

# MORAL REALISM AND THE PROBLEM OF MORAL ALIENS

Thomas GRUNDMANN

ABSTRACT: In this paper, I discuss a new problem for moral realism, the problem of moral aliens. In the first section, I introduce this problem. Moral aliens are people who radically disagree with us concerning moral matters. Moral aliens are neither obviously incoherent nor do they seem to lack rational support from their own perspective. On the one hand, moral realists claim that we should stick to our guns when we encounter moral aliens. On the other hand, moral realists, in contrast to anti-realists, seem to be committed to an epistemic symmetry between us and our moral aliens that forces us into rational suspension of our moral beliefs. Unless one disputes the very possibility of moral aliens, this poses a severe challenge to the moral realist. In the second section, I will address this problem. It will turn out that, on closer scrutiny, we cannot make any sense of the idea that moral aliens should be taken as our epistemic peers. Consequently, there is no way to argue that encountering moral aliens gives us any reason to revise our moral beliefs. If my argument is correct, the possibility of encountering moral aliens poses no real threat to moral realism.

KEYWORDS: disagreement, conciliatory view, steadfast view, moral realism, relativism

Robust moral realists believe that moral judgments have truth-values (cognitivism), that many of them are in fact true (denial of error theories) and that they are made true by facts that do not depend on moral attitudes of the subject or on moral codes of the subject's community (metaphysical realism). According to them, moral facts are completely objective and mind-independent. Moral realists can accept that there is moral disagreement of the everyday kind. Some of it is rooted in disagreement about non-moral facts and can be resolved by supplying further information about the case at hand. For example, two people may disagree about whether it is morally permissible for an agent to take home some item simply because they disagree about whether the agent owns this item. Another example: Descartes would disagree with many of us today about whether it is permissible to treat animals like things because he, in contrast to us, believes that animals do not have a mind and cannot suffer. Apart from this kind of disagreement, there is also genuine moral disagreement about moral values and duties or about how they apply to particular cases. Some such disagreements can be resolved by exchanging reasons. In other cases, the disagreement survives persistent exchange of reasons. Whether abortion is morally permissible, whether the air force is morally

permitted to shoot down an aircraft that has apart from its many innocent passengers a number of terrorists on board who are planning to steer the plane into a building with a huge number of residents, or whether we are morally permitted to perform active euthanasia in specific cases, seems to remain highly controversial even after all relevant arguments have been exchanged. In these rather local cases, moral realists may be tempted to suspend judgment.

As I said, moral realists can accommodate all these cases. But what happens when they encounter people who disagree with them in a more radical way? What should our rational reaction be when we meet people who persistently disagree with us not only about local moral issues but more broadly about the moral assessment of nearly every particular case and, more fundamentally, about moral principles as well? If these people appear to us as being intelligent, thoughtful, well-informed and if their views are without any obvious mark of incoherence, I will refer to them as “moral aliens.” One is inclined to say that one should not simply give in to those moral aliens but stick to one’s guns. However, how can the moral realist accommodate this intuition? From the moral realist’s perspective, only one of the disagreeing parties can be correct. If the moral alien can reasonably be taken to share all of our information about the non-moral facts and appears as internally rational and coherent as we are, then it seems that we must attribute similar weight to both perspectives. The resulting massive belief revision or even moral skepticism is in tension with our intuition that we should rationally stick to our guns when we encounter moral aliens. This seems to raise a severe problem for moral realism.

In this paper, I will argue that moral realists can solve this problem. More specifically, I will explain why moral aliens do not put any rational pressure on moral realists to revise their meta-ethical beliefs. Although this solution is motivated by deeply epistemological considerations which also apply to non-moral domains, it will turn out that it neither depends on a specific meta-epistemological position nor on a specific account of the epistemic significance of disagreement. In section I, I will explain and motivate the problem of moral aliens (for moral realism). In section II, I will argue that this problem dissolves when we look more carefully into its underlying epistemology. I will conclude with some general remarks about my solution to the problem.

## 1. The Problem of Moral Aliens for Moral Realists: The Deeper Motivation

The above sketched problem for moral realism can be articulated by the following inconsistent quartet of prima facie plausible assumptions:<sup>1</sup>

- I. Moral aliens are possible.
- II. If one were to encounter (a significant number of) moral aliens, one would not be rationally required to revise one's moral judgments.
- III. If moral realism is correct, one would be rationally required to revise one's moral judgments significantly, if one were to encounter (a significant number of) moral aliens.
- IV. Moral realism is correct.

To put it in a nutshell: there is a genuine possibility with respect to which our intuitive judgment runs counter to what moral realism requires. This raises a severe problem for moral realism.

In what follows, I will explain and motivate each of the above assumptions. Let me start with (1). What are moral aliens? Being a moral alien is a relative property. Person A is a moral alien with respect to person B if A's moral perspective conflicts radically with B's moral perspective. Being such an alien does not imply anything about who is right and who is wrong.<sup>2</sup> It simply involves global (or at least widespread) moral disagreement at the level of particular moral beliefs and fundamental moral disagreement at the level of moral principles.<sup>3</sup> However, this radical moral disagreement is not sufficient to constitute moral alienness. The

---

<sup>1</sup> Kieran Setiya, *Knowing Right From Wrong* (Oxford: Oxford University Press 2012) and Katia Vavova, "Moral Disagreement and Moral Skepticism," *Philosophical Perspectives* 28 (2014): 302-333, have also addressed the ethical significance of moral aliens. But they take a different stance on it. Setiya also starts with the intuition that we should not give in to what I call 'moral aliens'. But he derives very different metaethical consequence from my. He argues for some kind of epistemic externalism about moral evidence and disputes conciliatory views as well as a methodology of reflective equilibrium for ethics. Vavova argues that conciliationists would not be forced into moral skepticism if they were to encounter moral aliens since they could not reasonably assess them as epistemic peers. Although my own argument is close to hers, it differs in two significant respects. First, Vavova does not take the skeptical worry as a challenge to robust moral realism. Second, I fend off the skeptical challenge from moral aliens much more broadly than she does.

<sup>2</sup> The creature which I call "moral alien" is designed along the lines of what Setiya (*Knowing Right From Wrong*, 19-20) calls a "moral monster." However, while Setiya's terminology already involves an assessment of who is right and who is wrong, mine is neutral.

<sup>3</sup> If only either the particular judgments or the accepted moral principles differ radically, the opponent will have less coherent beliefs than we have.

disagreement also has to persist under full disclosure of reasons and arguments to both parties. Moreover, the moral alien has to share all her non-moral beliefs (and evidence) with the person she is alien to. Finally, the alien must have a seemingly coherent system of moral beliefs that at least appears rational to her opponent. Here is then a definition of being a moral alien:

A is a moral alien with respect to B if and only if

- (i) A's moral beliefs conflict widely and fundamentally with B's moral beliefs,
- (ii) the disagreement between A and B persists through full disclosure of relevant evidence and relevant beliefs,
- (iii) A and B share their evidence and non-moral beliefs,
- (iv) A's moral position is without any obvious incoherence and appears rational to herself.

What kind of moral perspective would a moral alien have? Of course, a moral perspective could radically diverge from ours in many different respects. Let us start with a minimal characterization of the common core of "our" moral perspective: We take it that there is at least a *prima facie* moral duty (or a moral value) of minimizing suffering, of treating all humans equally, of not killing other people, of saving the survival of humanity etc. We also believe that justice matters morally. These seem to be some platitudes of our common moral perspective. Moral aliens would not only disagree with us about them but also consistently apply their shocking platitudes to particular cases. Moral aliens would, e.g., claim "Suffering is morally irrelevant" or "The white race has a supremacy over all other human races" or "Justice is morally irrelevant" or "We need not care about the survival of humanity" etc. Even more radical deviations from our moral point of view are intelligible. We might, e.g., encounter a moral alien who claims:

(B1) It is morally required to maximize human suffering.

Of course, one might be worried that this alien does not talk about what is morally required *in our sense of the word* when her moral platitudes are deviant to such an extent from ours. According to one popular view of moral concepts,<sup>4</sup> they are determined by the moral platitudes which the subject associates with them. If two subjects associate two radically different sets of platitudes with the same linguistic or mental vehicle, it cannot instantiate the same concept. Hence, the creature we take to be a moral alien is not really disagreeing with us about moral

---

<sup>4</sup> See Michael Smith, *The Moral Problem* (Oxford: Blackwell Publishing, 1996); Frank Jackson, *From Metaphysics to Ethics: A Defense of Conceptual Analysis* (Oxford: Oxford University Press, 1998).

issues, but disagrees with us merely verbally. Here is a passage from Jackson (*From Metaphysics to Ethics*, 132) along these lines:

Genuine moral disagreement, as opposed to mere talking past one another, requires a background of shared moral opinion to fix a common, or near enough common, set of meanings for our moral terms. We can think of the rather general principles that we share as commonplaces or platitudes or constitutive principles that make up the core we need to share in order to count as speaking a common moral language.

In this paper, I don't want to engage with the intricate debate about the nature of moral concepts, i.e., with the question whether these concepts should be understood in an internalist, externalist, or primitivist way.<sup>5</sup> For the aim of the paper, this is not required. One can make moral aliens intelligible by telling a story about what motivates their diverging perspective. Recall (B1) which claims that it is morally required to maximize human suffering. Suppose that Theophilus believes that an act is morally required if and only if God commands us to do this thing.<sup>6</sup> Theophilus also believes that God has told him to do his best to increase human suffering in order to punish humans for their constant sinning. Given this background and the rationality of Theophilus' non-moral beliefs, it seems even rational for Theophilus to believe (B1). One might object that this is not a persuasive example of a moral alien, because Theophilus has beliefs about non-moral facts (e.g., about the revelation of God's will) that conflict with ours and our related evidence. Here is another case that may better fit this requirement: Physiophilos has the strong moral intuition that the preservation of biological diversity on earth is an absolute moral obligation. Physiophilos knows that human actions have caused a massive reduction of this diversity in the past and also has good reason to believe that humans will continue with this if their population is not massively reduced. He believes that killing people is the most effective means to stop this practice. Physiophilos therefore concludes (B2) that there is the moral duty to kill the majority of humans. These cases may illustrate that moral aliens holding to moral platitudes that radically differ from ours are indeed possible.<sup>7</sup>

---

<sup>5</sup> Whereas semantic internalists claim that the reference of moral concepts is determined by the platitudes related to them (e.g., Smith, *The Moral Problem*), semantic externalists claim that the reference of moral concepts is determined by their causal regulation (e.g., David Brink, *Moral Realism and the Foundations of Ethics*, Cambridge, MA: Cambridge University Press 1989). Primitivists (e.g., G.E. Moore, *Principia Ethica*, Cambridge, UK: Cambridge University Press 1903) claim that moral concepts such as "good" are not analyzable.

<sup>6</sup> See G.E.M. Anscombe, "Modern Moral Philosophy," *Philosophy* 33 (1958): 1-19, for the modern debate.

<sup>7</sup> Even if dispositions to believe determine meanings, these dispositions can be blocked to

The problem for moral realism does not depend on the actuality of moral aliens. Surely, we do not encounter them in everyday life. People whose behavior radically deviates from what we take to be morally permissible or required are typically either immoral or driven by their emotions and affects. But there seem to exist rare examples of people who at least approximate moral aliens. Adolf Hitler was an at least partly intellectually motivated racist. However, he relied on views about human races that are empirically false. In his infamous 1943 Posen speech, Heinrich Himmler encouraged SS-officers to fulfill their moral duty of annihilating the Jewish people in a morally decent way. His speech seems to express a moral view that radically differs from ours.

What about the second assumption of the inconsistent quartet?

(2) If one were to encounter (a significant number of) moral aliens, one would *not* be rationally required to revise one's moral judgments.

(2) expresses a strong intuition about what is rational when we encounter people with radically different moral views. Kieran Setiya (*Knowing Right From Wrong*, 19-20, my italics) articulates this intuition in the following way:

For the first time, you meet a stranger. He agrees with you outside of ethics, but when it comes to practical reason, his beliefs are shocking. Fill in the details as you like. Perhaps he thinks we should be utterly selfish, that we should maximize aggregate happiness, no matter who is trampled on the way. It turns out that he, too, belongs to a homogeneous community, exactly as numerous as your own. What should you now believe? (...) *We should not defer to moral monsters but condemn them, however numerous they are.*

There are some uncontroversial truisms about situations like the one described: When we encounter someone who believes that we should maximize human suffering or that it is morally required to kill the majority of the human population, this is not only shocking news but also disgusting and repulsive for us. Moreover, our conflicting moral views might be so deeply entrenched in our perspective and so tightly related to our personality that our immediate reaction to the moral alien (whom Setiya—pejoratively—calls “moral monster”) is outright rejection. This much should be uncontroversial. What Setiya claims about these cases is much stronger. According to him, we do not only react in an adverse manner but are also *rationally required* to behave in this way. Setiya maintains

---

manifest in linguistic behavior under stimulus conditions by other beliefs. I take it that Williamson-like cases (Timothy Williamson, *The Philosophy of Philosophy*, Oxford: Blackwell Publishing, 2007) can establish at least this much. For example, Williamson argues that one might even reject the Proposition *All vixens are vixens*, if one believes that there are no vixens and that sentences with universal quantifier phrases have existential implications.

that “we *should* not defer to moral monsters but condemn them” (my italics). I agree that we have the intuition that it would be irrational for us either to give in to the moral alien or to suspend judgment upon encountering her.<sup>8</sup>

Let us move on to the third proposition of the quartet:

- (3) If moral realism is correct, one would be rationally required to revise one’s moral judgments significantly, if one were to encounter (a significant number of) moral aliens.

What is the motivation behind this assumption? From the moral realist’s perspective, one of us—either we or the moral alien—has formed a false moral belief. However, when we reflect on the epistemic situation, each party has something similar going for it: for us, our moral point of view looks rational; the same is true for them, from their perspective. Given this epistemic symmetry, belief revision towards the middle seems to be the only rationally permissible reaction for both parties. And this reaction is in tension with what (2) claims.

Let me present this motivation in more detail by presenting two complementary arguments that call for the realist’s revision of moral judgment. When we are confronting (a group of) moral aliens, they either share all our morally relevant evidence or they possess evidence different from ours on their side. First, consider the case in which our shared non-moral evidence comprises the morally relevant evidence. We can make sense of such a case by assuming that the only evidence for our moral beliefs is evidence of mundane descriptive facts. For example, we justify our belief that a particular action is morally wrong by referring to nothing but the fact that this action causes suffering. Or, to choose another example, we justify our belief that it is morally wrong to keep a particular book by arguing that this book is the property of someone else. It is obvious that on this conception of evidence for moral beliefs the beliefs cannot simply be deduced from our evidence. This is so because in contrast to the moral conclusions the evidence has no moral content. We cannot deduce moral propositions from non-moral ones. Or, to put the same point differently, we cannot logically derive what ought to be done from what is. Nevertheless, either background standards or a certain kind of moral sensibility may facilitate our ability to form moral beliefs on the basis of evidence that has no moral content of its own. According to this view, the total evidence (which is shared by us and our moral aliens) *uniquely* determines which moral attitudes, i.e. either belief, disbelief or suspension, are

---

<sup>8</sup> However, there may be some room for explaining away this intuition such that our hostile attitude towards the moral alien is a matter of fact rather than what is rationally required. One also might argue that we cannot suspend acting such that there is in practice no room for being agnostic.

rational. This uniqueness thesis is supported by the well-known claim that moral facts supervene on non-moral facts.

This leads to the following *argument from undercutting defeat*. We and our moral aliens share all the evidence that is relevant for the justification of moral beliefs. This is true because we assume that all the evidence for moral beliefs is evidence *without moral content*. By stipulation, this evidence is fully shared between us and our moral aliens. At most one of us can have justified moral beliefs because the relevant evidence is shared and uniquely determines which moral beliefs are justified. Since our epistemic situation is fully disclosed to us, we also know that at most one of us—either the moral aliens or us—has justified moral beliefs. Since our moral aliens are as intelligent and thoughtful as we are and seems to have as internally coherent moral beliefs as we have, we as moral realists have no reason to privilege our own perspective. We therefore must conclude that there is a significant chance that we do not adequately respond to the shared evidence. This undermines our justification for our moral beliefs even if we are in fact assessing the evidence correctly. Hence, we must significantly downgrade our moral confidence.

Note that although this argument is conciliationist in spirit, it does not rely on any specific principle from the epistemology of disagreement.<sup>9</sup> It just uses general insights into the dynamics of rational revision of beliefs under defeating evidence. The situation above is similar to the Restaurant Case in which I learn that someone whom I reasonably take to be as competent in mental calculation as myself comes to a different solution to a simple mathematical problem.<sup>10</sup> In that case, my initial justification is undermined even if I, in fact, calculated correctly. The above argument is also neutral with respect to epistemic internalism and externalism since both are committed to acknowledge the defeasibility of justification.

So far, I have argued that there is rational pressure on the moral realist to suspend her moral beliefs upon encountering moral aliens, *if the evidence for their moral beliefs is fully shared*. However, we can also construe the epistemic situation in such a way that both parties have very different bodies of *moral* evidence. Here

---

<sup>9</sup> However, it is not fully neutral. For example, it is incompatible with Kelly's Right Reason View according to which we are fully rational in remaining steadfast in the face of disagreement if we reasoned correctly (Thomas Kelly, "The Epistemic Significance of Disagreement," *Oxford Studies in Epistemology* 1 (2005): 167–196). Kelly has given up this view in the meantime because he thinks that higher-order evidence has at least some epistemic significance.

<sup>10</sup> See David Christensen, "Epistemology of Disagreement: The Good News," *Philosophical Review* 116 (2007): 187–217.

are two different ways of spelling this out. First, we may assume that moral *intuitions* about particular cases or about moral principles constitute our moral evidence.<sup>11</sup> In this case, it seems plausible that the moral alien possesses a body of evidence that is very different from ours since she has different moral intuitions. For example, she may find it intuitively correct that suffering is morally irrelevant. Alternatively, we may assume that moral *perception* constitutes moral evidence.<sup>12</sup> This would mean that people directly observe whether an action is morally right or morally wrong. It then seems plausible that the moral alien would have moral perceptions that are very different from ours for much the same reasons as with respect to intuition. In both cases, the body of moral evidence is not shared between the two parties, even if they learn that they have different moral intuitions or moral perceptions. Why is that? When I learn what someone morally intuits (or perceives) and in this way learn that she intuits and perceives differently, I do not thereby acquire her intuitions or perceptions. Hence, even if I know of the alien's radically different body of moral evidence, I do not thereby share this evidence. We may call evidence that is not shareable through communication "private evidence."<sup>13</sup>

This gives us all that is required to run the *argument from rebutting defeat*: We have our private moral evidence that supports our own moral perspective. We also have reason to believe that the moral aliens have their own private evidence that, since it is radically different from ours, sufficiently supports their alien moral perspective. According to moral realism, both moral perspectives are in conflict with each other. The truth of one excludes the truth of the other. When we realize all of this, we come to know that the moral alien is sufficiently justified in disbelieving what I believe. But then we must accept that there are also strong reasons against our own moral beliefs and this rationally requires us to revise all our moral beliefs.

Taken together, the two arguments robustly support (3): No matter whether we attribute shared morally relevant evidence to the alien or not, the realist seems to be committed to the view that by encountering the moral alien, we acquire evidence that defeats the *prima facie* justification of our moral beliefs in one way

---

<sup>11</sup> Robert Audi, *The Good in the Right: A Theory of Intuition and Normative Value* (Princeton: Princeton University Press, 2005); Michael Huemer, *Ethical Intuitionism* (Houndmills and New York: Palgrave Macmillan, 2005).

<sup>12</sup> See Robert Audi, *Moral Perception* (Princeton: Princeton University Press, 2015).

<sup>13</sup> See Richard Feldman, "Reasonable Religious Disagreement," in *Philosophers Without Good*, ed. Louise Antony (Oxford: Oxford University Press 2007), 194-214.

or another and thus leads to rational revision or even suspension of judgment. This result is in conflict with (2).

By contrast, moral anti-realists clearly have the resources to explain (2) in a pretty straightforward manner. Emotivists may claim that the speakers' moral judgments express their own affective attitudes. It is obvious that, in light of our own affective attitudes, the moral alien performs badly. Other moral anti-realists understand moral utterances as propositional, but analyze the propositional content of the utterance as dependent on the speaker's attitudes or the moral norms of her community. On these (indexical) relativist views, it is obvious why we should not give in to the moral aliens. When asserting "Actions that cause suffering are morally wrong" we mean to express our belief that *according to our moral code, actions that cause suffering are morally wrong*.<sup>14</sup> We should not revise the belief when we encounter a moral alien who, by asserting "Actions that cause suffering are not morally wrong," expresses the belief that *according to her moral code, actions that cause suffering are not morally wrong*. Both beliefs are compatible and thus there is no reason to revise one's judgment. The two parties are simply talking past each other. Finally, there is the view that our moral assertions genuinely conflict with the alien's moral assertions but that their truth can be assessed relative to different perspectives of assessment (ours and theirs). This (genuine) truth-relativism can nicely accommodate the intuition that there is a genuine disagreement between us and moral aliens.<sup>15</sup> It also explains why we rely on our perspective rather than the alien's when we assess the truth of our moral judgments. According to this perspective, we get it right and they get it wrong. This is exactly, what our intuition (2) claims about the case.

## 2. The Problem of Moral Aliens: The Realist's Solution

Given the inconsistent quartet I introduced in section I, moral realism can only be maintained if we give up (1), (2) or (3). Since I find (1) and (2) extremely plausible, I will focus my criticism on (3). In this section, I will argue that although (3) looks convincing in the light of the two arguments given above, it is ultimately false for reasons that are independent of one's views concerning epistemological internalism and externalism or concerning the epistemic significance of disagreement. There is simply no way of interpreting the recognition of moral aliens in such a way that it generates defeaters that remove or reduce the prima facie justification of our moral beliefs. Encountering moral aliens does not put any

---

<sup>14</sup> See Max Kölbel, "Indexical Relativism versus Genuine Relativism," *International Journal of Philosophical Studies* 12 (2004): 300-303.

<sup>15</sup> Kölbel, "Indexical Relativism versus Genuine Relativism," 306.

rational pressure on the moral realist to revise her moral beliefs. Or so I will argue in this section.

First, I will discuss the argument from undercutting defeat:

- (1) The moral alien and I have radically conflicting moral beliefs such that at most one of us can be right. (from moral realism)
- (2) The moral alien and I base our moral beliefs on fully shared non-moral first order evidence. (by stipulation)
- (3) It is impossible that shared non-moral evidence supports conflicting moral beliefs for different agents. (from moral supervenience)
- (4) At least one of us holds moral beliefs that are not supported by the shared evidence. (from 1, 2, 3)
- (5) If at least one of us holds moral beliefs that are unsupported by the evidence, the likelihood that it is me is sufficiently high. (from epistemic peerhood)
- (6) The likelihood of my moral beliefs being unsupported by my first order evidence is sufficiently high. (from 4, 5)
- (7) I sufficiently justify (6) by deducing it from the justified premises (1), (2), (3) and (5). (by assumption)
- (8) If I am sufficiently justified in believing that the likelihood of my beliefs being unsupported by my first order evidence is sufficiently high, I can no longer rationally use my first order evidence to justify my beliefs. (undercutting defeat)
- (9) I can no longer rationally use my first order evidence to justify my moral beliefs. (from 7, 8)

I already said a bit about the motivation of the premises used in this argument before. Let me here add only a few remarks. (Ad1) If moral realism is correct, there is substantial disagreement between me and the moral alien such

that at most one of us can be right (if we hold contrary but not contradictory views, both of us may be mistaken). (Ad2) As I understand moral aliens, they share all our non-moral evidence by definition. As I already indicated above, this argument relies on the assumption that all the morally relevant evidence has non-moral content. Together, this motivates (2). (Ad3) In order to be rational, the formation of moral judgments on the basis of non-moral evidence has to respect that moral facts supervene on or are grounded in non-moral facts, i.e. that there cannot be a difference in the moral value/normative status of an act without there being a difference in the non-moral properties of that act. (Ad5) This premise articulates the idea that moral aliens have a significant epistemic weight relative to me (epistemic peers). This is motivated by the stipulation that my alien is approximately as intelligent and thoughtful in her judgments as I am and that her system of moral beliefs is as coherent as mine. (Ad7) This premise makes sure that the premises of the argument from undercutting defeat are *justified*. This is needed to produce a defeater since defeaters have to be justified in order to do their job of removing prima facie justification. I cannot give a full defense of this assumption here. What is crucial is that defeaters play the role of reasons against beliefs; and naked (unjustified) beliefs do not constitute reasons.<sup>16</sup> As we will see below, the fact that all the premises must be justified rather than true is the crucial weak point of this argument. (Ad8) This premise simply expresses the epistemic principle of undercutting defeat.

Although the argument from undercutting defeat does not explicitly rely on any assumptions about the epistemology of disagreement, there is an easy way out for non-conciliationists. They would object either to premise (5) or (8). The proponents of a non-conciliatory or steadfast position typically believe that in evaluating the justificatory status of beliefs in the face of disagreement we should not fully ignore our first order evidence.<sup>17</sup> Accordingly, they may argue that (5) is false if we have in fact correctly reasoned from our non-moral evidence. In this case, the majority of evidence would be on our rather than on the alien's side. Alternatively, proponents of a steadfast position might also argue that (8) is false because suggesting that there is a realistic chance of having reasoned improperly

---

<sup>16</sup> See William Alston, "Philosophy, Naturalism, and Defeat," in *Naturalism Defeated?*, ed. James K. Beilby (Ithaca and New York: Cornell University Press, 2002), Timothy Loughrist, *Reasons against Belief. A Theory of Epistemic Defeat* (Pro Quest Dissertations & Theses A&I, 2015), and David Alexander, "Unjustified Defeaters," *Erkenntnis* 82 (2017): 891-912, for a defense of the justificatory requirement for epistemic defeaters.

<sup>17</sup> Thomas Kelly, "Peer Disagreement and Higher Order Evidence," in *Social Epistemology: Essential Readings*, eds. Alvin Goldman and Dennis Whitcomb (Oxford: Oxford University Press, 2011), 201.

does not completely screen off the significance of the first order evidence but only reduces it within in the body of total evidence.

I do not take this route since I want to argue that the argument from undercutting defeat is unconvincing no matter what the correct view on the epistemology of disagreement is. For this reason, my argument continues under the less comfortable assumption that conciliationism is the correct view. As I indicated above, the argument from undercutting defeat succeeds only if premise (5) is justified. But can we justify the premise that my peer and I are epistemic peers, i.e. that we are equally competent in reasoning on the basis of the morally relevant evidence? So far, I just assumed that equal intelligence, thoughtfulness and coherence on both sides are sufficient to justify epistemic symmetry. But in fact, we need to establish more than this. In principle, there are three different ways of establishing epistemic peerness. First, we may use a *track-record argument*. When using this kind of argument, we rely on our comparative assessment of past performances of agents or instruments. If they produced roughly the same ratio of correct results in the past, we infer inductively that they are performing equally well in general. Elo ratings in chess play exactly this role. Secondly, we can justify that two people are peers in a specific domain if there are *good indirect indicators* for them being equally competent in their judgments about the domain. We use this source when we attribute peerness to people who do mental calculation or report what they see. In these cases, we assume that *normal* people are equally good at mental math or perceptual judgments. As long as there is no further sign of radical inequality, we are justified in believing that agents who appear to be normal are equally competent. Thirdly, one might think that we are justified by default to assume epistemic equality unless we have positive evidence that one of us is superior. Since Henry Sidgwick was the first to endorse this principle explicitly in his argument for suspending judgment in the face of disagreement, I will refer to it as *Sidgwick's principle*.<sup>18</sup>

I will now argue that we cannot use any of these routes to justify the epistemic peerness of moral aliens. First, we cannot use a track record argument to justify the peerness of moral aliens. This is so because, from our perspective, the moral reasoning of moral aliens looks terribly poor. From our perspective, they have always come to ridiculous moral conclusions when they rely on the same non-moral evidence as we do. Hence, there is no basis for supporting the attribution of equal competence to moral aliens by track record arguments.

---

<sup>18</sup> See Henry Sidgwick, *The Methods of Ethics* (Indianapolis and Cambridge: Hackett Publishing Co., 7th edition, 1907), 342; Sarah McGrath, "Moral Disagreement and Moral Expertise," *Oxford Studies in Metaethics* 4 (2007): 91-92.

Secondly, what about indirect indicators of epistemic peerhood with respect to moral beliefs? We know about the moral alien that she is as intelligent and thoughtful as we are. Moreover, her moral belief system seems to be coherent to the same degree as our moral belief system. Are these good indirect indicators of peerhood? I don't think so. On the one hand, intelligence and thoughtfulness are general abilities that are not significantly correlated with highly domain-specific competences like moral reasoning. On the other hand, the degree of coherence among moral conclusions does not systematically reflect the degree to which they respect the evidence. For example, one might be strongly and robustly biased in one's reasoning towards a certain direction. This may lead to mutually coherent conclusions that are completely inadequate responses to one's evidence. We cannot argue that moral aliens are normal moral thinkers by reference to their internal rationality because they are not. Any *prima facie* justified assumption of normality is defeated as soon as we discover that moral aliens have radically different moral beliefs. At least one of us has radically mistaken moral beliefs. Therefore, it is not possible to argue for the epistemic peerhood of moral aliens in this way either.

Thirdly, one might use Sidgwick's principle to defend the peerhood-assumption concerning moral aliens and thereby motivate our suspension of moral judgment. One might argue as follows: when we encounter a moral alien, we realize that there are two radically different moral perspectives. One of them is ours, the other is the moral alien's. We know that they cannot both be correct, but we have no reason to privilege one of them. Of course, when we rely on our own perspective, the alien's judgments look terrible. But the same is true about us from her perspective. There is simply no perspective accessible to us that permits an impartial assessment of both points of view. Without having any reason to prefer one point of view to the other, we should treat both as equally weighty. For this reason, we should suspend our moral judgment after all. This is the line of reasoning suggested by Sidgwick's principle according to which one ought to treat points of view as epistemically equal unless one has reasons to believe that one is superior to the other. However, this principle licenses verdicts that are intuitively too strong.<sup>19</sup> Consider the following case of an *Epistemic Troublemaker*. Suppose you are giving a public lecture at a foreign university to a very diverse audience. Students sit with faculty and interested laypeople from town. With the exception of the colleague who invited and introduced you, none of the people are known to you. In your talk you defend the proposition that *p* on a topic in normative ethics. At the end of your talk, someone unknown to you stands up and asserts with a

---

<sup>19</sup> See also David Christensen, "Disagreement, Question-Begging and Epistemic Self-Criticism," *Philosopher's Imprint* 11 (2011): 15-16; Vavova, "Moral Disagreement and Moral Skepticism."

serious tone “*p* is false.” Without saying anything else she leaves the room. If Sidgwick’s principle were correct, the troublemaker’s intervention would be sufficient to remove justification from your belief that *p*. This is so because you have no reason to regard the unknown opponent as epistemically inferior to you. But this consequence seems absurd. Justification cannot be lost so easily.

On closer inspection, the argument from undercutting defeat cannot be upheld. The proponent of a steadfast position will object to premise (5) or premise (8). Even on a conciliationist view, one must accept that there is no way to justify the peeriness-assumption concerning moral aliens as it is required by premise (5). In conclusion, the argument fails on both accounts. Note that this criticism does not depend on any commitment to either epistemological internalism or externalism. Either view must accept relevantly similar requirements for epistemic defeat.

Does the *argument from rebutting defeat* fare better? The core idea of the argument is the following: we assume that we and our moral aliens possess different but on both sides sufficiently good private moral evidence. As soon as we acknowledge this fact, we have to accept that there is strong evidence for and against our own moral position. However, if there are not only strong reasons in support of our position but also strong reason against it, a significant revision of our initial position is rationally required. Here is a semi-formal version of the *argument from rebutting defeat* that can be generalized to all moral propositions *p*:

- |  |                    |
|--|--------------------|
| (1) My moral belief that <i>p</i> is sufficiently justified.   | (by assumption)    |
| (2) I am justified in believing that my moral alien is sufficiently justified in believing that not- <i>p</i> .  | (premise)          |
| (3) If someone is sufficiently justified in believing that someone else is sufficiently justified in believing not- <i>p</i> , then she is sufficiently justified in believing not- <i>p</i> . | (premise)          |
| (4) I am sufficiently justified in believing not- <i>p</i> .   | (from 2,3)         |
| (5) I am sufficiently justified in believing <i>p</i> and believing not- <i>p</i> .  | (from 1,4)         |
| (6) If I am sufficiently justified in believing <i>p</i> and believing not- <i>p</i> ,   | (rebutting defeat) |

believing *p* is no longer sufficiently justified.

- (7) Believing *p* is no longer sufficiently justified. (from 5,6)

In what follows, I will argue that no matter whether we understand justification in an externalist (reliabilist) or internalist (mentalistic) way one of the argument's premises will be false—though not the same. Let me start with the reliabilist understanding of justification. Roughly speaking, on a reliabilist account, beliefs are justified only if they are produced by reliable processes that result predominantly in true beliefs. Given this understanding of epistemic justification, premise (2) cannot be true. When we encounter the moral alien, her moral beliefs appear to be widely mistaken and thus the underlying processes appear to be unreliable from our perspective. Hence, we have no reason to believe that they are sufficiently reliable.

By contrast, what does the epistemic situation look like if we understand epistemic justification along the lines of internalist mentalism? Then, justification does not require objective reliability. Hence, we need not attribute reliable mechanisms of forming moral beliefs to the moral alien in order to attribute justified moral beliefs to her. For having the justified belief that *p* it might be sufficient that it seems or appears (in some non-doxastic sense) true to the believer that *p*. In this sense, we can attribute justified moral beliefs to the moral alien, and we can be justified in doing so. We just need to attribute corresponding moral intuitions or moral perceptions to her. On this account, premise (2) is satisfied. But now premise (3) turns out to be problematic. Here is why: Replace “sufficiently justified” in (3) by “appears true.” Then we get: If it appears true to someone (say A) that it appears true to someone else (say B) that not-*p*, then it appears true to A that not-*p*. But why should that be true? Just by acknowledging the appearance of some proposition's truth to someone else, this proposition need not appear true to me. To ascribe justified beliefs to someone else need not justify these beliefs for me.<sup>20</sup> We need a bridging principle to arrive at the required connection. If we

---

<sup>20</sup> Alternatively, one might use Feldman's principle that evidence of evidence for *p* is evidence for *p* itself. However, there are many objections to this principle—in particular, if it is applied intersubjectively. See Feldman, “Reasonable Religious Disagreement,” 151, and the critical discussion in Branden Fitelson, “Evidence of Evidence is not (Necessarily) Evidence,” *Analysis* 72 (2012): 85-88; Juan Comesana and Eyal Tal, “Evidence of Evidence is Evidence (Trivially),” *Analysis* 75 (2015): 557-559; Luca Moretti, “Tal and Comesana on Evidence of Evidence,” *The Reasoner* 10 (2016): 38-39; William Roche, “Evidence of Evidence is Evidence under Screening-

attribute the appearance of  $p$ 's truth to the alien and if we also assume that the alien's appearances are reliable indicators of truths about the relevant domain, then it should appear true to us that  $p$ . However, we cannot add this further assumption in the case at hand for the simple reason that moral aliens do not appear to us as being reliable about the moral domain. We cannot avoid believing that they massively misrepresent moral reality by their appearances.

#### 4. Conclusion

In this paper, I started out with a problem for moral realism, the problem of moral aliens. On the one hand, moral realists want to claim that we should stick to our guns when we encounter radically different moral views. On the other hand, moral realists seem to be committed to the view that there can be a certain epistemic symmetry between us and our moral aliens that forces us into rational suspension of our moral beliefs. Unless one disputes the very possibility of moral aliens, this poses a severe challenge to the moral realist--a challenge that was articulated by the inconsistent quartet.

On closer scrutiny, it turned out that we cannot make any sense of the idea that the moral aliens should be taken as our epistemic peers. The epistemic asymmetry between us and them is inescapable. Interestingly, my argument does not rely on any meta-epistemological or methodological background assumptions. No matter whether one is an internalist or externalist, a steadfast or a proponent of conciliationism, there is simply no way to argue that encountering a moral alien gives us any reason to revise our moral beliefs. If this is correct, the possibility of meeting moral aliens poses no real challenge to moral realism.<sup>21</sup>

---

Off," *Episteme* 11 (2014): 119-124, Eyal Tal and Juan Comesana, "Is Evidence of Evidence Evidence?," *Nous* 51 (2017): 95-112.

<sup>21</sup> **Acknowledgements:** Early versions of this paper were presented at the Conference on Moral Disagreement, organized by the ACU at Rome in March 2018, and at the Concept Epistemology Brownbag at the University of Cologne in May 2018. I am particularly grateful to Robert Audi and Massimo Dell'Utri for their extremely valuable written comments on my paper. Further comments from and discussions with the following colleagues also helped me to work out the final version of this paper: Robert Adams, Mark Alfano, Dominik Balg, Sven Bernecker, Sofia Bokros, Jan Constantin, Roger Crisp, Amy Flowerree, Michael Moore, Jakob Ohlhorst, Francesco Praolini, Paul Silva. I am extremely grateful to all of them.