This is the Author's Original Manuscript (AOM). The Version of Record has been published online in *Inquiry: An Interdisciplinary Journal of Philosophy*, 31 December 2021, https://www.tandfonline.com/doi/full/10.1080/0020174X.2021.2014357

From Rational Self-Interest to Liberalism:

A Hole in Cofnas's Debunking Explanation of Moral Progress

Marcus Arvan

The University of Tampa

Abstract: Michael Huemer argues that cross-cultural convergence toward liberal moral values is evidence of objective moral progress, and by extension, evidence for moral realism. Nathan Cofnas claims to debunk Huemer's argument by contending that convergence toward liberal moral values can be better explained by 'two related non-truth-tracking processes': self-interest and its long-term tendency to result in social conditions conducive to greater empathy. This article argues that although Cofnas successfully debunks Huemer's convergence argument for one influential form of moral realism—*Robust Moral Realism*, which holds that moral facts are non-natural, stance-independent normative facts—Cofnas's debunking argument broadly supports a second type of moral realism: *Enlightened Self-Interest Realism*, the view that moral facts are reducible to stance-dependent requirements of instrumental ('means-end') rationality. Finally, this article argues that insofar as different Enlightened Self-Interest Realist theories make specific predictions about the intra- and inter-personal mechanisms behind moral convergence toward liberalism, empirical observations of cross-cultural convergence can provide independent support for Enlightened Self-Interest Realism. I conclude that this is an important mark in favor of Enlightened Self-Interest Realism over Robust Moral Realism.

Key words: ethics, evolution, metaethics, moral realism, naturalism, non-naturalism

Michael Huemer (2016) argues that cross-cultural convergence toward liberal values is evidence of the objective truth of liberalism, and by extension, evidence for moral realism. Specifically, Huemer contends that convergence toward liberal values cannot be adequately

explained in evolutionary or cultural terms, and that the best explanation of this convergence is that people across different cultures have progressively come to recognize the objective truth of liberalism. In response, Nathan Cofnas (2020, p. 3171) contends that Huemer's convergence argument can be debunked by 'two related non-truth-tracking processes':

> First, large numbers of people gravitate to liberal values for reasons of self-interest. Second, as societies become more prosperous and advanced, they become more effective at suppressing violence, and they create conditions where people are more likely to empathize with others, which encourages liberalism. (ibid.)

The present paper contends that Cofnas's debunking explanation at most undermines a convergence argument for one influential form of moral realism: *Robust Moral Realism*, the view that moral truths are 'stance-independent' (or *sui generis*) non-natural normative facts (see Kant 1785, Brink 1989, Cuneo 2007, Enoch 2011, FitzPatrick 2008, Huemer 2005, Moore 1903, Parfit 2011, and Shafer-Landau 2003). I show that although Cofnas's debunking story undermines Huemer's convergence argument for Robust Moral Realism, Cofnas's debunking story broadly supports another form of moral realism: *Enlightened Self-Interest Realism*, the view that moral truths are reducible to stance-dependent requirements of instrumental, means-end rationality. While Enlightened Self-Interest Realism does not enjoy the same level of popularity today as Robust Moral Realism, it has a long and venerable history[1], as well as contemporary proponents, many of whom argue that instrumental rationality requires adopting liberal values (see Gauthier 1986; Arvan 2016, 2020; and Vanderschraaf 2019. See

[1] See Hobbes (1651), Cudd & Eftekhari (2021), and Sidgwick (1874), who notes, examining an array of theorists, that 'it has been widely held by even orthodox moralists that all morality rests ultimately on the basis of "reasonable self-love"; i.e. that its rules are ultimately binding on any individual only so far as it is his interest on the whole to observe them' (p. 7; see also pp-1. 120; Book I, Ch. VII; and Book II, Chs. III-VII).

also Buchanan 1975, Narveson 1988, and Vanberg 2014. Cf. Railton 1986 and Luco 2019).[2] Although I do not provide a full defense of Enlightened Self-Interest Realism here, I suggest that its resistance to Cofnas's debunking story is a mark in its favor over Robust Moral Realism, and furthermore, that insofar as Enlightened Self-Interest Realist theories make detailed inter- and intra-personal empirical predictions about *why and how* individuals and societies should converge toward liberal moral values, specific forms of Enlightened Self-Interest Realism can receive empirical support or disconfirmation by the examination of actual processes of convergence toward liberal values.

1. Huemer v. Cofnas: The State of Play

Huemer argues that the best explanation of 'the spread of liberalism across the world over the course of human history, especially recent history' is that liberalism constitutes an objectively

[2] Earlier precursors to the idea that rational self-interest supports the adoption of liberal-democratic values can be found in Aristotle and Hobbes. In *Nicomachean Ethics* (Books I and II), Aristotle argues that moral virtue is necessary for achieving our highest end: a happy/*eudaimon* life. Then, in the *Politics*, Aristotle argues that the best realistic form of government for achieving *eudaimonia* is one where citizens *freely* pursue the good life together (1278^{b}15, 1260^{b}27), vote for representatives (Book III, Chapter 11), and where wealth and political power are distributed broadly equitably, favoring neither the rich nor the poor (Book IV). Accordingly, although Aristotle notoriously defended slavery (Book I, Part IV), a meritocratic account of justice (1281a5-6), and perfectionist political ideals (Book VI), there is also clear sense in which Aristotle argued that *some* broadly liberal political values and practices (voting, political participation, and otherwise equitable distributions of wealth and power) are requirements of enlightened self-interest. Similarly, although Hobbes [1651] argues for near-absolute submission to government (Chapter XVIII) and the practical superiority of monarchy (Chapter XIX), Hobbes also argues that self-interest requires obeying the 'Law of the Gospel': the Golden Rule do unto others as we think they should do unto us (Chapter XIV). Because the Golden Rule can be plausibly appealed to in defense of liberal norms, values, and practices (viz. we should confer liberal rights, opportunities, etc. on others because we would want them conferred upon us—see Locke 1689, esp. Chapter 2), one can plausibly argue that although Hobbes thought people should submit to whatever government they have, Hobbes *should* have defended the superiority of liberal values (for an argument to this effect, see Gauthier 1986).

true set of moral requirements (Huemer 2016, p. 2007). Here, Huemer understands liberalism as a 'broad ethical orientation [that] (1) recognizes the moral equality of persons, (2) promotes respect for the dignity of the individual, and (3) opposes gratuitous coercion and violence' (ibid., p. 1987). The crux of Huemer's argument, as noted earlier, is that neither evolutionary history nor purely cultural processes can explain this convergence. Here, Huemer claims that the convergence toward liberalism has occurred too rapidly to be explainable by evolution, and that purely cultural explanations would require positing an implausible set of 'large coincidences' across cultures (ibid., p. 2007).

Cofnas argues, to the contrary, that convergence on liberalism can be plausibly explained on cultural and evolutionary grounds, without positing any large or implausible coincidences. Specifically, he contends that the following naturalistic story explains convergence on liberalism in terms of plausible claims concerning evolved human psychology and cultural dynamics:

1. Earlier human societies were structured like chimpanzee communities, dominated by alpha males.
2. Several hundred thousand years ago, hunter-gatherers overthrew the alpha males to establish 'strong egalitarian norms (at least among adult men).'
3. Social hierarchies reemerged with the advent of agriculture, with hierarchical and militaristic social systems exterminating or absorbing more egalitarian ones.
4. *For reasons of self-interest*, most people in hierarchical societies did not enjoy being abused by those in power and resisted hierarchical mistreatment and violence, demanding more liberal treatment.
5. As societies become more liberal, *social conditions reinforce* psychological dispositions such as empathy and aversion to violence that in turn drive further liberalism.

6. 'The main drivers of this trend, *the pursuit of self-interest and empathetic concern for others, do not track objective moral truth*.'
7. Thus, Huemer's argument for moral realism is blocked. (Cofnas 2020, p. 3175)

To be clear, Cofnas does not contend that this argument refutes moral realism. His claim is merely that, 'If both realism and antirealism predict convergence [toward liberalism] then the fact of convergence per se does not support either ethical view' (ibid., p. 3189). Notice, however, that Cofnas's argument hangs, across premises (4)-(6), on several claims about the nature of moral truth: namely, that self-interest and socially reinforced empathetic concern are 'wholly naturalistic' and do not track moral truth. I will now argue that Robust Moral Realism cannot deny (6) without begging the question, assuming what Huemer's convergence argument is supposed to show. Robust Moral Realists can claim that liberalism is objectively true in a sui generis, stance-independent sense, contending in addition that self-interest and empathy also converge upon liberalism, as well. However, as we will see, Robust Moral Realists can only do this by begging the question, presupposing that this is a better explanation of cross-cultural convergence than the alternative hypothesis that sui generis, objective moral truths do not exist and cross-cultural convergence is *only* the result of self-interest. I will then argue that Enlightened Self-Interest Realism can deny (6) without similarly begging the question. Cofnas's story—if it is broadly empirically correct—constitutes a set of empirical facts that can be understood as providing independent confirmation of the normative predictions of forms of Enlightened Self-Interest Realism that defend liberal values: namely, their prediction that persons and cultures *should* progressively converge on liberalism due to specific intra- and inter-personal mechanisms. If this is right, then Cofnas's story does debunk a Huemer-style convergence argument for Robust Moral Realism, but not for Enlightened Self-Interest Realism. On the contrary, if Cofnas's story is empirically accurate, then further empirical study of the

mechanisms by which cross-cultural convergence toward liberal values occurs can serve as independent confirmation of some form of Enlightened Self-Interest Realism.

2. The Hole in Cofnas's Argument: Two Forms of Moral Realism

Let us begin by clearly defining each form of moral realism at play. Some moral realists contend that it is the doctrine that there exist objective, *stance-independent* moral facts—that is, that moral facts are *sui generis*, non-natural normative facts that do not reduce to facts of human psychology, such as what people want or prefer (Cuneo 2007, Enoch 2011, FitzPatrick 2008, Huemer 2005, Moore 1903). However, this way of defining moral realism is controversial and argued by critics to be artificially narrow (Brink 1989, Boyd 1988, Copp 2007, Gauthier 1986, Luco 2019, and Railton 1986). After all, moral realism can be less controversially understood simply as the position that 'moral claims do purport to report facts and are true if they get the facts right', where it is a further substantive question whether moral facts are stance-independent or stance-dependent (Sayre-McCord 2021). Once we adopt this less-controversial definition, we can see that there are at least two potential types of moral realism to examine in the context of the debate between Huemer and Cofnas:

> **Robust Moral Realism:** objective moral truths exist, but are non-natural, stance-independent *sui generis* normative requirements that are neither identical to nor reducible to normative requirements of instrumental, 'means-end' rationality.

> **Enlightened Self-Interest Realism:** objective moral truths exist, but *are* reducible to stance-dependent facts regarding 'rational self-interest', where rational self-interest is broadly understood in terms of *means-end rationality*, or the instrumentally rational pursuit of the agent's ends, whatever they may be, including ends rooted in natural or cultivated sympathy, empathy, etc.

Robust Moral Realism currently appears to be dominant among moral realists today. Kantians, for example, hold that moral truths are categorical imperatives, deeming requirements of means-end rationality mere hypothetical imperatives, or imperatives of prudence. Similarly, externalist moral realists who subscribe to 'normative reasons primitivism'—the view that moral reasons are irreducible to *anything* more basic—hold that moral truths are 'out there' in the world to be perceived, similar to how we perceive tables or chairs (but which, in the case of moral truths, we are said to perceive in the first instance via moral intuition). Enlightened Self-Interest Realists, however, envision moral facts very differently. Hobbes, for example, famously argues that moral facts are identical to requirements of rational self-interest, claiming that his proposed moral 'Laws of Nature' are but 'conclusions or theorems concerning what conduceth to [people's] conservation and defense of themselves' (Hobbes 1651, p. 117). More recently, other Enlightened Self-Interest Realists have argued that moral truths are reducible to higher-order, instrumentally rational constraints on the instrumentally rational pursuit of our first-order ends: specifically, as means to stable, productive, and mutual beneficial social cooperation (Gauthier 1986, Arvan 2016, 2020).

Of course, Robust Moral Realists and Enlightened Self-Interest Realists have arguments against each other's positions. For example, Robust Moral Realists sometimes argue that it is a conceptual mistake to identify or reduce moral truths to truths about rational self-interest, arguing that enlightened self-interest theories are at most theories of prudence, not morality (Kant 1785, Joyce 2001). Other Robust Realists argue that as stance-dependent normative requirements, instrumental requirements cannot play an intuitively suitable role in justifying actions (Scanlon 2014. Cf. Luco 2016). However, Enlightened Self-Interest Realists typically reply that there are compelling metaphilosophical grounds to identify moral truths with requirements of instrumental rationality, even if this requires revising our conceptual scheme

about morality and our views about what 'suitable' justification involves (Hobbes 1651, Chapter IV; Gauthier 1986, Chapter 1; and Arvan 2016, Chapter 1).

We cannot settle these debates here. Instead, the relevant points for our purposes are two-fold. First, because Enlightened Self-Interest Realism in the Hobbesian contractarian tradition has a number of contemporary adherents, it is worth examining whether the view withstands Cofnas's debunking argument in a way that Robust Moral Realism does not. Second, if (as I shall maintain) Cofnas's story coheres with Enlightened Self-Interest Realism—drawing attention to ways in which liberalism in fact appears to be in people's long-term rational interests—then observed convergence toward liberal values may provide some additional grounds to favor Enlightened Self-Interest Realism over Sui Generis Realism: namely, on the grounds that Enlightened Self-Interest Realism is a more parsimonious and unified *natural* explanation of why convergence toward liberalism is convergence upon moral facts. More on this later.

Let us examine, first, whether Cofnas's story plausibly debunks a convergence argument for Robust Moral Realism. Cofnas recognizes that it is possible that convergence toward liberalism might be overdetermined by moral reasons and reasons of self-interest and empathy. That is, cultures might converge on liberalism both because liberalism is a *sui generis*, non-natural moral truth, as well as for reasons of self-interest and cultivated empathy. Cofnas's point, however, is that insofar as such convergence can be explained purely in terms of self-interest, empathy, and sociocultural dynamics—without appeal to any sui generis, non-natural moral facts—such convergence provides no evidence for the latter. For all the convergence itself shows, there are no (sui generis) moral facts, and it is only reasons of self-interest, empathy, and cultural dynamics that have driven the convergence. This line of argument seems to me sound. Because the Robust Moral Realism takes moral facts to be fundamentally distinct

from instrumental requirements of rational self-interest, the only way for the Robust Realist to respond to Cofnas's debunking story is to assume what Huemer's convergence argument is supposed to show: that *sui generis*, non-natural moral facts are a *better* explanation of convergence on liberalism than any purely naturalistic explanation. But that would not only be question-begging. If, *qua* Cofnas's debunking story, we can wholly explain convergence on liberalism naturalistically, without positing *any* non-naturalistic, sui generis moral facts, then the simplest and most parsimonious explanation is the naturalistic one, not the Robust Moral Realist explanation. The lesson here, as Cofnas puts it, is not that Robust Moral Realism is false. The point is simply that if moral facts are supposed to be sui generis, non-natural entities, and there is a plausible naturalistic explanation of convergence toward liberalism that involves appeal to *no* such facts, then such convergence by itself is *no* good evidence for their existence. Cofnas's debunking argument thus seems sound, at least when directed at Robust Moral Realism.

Now turn to Enlightened Self-Interest Realism. A central point of Cofnas's story is that in the long run, illiberal values and practices are neither in the self-interest of those lower in social hierarchies nor those in positions of power. First, Cofnas implies that those who are oppressed by illiberal hierarchies have clear grounds of rational self-interest, *at least in the long run*, to rebel against and overthrow those hierarchies in favor of more liberal ones (Cofnas 2020, p. 3181-2 and §7.2). Sure, rebelling against illiberal values and practices can have immense costs (up to and including death). But notice: this is consistent with the idea that as ideals to strive toward, liberal values and practices that are genuinely in the long-term self-interest of the oppressed. In the long run, slaves are better off becoming *not slaves*; people who are oppressed by race, gender, sexuality, and so on, better off becoming *not oppressed*; and so on. Second, Cofnas also implies an analogous argument that, at least in the long-run, it is

plausibly in the interest oppressors in illiberal systems to afford those lower in the hierarchy greater rights, freedoms, equal treatment, and so on (ibid., §7). Here again, the reasons why are fairly obvious: those who are oppressed by illiberal values and practices tend to revolt (at least in the long run) against illiberal hierarchies, seeking to *make* it in the interest of their illiberal oppressors to change their society and moral views in a more liberal direction. This rarely comes easily and without great cost, of course. For example, a long and bloody Civil War had to be fought in the United States for slaves to be emancipated. Similarly, blacks, women, and other minorities (including sexual minorities) have had to organize, protest, and rise to positions of economic and political power over countless generations to progressively combat racism, sexism, and other forms of illiberal bigotry (and indeed, these battles continue). The point is simply that, as Cofnas's story illustrates, this in fact appears to be how social convergence toward liberalism has broadly occurred (and continues to occur): through (1) the oppressed seeing it as in their own long-term self-interest to live under more liberal conditions, and (2) the oppressed and their empathetic allies to fight to make it in the self-interest of their oppressors to change and become more liberal as well. Railton (1986, pp. 191-2), in defending a naturalistic form of Moral Realism, made similar points many years ago: that repressive moral systems tend to generate unrest, undermining their own continued existence.

But this is not all. Cofnas's story has a second part, which is that from the standpoint of self-interest, liberal values and practices have strong tendency to become self-reinforcing over time. Specifically, Cofnas claims that once liberal values take root in a given social system, social incentives tend to arise to induce people to engage in even greater empathy and aversion to violence. We can see how plausible this is by considering one representative case: feminist

movements to overthrow sexist, patriarchal values and social practices.[3] As #MeToo and other social movements vividly illustrate, for much of American history until quite recently, American men (particularly white men) were able to treat women and minorities with broad impunity. Women and disadvantaged minorities had little power or social standing to incentivize such men to act empathetically and fairly. Yet, across the many decades to follow, women and other oppressed populations have fought—slowly but progressively across many cultures—to hold privileged men to account. In exposing perpetrators of sexual violence and misconduct, #MeToo has served to incentivize and reinforce more liberal and equal treatment of women. This is not to say that these battles (against sexism, racism, etc.) have been won. What it is to say is that we clearly do see—both in the United States, and in countries around the world that have converged toward liberalism[4]—the kind of self-reinforcing character that Cofnas ascribes to liberal values, norms, and practices.

Now, it is one thing to say that it is in the long-term interest of those oppressed by illiberal values and practices to fight for more liberal ones, and for them seek to make it in the interest of others (particularly oppressors) to change. It is another thing to explain how these facts about long-term self-interest constitute *moral truths*. Gauthier and Arvan both defend this step on methodological grounds. First, whereas many people are skeptical of the existence of *sui generis,* non-natural moral truths, instrumental means-end rationality enjoys widespread recognition (Arvan 2016, pp. 24-35; Cf. Gauthier 1986, pp. 8, 17; Huemer 2021). Second, reducing moral facts to requirements of instrumental rationality arguably has a variety of theoretical and practical advantages, including ontological parsimony, explanatory power,

[3] I do not mean to imply that this is all feminism aims to do, as I recognize that feminism is increasingly intersectional. I simply focus on sexism and patriarchy for simplicity.

[4] See, for example, the Women's Rights Movement in Islam, liberation movements in Asia, etc.

unity, and engagement with people's motives to explain moral motivation (Arvan 2016, pp. 29-30, 218-29; Arvan 2020, Chapter 4. Cf. Prichard 1912, pp. 22-3). Finally, in addition to providing arguments for why we should be willing to identify moral truths with truths regarding self-interest, different Enlightened Self-Interest theorists give different theories of why individuals and societies should gravitate toward liberal values, and via which particular mechanisms. For example, Gauthier argues that under the kinds of conditions in which we normally find ourselves—conditions of scarcity where *laissez faire* market interactions would give rise to market failures (Gauthier 1986, Chapters IV-V)—it is instrumentally rational for individuals to constrain their actions according to an impartial, hypothetical moral agreement. Gauthier's basic ideas here are straightforward. Following Hume (1888, Book III, Part II, Sections I-III) and Rawls (1971, p. 126), Gauthier (1986, pp. 113-4, 333-5) notes that normally find ourselves living among other people under conditions where it is difficult to obtain everything we want or need—such as food, water, shelter, but also social goods such as wealth. Gauthier then notes (ibid., p. 114) that 'we become aware of each other as competitors for scarce goods', and that market interactions can both increase the availability of many scarce goods (viz. farming, commerce, etc.) as well as provide 'new benefits', such as technology. However, market interactions generate *externalities* (such as pollution or the concentration of wealth and power), which impose new costs upon people: a kind of 'market failure' (ibid., p. 116). Consequently, Gauthier argues, an instrumentally rational person should want to protect themselves against such market failures—which we can do by acting on a 'joint cooperative strategy' that we and others can freely accept as an interpersonal bargain to correct for externalities, dividing them between themselves and others (ibid., p. 128). Finally, Gauthier contends, the only principles we can expect others to freely accept as a joint cooperative strategy are ones that treat every individual fairly (ibid., Chapters VII-VIII, p. 338). So,

enlightened rational self-interest requires *becoming liberal individuals*: people who value others as equals (ibid., pp. 338, 347 and Chapter XI).

Other Enlightened Self-Interest Realists defend the rationality of liberal values via similar, though distinct, mechanisms. For example, Arvan (2020, pp. 26-8) argues that prudent, instrumentally rational individuals ought to aim to maximize their own expected lifetime utility. However, due to the profoundly uncertain nature of the future, Arvan argues (following Donald Bruckner) that prudence requires adopting a standpoint of *radical diachronic uncertainty* modeled via a Prudential Original Position: a hypothetical model wherein one acts behind a veil of ignorance applied to one's own possible future selves, withholding from oneself any knowledge of which future selves one is likely to be (ibid., pp. 28-32). To simplify greatly, Arvan then argues that *moral risk-aversion*—very roughly, aversion to violating social norms against murder, theft, lying, infidelity, etc.—is prudentially rational given radical diachronic uncertainty. This is because violating such norms routinely results in prudential disaster for violators, both for individuals and groups they comprise (ibid., pp. 32-52). Finally, Arvan argues that moral risk-aversion makes it prudentially rational to act in ways that approximate a *fair balance* between self-interest and other-regard—which Arvan argues is best modeled by a series of original positions: a Moral Original Position for selecting moral principles justifiable to oneself and others, followed by a series of Rawlsian Social and Political Original Positions for applying moral principles to society (ibid., Chapter 3. See also Arvan 2016). Importantly, much as Gauthier argues that instrumental rationality requires becoming a liberal individual, Arvan argues that the above variants of the original position justify liberal requirements of fairness: prudence being *fairness to oneself* across time, morality being *fairness to others*, and justice being *fairness in society and politics* (Arvan 2020, pp. 83-7. Cf. Gauthier 1986, pp. 343, 348). Other Enlightened Self-Interest theorists give similar, albeit distinct accounts of why rational

self-interest should lead to convergence on liberal values. For example, Vanderschraaf argues that society itself can be understood as a solution to bargaining problems, and 'inductive learning models applied to several well-known bargaining problems yield evolved distributions of bargaining conventions that are centered around the *egalitarian solution*…' (Vanderschraaf 2021, p. 1703. See also Vanderschraaf 2019, Chapter 5).

Notice how well each of these forms of Enlightened Self-Interest Realism cohere with Cofnas's 'debunking story' of moral progress. Cofnas's basic idea is that in the longer run, illiberal conditions are neither in the interest of the oppressed nor their oppressors. The oppressed have self-interested grounds to seek more liberal conditions, ones that treat them as equals with the same rights and freedoms as others. Then, because it is in the interest for oppressed to seek more liberal conditions, it is also in the long-term interest of oppressors to seek liberal values as well—since, as we see in civil wars and social unrest, illiberal forms of oppression tend to lead, at least in the long run, to retributive actions against oppressors and in favor of the promotion of more liberal conditions. But this, as we see above, is just what contemporary Enlightened Self-Interest Realists generally hold: that illiberal moral values are socially unstable because they tend not to be in people's long-term interests, particularly over generations. To be clear, some individuals—such as kings, tyrants, slaveholders, and others—may admittedly benefit from illiberal values and social systems over the course of their lives. Arvan (2020, pp. 128-30) suggests that this raises interesting questions for Enlightened Self-Interest Realism about the normative scope of moral requirements—specifically, about whether moral norms normatively apply to all individuals in all circumstances. While these potential implications of Enlightened Self-Interest Realism have been long controversial (see Cudd & Eftekhari 2021, §§3-4 for an overview. Cf. Robson 2015), the relevant point for our purposes is that Enlightened Self-Interest theories of morality provide a ready explanation for

long-term, cross-cultural convergence toward liberalism over generations. Thus, unlike Robust Moral Realism—which Cofnas's story does debunk a Huemer-style convergence argument for—contemporary forms of Enlightened Self-Interest Realism predict cross-cultural convergence toward liberal moral values, specifying *specific* intra- and interpersonal mechanisms to how and why convergence occurs. Consequently, to the extent that Enlightened Self-Interest theories posit specific mechanisms by which means-end rationality should result in cross-cultural convergence toward liberal values, observed facts about cross-cultural convergence—and specifically, observation of the mechanisms by which societies converge toward liberal moral values—can be understood as independent confirmation of the normative and empirical predictions of these theories, and hence, of their accounts of which moral values are objectively correct and why (for an argument that Enlightened Self-Interest theories imply testable normative and empirical predictions, see Arvan 2020, Chapter 4).

3. Replies to Potential Objections

Objection 1: 'Even if we grant that there are long-term prudential reasons for individuals to have or pursue liberal ideals, these again are at most prudential facts regarding rational self-interest, not moral facts (which concern very different kinds of reasons).'

Reply: This objection is based on the relatively common assumption—at least in contemporary metaethics—that moral truths must be categorical and 'stance-independent' (Joyce 2001, 2016). However, this conception of morality has received numerous philosophical critiques (see e.g. Anscombe 1968, Foot 1972, Forcehimes & Semrau 2018, and Velleman 2013). Enlightened Self-Interest theorists often argue that it is overly restrictive conceptually and not a well-supported assumption methodologically (Gauther 1986, Chapter 1; Arvan 2016, Chapter 1). Further, there is also ample empirical evidence that ordinary laypeople are *not*

"unrestricted objectivists" about morality in the sense that the 'categorical' conception of morality seems to presuppose (Beebe & Sackris 2016); and indeed, ordinary laypeople tend to be *metaethical pluralists* willing to recognize a number of things as 'morality' (Wright *et al.* 2013. See also Davis 2021, Goodwin & Darley 2008, Pölzler & Wright 2020). Finally, Huemer (2021) himself recognizes that many laypeople are willing to take seriously the idea that morality is a matter of self-interest or otherwise stance-dependent. We cannot settle these debates here. The relevant point is that because these are open debates, this paper's line of argument is of interest. It should be of interest, in particular, to anyone who is skeptical of Robust Moral Realism, but who is otherwise broadly amenable to Enlightened Self-Interest theories of morality—theories that again (insofar as they plausibly include theories ranging from Aristotelian virtue ethics to contractarianism) have numerous historical and contemporary proponents.

Objection 2: 'The defense of Enlightened Self-Interest Realism against Cofnas's debunking story has not dealt with the real thrust of the argument, which is that convergence toward liberalism can be explained in wholly naturalistic terms, without appealing to *any* normative facts at all, including normative facts regarding enlightened self-interest.'

Reply: Cofnas's argument is that his debunking story explains convergence on liberalism by reference to 'two related non-truth-tracking processes.' This is important, because if Enlightened Self-Interest Realism is correct, then normative moral truths supervene on (and hence track) the kinds of naturalistic facts that Cofnas's story appeals to (Railton 1986). Indeed, Enlightened Self-Interest theorists often argue—by reference to our everyday conception of normative means-ends rationality (including such commonplaces as that if one is hungry, then

one *ought* to eat to something)—that there are *clearly* normative facts about what a person ought to do in a means-end sense (Gauthier 1986, Chapter 1; Arvan 2016, Chapter 1). Further, a number of naturalistic philosophers have defended what is known as a Humean reduction of normative propositions—a reduction according to which the truth-conditions and truth-makers of normative propositions (such as the obviously true proposition that if you want to win at tennis, then you ought to hit the ball over the net) are entirely reducible to *descriptive, naturalistic facts* (see e.g. Jackson 1988, Arvan 2021). While Humean reductions of this sort are philosophically controversial, the relevant points for our purposes are that they may be correct, and in any case it is widely recognized in everyday speech that instrumental normative facts (such as what one ought to do in order to win a game of tennis) *exist*, supervening on the kinds of naturalistic facts that Cofnas's debunking story affirms.

Objection 3: 'Earlier it was claimed that the only way to defend Robust Moral Realism against Cofnas's debunking story is to beg the question in favor of the Robust Moral Realism, asserting on the basis of normative arguments in favor of Robust Moral Realism that *sui generis*, non-natural moral facts are the best explanation of cross-cultural convergence toward liberalism. You rejected this as question-begging, asserting that it reveals the basic flaw in a cross-cultural convergence argument for Robust Moral Realism. However, how is your defense of Enlightened Self-Interest Realism any less question-begging? Your defense of the view involved appeal to independent normative arguments for Enlightened Self-Interest Realism (viz. Hobbes, Gauthier, Arvan, Vanderschraaf, etc.). This means those normative arguments that are doing all of the philosophical work, and that Cofnas is still right that convergence toward liberalism is not in itself *any* independent evidence for moral realism, even Enlightened Self-Interest Realism.'

Reply: Consider the nature of theory confirmation from the philosophy of science. Although there are complex issues here, take a simple case: the theory of universal gravitation that all massive objects attract each other in proportion their respective masses and the inverse square of their distance. Every observation we make of objects in our world cohering with this hypothesis is a kind of independent confirmation of the theory (Crupi 2021, §1). Conversely, if we observed objects violating the predictions of the inverse-square law, this would disconfirm current theories of gravitation. Something similar is true of Enlightened Self-Interest Realism. Notice, next, that Gauthier and others in the Enlightened Self-Interest tradition defend it as a theory of morality by reference to speculations about individual-level rationality and human psychology. Specifically, Gauthier defends the rationality of being a liberal individual by reference to the nature of instrumental rationality, resource scarcity, and market failures. Arvan, in contrast, defends the rationality of liberal values by reference to radical uncertainty about long-term outcomes over the case of a lifetime, and to a specific form of risk-aversion as a rational solution to that uncertainty. Vanderschraaf provides yet another explanation, holding that fairness—a quintessential liberal moral value (Rawls 1971)—is a solution to particular bargaining problems.

As we see here, each of these versions of Enlightened Self-Interest Realism predict that rational individuals (and by extension, groups of thereof) should progressively converge toward liberal values on the basis of very specific intra- and interpersonal mechanisms (importantly, the theories differ over what exactly these mechanisms are). These are, in essence, both normative predictions about how individuals ought to behave, but also—to the extent that individuals in the world in fact behave rationally—predictions about how individuals in the world *will* behave, and how societies thereof are likely to evolve over time (Arvan 2020, pp. 95-7). Enlightened Self-Interest Theories thus not only predict, then, that

(insofar as human beings have some propensity to behave rationally) convergence toward liberal values should *and will* occur; they make detailed empirical predictions about *how and why* such convergence occurs—predictions that can be independently confirmed or falsified by observations of individuals and societies. Consequently, on a standard understanding of theory confirmation, particular facts of cross-cultural convergence—depending on whether they confirm or disconfirm the particular mechanisms for converging toward liberal values posited by any given Enlightened Self-Interest theory—may indeed serve as *independent evidence* confirming or disconfirming the theory.

The problem with Robust Moral Realism is that the same is simply not true. Robust Realists can of course 'predict' convergence toward liberal values, holding that sui generis, non-natural moral truths are liberal in nature. The problem, though, is Cofnas's story *does* debunk this explanation, as his point is that we can explain the convergence without *any* appeal to sui generis moral facts. Consequently, Enlightened Self-Interest Realism and Robust Moral Realism are not on a par as explanations of convergence toward liberalism. Enlightened Self-Interest Realist theories offer *naturalistic* explanations of convergence toward liberal values that can be confirmed or disconfirmed. Robust Moral Realism does not, as it takes moral truths to be fundamentally *non-naturalistic*.

Objection 4: 'Recently, Luco has given a theory about how objective moral facts can cause moral progress—and his account seems about as consistent with Cofnas's debunking story as the Enlightened Self-Interest theories discussed in this paper. Specifically, Luco argues that moral cognition was selected for in evolutionary history to facilitate social cooperation (Luco 2019, p. 435); that moral cognition as such involves judging what impartially promotes well-being within indefinitely extended populations of interacting agents (ibid., §3, esp. pp. 438-9); and

that moral cognition as such is conducive to *emancipatory* (i.e. liberal) values (ibid., §4). Yet, Luco's account does not seem to fit cleanly into either conception of moral realism examined in this paper. Luco defends Naturalist Moral Realism (NMR), the view that objective moral facts are either identical to or constituted by natural facts (ibid., p. 430). NMR does not appear to be a form of Robust Moral Realism, as it does not posit *sui generis* non-natural facts; and it is not obviously a form of Enlightened Self-Interest Realism, as Luco argues that the etiological (or evolutionary) function of moral cognition is not rational self-interest but rather a form of impartial selflessness (ibid., p. 436). So, it seems, there is a third form of moral realism, Naturalistic Moral Realism, that also predicts cross-cultural convergence toward liberal moral values via specific mechanisms, and hence, withstands Cofnas's debunking argument.'

Reply: If Naturalistic Moral Realism truly is distinct from the two forms of moral realism examined in this article, then I have no qualms with the idea that it too withstands Cofnas's debunking argument of moral progress. Still, I demur for the following reasons. First, moral facts, whatever else they are, are intuitively *normative* facts purporting to express truths about how people morally ought and ought not to behave. Assuming this is true, then Luco's claim that NMR takes objective moral facts to be either identical to or constituted by natural facts runs into an obvious dilemma. Natural facts—such as facts about protons, electrons, what is conducive to social cooperation, and how moral cognition functions—all appear in the first instance to be descriptive facts about what is. For natural facts, as such, to be identical to or constitute normative moral facts, normativity must enter the picture somehow. That is, the proponent of NMR must explain how, say, the fact that moral cognition emerged to impartially promote well-being makes it objectively true that people *ought* to impartially promote well-being. But, it seems, there are only two plausible ways to do this: it can either be taken to be a

sui generis, non-natural normative fact that people ought to conform to moral cognition's etiological function (rather than, say, behave selfishly), or this can be taken to be requirement of normative, means-end rationality: that is, that people ought to impartially promote well-being because it is in our *interest* to do so, given the nature of our evolved psychological constitution, which (it seems, on Luco's account) makes human beings tend to take an interest *in* impartially advancing well-being as an end. At least on the surface, Luco's account seems to assume something like the latter picture, as Luco's account of how moral cognition evolved—focusing on how moral cognition facilitates social cooperation—is, broadly speaking, an account of how moral cognition advances the interests of individuals and the groups we comprise. But in any case, I submit that NMR faces the following dilemma: it either amounts to a form of Robust Moral Realism, holding that we have *sui generis*, non-natural moral reasons to act according to moral cognition's etiological function; or alternatively, it is a form of Enlightened Self-Interest Realism, holding that we ought to conform to moral cognition's etiological function (impartially advancing well-being) on instrumental, means-end grounds. But, if NMR is interpreted as a form of Robust Moral Realism, then it runs straight into Cofnas's debunking argument. For if Luco's naturalistic evolutionary story is correct, then we can account for cross-cultural convergence toward liberalism without positing *any* non-natural moral facts—leaving NMR's contention that moral facts are identical to or constituted by natural facts not supported by the mere fact of cross-cultural convergence (since, by hypothesis, the same convergence toward liberalism would occur even if no non-natural moral facts existed). On the other hand, if NMR is just a form of Enlightened Self-Interest Realism (which I suspect it is), then it does withstand Cofnas's debunking story just as other forms of Enlightened Self-Interest Realism do, and it is a further empirical question which particular

Enlightened Self-Interest theory (Luco's or some other alternative) best explains facts of cross-cultural convergence.

4. Conclusion

Nathan Cofnas claims to debunk Michael Huemer's convergence argument for moral realism. This article argued that Cofnas's naturalistic story does appear to debunk a convergence argument for one influential form of moral realism: Robust Moral Realism. However, we saw that Cofnas's story fails to debunk a Huemer-style convergence argument for another kind of moral realism: Enlightened Self-Interest Realism. Many forms of Enlightened Self-Interest Realism not only predict that individuals and societies should converge toward liberal values. These theories entail detailed naturalistic predictions for how this convergence should occur and will occur to the extent that people in fact behave rationally. Insofar as these predictions can be confirmed or disconfirmed by observation of facts about cross-cultural convergence, empirical observation of cross-cultural convergence can indeed provide independent support for Enlightened Self-Interest Realism, and for some specific Enlightened Self-Interest account(s) of why liberal moral values are *objectively* better (viz. rational self-interest) than illiberal values. None of this is say that Enlightened Self-Interest Realism is true and Robust Moral Realism false. It is to say that Enlightened Self-Interest Realism has at least one important mark in its favor—empirical testability via observation of mechanisms of cross-cultural moral convergence—that Robust Moral Realism does not.[5]

[5] [Acknowledgments redacted for anonymized review].

References

Anscombe, G.E.M. (1958). Modern Moral Philosophy. *Philosophy*, *33*(124): 1-19.

Aristotle [1984]. *The Complete Works of Aristotle: The Revised Oxford Translation*. J. Barnes (Ed.). Princeton: Princeton University Press.

Arvan, M. (2021). The Normative Stance. *Philosophical Forum* 52(1): 79-89.

----- (2020). *Neurofunctional Prudence and Morality: A Philosophical Theory*. New York, USA: Routledge.

----- (2016). *Rightness as Fairness: A Moral and Political Theory*. London: Palgrave MacMillan.

Beebe, J.R., & Sackris, D. (2016). Moral objectivism across the lifespan. *Philosophical Psychology*, 29(6): 912-29.

Boyd, R.N. (1988). How to Be a Moral Realist. In G. Sayre-McCord (ed.), *Essays on Moral Realism* Ithaca, NY: Cornell University Press: 181-227.

Brink, D. (1989). *Moral Realism and the Foundations of Ethics*. Cambridge and New York: Cambridge University Press.

Buchanan, J.M. (1975). *The Limits of Liberty – Between Anarchy and Leviathan*. Chicago: The University of Chicago Press.

Cofnas, N. (2020). A debunking explanation for moral progress. *Philosophical Studies*, 177: 3171-91.

Copp, David. (2007). *Morality in a Natural World: Selected Essays in Meta-ethics*. Cambridge: Cambridge University Press.

Crupi, V. (2021). Confirmation. In E.N. Zalta (ed.) *The Stanford Encyclopedia of Philosophy* (Spring 2021 Edition), https://plato.stanford.edu/archives/spr2021/entries/confirmation/.

Cudd, A. & Eftekhari, S. (2021). Contractarianism, The Stanford Encyclopedia of Philosophy (Winter 2021 Edition), Edward N. Zalta (ed.), https://plato.stanford.edu/archives/win2021/entries/contractarianism/.

Cuneo, T. (2007). *The Normative Web: An Argument for Moral Realism*. Oxford: Oxford University Press.

Davis, T. (2021). Beyond objectivism: new methods for studying metaethical intuitions. Philosophical Psychology, 34(1): 125-153.

Enoch, D. (2011). *Taking Morality Seriously: A Defense of Robust Realism*. Oxford: Oxford University Press.

FitzPatrick, W. (2008). Robust ethical realism, non-naturalism, and normativity. *Oxford Studies in Metaethics*, 3: 159-205.

Foot, P. (1972). Morality as a system of hypothetical imperatives. *The Philosophical Review*, *81*(3): 305-316.

Forcehimes, A.T. & Semrau, L. (2018). Are There Distinctively Moral Reasons? *Ethical Theory and Moral Practice* 21(3): 699-717.

Gauthier, D. (1986). *Morals by Agreement*. Oxford: Oxford University Press.

Goodwin, G.P., & Darley, J.M. (2008). The psychology of meta-ethics: Exploring objectivism. Cognition, 106(3): 1339-1366.

Hobbes, T. [1651]. *Leviathan*. in Sir W. Molesworth (ed.), *The English Works of Thomas Hobbes*, Vol III, London: John Bohn, 1839.

Huemer, M. (2021). Are Humans Amoral? *Fake Nous*, https://fakenous.net/?p=2398, retrieved 21 June 2021.

----- (2016). A liberal realist answer to debunking skeptics: The empirical case for realism. *Philosophical Studies*, 173: 1983–2010.

----- (2005). *Ethical Intuitionism*. Basingstoke: Palgrave Macmillan.

Hume, D. [1888]. *A Treatise of Human Nature*. L.A. Selby-Bigge (ed.), Oxford: Clarendon Press, 1896.

Jackson, F. (1998). *From Metaphysics to Ethics: A Defence of Conceptual Analysis*. Oxford: Clarendon.

Joyce, R. (2001). *The Myth of Morality*. Cambridge, UK: Cambridge University Press.

Kant, I. [1785]. *Groundwork of the Metaphysics of Morals*, in M.J. Gregor (ed.), *The Cambridge Edition of the Works of Immanuel Kant: Practical Philosophy*. Cambridge: Cambridge University Press, 1996: 38-108.

Locke, J. (1689). *Two Treatises of Government.* P. Laslett (ed.), New York: Cambridge University Press, 1988.

Luco, A.C. (2019). How Moral Facts Cause Moral Progress. *Journal of the American Philosophical Association* 5(4): 429-448.

----- (2016). Non-negotiable: Why moral naturalism cannot do away with categorical reasons. *Philosophical Studies*, 173(9): 2511-28.

Moore, G.E. (1903). *Principia Ethica*, Cambridge: Cambridge University Press.

Narveson, J. (1988). *The Libertarian Idea*. Philadelphia: Temple University Press.

Parfit, D. (2011). *On What Matters*: Two-Volume Set. Oxford: Oxford University Press.

Pölzler, T., & Wright, J.C. (2020). Anti-realist pluralism: A new approach to folk metaethics. *Review of Philosophy and Psychology*, 11(1): 53-82.

Prichard, H.A. (1912). Does moral philosophy rest on a mistake? *Mind* 21(81): 21-37.

Railton, P. (1986). Moral realism. *The Philosophical Review* 95(2): 163-207.

Rawls, J. (1971). *A Theory of Justice*. Cambridge: The Belknap Press of Harvard University Press.

Robson, G.J. (2015). Two Psychological Defenses of Hobbes's Claim Against the "Fool". *Hobbes Studies* 28(2): 132-148.

Sayre-McCord, G. (2021). Moral Realism. In E.N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Summer 2021 Edition), https://plato.stanford.edu/archives/sum2021/entries/moral-realism/.

Scanlon, T.M. (2014). *Being Realistic About Reasons*. Oxford: Oxford University Press.

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

Sidgwick, H. (1874). *The Methods of Ethics.* Seventh Edition. Indianapolis: Hackett Publishing Company, Inc.

Vanberg, V.J. (2014). James M. Buchanan's contractarianism and modern liberalism. *Constitutional Political Economy* 25(1): 18-38.

Vanderschraaf, P. (2021). Precis of Strategic justice: convention and problems of balancing divergent interests. *Philosophical Studies* 178(5): 1701-1705.

----- (2019). *Strategic Justice: Convention and Problems of Balancing Divergent Interests.* New York: Oxford University Press.

Velleman, J.D. (2013). *Foundations for Moral Relativism*. OpenBook Publishers.

Wright, J.C., Grandjean, P.T., & McWhite, C.B. (2013). The meta-ethical grounding of our moral beliefs: Evidence for meta-ethical pluralism. *Philosophical Psychology*, *26*(3): 336-361.