The Cognitive Science and Metaethics of Evolutionary Debunking Arguments
Matt Bedke

Introduction: Two Impressions from Debunking Arguments

When you read some evolutionary debunking arguments, you get the impression that the cognitive scientific details do not matter. Regardless of those details, this concern remains: whatever forces pushed our evaluative attitudes around, and whereto they pushed them, those processes were not trying to track evaluative truth. If we nonetheless believe that our evaluative attitudes largely track evaluative truth, we need to explain how so. For Sharon Street, this is a great difficulty for theories that make the evaluative truths attitude-independent, while attitude-dependent theories, including her favored Humean Constructivism, bear the explanatory burden fairly easily (Street 2006; 2008a; 2011). For under the assumption that truth is independent from attitude, it looks like there is no explanation for how the forces responsible for shaping our attitudes pushed them toward alignment with truth, at least no explanation that does not simply assume alignment for a large swath of our evaluative attitudes. Under the assumption that truth is constructed from attitude, however, it looks like the sources of our attitudes are incapable of pushing them out of alignment with truth, and any composition of attitudes is guaranteed to track truth. So the detailed cognitive science concerning etiology and moral psychology does not matter – whatever those details, there is an explanatory burden that attitude-independence theories cannot bear but attitude-dependence theories can.

Another impression you get from some evolutionary debunking arguments is that they present the heaviest burden for certain metaethical views, and they apply regardless of the first-order positions paired with those vulnerable metaethical views. Street, for

---

1 I set to one side Street’s arguments about epistemic normativity.
instance, has been broadly concerned with attitude-independent views as incarnated in realism and quasi-realism.

Here, I challenge these impressions. If the burden is to explain how evaluative attitude tracks evaluative truth without helping oneself to many evaluative truths, whether an attitude-dependent theory like Humean Constructivism can bear the burden very much depends on cognitive scientific details, including details about the etiology and composition of evaluative thought. Also, if an attitude-dependent theory like Humean Constructivism can explain tracking it is by adopting (or assuming) certain first-order positions that are absolutely neutral metaethically. In short, the metaethics matters not at all when it comes to answering a certain sort of debunking challenge. As we shall see, the claim is not that theorists fond of Humean Constructivism and its ilk need to take for granted substantive evaluative truths about cases in context (e.g., that an ideally coherent Caligula has reason to torture others for fun), or mid-level generalizations (that pain is bad); rather, this theorist needs to take for granted a certain bit of abstract ethical theory, ethical theory that can be paired with an attitude-independent metaethics.

These two points in hand, I will reinforce the first. Whether rivals to attitude-dependent theories like Humean Constructivism (now clearly reading the dependency as a first-order position) can bear the explanatory burden also very much depends on the cognitive scientific details. Indeed, first-order attitude-independent theories just might bear the burden more easily than first-order attitude-dependent theories, depending on the etiology and composition of our attitudes.

In closing, I will explain how my own debunking argument (Bedke 2009; forthcoming) differs from Street’s. I do not think the burden is to explain tracking
without assuming a large swath of evaluative truths (about cases in context or mid-level generalizations). That way lies general skepticism. For me, debunking threatens after one initially helps oneself (justifiably) to a large swath of evaluative attitudes. What threatens is a defeater generated by and for those who additionally adopt a non-naturalist, realist metaethics. That second-order position makes one’s heretofore justified evaluative attitudes insensitive to evaluative truth in a rather robust way – something I call obliviousness (forthcoming). This is not a sort of sensitivity that needs to be secured prior to having justified evaluative attitudes, but rather a sort of insensitivity that, when shown to apply to one’s own attitudes, generates a defeater for those attitudes. Thus, for my own debunking argument, the cognitive scientific details really are irrelevant, while the action is in the metaethics, not the first-order normative theory.

1 Debunking: Yet Another Analogy

Street-style debunking highlights an explanatory burden. To get a feel for the weight of it, consider the following analogy. Suppose you have one board with holes in it, and another board with pegs sticking out of it. The pegs are moveable on their board. Your job is to move the pegs around so that each peg lines up with a hole on the other board as best you can. Now consider three cases.

Under case A, we assume the holes take a fixed pattern on their board – a pattern that is independent of the arrangement of the pegs on the other board. And we assume that you have some process for moving the pegs around that is insensitive to the fixed pattern of holes you are trying to match. If so, after the process for moving the pegs around plays out, the chance that the pegs will align with the holes is very low. If you
think there is alignment you have some explaining to do. And the explanation cannot rest on the assumption that the pegs are a good indication of where the holes are.

Under case B, we assume that the positions of the pegs depend on the positions of the holes in the right way. If so, we can get alignment. Just imagine this process: someone observes the holes and moves the pegs into alignment. That is a pretty good explanation for how the pegs get to line up with the holes (whatever pattern the holes take).

Under case C, we reverse the dependence and assume the holes depend on the positions of the pegs in the right way. So assume that the positions of the holes are moveable but only by being magnetically attracted to the pegs. The pegs you can still move around as you wish. Use any process you like for moving around the pegs and notice that they align with the holes. Again, we have a good explanation for how the pegs got to line up with the holes, this time regardless of the initial pattern of the holes and regardless of the process used to move around the pegs. We need not assume any discrete locations of pegs and holes to get this explanation up and running – it works regardless of how these things are arranged.

Now, case A is definitely the one to avoid. B will get you alignment, but it is unattractive when we take the pegs to be evaluative attitudes and the holes to be the target evaluative truths. There is simply no way to get outside our attitudes to figure out what the truths are, and it would be unjustifiably presumptuous to take for granted that one’s attitudes are a good guide to the truths.

2 The Cognitive Science
Case C is the best-case scenario for tracking. It is, roughly, the sort of case Street creates by embracing “antirealist” views like Humean Constructivism. Here is how I will formulate that view (cf. 2006, p. 153; Street 2008b, pp. 223-27; 2010, passim):

Humean Constructivism (HC): A has reason to phi (/A’s judgment that she has reason to phi is correct) iff and because (/by being grounded in the fact that/by being a function of the fact that) A would have a positive evaluative attitude toward her phiing were all her evaluative attitudes subject to their own scrutiny and modified accordingly.²

Assuming HC is true,³ as evolutionary forces push around our other evaluative attitudes, so too do they push around evaluative facts, for the evaluative facts depend on our attitudes. According to Street, “Whatever explanation the natural and social scientists ultimately arrive at is granted, and then evaluative truth is understood as a function of the evaluative attitudes we have, however we originally came to have them” (Street 2006, p. 153). If so, it certainly seems as though the exact details of the etiological forces at play are immaterial, as is the exact composition of our evaluative attitudes. Whatever the details, HC has a ready explanation for why evaluative attitudes (about particular case, mid-level generalizations, or positions in between) track truth. History could not have thrown us off track, for our attitudes lay the track as we go. Others besides Street feel the force of this thought (cf. Copp 2008, p. 192; Dreier 2012, p. 272).

² For Street, this might capture a broader notion of constructivism, where Humean versions add that there are no evaluative attitudes that are essential to or necessarily follow from valuing anything (Street 2012). I will leave those details to one side.
³ I will address whether this is as illicit an assumption as assuming a large swath of substantive judgments about cases or mid-level generalizations in section 3.3 and 3.4.
I feel it but weakly. I will explain my concerns by separately discussing some agent’s judgments about the reasons she has (first-personal), and her judgments about the reasons others have (second- and third-personal) (cf. Berker forthcoming).

2.1 First-Personal Evaluative Attitudes

Here are the considerations to focus on: actual evaluative attitudes, etiological forces that shaped them, scrutiny and modification of them (something like reflective equilibrium), and the evaluative attitudes that would be the outputs of scrutiny and modification. My first worry here is that Humean Constructivism introduces the real possibility that one’s current attitudes do not align with the outputs of scrutiny and modification of one’s attitudes. For the view does not offer a completely tight connection between one’s current evaluative attitudes and the outputs of scrutiny and modification.

Suppose A has a set of evaluative attitudes, which includes evaluative judgment J with the following content P: That someone is my child is a(n independent) reason to promote their welfare. How shall we think of the possibilities for what A’s judgment could be under scrutiny and modification? Is it [J(P), J(¬P)]? Or is it [J(P), J(¬P), ¬(J(P) or J(¬P))]? More importantly, are we to understand the space of possibilities regarding content P, given Humean Constructivism, as less than the space for mind-independence views? It looks like the spaces of possibilities canvassed above are just the ones realists should consider.

Perhaps we should not focus on the space of possibilities, but on the chances of alignment. Perhaps the point is, whatever the relevant set of possibilities, the chances that my current J(P) would survive scrutiny and modification is high enough – at least higher than the chances that my current J(P) aligns with some mind-independent fact of the
Further, maybe we should not be focusing on individual judgments like J(P), but rather one’s whole set of evaluative judgments. Perhaps the chance that *a respectable number of my judgments will survive scrutiny and modification* is high enough – at least higher than the chances that a respectable number of my judgments will align with some set of mind-independent truths.

However we understand the relevant chance, it looks like the chance of alignment with the outputs of scrutiny and modification really depends on some additional details. A high probability of aligning with the outputs of scrutiny and modification is plausible *if* scrutiny and modification cannot (or is not likely to) affect a sea change in our evaluative attitudes. But this is not a foregone conclusion. Some starting funds of evaluative attitudes make it plausible that most of them would perish under scrutiny and modification.

Consider a psychology full of common-sense moral thoughts, but where consequentialist commitments are more deeply engrained, and there is a strong preference for minimizing basic principles. That is, consider Sidgwick. We can assume he initially had many judgments that comported with common sense or intuitive morality, but where some of his deepest convictions were more consequentialist, and where he strongly preferred a ‘rational synthesis’ of intuitive principles over leaving them ‘unsystematized’ (Sidgwick 1907, especially IV.2-IV.3). Under scrutiny and modification many of his judgments are likely to change (to be brought in line with consequentialism).

---

4 I am simplifying things by omitting a credence level. If we include it, the thought would be that the chances that my J(P) with credence C is close enough to the output of scrutiny and modification.
Indeed, depending on how we count attitudes, arguably most of his pre-scrutiny evaluative attitudes would not survive scrutiny and modification.

This is a delicate issue, for it is not clear what sorts of evaluative attitudes we need to consider when asking whether they track truth. If it is mid-level generalizations like Ross’s list of prima facie duties, Sidgwick would come to reject most of those as genuine prima facie duties. If it is judgments about actual and hypothetical cases, whether Sidgwick would reject those depends on how well the consequentialist calculations in those cases deliver the same judgments as intuitive or common-sense morality. That is hard to say. But it is not hard to say that whether his evaluative attitudes track those he would have under scrutiny and modification depends on some details about the nature of his starting fund of evaluative attitudes (in addition to depending on the nature of scrutiny and modification).

Contrast that psychology with one that did not have such deep convictions about consequences, or any other single morally relevant consideration, and/or was not so fastidious about systematizing convictions to subsume them under one master principle. One with this psychology might well end up with a pluralistic moral sensibility under scrutiny and modification, one that is not such a radical departure from the initial starting points. Here, we might think of W. D. Ross, Bernard Williams, or Jonathan Dancy and their versions of anti-theory. Some starting funds of evaluative attitudes are more likely to align with the outputs of the constructive function than others.

The point here is that we are owed more of an explanation for how getting things right under Humean Constructivism is likely, which depends on some details about what scrutiny and modification is, and also depends on details about what our actual evaluative
psychologies are like. To emphasize the point it might help to consider a view that would not have more explaining to do.

Crude Constructivism: A has reason to phi iff and because (/by being grounded in the fact that/by being a function of the fact that) A has a positive evaluative attitude toward her phiing.

Crude, but assuming it is true there is no chance of being off track. We do not need to know what scrutiny and modification are, for they are not helping to fix the truths, and we do not need to know what A’s actual psychology is, for whatever its composition it tracks the truth about what A has reason to do.

What of etiology? Can it throw the Humean Constructivist off track? As we have seen, depending on just what the function of scrutiny and modification is, certain starting points are more likely to be off track than others. For some starting points are more likely to undergo radical revision under scrutiny and modification than others. Now, if certain etiologies are more likely than others to produce the questionable starting points (the ones likely to misalign with the outputs of scrutiny and modification), then the actual etiological facts are quite relevant if we are wondering how likely it is that we are currently on track with the outputs of scrutiny and modification.

Here is an example of what I have in mind. Brian Skyrms has run some computer simulations of the evolutionary dynamics of simple ultimatum games. In one particular game, he pitted a Utilitarian Strategy against two other popular strategies, a Nash Strategy and a Kalai-Smordinski, and found that the Utilitarian strategy won (Skyrms 1996, pp. 106-108). It took over the population. Now, the details of the game he set up
are crucial for the result. This is not the place to explicate those details.\footnote{5} My point is this: if there are models that approximate the historical conditions of our interactions, and the models predict the success of Utilitarianism, we might expect our own evaluative attitudes to have some Utilitarian tendencies as well. Our own histories are messy and complicated, but we only need to be pushed in that direction for the effects of scrutiny and modification to be rather transformative. And Utilitarianism is just one example. Further study could reveal other funds of evaluative attitudes that are likely to transform a great deal under scrutiny and modification, and cognitive science could reveal how likely it is that our history has pushed us toward attitudes of those types.

This is an interesting role for etiology, one where cognitive scientific detail can matter, if we want to know whether each person has evaluative attitudes that roughly rack what they themselves have reason to do.

2.2 Second- and Third-Person Evaluative Attitudes

Now consider one’s judgments about what other people have reason to do. Here, the threat of a truth-blind etiological process is more direct than in the case of first-person judgments. Consider two individuals, A and B. Under Humean Constructivism, B’s reasons depend on B’s evaluative attitudes, not A’s. So for A to get things roughly right

\footnote{5 The ultimatum game involved the division of a cake between two players, where there are two roles for the game, defined in terms of differing payoffs for various amount of cake. The game was iterated, and in each round each player plays both roles. The distribution of strategies in subsequent iterations of the game were influenced by the relative successes of the strategies in the previous iterations. And the encounters between players were highly correlated (you only play with others who adopt your strategy). The three strategies were these: divide the cake in a way that the sum of the payoffs is maximized (Utilitarian), divide it in a way that the product of the payoffs is maximized (Nash), and divide it in a way that makes the payoffs proportional to the ratio of payoffs you get when you divide the payoff one player gets when she receives the whole cake by the payoff the other player gets when she gets the whole cake (Kalai-Smordinski).}
with respect to B’s reasons we need her evaluative attitudes to track the ones B would have under scrutiny and modification (cf. Berker forthcoming).

The etiological details clearly matter in this case. For the etiology of A’s evaluative attitudes can indeed throw her off track when it comes to tracking B’s attitudes (under scrutiny and modification). We must consider what the possible evaluative facts are concerning B’s reasons, and the chances that the forces responsible for A’s attitudes about what B has reason to do pushed those attitudes toward alignment with B’s reasons. The etiologies of my beliefs about what you should be doing are not guaranteed to track your evaluative attitudes about what you should be doing.

This issue is apt to be overlooked if we focus on traditions that emphasize the first-personal role of practical reason, such as the Kantian tradition that emphasizes agency, freedom, and determination of the will. I myself would emphasize the social aspects of normative discourse, especially its use in communicating to others expectations of pro-social behaviors, cooperative behaviors, the grounds of blame and censure, and so forth, as well as the usefulness of certain evaluative thoughts when it comes to inducing the punishment of norm violators.

In any event, the best-case scenario for alignment in the second- and third-personal cases is where the forces that pushed around B’s attitudes were roughly the same as the forces that pushed around A’s attitudes (and scrutiny and modification is unlikely to cause a sea-change in attitudes); that is, their attitudes were caused to have roughly the same contents. This can happen if both sets of attitudes were largely shaped by selective pressures operating universally across the selective contexts for A and B. As Street intimates, perhaps attitudes about special obligations to take care of one’s children rather
than strangers or distant relatives are explained by such universal forces (Street 2006, pp. 126, 132, 153).

Even if that is so for that attitude, it is not clear how well it delivers alignment. For we might expect different adaptive pressures to apply to attitudes about our own conduct (the first-personal ones) versus attitudes about other people’s conduct (the second- and third-personal ones). This could give rise to different contents for the two sorts of judgment such that my evaluative attitudes about what you have reason to do are not so likely to match your evaluative attitudes about what you have reason to do.

Consider the hypothesis that, at least in some circumstances, evolutionary pressures would have favored people who defect from, or free ride upon, cooperative ventures. If there is a way to reap the benefits of everyone else abstaining from overfishing, for example, while you yourself fish as much as you like, you will be better off than the cooperators. There are models that show some pressure to not defect or free ride, to be sure, but those pressures could be operating against an independent pressure toward defection or free-riding. Again, if this picture is roughly right, we might expect my attitudes about what you have reason to do (unconditional cooperation) to differ from your attitudes about what you have reason to do (defect if you can). So whether I roughly track what you have reason to do with my second- and third-personal evaluative attitudes depends on the details of the etiological forces operating on both our attitudes and more proximally the current composition of those attitudes.

We can get these first-personal and second-/third-personal contrasts even if there are universal selective pressures, so long as there-pressure differ between first- and second-/third-personal judgments. Now consider the real possibility that the pressures
were not so universal after all. Take attitudes about what counts as fair
treatment when distributing some good, whether one ought to share one’s money, and if so how much, and whether those on the receiving end of transactions, or third party others, should engage in costly punishment when others fail to meet the standards of fairness and altruism. There is mounting evidence in support of a diversity of these norms across cultures, including a diversity in the norms of costly punishment (where researchers again use various games from the decision theory literature to test norms about these things) (Henrich et al. 2010; 2006; House et al. 2013). There has been some evidence that evolutionary pressures operating on cultural norms selected for more fairness and altruism, as well as greater levels of punishment for unfair behavior, in large-scale societies with many anonymous, market-based interactions (Henrich et al. 2010). For in those settings one cannot rely on reciprocal altruism and kin selection to evolve large-scale cooperative behaviors. There is also some evidence that norms for engaging in prosocial behavior diverge as children develop, with children tracking toward the norms of adults in their society, suggesting that attitudes about fairness are not a fixed part of our psychology (House et al. 2013).

I mention these particular studies because they support the theory that some of our deepest moral attitudes—about fairness and punishment—are at least partly formed by evolutionary forces operating on cultural norms. This makes variability between my attitudes about someone’s reasons and their evaluative attitudes (under scrutiny and modification) a real possibility. It introduces another source of misalignment that threatens Humean Constructivism.
Note here that the evolutionary pressures and cultural variations need not be so separate in time and space. Jonathan Haidt has argued that there are at least six moral foundations, or morally significant dimensions. Democrats, he claims, emphasize two of them—ones concerning care/harm and fairness/cheating. Republicans, he claims, more evenly emphasize all foundations, including a distinctive sort of fairness/cheating—which has more to do with reaping what you sow than equality—loyalty/betrayal, authority/subversion and sanctity/degradation (Haidt 2013, chapters 6 and 7). This general point about variability in evaluative attitude likely applies *intra*-culturally, for each person’s learning history could deliver a unique fund of starting evaluative attitudes. Just think of the complaint that students who learn from consequentialists become consequentialists, while students who learn from Kantians become Kantians.

Now, none of this material from the cognitive science of morality is conclusive evidence that we fail to track the reasons of others. I simply point out that, assuming Humean Constructivism is right about what reasons there are, whether I am tracking the reasons of others (with my judgments about cases and mid-level generalizations) depends on how well my evaluative attitudes about the reasons of others track their evaluative attitudes under scrutiny and modification. The etiology behind my attitudes might well throw me off track. So if Humean Constructivism is to avoid debunking concerns when it comes to others’ reasons, it must hang its hat on the hope that the correct etiological story is one that puts your attitudes into alignment with others’ attitudes under scrutiny and modification. As with tracking first-personal reasons, the cognitive science matters.

To thoroughly explain coincidence of attitude and fact without any help from cognitive science one must adopt a radically perspectivalist view. Consider this:
Debunk-Proof: There are no such facts as A having reason to phi simpliciter. There are only facts about whether, for some agent X, A has reason to phi. And for X, A has reason to phi iff and because X judges A has reason to phi.

Assuming this abstract principle is true, there is no chance that etiology could push one’s reason judgments out of alignment with fact, and no chance the constructive function (basic as it is) could output reasons for A (relative to X) that deviate from X’s current attitudes.

3  The Metaethics

In all of this, I have been granting the Humean Constructivist the truth of her principles and examining whether she can explain tracking without the help of cognitive science. She cannot. Note that we could do this with any competing theory of reasons T. Just assume T, and examine whether T can explain tracking without the help of cognitive science.

You might think, however, that attitude-dependent theories like Street’s still fare better than certain attitude-independent metaethical theories. The thought is something like this: Making evaluative truth depend on attitude at least ensures that the truth will not stray too far from attitude. Alternatively, if evaluative truth is independent of attitude, the truth could be quite distant from attitude. In terms of our peg analogy, if each hole is magnetically attracted to a peg but encounters some resistance to moving (say, via friction), you might not get perfect alignment after moving the pegs around, but for each peg you can be sure that a hole will be somewhere in the neighborhood. By contrast, if
the locations of the holes are independent of the pegs there is no assurance that each peg will be anywhere near a hole after moving the pegs around.

What shall we make of this analogy? First, I want to argue that Humean Constructivism discharges the burden of explaining alignment (if it does) not via a metaethical or second-order position (I will use these terms interchangeably), but via a certain first-order position compatible with any metathetical view (save error theory or skepticism), *realism and quasi-realism included*. More broadly, at least for Street-style debunking, how easy or difficult it is to explain alignment depends on what first-order positions one adopts, not on one’s metaethics.

Second, some first-order positions that compete with Humean Constructivism, such as Cultural Relativism, might be *less susceptible* to Street-style debunking than Humean Constructivism. That is, they might be more able to explain tracking that Humean Constructivism is (without relying on judgments about cases or mid-level generalizations). Again, it depends on the etiological details. Let me take these points in turn.

### 3.1 *Humean Constructivism: How Metaethical?*

The interpretive issues of how Street understands her argument are not entirely clear to me. I think the preponderance of evidence indicates that Street takes metaethics to matter a great deal for her debunking argument. At one point Street indicates that the kind of (in)dependency she has in mind is stance (in)dependency in the way Ron Milo articulated it (Street 2006, p. 111 n. 3; Milo 1995, pp. 191-92). As I understand that view, a fact is stance dependent if (defn) it is *identical to* or *fully constituted by* or *fully realized by* facts about stances. Otherwise it is stance independent. That would be a heavily
metaethical characterization of the dependency at issue. Similarly, Street bills her view as antirealist, which certainly sounds metaethical. And she explicitly says that antirealism explains alignment with a metaethical hypothesis: “Antirealism explains the overlap not with any scientific hypothesis such as the tracking account, but rather with the metaethical hypothesis that value is something that arises as a function of the evaluative attitudes of valuing creatures – attitudes the content of which happened to be shaped by natural selection” (Street 2006, p. 154).

There is other evidence, however, that her debunking concern hinges more on first-order matters. For instance, she elaborates on independency with counterfactuals: If we were to alter our evaluative attitudes, that would not alter the evaluative facts (Street 2006, pp. 136-37). I take it that counterfactuals are secured (or not) by biconditional first-order views (more on this shortly). But because these first-order views can be entailed by second-order views, this evidence is less clear on the interpretive question of whether Street takes the issues to be metaethical or not.

Better evidence comes from her treatment of quasi-realists (Street 2011). Street thinks the quasi-realist is in trouble because he thinks there are attitude-independent normative truths, which is understood as an existential generalization of at least one independence counterfactual. A good example of an independence counterfactual is this: ‘Even if we all approved of kicking dogs, kicking dogs would still be wrong’. Street would allow the quasi-realist to avoid her argument by rejecting independence counterfactuals in favor of attitude-dependent claims, like this: ‘If we all approved of kicking dogs, kicking dogs would not be wrong’ (Street 2011, pp. 29-30). Crucially, either counterfactual can be paired with a quasi-realist metaethic, which suggests that the
explanatory burden is bearable or not depending on one’s first-order commitments, not one’s metaethics.

For the most part, however, Street sticks to discussion of whether evaluative facts depend on, are grounded in, or are a function of evaluative attitudes, which I take to be ambiguous between first- and second-order issues. Interpretation aside, then, I think the best version of her debunking argument targets certain first-order positions, not metaethical ones.

I’ll try to explain what I mean here by distinguishing first-order attitude-(in)dependency theses from second-order attitude-(in)dependency theses. Recall first Humean Constructivism.

Humean Constructivism (HC): A has reason to phi (/A’s judgment that she has reason to phi is correct) iff and because (/by being grounded in the fact that/by being a function of the fact that) A would have a positive evaluative attitude toward her phiing were all her evaluative attitudes subject to their own scrutiny and modified accordingly.

This disjunctive formulation was meant to be a neutral as possible about whether it is offered as a first-order position or a metaethical position, or some mix between the two.

---

6 When spelling out constructivism as antirealist, Street refers to the Euthyphro dilemma and it’s contrasting orders of explanation (Street 2010, p. 370; 2012, p. 41), which again I take to be ambiguous between first- and second-order issues.

7 Selim Berker also struggles to figure out the extent to which Street’s issues are first-order or second-order, and concludes that “what grounds grounding facts is where most of the philosophical action is” (Berker forthcoming). This is a thoroughly metaphysical and so metaethical interpretation of what is going on, in contrast to my own. Berker is primarily focused on the ‘because’ clause in something like my Explanatory HC, which he takes to pick out an in-virtue-of or grounding relation, whereas I think Street’s issues are settled with extensional theory and without need of these asymmetric metaphysical relations.
Granted, there are difficulties separating out first-order issues from second-order or metaethical issues, but I hope that the following shows that the distinction works for our purposes.

The biconditional helps to identify some first-order content in HC. First-order theories are in the business of extensional adequacy at least. So here is Extensional Humean Constructivism (and allow me to drop parenthetical alternative formulations).

Extensional HC: A has reason to phi iff A would have a positive evaluative attitude toward her phiing were all her evaluative attitudes subject to their own scrutiny and modified accordingly.

This is an attitude-dependent theory. What makes an attitude-dependent theory attitude dependent is that the evaluative status of the item of evaluation, O, depends on facts about attitudes toward O. Here, having reason to phi depends on attitudes toward phiing.

Further, I take Extensional HC to be wholly first-order and compatible with a number of second-order positions, some of which are attitude-independent. More on this shortly. In the meantime, what we have here is an attitude dependent bit of first-order theory.

Another task often associated with first-order theories is explanatory power. In particular, one wants the evaluative statuses identified on the left hand sides of the biconditionals—here, having a reason—to be explained by the contents of the right hand sides—here, evaluative attitudes and scrutiny and modification. So let us add in a ‘because’ clause to do this explanatory work.

\[8\] If you say that any theory with a fact about attitudes on the right hand side counts as attitude dependent, you lose the main reason for distinguishing attitude-dependent views from attitude-independent views, and you get wonky taxonomies. We would need some good reason for counting hedonic utilitarianism as attitude dependent (see Milo 1995, pp. 191-92). Moreover, I do not think Street would count hedonic utilitarianism as attitude dependent in her sense, and so a view that easily bears her explanatory burden.
Explanatory HC: A has reason to phi iff and because A would have a positive evaluative attitude toward her phiing were all her evaluative attitudes subject to their own scrutiny and modified accordingly.

It is debatable whether the ‘because’ here makes the principle partly metaethical. You might think that it stands for a metaphysical relation such as identity, or constitution, or some sort of necessitation relation between distinct existences (as some talk of grounding has it, an in-the-world, asymmetrical, explanatory, robustly metaphysical relation).

Though debatable, I am inclined to think that there is some explanatory work that first-order theory does, signifiable by a ‘because’ clause, that is entirely neutral on second-order issues. That is, it does not commit the principle to particular metaphysical doctrines, and so remains compatible with further metaethical theses about the relationship between the left and right hand sides. Consider that quasi-realists can happily talk about evaluative statuses holding because of some consideration or another. It is wrong to kick dogs because it hurts them, for example. They say this in apparent agreement with some metaethical non-naturalist realists. So I do not think the explanatory clause takes a metaethical stand. Rather, it helps to make the evaluation intelligible to people with certain sorts of attitudes.

Fortunately, I do not have to take a stand on this issue because I think explaining alignment between truth and value ultimately depends on nothing stronger than extensional theories. Before getting to that, let me briefly discuss some clearly metaethical positions one might take in relation to HC. It is natural to pair Extensional or Explanatory HC with a reductive metaethical view like this.
Reductive HC: *What it is* for A to have reason to phi, or what A’s having reason to phi *consists in or is constituted by*, or what the correctness of A’s judgment that she has reason to phi *consists in or is constituted by*, is that A would have a positive evaluative attitude toward her phiing were all her evaluative attitudes subject to their own scrutiny and modified accordingly.

That is probably Street’s view. It is clearly a second-order, attitude-dependent view in that it reduces the evaluative status of an item of evaluation, O, to attitudes toward O. But we could also pair Extensional or Explanatory HC with a noncognitivist metaethical view like this.

Non-Cognitive HC: Extensional or Explanatory HC. Further, Extensional and Explanatory HC do not purport to represent normative ways the world might be. Rather, they are used to express certain basic pro- and con-attitudes toward people doing things they themselves would positively evaluate after bringing their attitudes under scrutiny and modification.

This could be called an attitude-dependent second-order view, but it is clearly different from the reductive one above, and unlike the reductive one it can be paired with attitude-independent first-order theses. That is, it can be paired with conditionals or biconditional analyses of evaluative statuses that do not mention attitudes toward the item of evaluation. So I think it best to call this an attitude-independent second-order view.

Perhaps surprisingly, there are robustly realist attitude-independent metaethical views that can be paired with the attitude-dependent first-order views of Extensional or Explanatory HC.
Non-Naturalist HC: A’s having reason to phi supervenes on the fact that A would have a positive evaluative attitude toward her phiing were all her evaluative attitudes subject to their own scrutiny and modified accordingly, where having the reason is a *sui generis* relation distinct from having the attitudes, and not fully constituted by having them. Further, the supervenience is secured by metaphysical (or normative) necessity relations between attitudes (hypothetical ones under scrutiny and modification) and reasons.

Again, you might call this an attitude-dependent metaethic simply because the subvening base is comprised by facts about attitudes toward the items of evaluation. But that focuses on the first-order commitments of this particular realist view, while obscuring its more central metaethical characteristics. Importantly, the view makes the evaluative facts *sui generis* yet towed along by the subvening base thanks to brute necessitation relations. This is enough to make it an attitude-independent metaethic in my book. As with Non-Cognitive HC, this view can be paired with attitude-independent first-order views. At the very least, this is a *realist* view, and it should come in for the same metaphysical and epistemological troubles as more standard versions of non-naturalist realism (i.e., versions that do not have attitudinal facts about the items of evaluation in the subvening base).

Notice that each explicitly metaethical position, including the one Street would likely favor, entails the first-order position of Extensional HC, and arguably entails Explanatory HC. The big question is whether any version of Humean Constructivism *epistemically relies* on metaethical content to help answer the debunking challenge, or
instead relies only on first-order content. If the second, it looks like that first-order content is available to other metaethical positions outlined above so that they, too, can answer the debunking challenge as well as Street can.

3.2 All the Work Done at the First Order

The first thing to note is that the first-order theory Extensional HC is sufficient to answer the debunking concerns as well as Street can. All we need is for the evaluative truths to depend on evaluative attitudes in the right way – therein lies the hope of explaining how attitude tracks truth. The biconditional gets you that. It is a way to ‘construct’ truths out of attitudes (or ‘ground’ them, or have a ‘function’ from one to the other) that remains metaethically neutral. If there is any reason to believe Extensional HC, presumably all the metaethical positions canvassed above can absorb the lesson. If they do so, they can answer Street’s debunking argument as well as she does. I say ‘as well as she does’ because I have noted above some of the ongoing difficulties HC has, and the need from premises from cognitive science, to really explain tracking.

Then again, perhaps adopting Extensional HC is unjustifiable unless one is a metaethical constructivist. I think this is what Street has in mind. On this view, the metaethical position of something like Reductive HC is justifiable, and Extensional HC is entailed by it, so that is how Extensional HC is justified (otherwise, it is unjustified). If so, even though Extensional HC is sufficient to answer the Street-style debunking concerns as well as Street can, it is out of justificatory reach for everyone except the metaethical constructivist.

I do not think that is right. To explain, let me extract the first-order content out of Reductive HC and just state its metaethical content.
Reductive Attitude-Dependency: *What it is for A to have reason to phi, or what A’s having reason to phi consists in or is constituted by,* or what the correctness of A’s judgment that she has reason to phi *consists in or is constituted by,* is some fact(s) about attitude(s).

There are lots of views that fill this bill. You could reduce facts about reasons to facts about God’s attitudes, some ideal observer’s attitudes, cultural attitudes, the attitudes of the target of evaluation, my attitudes, etc. If we settle on reducing A’s reasons to facts about A’s attitudes, still many options abound. You could say A has a reason to phi if A is against phiing, if he is for phiing, if he would be for his phiing after process P1, if he would be for his phiing after process P2, etc.

You get the idea. If all we had were Reductive Attitude-Dependency, we could not meet the debunking argument and its explanatory challenge. The evaluative facts could be any of the many ones just alluded to. And etiological forces are blind to which of these attitude-dependent theories is the right one. So, what are the chances that the forces pushed us toward the right truths about attitudes? I have the evaluative attitudes I have because of some causal forces (perhaps they were adaptive), and I would have those attitudes regardless of whether they track the right facts about attitudes (where the domain of possible attitude-dependent facts is rather large – facts about my attitudes? God’s? My societies? . . .).

The problem is perhaps most acute if the attitude-dependent normative facts have no etiology, and so do not shift around with shifts in our attitudes. Divine command theories and ideal observer theories, for instance, fix evaluative fact independently of *any actual set of* attitudes, but not independently of attitude. So, if all we have to work with is
the metaethical hypothesis of Reductive Attitude-Dependency, we need to explain how it is that beliefs or processes with a truth-blind etiology managed to glom on to whatever attitude-dependent truths there are.

Having a generic attitude-dependent metaethics is not enough. It would be great if we could help ourselves to a particular metaethical view like Reductive HC, for that particular metaethical view will entail the needed first-order content. But how do we justifiably settle on, say, Reductive HC among all the attitude-dependent metaethical views? I do not see how to justifiably come to that conclusion except by justifiably believing that Reductive HC is extensionally superior to the other views. That is, on first-order grounds. The very same grounds that one would use settle on Extensional HC over competing first-order views and regardless of one’s metaethical leanings. So, when it comes to answering Street’s debunking argument, the action is at the first-order.

This is not how Street sees things. She thinks there an argument for Humean Constructivism, or at least constructivism of some sort, that does not rely on ‘substantive’ judgments (Street 2006, p. 163, note 7; 2010, pp. 381-82, note 16, p. 383, note 60), by which I think she means to preclude judgments about cases and mid-level generalizations about evaluative statuses. The most promising remarks in that vein suggest there is something of a transcendental argument for Reductive HC that does not proceed on first-order grounds (Street 2010, p. 383, note 60). It is based on something like this thought: A condition on the possibility of explaining how evaluative attitude tracks evaluative truth is Reductive HC. There are several problems with this. First, it is not true. As we have already seen, Extensional HC can explain tracking if Reductive HC can, and Extensional HC is weaker.
Second, the transcendental premise does not even support Extensional HC. If it did, it would provide even greater support for the extensional components of Crude Constructivism and Debunk-Proof. For they have an even easier time explaining the coincidence of attitude and truth than does Extensional HC. Yet none of this speaks in favor of Extensional Crude Constructivism or Extensional Debunk-Proof. So the fact that Extensional HC has an easier time explaining coincidence than first-order stance-independent views (something I will question in the next section) does not speak in favor of Extensional HC either.

Third, even robustly attitude-independent first-order views make alignment possible. The concern must be that they do not make it sufficiently probable. Or, as Street puts it, we need a theory that is not hopeless: “[M]y argument is that we must give up the normative position that value is mind-independent in favor of the normative position that value is mind-dependent because the former view, when combined with an evolutionary picture, yields the unacceptable normative conclusion that we are hopeless at judging what to do” (Street 2011, p. 29). We need a slightly different transcendental premise, then: A condition on the probability of alignment is Extensional HC. Again, however, this would benefit Extensional Crude Constructivism and Extensional Debunk-Proof if it benefits any theory. Yet they are not beneficiaries of Street’s explanatory burden (are they?). Further, I argue below that some robustly attitude-independent first-order theories might make tracking more probable than Extensional HC, again depending on the cognitive science.

So I am not optimistic about a transcendental argument for some attitude-dependent position, at least not one based on either 1) the possibility or probability of
explaining alignment with truth, or 2) assuming a large swath of truths about cases or mid-level generalizations.

Let me stop there, for we have left metaethical issues far behind. I conclude that metaethics is really an idle wheel when it comes to meeting Street’s debunking concerns. She wants tracking, and you get it via first-order biconditionals that are available to just about any metaethical position you like, realist and quasi-realist included.

3.3 Playing the Game

I have tried to make two points so far. First, whether an attitude-dependent theory like Humean Constructivism explains alignment depends on the details of some cognitive science. Second, such a theory explains alignment (if it does) only via its (attitude-dependent) first-order content, which is as available to realist and quasi-realist metaethical views as it is to attitude-dependent metaethical views. When we put these points together, we can see that some other first-order ethical theory might do a better job explaining alignment than Extensional HC.

Recall that the chance that our evaluative attitudes about cases track the outputs of scrutiny and modification, both for ourselves and for others, depends on the cognitive science. Favorable empirical details could make tracking likely, but unfavorable empirical details could make tracking unlikely. I have already indicated that Crude Constructivism and Debunk-Proof make tracking more likely than Extensional HC. But they are still attitude-dependent first-order views.

Now consider Rossian pluralism, which offers an attitude-independent first-order theory. What are the chances that our evaluative attitudes about cases track the evaluative truths, assuming Rossian pluralism is the most systematic theory of evaluative truth?
Well, if the cognitive science is such that causal forces would push us toward having
particular judgments about reasons that accord with the deliverances of Rossian
pluralism, then we will track truth pretty well. Whether we would do better at tracking
assuming Rossian pluralism rather than assuming Humean Constructivism depends on
those details about etiology and the current composition of our attitudes. The point is that
Humean Constructivism does not hold the high ground before working out those details.

Street worries about assuming a swath of substantive evaluative truths to help
explain tracking (see especially Street 2008a, pp. 214-17). I am not certain if Rossian
pluralism does so any more than Humean Constructivism, for I am not certain which
evaluative attitudes are the illicit ones. As previously indicated, I think she would rather
avoid relying on judgments about particular cases (both actual and hypothetical) and mid-
level generalizations, while she is less concerned about relying on fairly abstract
principles that are not themselves judgments about cases in context, but which can be
paired with some supplementary premises to derive conclusions about cases in context.
Granted, Humean Constructivism is a rather abstract principle while the Rossian duties
are more mid-level, but it is not clear to me if this difference matters.

Perhaps the more abstract the principle, the more palatable it is to assume it when
explaining tracking. If so, I would guess that assuming Cultural Relativism—a principle
as abstract as Humean Constructivism—would actually make the likelihood of tracking
truth with particular judgments more likely, given the cognitive science of our attitudes.
At any rate, if tracking depends on first-order content plus cognitive science, it is not
clear that Humean Constructivism fares better than the competition. Maybe Hedonic
Utilitarianism or Kantian theory or some kind of virtue ethical analysis of reasons
explains tracking pretty well, and one of these might well explain tracking better than Humean Constructivism when paired with the truths of cognitive science.

What if I am wrong? What if Street is right that some constructive view like HC is the only non-hopeless theory when it comes to explaining truth-tracking without assuming the truth of substantive judgments about particular cases? Is the rational response to restrict our normative musings to the attitude-dependent ones? Should we blind ourselves to the arguments for principles of the form “A has reason to phi iff [fill in favored fact that is not about attitudes toward the items of evaluation]”? Should we abandon the deeply held judgment that an action is wrong if and because it causes pain?9

Speaking for myself, I am more confident that attitude-dependent first-order views are wrong than I am that one needs to earn one’s justificatory stripes by first explaining tracking as such, without relying on any substantive evaluative judgments about cases or mid-level generalizations. But I suppose others might come to a different conclusion after critical scrutiny of the issues.

4 Conclusion: Alternative Conceptions of Debunking

To wrap up, it might help to contrast Street’s debunking argument with my own (Bedke 2009; forthcoming). I take my own argument to offer up a defeater for anyone who wishes to blend his or her normative beliefs with non-naturalist realism. The defeater is closely related to an insensitivity concern. Roughly, the thought is that if you believe non-naturalist realism, then you should realize that your normative beliefs are oblivious to the truth. For under various hypotheses about what the actual normative truths are, you

---

9 Pain is a mental state. But this does not make the position attitude-dependent in the relevant sense. As previously stated, the evaluative status of some item O must depend on facts about attitudes toward O for the evaluative status to be attitude-dependent. See note 8.
realize that whatever they turn out to be you would have the same normative beliefs and for the same causal reasons. It is a bit like the classic cases of defeat where you might come to believe that your experience as of a red wall is caused by a red light shining on that wall, or caused by a pill that will induce you to see the wall as red, so that under various hypotheses about the actual color of the wall you would believe the wall is red and for the same reasons (the light or the pill) regardless of what color the wall actually is. Analogous to these classic cases, the belief in non-naturalism helps to defeat your otherwise justified normative beliefs. It is akin to believing a pill affects your color vision.

Note that, seen in the light of classic cases of defeat, it is not that we need to explain reliability or tracking or alignment before we justifiably believe a wall is red based on the way it looks. But certain explanations of the grounds of our beliefs can defeat any pro tanto justification.

So, according to my debunking concern, one’s metaethics is quite important while one’s first-order commitments are not. You will not get obliviousness regardless of your metaethical proclivities. But accept non-naturalist realism and the defeater applies. Moreover, it applies regardless of your first-order views, be they Rossian, Humean Constructivist, Kantian, a heap of substantive judgments about cases in context, or what have you.

As for the cognitive scientific details concerning your normative attitudes, they are irrelevant to my debunking concern. So long as there is some complete causal explanation for why you have the attitudes you have your beliefs will be oblivious to the
(non-natural) normative truths. Again, this is a defeater, not a condition that must be rejected prior to having justification in the first place.

So, if you think there is a real debunking concern in the neighborhood that besets attitude-independent views, consider whether your concerns are best captured as a defeater for certain metaethical positions, or alternatively whether they are best captured as concerns that certain first-order positions cannot explain alignment with truth. If the concerns are of the second sort, as I think they are with Street, I think we need to wait on the cognitive science concerning the origins of evaluative thought and the compositions of our attitudes before concluding that Humean Constructivism or some alternative first-order theory can explain tracking. We should stop thinking that this is an independent argument against certain metaethical views.

If the concerns are more my style, certain metaethical views are in trouble, and we need not wait on the cognitive science to see this is so.

Bibliography


