a life with her violent partner in part because she believes, at some semi-conscious level, that a man has the right to control her in this way. One day, after a particularly violent outburst, her urge to simply get away from the suffering caused by her partner becomes overwhelming and she flees. In the relative peace and security provided by comforting friends, she reflects on her life in light of her powerful urges to avoid any further violence. Suddenly, she arrives at a powerfully resonant moral epiphany: *he doesn't have the right to do this to me*.

I imagine that we are all inclined to say that her newfound belief is true, and I would suggest that she has clearly formed the epiphany in a reliable or trustworthy fashion. Moreover, it is plausible to say that she would not have come to believe this new truth unless she were under the rather powerful influence of self-interest. Her desire to get away from a very bad situation is, in this case, driving her non-accidentally toward the moral truth.

Now, at this point, the defender of DSI may be tempted to say that this moral epiphany can only count as a justified belief if it is derived from the more impartial claim that no person has the right to abuse *anyone*. They may thus hold the line and say that until this woman realizes that her rights are merely tokens of a much more general type, she cannot really know that she deserves better. This, I want to stress, is an enormous bullet to bite. To repeat, not only will we will be forced to say that this woman *does not know* that she ought not to be treated in this way, but also DSI forces us to say that she has acquired *no* justification for her belief whatsoever. While a

philosopher might try to claim that she lacks full-blown knowledge in this case, the idea that she has not acquired any justification at all simply looks bizarre, and all merely because she has not derived this particular claim from a more general principle. Indeed, if she never subsumes this epiphany under some more general principle, it is never actually an epiphany; she will spend her life not knowing that she ought not to be treated this way, even though she firmly believes it on the basis of what looks like eminently relevant experience.

It might be said that while this woman does not need a general principle, she still needs to be able to generalize her right to others. It is clear enough that she would acquire more knowledge if she could go on to think: no man has the right to do this to *anyone*. However, again, it is important to focus on what the defender of DSI must say here. In the absence of this generalization, they are forced to say that she has not even acquired *any* justification for her particular belief about herself. Indeed, I am being generous here, since technically they have to say that her belief is *distorted* by self-interest. While this all comes down to intuitions, I suggest that this is about as firm as intuitions about justification get: this is *not* a badly formed belief.

It might also be suggested that there is something of a paradox here: *in* judging that this agent has acquired moral knowledge, are not we applying our own highly impartial principle, one which says that no one has the right to abuse anyone? After all, if we were each only committed to the principle "no one has the right to abuse *me*," then we would not see this as a case of moral learning. Quite right, but there is no paradox here. It is entirely possible that we too have arrived at this more general principle in part because we have been influenced by self-interest, by experiences of being disregarded or abused which fueled a search for a more comprehensive moral worldview, but this sort of etiological fact never entails that the principle itself cannot have fully impartial content. As I will soon argue, we should take great care to distinguish between causal history and propositional content, and so *in* deploying the impartial claim about universal rights we in no way impugn the epistemological value of self-interest.

In sum, I think it should be clear that many more cases can be generated which have this basic structure. When a person is subjected to awful treatment, it is very often their own perception of a better life which motivates their defiant assertion of a newly grasped moral truth. They may of course go on to construct more general principles out of these particular acts of moral perception, but it seems clear enough that the defiant assertion itself is a case of moral learning.

4 | EUDAIMONISM

Jones rightly points out that when we are identifying filters in moral epistemology, we ought to obey a fairly simple requirement. "An answer to the filtering problem," she writes, "must avoid appeal to ethically substantive constraints that are contested by currently live competitor theories."(Jones, 2005, 75) The reason is not hard to see: there is something pointlessly self-congratulatory about assuming that one's preferred moral theory is correct in order to impose an epistemological principle which mainly serves to rule out competitor theories. In moral epistemology, then, most philosophers want to develop theories of justified belief which do not depend on the truth or falsity of substantive first-order moral beliefs, particularly when those beliefs are hotly contested. The whole point of filtering in epistemology is to help us to resolve debates by screening off potentially bad beliefs, not to declare those debates settled by fiat and to conclude that trustworthy belief-formation processes are just those displayed by the disputants we have dogmatically crowned the winners. This is, I think, what lies behind Jones' requirement, and it is

why canonical moral epistemologies such as reflective-equilibrium theory do not presuppose the truth of any contested normative framework.

Yet, it is now easy to see that the prohibition against egoistic and parochial influence probably violates Jones' requirement. Once we widen our gaze away from certain strands of post-Kantian Western moral philosophy, we encounter a startling fact: Our moral-philosophical tradition is largely dominated by *Eudaimonism*, the notion that right action just *is* action which is reliably productive or constitutive of the agent's well-being.[6] Plato and Aristotle, for example, each embraced this view. Some Eudaimonists such as Epicurus famously identified well-being with pleasure, while the Stoics tended to describe it as a kind of harmony with nature. For Augustine, happiness was our supreme practical end, and he defined it as genuine possession of the virtues which could only come about via God's grace. The Confucian philosopher Mencius argued that virtue is a means to happiness, or that happiness is the key incentive to pursue virtue. Moreover, as Owen Flanagan has convincingly argued, much Buddhist philosophy can be straightforwardly read in Eudaimonistic terms, since the various forms of right thought and action recommended by that tradition are meant to lead to a kind of flourishing or contentment.[7]

Now, strictly speaking, none of this *entails* that self-interest is not a distorting factor. As I emphasized at the outset of this paper, normative ethics is not moral epistemology, and it is conceivable that the true moral theory is egoistic while the true moral epistemology is impartialist. Nonetheless, if any of these live theories is correct, then it would be very strange if egoistic influence was *necessarily* distorting in moral epistemology. After all, if right action is definitionally productive of individual self-interest, how can moral belief always be distorted by the influence of self-interest? This would be as strange as holding that objective consequentialism is true but that reflection on the consequences of our actions was necessarily an unreliable way to arrive at moral truths. While this is not logically incoherent, it would be a strange result indeed. This is why a defensible moral epistemology—a defensible set of filters—should not forbid us from being under the influence of factors that lie at the heart of live normative theories.

The point is not just about egoistic influence, for there are also parochial theories, and not all of them are so baldly implausible as Thrasymachus' infamous conception of justice. Most notably, there is the Ethics of Care, pioneered by Carol Gilligan, Nel Noddings, and Virginia Held, which holds that caring for particular others within one's "circle of care" is the foundation of all ethical life (Gilligan, 1993; Held, 2006; Noddings, 2013). This theory is plainly a live option in normative ethics, and if it (or something like it) is correct, then what I am calling Parochial Influence could easily count as epistemologically virtuous. We cannot simply assume that this influence is distorting; to do so is to very nearly preclude the truth of a live moral theory.

[6]William Frankena expressed some astonishment at such conceptions, calling them "basically nonsocial" and suggesting that they might not even count as "moralities" (Frankena, 1966, 692). But such declarations are mainly pointless. If theories of the best life embraced by Plato, Aristotle, Mencius, and Augustine do not count as "moral," then we may simply discard the term "moral" in favor of some broader term, such as the one suggested by the label meta-*ethics* (Williams, 1985).

[7]For discussions and elaboration, see (Annas, 1999; Kent, 2001; LeBar, 2018; McDowell, 1995; Rutherford, 2003). On Mencius see (Huff, 2015), and on Buddhism see (Flanagan, 2019).

In sum, there are huge number of live moral theories which have either egoistic or parochial interest at their core. Moreover, the methods their defenders use to motivate their theories are shot through with the influence of self-interest, since readers are meant to reflect on the ways in which various actions and dispositions would be good *for them*. This, at the very least, should make us suspicious of DSI, and of the fact that it is rarely argued for in any serious manner. After all, a very good explanation for this argumentative lacuna is that moral epistemologists are generally assuming that modern, impartialist, and non-eudaimonistic theories are generally correct, and that this leads them to assume that DSI must be correct, but this is exactly the move that Jones rightly prohibits.

However, as I will now argue, my case is even stronger than this. Suppose, we *do* reject the entire eudaimonist tradition on independent grounds, in favor of some broadly agent-neutral or impartialist moral theory that permits agents to give no intrinsic weight to their own welfare. As it turns out, it is entirely possible that even belief in *impartial* moral theories is motivated, in certain ways, by self-interest.

5 | THE ETIOLOGY OF IMPARTIAL EGALITARIANISM

Where do broadly impartial or egalitarian moral views come from? Here, we should strongly distinguish between the content of a belief and its etiology. Once we do, we can see that it is entirely possible that broadly impartialist moral beliefs often have egoistic or parochial sources. Somewhat counter intuitively, it is possible that they arise via a process that essentially involves the heavy influence of self-interest on moral belief. If this is so, then anyone committed to this kind of moral view must acknowledge that self-interest is at least sometimes positive influence in moral epistemology, since it is a necessary stage on the journey to the truth.

Before, I make this case, a quick distinction. The etiology of moral belief occurs at two explanatory levels, that of the individual and that of the social group. That is, there is the question of how an individual's moral beliefs have formed, and there is the separate (if related) question of how a social group's moral beliefs have been formed. In what follows, I will motivate the twin ideas that the influence of self-interest is crucial for the development of impartial egalitarianism, at both the individual and the social level.

Let us begin at the level of the individual. In a penetrating essay, Erich Fromm argues that a certain kind of self-love is absolutely necessary for the development of *loving* agency, the kind of open, expansive, giving agency that Singer and Unger prize so highly. He argues, persuasively, that our narrow focus on zero sum economic cases has obscured this more general truth: that human beings rarely become deeply altruistic unless they develop "strength, independence, and the integrity of the self". Therefore, taking important or necessary means to establish this integrity cannot constitute "selfishness," nor can it evidence a distorted moral outlook. In fact, he argues that those who are prepared to make the kind of extreme sacrifices required by utilitarian theory normally only do so because, for them, it is (Fromm, 1939).

> necessary to give one's life for the preservation of an idea which has become part of oneself... the sacrifice may be the extreme expression of self-affirmation... In this case the sacrifice is the price to be paid for the realization and affirmation of one's own self
>
> (Fromm, 1939).

This idea is echoed in much modern self-psychology, which sees our capacity to relate to others in a healthy way as essentially bound up with the development and eventual sublimation of our own initial narcissistic tendencies. Here is Heinz Kohut, speaking of our capacity for "object love" (roughly, love of others; Kohut, 2011):

> there are various forms of narcissism that must be considered as forerunners of object love... a number of complex and autonomous achievements of the mature personality are derived from transformations of narcissism, i.e., created by the ego's capacity to tame narcissistic cathexes and to employ them for its highest aims.
>
> (Kohut, 2011, 460).

Thus, even *if* some kind of impartial egalitarianism is true, this does not imply that belief *in* impartial egalitarianism will not require Egoistic and Parochial influence at various points of the whole belief-formation process. In fact, it seems entirely reasonable to think that the egalitarian has to start with a perspective that is somewhat self-aggrandizing, in order to develop a robust sense of their own value. They might then expand their moral concern to a parochial sphere, firmly valuing the lives and welfare of their family and friends. Finally, they might come to see humanity or sentience itself as valuable. And all along, as Fromm insists, they have to maintain a healthy respect for their own psychological integrity, perhaps prioritizing it over the health or well-being of strangers. Only then will they be able to embrace or fully accept the truth of some kind of universalist moral theory. This, I think, is almost certainly the normal developmental case. Moreover, this is not *just* a process that occurs in childhood; fully grown adult moral agents often lack this sense of self-worth and confidence, and some of them obviously learn to develop genuine care and concern for others by first developing this egoistic foundation.

So, in the emergence of expansive altruism or even full-blown impartialist utilitarianism, there is a good case to be made for the necessity of egoistic and parochial influences at the individual level. Moreover, while I lack the space to flesh this claim out, I believe that there is an equally strong case to be made at the social or group level, and once this case is made, we can see that very often the operation of self-interest leads to uncontroversial moral progress. Briefly, the history of human liberation movements and of human rights regimes is suffused with the same defiant perception of particular moral truths that was illustrated earlier in the domestic-violence case. These truths were often the basis for more general or universal frameworks, but nonetheless those frameworks depended, as a matter of pure etiology, on the prior influence of self-interest.[8]

To sum up, I have argued that DSI is false for three principal reasons. First, there are too many obvious counterexamples to the principle, second, it appears to rest on a dogmatic dismissal of the entire Eudaimonist tradition, and third, it stands in danger of rendering even strongly *impartialist* or egalitarian moral theories unjustified. Indeed, at this point it is hard to know which moral view would be left standing if we entirely prohibit the influence of self-interest in moral epistemology.

[8]For example, those who conducted the Haitian slave revolts created the first nation to successfully ban slavery in human history. But of course, they did not merely apply some reflectively endorsed impartiality to their situation and come to oppose slavery. Rather, the etiology in question essentially involved anger and vengefulness directed at their particular situation, both as individuals and as an oppressed group (Louverture, 1953). As historian Samuel Moyn notes, virtually none of the anti-colonial movements emerging from within colonized countries in the 20th century deployed any universalist or abstract notion of *human* rights. Rather, their rhetorical energies were directed firmly on far more Parochial notions. They wrote and spoke and fought "in favor of their *own* rights to sovereignty *as against* their oppressors" (Moyn 2012, ch 3.).

However, at this juncture, it will no doubt be insisted that there remain a large number of cases in which the influence of self-interest does look positively malignant. There are profit-hoarding CEOs who believe that they are morally entitled to their colossal wealth, and whose belief is causally regulated mainly by their enjoyment of having it. There are politicians who rationalize taking bribes, and the thought that "it's fine, there's no harm" is purely motivated by their enjoyment of financial reward. Such figures clearly seem to be led *by self-interest* to moral error. So now, our question—the one we should have been asking all along—is as follows: *under what conditions* is self-interest a distorting factor?

6 | TOWARD A BETTER MORAL EPISTEMOLOGY

Moral epistemologists sometimes operate under the assumption that what Jones calls "filters" are going to be relatively simple and contextually invariant, but it is probably true that no genuine epistemic filters are actually like this. Consider again the idea that we should filter out beliefs which are formed under the influence of prior false beliefs. This, it turns out, is not always the case. After all, successful scientists often operate under assumptions which are known to be false—indeed, they are sometimes not even *approximately* true (Cartwright, 1983). False beliefs often distort inquiry, but sometimes, they very clearly do not. The more general lesson is that legitimate epistemological filters are complex and contextually sensitive.

In the light of this, it is an unfortunate fact that when they are constructing their models, moral epistemologists often rely on simple lists of singular, non-contextual requirements on moral belief formation. Huemer, for example, proposes that we reject ethical intuitions when they are formed under the influence of (1) acculturation, (2) biological programming, (3) emotion, or (4) self-interest (Huemer, 2008, 374–378).[9] These simple, contextually insensitive factors are taken by Huemer to be unconditionally distorting, but a little reflection on any of them will reveal that there may be cases for which they are epistemologically benign or even virtuous. For example, the emotions of anger or sympathy certainly look distorting in a number of cases, but there are also cases where anger or sympathy can appropriately focus one's mind on injustice or on dire need.

In order to show where we might go instead, let me finish with a proposal, which I will not fully defend but which will hopefully point the way toward a more responsible kind of moral epistemology. When is self-interest distorting in morality? My idea, very roughly, is that self-interest is distorting when it justifies itself. Put another way, our suspicion of self-interest in morality derives from a more general suspicion of a certain vicious form of circularity, where a moral belief is caused by the very same thing that it licenses or justifies.

To begin, my normative assumption, which I will not defend here, is that the advantages of social power do not justify themselves. This assumption appears to be widely shared and not particular to any normative theory. It is worth noting that even Aristotle, a staunch defender of slavery even in his own day, did not think that the advantages of power which accrued to masters were somehow *self*-justifying. Quite the opposite, Aristotle was driven to articulate a series of justifications which made reference to the alleged advantage which also accrued to enslaved persons.[10]

[9]For more clear cases of this tendency, see (Nichols, 2014, 733; Sinnott-Armstrong 2006).

[10]"Those who are as different [from Greeks] as the soul from the body or man from beast—and they are in this state if their work is the use of the body, and if this is the best that can come from them—are slaves by nature. *For them it is better* to be ruled in accordance with this sort of rule." (*Politics* 1254b16–21, italics added)

Indeed, this tendency can be found throughout the history of this institution: while comparatively unreflective master classes may have simply proceeded as though the institution was natural and in need of no justification, as soon as the question of justification was raised, rarely did masters or social elites satisfy themselves with the dubious claim that the institution is justified *simply* in virtue of the advantage it brings to them and to their fellow slave owners. Rather, we hear of the slave's natural inferiority, the division of humanity into natural orders, or the lack of favor (for some) in the eyes of God.[11]

Yet, when a master's moral belief (that slavery is just) is significantly caused by the advantages this belief brings to him and to his family, the belief is subject to a cause which is not, at the same time, a justification. This is a classic debunking explanation, since a moral belief is influenced by factors which are not justifying (Sauer, 2018; Vavova, 2016).

What principle do these observations support? Begin with the notion of a *relative social advantage*, that is, of one agent's being comparatively better-off than some other agent. Now, notice that some moral beliefs *directly* rationalize a relative social advantage in the sense that they logically imply support for it. For example:

> Slavery is just

This logically implies that a relative social advantage is just. However, other beliefs do not directly support a relative social advantage, but they nonetheless offer *strong* support for it, given the way the world contingently is. Consider:

> I'm permitted to shout at this waitress for getting my order wrong

At the level of pure meaning, this belief rationalizes nothing other than the shouting. It does not *entail* that any social advantage of mine will be secured or reinforced, because it is possible that the waitress will simply ignore us and remain unscathed. However, given the way the world actually is, it is extremely likely that she will be hurt by our outburst and that we will temporarily enjoy a certain type of social domination over her. To think that we are allowed to shout in this case is to think that we are permitted to enjoy this domination, even if the shouting does not logically entail that domination.

Our moral beliefs rationalize much more than they explicitly say, and in this kind of case.

Bearing this in mind, the resulting moral-epistemological principle is this:

> **(F3)** If an agent holds (a) a belief that *directly* rationalizes a relative social advantage, or (b) a belief that strongly rationalizes a relative social advantage, then their belief is unjustified to the extent that it is explained by their either actually enjoying that same advantage or by their perception of that same advantage.[12]

This condition makes sense of the "obvious" cases mentioned above. The CEO who believes that it is permissible to enjoy his colossal wealth obviously believes in his right to have more social power

[11]For example, many defenders of the slave trade drew upon a theological idea, namely that enslaved populations were "sons of Ham," descendants of Noah's rebellious son who had been cursed by God (Adhikari 1992). Their differential status was (somehow) deserved for this reason.

[12]Combatants in the internalism/externalism wars may wonder whether agents have to have some kind of *access* to this explanatory influence or whether its mere existence debunks their moral belief. I remain entirely neutral on this question here: my view is that either epistemological position can adopt (F3), and that this is exactly as it should be.

than others; this is a case of relatively direct rationalization. There is nothing wrong with this as such; if the CEO has arrived at this belief through careful reflection, imagination, and wide experience; then, perhaps it is justified. But if this belief is mainly explained by his enjoyment of this very advantage, his belief is unjustified.

To repeat, however, not every case will concern direct rationalization. Suppose Jenny's restaurant order arrives and it is incorrect. She might think, "this waitress deserves a loud talking to, and as a paying customer I'm entitled to give her one." This strongly rationalizes, even if it does not directly entail, the further moral idea that she *ought* to enjoy a kind of dominion over people in such situations. That is, she is committed to her right to a relative social advantage, even if this commitment is not strictly one of logical entailment. Yet, suppose that her moral belief is caused by her actual or prospective enjoyment of this very advantage; perhaps she perceives how cathartic it would be, or perhaps she has enjoyed such catharsis before, and this alone drives her to think that she has this right in this situation. This perceived benefit is therefore morally distorting, not merely because it is influencing her moral belief, but because it is influencing the very beliefs which rationalize her enjoyment of the benefit itself.

Moreover, while this is not always the case, this sort of epistemically vicious feedback loop is often social, in a sense we can see whether we return to the case of antebellum slavery. In such a case, the *general* acceptance of slavery is beneficial for slave owners, and there exists a set of feedback mechanisms which ensure that this differential benefit, in turn, explains the general acceptance of the belief. This is a kind of *ideological* social system, one which contains distortions which are, in a way, self-perpetuating (Geuss, 1981; Haslanger, 2007; Shelby, 2003). My suggestion is that when a person's belief is definitively explained by the sort of feedback mechanism just described, that belief lacks justification. Ideology critique, on this proposal, is of crucial importance for moral epistemology.

But sometimes, self-interest will influence moral belief in a way which does not run afoul of (F3). In such cases, self-interest is not a distorting factor. For example, no such ideological influence exists in the case of *anti*-slavery beliefs developed by those leading slave rebellions. Their beliefs are *not* generally accepted by their society, indeed, that is precisely why they are rebelling. Moreover, they are not aiming at a state of affairs in which they become the masters and some other group becomes enslaved; their moral belief is unlikely to support or rationalize any institution with these differential power relations.

Now, it is well documented that persons under the boot of an oppressive regime can believe, sincerely, that the social arrangement is just. It can therefore be asked: what of *pro*-slavery beliefs held by enslaved persons? Such beliefs are often held because those in power have arranged things such that any opposition is extremely dangerous to them and their families. This is a paradigmatic case of a moral belief that is influenced by self-interest. Is this influence distorting? The answer is no. It is true that such a person enjoys the safety and security provided by their moral acquiescence, safety, and security that is not afforded to more rebellious members of the enslaved classes. So, there is some relative social benefit that they do enjoy. However, notice that this is *not the same benefit* that is rationalized by their moral belief. That belief, "our enslavement is just" only rationalizes the benefit enjoyed by the master classes. My condition only rules out moral convictions that both rationalize some relative social benefit and are caused by that same benefit.[13]

[13]This being said, it might be the case that these beliefs will be ensnared by another epistemological filter. While there is some debate about this, it seems reasonable to suspect that beliefs which are unjustified remain so when they are transmitted by testimony. So, given that the acceptance of oppression by the oppressed is very often going to take the form of (subtle, culturally mediated) testimony, and given the previous result—that such beliefs in the masters are unjustified—there is still an important distorting influence on most oppressed people who accept that their oppression is good, right or just. For a series of discussions of this sort of filter, see (Lackey & Sosa, 2006).

Finally, we can now see why the Eudaimonist tradition does not necessarily provide a distorting model of moral investigation. When Mencius advises us to seek virtue in contentment, or when Aristotle tells us to pursue individual *eudaimonia*, they are not telling us to achieve well-being *at the expense* of others, in a way which guarantees our own relative social advantage. While of course these historical views are embedded in aristocratic societies where only certain persons were permitted to flourish in these terms, neo-Aristotelian and neo-Confucian models have easily shed this historical baggage (Angle, 2009; Hursthouse, 1999). The kind of self-interest the Eudaimonist encourages us to pursue is not essentially tied to relative social inequality, and when we fall under its influence, we are therefore not falling afoul of (F3).

In sum, self-interest distorts moral belief when it positively influences the belief *that* an agent's own comparative self-interest ought to exist or persist. At the individual level, a person's belief *that* they deserve some relative social advantage is unjustified if it is principally explained by the advantage itself. Moreover, at the group level, our beliefs can be unjustified when we are caught up in larger ideological systems which have roughly the same functional structure.

7 | MORAL DISAGREEMENT AND THE GLOBAL POOR, AGAIN

Suppose this idea is correct, how can it be applied? Specifically, given that the total prohibition against self-interest arose in debates over our obligations to the global poor and over the significance of moral disagreement, does my principle have anything to say about those debates?

At first glance, the answer looks to be straightforward. Consider the debate over our obligations to distant, suffering strangers. As Enoch notes, many people do seem to believe that

> **(MB)** agents ought to be able to prioritize the welfare of themselves and their loved ones, and as such the well-off are not *required* to donate significant time, money, or resources to these persons.

And, in this case, it can look like:

1. Given the way the world normally is, (MB) rationalizes a certain social inequality; some have resources while others do not. And,
2. The wide acceptance of (MB) helps to maintain this same inequality, disproportionately benefiting those with resources.

Wilson suspected as much, and these were precisely the observations which drove her critique of demandingness arguments. So, my model might seem to support her critique, applying to this case in just the way that the broader principle, DSI, does. Moreover, we may notice that this pattern has repeated itself throughout history: widely held moral beliefs have often both rationalized and helped to maintain one and the same social inequality. Thus, when we turn to the problem of moral disagreement, it looks like objectivists can take equal comfort, since a great deal of moral belief looks like it *has* been distorted by self-interest. Has my refinement turned out to be of no real philosophical interest?

As you might guess, I do not think this is so, because in fact (F3) does *not* support these conclusions. Recall that (F3) requires that the relevant moral belief must directly or strongly rationalize

some relative social advantage. Of course, (MB) does not directly rationalize social inequality; it says nothing about it (compare "slavery is just," which obviously does). To say that agents ought to have certain protected "space" for personal projects, such that they are not required to donate massive amounts of time and resources to the global poor, is in no way to logically imply that some agents ought to have social power or advantage over other agents; one may coherently wish for a world where every agent enjoys a rich, full, project-laden life.

Does MB *strongly* imply support for relative social advantage? Might there be mechanisms, normally hidden from our view, which ensure that our egoistic and parochial pursuits necessarily lead to increased inequality? More generally, when *does* a moral belief strongly rationalize social domination, and when does it not? This, I think, is the extremely interesting question that (F3) leaves us with. Is there a clearly established, well-known mechanism which begins with inequality and domination and ends with the belief (MB)?

Many people are fond of saying that there is, but there are powerful reasons to be suspicious here. Social inequality is unlikely to be the primary cause of (MB). After all, while moral codes vary widely, almost every known human culture encourages its members to give *some* kind of priority to themselves and to their loved ones. Indeed, as Joseph Henrich has recently shown, the strong prioritization of close kin is the cultural and historical norm, and the gradual move away from that moral orientation is peculiar to certain Western cultures (Henrich, 2020). Moreover, this prioritization is currently predominant in many cultures that are at the bottom of the contemporary global economic ladder, and this is exactly the opposite of what you would expect to be the case if (MB) were best explained by the fact that people who accept it enjoy socio-economic advantage. In addition, it is possible that (MB) is probably common to any evolved social species, under the assumption that standard evolutionary mechanisms partly explain some of our basic social tendencies.[14] These considerations strongly suggest that our acceptance of (MB) is not explained in any useful way by our being the beneficiaries of unequal social arrangements. This moral belief may predate most forms of social organization with which we are at all familiar.

Moreover, even if such inequalities are somehow part of the current explanation, there is a clearly imaginable alternative: we can work toward a world where my right to give modest priority to myself in this way is not made possible by the denial of that right to anyone else. There is, of course, the old criticism of this ideal, which is that it cannot work and that it is itself part of a damaging social ideology. If a social system exists where individuals are encouraged to focus on their own self-interest, that system will always produce terrible inequality, or so the criticism goes. This, I think, is one claim at the heart of many Marxist critiques of liberal capitalism, and it is opposed by those who think, more optimistically, that personal projects and a meaningful individual life do not represent a zero-sum game.

For my purposes, I need only note that this question is both incredibly complex and unresolved. If we cannot definitively show that the Marxist critique is correct, then we cannot think that (MB) strongly rationalizes social inequality, and we must, I believe, reject the notion that (MB) has been distorted by self-interest. This is because affirming one's right to prioritize self-interest does not rationalize anyone's relative social advantage. Where there is no such rationalization, there is no epistemological vice, because this is not a case of social power both causing

[14]Evolutionary explanations for human sociability often rely heavily on the mechanism of *kin selection*, which selects-for creatures who will offer help to their genetic relatives before offering help to non-relatives (Smith, 1964). Moreover, the mechanism of *group selection* selects for those who believe that members of a perceived social grouping are more worthy of aid and protection (Sober & Wilson, 1998).

and justifying itself. That said, if the Marxist critique is correct, then (MB) does strongly rationalize the very social inequality that it is caused by, and it is epistemically bankrupt. This is one of the fascinating empirical issues that (F3) leaves us with.

8 | CONCLUSION

When does self-interest distort moral belief? My answer is an unconditional, unwavering "sometimes." More specifically, I have claimed that self-interest distorts when it is basically in the business of justifying itself, *qua* relative social advantage. This makes sense of what I believe to be commonly shared intuitions about who is operating under moral distortion, and it passes no judgment on agents engaged in certain emancipatory or egalitarian projects. However, once moral epistemology moves away from the comforting simplicity of an "always," it faces the bewildering social complexity of that "sometimes." My suggestion is that we learn to live with that, because an absolute prohibition against self-interest in moral epistemology is indefensible.

ORCID

Nicholas Smyth https://orcid.org/0000-0001-6155-1461

REFERENCES

Adhikari, M. (1992). The sons of ham: Slavery and the making of coloured identity. *South African Historical Journal*, *27*(1), 95–112. https://doi.org/10.1080/02582479208671739

Angle, S. C. (2009). *Sagehood: The contemporary significance of neo-Confucian philosophy*. Oxford University Press.

Annas, J. (1999). *Platonic ethics, old and new*, vol. *57*. Cornell University Press.

Ayer, A. J. (1936). *Language, truth and logic*, vol. *47*. V. Gollancz Ltd.

Berkey, B. (2016). The demandingness of morality: Toward a reflective equilibrium. *Philosophical Studies*, *173*(11), 3015–3035. https://doi.org/10.1007/s11098-016-0648-9

Blackburn, S., & Sturgeon, N. L. (1991). Just causes. *Philosophical Studies*, *61*(1/2), 3–17. https://doi.org/10.1007/BF00385831

Boyd, R. (1988). How to be a moral realist. In G. Sayre-McCord (Ed.), *Essays on moral realism* (pp. 181–228). Cornell University Press.

Brink, D. (1984). Moral realism and the sceptical arguments from disagreement and queerness. *Australasian Journal of Philosophy*, *62*(2), 111–125. https://doi.org/10.1080/00048408412341311

Brink, D. (1989). *Moral realism and the foundations of ethics*. Cambridge University Press.

Cartwright, N. (1983). *How the laws of physics lie*. Oxford University Press.

Doris, J., & Plakias, A. (2008). How to argue about disagreement: Evaluative diversity and moral realism. In W. Sinnott-Armstrong (Ed.), *Moral psychology* (pp. 303–331). MIT Press.

Enoch, D. (2009). How is moral disagreement a problem for realism? *Journal of Ethics*, *13*(1), 15–50. https://doi.org/10.1007/s10892-008-9041-z

Enoch, D. (2011). *Taking morality seriously: A defense of robust realism*. Oxford University Press.

Fitzpatrick, S. (2014). Moral realism, moral disagreement, and moral psychology, *Philosophical Papers*, *43*(2), 161–190. https://doi.org/10.1080/05568641.2014.932953

Flanagan, O. (2019). Buddhist Persons and Eudaimonia. *The Routledge Companion to Philosophy of Psychology*, 173–186. Routledge.

Frankena, W. K. (1966). The concept of morality. *The Journal of Philosophy*, *63*(21), 688–696. https://doi.org/10.2307/2024163

Fromm, E. (1939). Selfishness and self-love. *Psychiatry*, *2*, 507–523. https://doi.org/10.1080/00332747.1939.11022262

Geuss, R. (1981). *The idea of a critical theory: Habermas and the Frankfurt School*. Cambridge University Press.

Gilligan, C. (1993). *In a different voice*. Harvard University Press.

Golub, C. (2017). Expressivism and the reliability challenge. *Ethical Theory and Moral Practice*, *20*(4), 797–811. https://doi.org/10.1007/s10677-017-9794-1

Haslanger, S. (2007). "But mom, crop-tops are cute!" Social knowledge, social structure and ideology critique. *Philosophical Issues*, *17*(1), 70–91.

Held, V. (2006). *The ethics of care: Personal, political, and global*. Oxford University Press.

Henrich, J. (2020). *The WEIRDest people in the world: How the West became psychologically peculiar and particularly prosperous*. Farrar, Straus and Giroux.

Huemer, M. (2008). Revisionary intuitionism. *Social Philosophy and Policy*, *25*(1), 368–392. https://doi.org/10.1017/S026505250808014X

Huff, B. I. (2015). Eudaimonism in the Mencius: Fulfilling the Heart. *Dao*, *14*(3), 403–431. http://doi.org/10.1007/s11712-015-9444-z

Hursthouse, R. (1999). Virtue ethics and human nature. *Hume Studies*, *25*(1/2), 67–82.

Jones, K. (2005). Moral epistemology. In F. Jackson, & M. Smith (Eds.), *The Oxford handbook of contemporary philosophy*. Oxford University Press.

Joyce, R. (2006). *The evolution of morality*. MIT Press.

Kent, B. (2001). Augustine's ethics. In E. Stump (Eds.), *The Cambridge companion to Augustine*, pp. 205–233. Cambridge University Press.

Kohut, H. (2011). *The search for the self: Selected writings of Heinz Kohut. 4. 1978–1981*, Vol. *4*. Karnac Books.

Lackey, J., & Sosa, E. (2006). *The epistemology of testimony*. Oxford University Press.

LeBar, M. (2018). Eudaimonism. In N. Snow (Eds.), *The Oxford handbook of virtue*. Oxford University Press.

Louverture, T. (1953). *Toussaint Louverture à travers sa correspondance, 1794–1798*. Industrias Graficas Espana.

Mackie, J. L. (1977). *Ethics: Inventing right and wrong*. Penguin.

McDowell, J. (1995). Eudaimonism and realism in Aristotle's ethics. In R. Heinaman (Ed.), *Aristotle and moral realism* (pp. 201–218). Westview Press.

Moyn, S. (2012). *The last utopia*. Harvard University Press.

Nagel, T. (1986). *The view from nowhere*. Oxford University Press.

Nichols, S. (2014). Process debunking and ethics. *Ethics*, *124*(4), 727–749. https://doi.org/10.1086/675877

Noddings, N. (2013). *Caring: A relational approach to ethics and moral education*. University of California Press.

Parfit, D. (2011). *On what matters*. Oxford University Press.

Rawls, J. (1951). Outline of a decision procedure for ethics. *Philosophical Review*, *60*(2), 177–197. https://doi.org/10.2307/2181696

Rutherford, D. (2003). In pursuit of happiness: Hobbes's new science of ethics. *Philosophical Topics*, *31*(1/2), 369–393. https://doi.org/10.5840/philtopics2003311/26

Sauer, H. (2018). *Debunking arguments in ethics*. Cambridge University Press.

Seipel, P. (2019). Famine, affluence, and philosophers' biases. *Philosophical Studies*, *177*, 2907–2926. https://doi.org/10.1007/s11098-019-01352-7

Shafer-Landau, R. (2003). *Moral realism: A defence*. Oxford University Press.

Shelby, T. (2003). Ideology, racism, and critical social theory. *The Philosophical Forum*, *34*, 153–188. https://doi.org/10.1111/1467-9191.00132

Singer, P. (2010). *The life you can save: How to do your part to end world poverty*. Random House Incorporated.

Sinnott-Armstrong, W. (1996). Moral skepticism and justification. In W. Sinnott-Armstrong, & M. Timmons (Eds.), *Moral knowledge? New readings in moral epistemology*. Oxford University Press.

Sinnott-Armstrong, W. (2006). *Moral skepticisms*. Oxford University Press.

Smith, J. M. (1964). Group selection and kin selection. *Nature*, *201*(4924), 1145–1147. https://doi.org/10.1038/201114 5a0

Smyth, N. (2021). Moral disagreement and non-moral ignorance. *Synthese*, *198*(2), 1089–1108. http://doi.org/10.1007/s11229-019-02084-1

Sober, E., & Wilson, D. S. (1998). *Unto others: The evolution and psychology of unselfish behavior*. Harvard University Press.

Street, S. (2010). What is constructivism in ethics and metaethics? *Philosophy Compass*, *5*(5), 363–384. https://doi.org/10.1111/j.1747-9991.2009.00280.x

Sturgeon, N. (1998). Moral Explanations. In J. Rachels (Ed.), *Ethical theory 1: The question of objectivity*. Oxford University Press.

Unger, P. K. (1996). *Living high and letting die: Our illusion of innocence*. Oxford University Press.
Vavova, K. (2016). Irrelevant influences. *Philosophy and Phenomenological Research*, *96*(1), 134–152. https://doi.org/10.1111/phpr.12297
Wedgwood, R. (2014). Moral disagreement among philosophers. In M. Bergmann, & P. Kain (Eds.), *Challenges to moral and religious belief: Disagreement and evolution* (pp. 23–39). Oxford University Press.
Williams, B. A. O. (1985). *Ethics and the limits of philosophy*, vol. *83*. Harvard University Press.
Wilson, C. (1984). On some alleged limitations to moral Endeavour. *The Journal of Philosophy*, *90*(6), 275–289.
Zimmerman, A. (2010). *Moral epistemology*. Routledge.

How to cite this article: Smyth, N. (2023). When does self-interest distort moral belief? *Analytic Philosophy*, 64, 392–408. https://doi.org/10.1111/phib.12261