

University of Chicago Law School

Chicago Unbound

Public Law and Legal Theory Working Papers

Working Papers

2018

Disagreement, Anti-Realism About Reasons, and Inference to the Best Explanation

Brian Leiter

Follow this and additional works at: https://chicagounbound.uchicago.edu/public_law_and_legal_theory



Part of the [Law Commons](#)

Chicago Unbound includes both works in progress and final versions of articles. Please be aware that a more recent version of this article may be available on Chicago Unbound, SSRN or elsewhere.

Recommended Citation

Leiter, Brian, "Disagreement, Anti-Realism About Reasons, and Inference to the Best Explanation" (2018). *Public Law and Legal Theory Working Papers*. 697.

https://chicagounbound.uchicago.edu/public_law_and_legal_theory/707

This Working Paper is brought to you for free and open access by the Working Papers at Chicago Unbound. It has been accepted for inclusion in Public Law and Legal Theory Working Papers by an authorized administrator of Chicago Unbound. For more information, please contact unbound@law.uchicago.edu.

DISAGREEMENT, ANTI-REALISM ABOUT REASONS, AND INFERENCE TO THE BEST EXPLANATION

Brian Leiter

bleiter@uchicago.edu

draft of August 6, 2018

for presentation at the Australian Catholic University conference in Rome, September 2018

please do not cite or quote without permission

I will argue that the best explanation of one important kind of intractable disagreement about reasons for action—namely, the disagreement that defines the entire history of moral philosophy—includes the fact that realism about reasons is false. By realism about reasons, I will mean the view that what an agent has reason to do is “mind-independent,” that is, it does not depend on anyone’s beliefs or attitudes about the reasons for acting, even under ideal epistemic conditions.¹ I here revisit some earlier arguments (Leiter 2001; Leiter 2014; Leiter 2015), trying to refine them and respond to possible objections.

A threshold objection to this strategy of argument deserves brief consideration. Enoch (2009: 23) notes that “[n]ot every phenomenon calls for explanation—we are inclined to take some facts as brute.” That seems right, but is disagreement a “brute fact” that does not “call” for explanation? Enoch notes that a “full answer...would have to determine first what makes a phenomenon explanatorily interesting,” and confesses, “I know of no satisfactory answer to this question” (2009: 23). We do not need, of course, a *full* answer to that question, just an answer to the question whether intractable disagreements should be treated as “brute.” Enoch admits that “declaring all cases of moral disagreement as explanatorily uninteresting is a rather desperate (and dogmatic) move” (2009: 23).

Let us distinguish between the metaphysical claim that some facts are brute, i.e., that they do not admit of further explanation, and the epistemological or practical claim that a phenomenon does not “call” for or warrant an explanation. Many phenomena are brute in the latter sense, but not in the

¹Street, cf. Enoch (2009: 16) for ‘robust realism’

former: e.g., that my bread toasts in the toaster each morning, that the car starts when I turn it on, that my daughter Celia is a terrific cook. None of these phenomena are metaphysically brute, of course, and they all *can* be explained, but except in unusual circumstances we have no reason, epistemic or practical, to seek out such an explanation: they do not “call for” explanation.

Metaphysically brute facts simply do not admit of any further explanation. There may not be many such facts, though this will likely depend on one’s metaphysics. That we experience time as moving forward not backwards, that fundamental physics gives us the best account of matter in motion, that death awaits all living organisms, may all be metaphysically brute facts. The phenomenon of disagreement is obviously not metaphysically brute in any similar sense.

So Enoch’s challenge should be understood as the demand that we need to articulate when, for epistemic or practical reasons, we should seek an explanation of a phenomenon. We do not, as noted, need a general account of “what makes a phenomenon explanatorily interesting,” since it will suffice if we can identify some features that make a phenomenon explanatorily interesting, and show that disagreement has those features. I here propose one such feature: when people take themselves to be trying to discover some objective fact² about the world, but fail to converge on an answer despite trying for a long time, it is natural to wonder what explains their failure.

Imagine: you and your neighbors spend hours looking in the backyard for a lost ball; you all think you know what the ball looks like, and you all profess sincerely wanting to find it. Yet, after fruitless hours of searching, there is no agreement that it has been found. While some neighbors produce ball-like objects, others disagree that that is the ball; still others deny that any candidate “balls” have even been found. Is it not quite appropriate to ask what the explanation is for this strange state of affairs? Perhaps the ball is not in the yard; or perhaps there is no ball at all; or perhaps some of the ball hunters are blind; or perhaps some of the ball hunters have been bribed by those who hid the ball to assist in

²An “objective fact” is a mind-independent fact in the following sense: its existence and character does not depend on the epistemic states of other persons, even under epistemically ideal conditions.

their scheme. All of these could be possible explanations for the disagreement among the ball searchers, but what seems clear is that some explanation is “called for.”

The situation is the same, I suggest, with regard to trying to figure out the reasons for action we have. “Reasons for action” may be the *au courant* philosophical lingo,³ but the entire history of Western philosophy has been devoted to this question under various descriptions. Consider:

- (1) Are reasons for acting culturally relative (the Sophists) or absolute (Plato)?
- (2) Are reasons for acting deontological (Kant) or consequentialist (too many to list) or derived from some other source (Aristotle et al.)?
- (3) If reasons for acting are consequentialist in justificatory structure, do they appeal to what maximizes pleasure (Bentham), happiness (Mill), informed desire-satisfaction (Brandt, Railton) or something else?
- (4) If reasons for acting are deontological in justificatory structure, do they require acting from the motive of duty, or acting in accordance with the categorical imperative, or showing respect for persons (a vast literature of intra-sectarian disputes among Kantians)?
- (5) Are moral reasons for acting overriding, or do non-moral reasons sometimes (or often) trump moral reasons (Foot, Williams, Nietzsche, maybe Wolf) or are moral and egoistic reasons for acting locked in irresolvable conflict (Sidgwick)?

After more than two thousand years, philosophers still cannot find the “ball,” as it were: they do not agree about what our reasons for acting really are. That is our explanandum. This calls for an explanation: more precisely, we want to know what is the *best* explanation?

The “best” explanation of some phenomenon satisfies a variety of epistemic and pragmatic desiderata,⁴ ones derived from the successful empirical sciences, which remain our paradigms of fruitful ways of explaining observable phenomena. These desiderata include: (1) simplicity (especially

³Derek Parfit and Joseph Raz seem to be most responsible for this way of talking.

⁴See Leiter 2001 for an overview; cf. Quine & Ullian, Thagard, Harman

ontological simplicity), as opposed to *gratuitous* complexity (i.e., the gratuitous positing of entities that satisfy no other epistemic or pragmatic desiderata); (2) consilience, that is, the ability to show how seemingly disparate phenomena have common causes and thus are not really independent, but causally dependent on the same underlying facts; and (3) conservatism, that is, being consistent with, and thus not unsettling, other beliefs about the world that have been well-confirmed. This is a non-exhaustive list, to be sure, but it is on these that I will focus.

Consider: sitting at home, watching a movie on your computer, suddenly the computer goes off, as do all the lights in the house. What is the best explanation of this phenomenon? Here's a first hypothesis: mischievous leprechuans have invaded the house, discovered the circuit breakers, and flipped them all simultaneously, cutting power to all electrical devices. This explanation does complicate our ontology: it posits leprechuans, beings who are capable of mounting mischievous attacks, for example. That would not be objectionable if, for example, there were other observable phenomena that were satisfactorily explained by such an expansion of our ontology: for example, finding little green hats in the backyard, or pots of gold at the end of rainbows, or, even better, the corpses of very small persons dressed in green. But absent that, this first hypothesis would be quite radical: it would add to our ontology a class of little creatures (are they humans? they seem to have complex cooperative planning capacities!) that, heretofore, we had no reason to think existed.

Our first hypothesis, in short, is not a good candidate for the best explanation of the phenomenon that demands explanation. So let us turn to a second hypothesis: there has been an electrical power failure in the neighborhood. This hypothesis would win additional support from observable facts like: (1) all items powered by electricity went off simultaneously; perhaps (2) lights and other electrically powered items went off in neighboring properties; and perhaps (3) there was a thunderstorm or an electrical explosion that would explain the sudden cessation of electricity to homes in the area. This explanation, assuming the observable evidence confirms it, would be simple (no new

ontological entities), consilient (it would explain what happened in the house and in neighboring properties), and conservative (electricity, power outages, the role of storms in causing them, are all part of our familiar and going theory of how the world works).

I use this silly example in order to motivate an answer to the question: what is the best explanation of why philosophers cannot agree on our reasons for action? The analogy is not exact in lots of ways. But the same kind of “best explanation” phenomenon presents itself. There is in both cases a fact that everyone agrees upon: in the power outage case, that there is suddenly no electrical power; in the reasons for action case, that there is no agreement among philosophers about what our reasons for action *really* are. What is the “best explanation” of that agreed-upon phenomenon? The power outage case illustrates, in simple form, some of the relevant considerations.

But let us take a different hypothetical case, with even more structural analogies to our real topic. A parent takes a group of young children—four or five years old, suppose—to a restaurant to celebrate a child’s birthday. The children are served a cake, with a rich vanilla icing, and strawberry and chocolate filling. Some children think it delicious, the best cake they have ever had; others think it disgusting, indeed inedible; still others eat it, without enthusiasm. Now these are children, prone to stubborn opinions, convinced of their rectitude and good sense (not like philosophers, of course). The enthusiastic cake eaters are denounced by their repulsed cake-eating compatriots: “that’s disgusting, how could you eat that”? The cake enthusiasts are appalled, “What’s wrong with you? This is a great cake!” The “neutral” consumers are mystified: “Don’t fight,” they say.

I have hosted children’s birthday parties, and I exaggerate only a bit the extremities of reaction, though that is to be chalked up to bourgeois social norms of politeness that are often inculcated even in younger children. But the “best explanation” of the birthday party cake is easy: none of the children are “wrong,” tastes differ, and add to that oddities of child psychology, and the drama is easy to explain.

There is no objective fact about whether or not the birthday cake is good; there are different gustatory sensibilities, together with facts about child psychology, that explain the conflagration that ensues.

Now let us return to the philosophers. Everyone knows they do not agree about reasons for action, in the sense that there is no consensus about propositions 1-5 on page three.⁵ One contemporary, the late Derek Parfit, thought this a scandal, an indictment of the *bona fides* of all moral philosophy, including his own. (Here I agree with him.) His final work was devoted to demonstrating that all “major” moral philosophers (or at least his friends, plus Nietzsche!⁶) agreed with his views (or were not *really* disagreeing), appearances notwithstanding. No one was convinced,⁷ though the details do not matter for my purposes. What I want to emphasize is that even some realists about reasons think that the failure of philosophers to agree impugns the objectivity of the results.

But why should it? Here is where we come back to inference to the best explanation (IBE) considerations.⁸ If two epistemic agents are equally well-situated to cognize certain objective facts, yet neither one can agree on what those facts are, we may appropriately ask: why do they not agree? One possible (anti-realist) explanation is that (1) there is no objective fact to be cognized, and (2) judgments about that putative fact are better-explained by non-epistemic features of agents. (I return to possible realist hypotheses, below.)

That seems to describe, nicely, the situation in philosophy regarding our real reasons for acting. Sentimentalists—Hume, Herder, Nietzsche, for example—claim that moral judgments result from the arousal of emotions in response to certain states of affairs, and that the emotions aroused are not to be

⁵Philosophers do mostly agree about trivial applied propositions that no one has ever disputed, e.g., “No one has a reason to torture babies for fun.” (As an aside, it’s not clear that certain kinds of internalists about reasons, like Sharon Street, do agree with this, but put that worry to one side.) Accidental convergence on applied claims about reasons for action typically belies deeper disagreements about *why*, say, torturing babies for fun is wrong.

⁶Parfit’s chapter on Nietzsche is incompetent, but as a matter of academic sociology, those of us in the boring Anglophone world interested in Nietzsche can only be grateful that he thought it worth the bother. He sent me an early version, I sent him detailed comments, but it did not salvage the final product alas. For the problems, cf. Janaway, *Inquiry* (2016).

⁷Reviews by Schroeder, Blackburn...

⁸This is a version of option #4 in Enoch (2009: 21-29); I’ll have more to say about some of his objections in the text.

explained by their “cognitive detective” capacities (i.e., their ability to pick out features of the world warranting a particular response) but by psychological facts about persons that are epistemically insensitive to the “real” moral (or reason-giving) qualities of states of affairs. When sentimentalists observe a two-thousand-year-long intractable debate among moral philosophers they are hardly surprised: even moral philosophers are hostage to their psychology, which determines their affective responses, which determines their intuitive moral judgments, which ultimately determines their moral philosophies.

Let us now state the argument more precisely:

1. If there were objective reasons for action, epistemically-well-situated observers would eventually converge upon them after two thousand years.
2. Contemporary philosophers, as the beneficiaries of two thousand years of philosophy, are epistemically well-situated observers.
3. Contemporary philosophers have not converged upon reasons for action.
4. Conclusion: there are no objective reasons for action (IBE from the first three premises; the details of the IBE will be spelled out below).

This argument makes two key assumptions about epistemic access and epistemic capacities: first, it assumes that *if* there were objective facts about reasons for action, they would be knowable in principle by epistemically well-situated agents; and second, it assumes that two thousand years of moral philosophy should have rendered us epistemically well-situated observers (if it has not, why not?).

An easy way for the realist about reasons to defeat the argument is to deny one or both of the epistemic assumptions. There are, of course, serious costs to doing so. If one denies the first assumption, and allows that *real* reasons for action can be wholly evidence-transcendent, even under ideal conditions, then one ends up in the same practical place as the anti-realist about reasons. The anti-realist says there is no such thing as an objective mistake about reasons for acting so one can never

criticize someone's reasons as involving an objective mistake; the realist say there is such a thing as an objective mistake, but we will never know whether anyone is making it so one can never, with justification, criticize someone's reasons on those grounds. One need not agree with Marx's Second Thesis on Feuerbach ("The dispute over the reality or non-reality of thinking which is isolated from practice is a purely scholastic question") to nonetheless agree that it is a high cost for the realist to allow that there are practical truths, but, alas, they remain beyond our ken.⁹

If the realist denies the second assumption, then the realist must consign to irrelevance some (or perhaps all) of two thousand years of moral philosophy as epistemically inadequate to what our reasons for action really are. But what could justify doing so? Is there some other method for discovering real reasons for acting? Holy texts, divine revelations, the words of prophets, the advice of poets appear to be as deeply in conflict as the views of moral philosophers, but even if they were not, there is so much other evidence that these sources are epistemically unreliable that they seem poor contenders as alternatives. If that is right, then this second realist response appears simply *ad hoc*.

With respect to the possible realist repudiation of the first assumption, we may also ask: what could justify it, beyond the desire to salvage realism about reasons? Put differently, we may ask whether the existence of evidence-transcendent practical truths would survive IBE?¹⁰ Is there an IBE of intractable disagreement about reasons for action among philosophers that would include as a premise the existence of evidence-transcendent practical truths? I am not aware of any, but I invite the realist who wants to opt for this route to supply one.

This is a natural point at which to spell out the IBE involved in the fourth proposition above: why is the non-existence of objective reasons for action part of the best explanation of intractable

⁹Cf. Tersman (2006: Ch. 4) discussing arguments in Wright (1992)...

¹⁰Rowland (2017) offers one argument why there cannot be such practical truths. Arguably Nietzsche does too with his quip about Kant: "how could something unknown obligate us?" (*Twilight of the Idols*). My argument can proceed without defending either of theirs.

disagreement among philosophers? Answering this questions requires proffering a full explanation for this disagreement, which I now set out in terms of the premises to which it is committed.

First Premise: Sentimentalism. Basic intuitive moral judgments (e.g., about the right-making or good-making features of situations or actions) are caused by basic emotional or affective responses.

Second Premise: Non-cognitivism about affects. The affects involved in basic intuitive moral judgments have no cognitive component: they have no cognitively evaluable belief component, but rather a distinctive phenomenal feel. (I will assume, following Nietzsche [Leiter 2013, 2019], that these affects are primarily ones of *inclination* and *aversion*; more on that, below)

Third premise: Philosophical arguments for what our reasons for action are always involve a premise that depends on a basic intuitive moral judgment.

Fourth premise: Philosophers have not converged on our real reasons for acting.

Conclusion (by IBE): The best explanation for the views philosophers defend are the psychological and sociological facts that explain their basic affects.

But why is this the *best* explanation? IBE is a comparative test, so we must have some competitor explanation for moral disagreement on offer if we are to evaluate the relative merits. Here is moral realist's David Enoch explanation of non-philosophical moral disagreement¹¹ which I quote at some length:

Many moral matters are complex, and not at all straightforward; people are the victims of any number of cognitive shortcomings...and so that some may be more likely to make moral mistakes than others; many find it hard—or do not want—to sympathize and imagine what it is like to occupy a different position in the relevant interaction, and different people are sensitive to the feelings of others to different degrees; we let our interests influence our beliefs (moral

¹¹I agree with Enoch (2009:) and many others (e.g., Brink 1989: __-__) that some moral disagreements really mask empirical disagreements, but I do not see that as being at issue in the case I am interested in, i.e., intractable disagreement among moral philosophers. Benthamites and Kantians disagree, but they are not disagreeing about any empirical facts.

and otherwise); we are subject to the manipulation of others, and so to the distorting effects also of their self-interests....

Let me draw special attention to one of these kinds of alternative explanation: that in terms of the distorting effects of self-interest. This kind of explanation is especially important for at least two reasons: It is extremely powerful, and it helps explain the difference in the scope of disagreement in morality and in other discourses. (Enoch 2009: 25-26)

Since a central part of what is in dispute in the history of philosophy is whether our real reasons for action are self- or other-regarding, Enoch's last point is question-begging, so I will put it to one side. But even if we discount Sidgwick (who thought the conflict between self-regarding and other-regarding reasons intractable), as well as the various opponents of Socrates who defended self-regarding reasons for acting, how would the considerations Enoch adduces fare as an explanation of the disagreement among *the other philosophers* for two thousand years about our reasons for action? What cognitive shortcomings, for example, did Bentham suffer from that Kant did not? What "interests" influenced Kant that led him to disagree with Bentham? Who "manipulated" Mill and Aristotle? (And, as Kant would instruct us, "sympathizing" with someone is not a sound basis for determining one's moral obligations anyway!)

Enoch's points seem unlikely candidates to explain away the intractable disagreement about reasons for action that characterize the history of Western philosophy (even if they have some purchase in the case of non-philosophical disagreement). But does the anti-realist explanation sketched above fare better? I shall argue it does.

First, the empirical evidence in support of sentimentalism (the first premise) is very strong. As a recent literature survey notes, individuals "with selective deficits in emotional processing" due to disease or injury to the brain render different moral judgments about hypothetical situation like the Trolley cases, than most emotionally normal subjects to hypothetical situations (Cushman et al. 2010:

53-54), suggesting that the affective responses are causes of the moral judgments. Psychologist and philosopher Joshua Greene (2007) has argued that emotional responses loom larger in deontological than consequentialist moral judgments, the latter demanding more “controlled cognition” (Cushman et al. 2010: 54), but in more recent work even Greene has acknowledged that “affect supplies the primary motivation to view harm as a bad thing” in the first place, so that even consequentialist reasoning has “an affective basis” (Cushman et al. 2010: 62).¹² In another recent review of the empirical literature, Timothy Schroeder, Adina Roskies and Shaun Nichols found that the view they dub “sentimentalism”—namely, the view that “the emotions typically play a key causal role in motivating moral behavior” (Schroeder et al. 2010: 77)—is well-supported by the “evidence from psychology and neuroscience” (Schroeder et al. 2010: 98), and that while “motivation derived [exclusively] from higher cognitive centers independently of desire is possible...the only known model of it is pathological” involving Tourette syndrome (Schroeder et al. 2010: 94). In sum, the empirical evidence supports the claim that sentimentalism is part of a consistent explanation of the phenomenon of moral judgment.¹³

Second, non-cognitivism about basic affects (the second premise) is also highly plausible, psychologically and philosophically. Here let me focus on Nietzsche’s version of this view. In *Daybreak*, Nietzsche says “moral feelings” (*Moralische Gefühle*) are inculcated when “children observe in adults inclinations for and aversions to certain actions and, as born apes, *imitate* these inclinations [*Neigungen*] and aversions [*Abneigungen*]; in later life they find themselves full of these acquired and well-exercised affects [*Affekten*] and consider it only decent to try to account for and justify them” (D 34). We may bracket the astute concluding observation—about the impulse to supply post-hoc

¹²More precisely, “affect supplies the primary motivation to regard harm as bad. Once this primary motivation is supplied, reasoning proceeds in a currency-like manner [“currency emotions are designed to participate in the process of practical reasoning”]” (Cushman et al. 2010: 63). “[A]larm-bell emotions are designed to circumvent reasoning” (*id.* at 62) and, arguably, this is “the origin of the welfare principle,” namely “in “Parkinson disease appears to show that intrinsic desires are necessary to the production of motivation in normal human beings, and this would seem to put serious pressure on the cognitivist position” (93)

¹³Note that sentimentalism could be true and well-supported by empirical evidence, but that could be compatible with moral realism.

rationalizations for evaluative judgments produced by a non-rational mechanism—in order to focus on its picture of basic affects. Moral feelings or affects are equated with “inclinations for and aversions to certain actions,” or, more precisely, with the mental state, whatever it is, that motivates one to perform certain actions or avoid certain other actions. It is surely plausible that any credible ontology of the mind includes mental states that are *motivationally effective*, at least in the sense that they produce what I will call henceforth *motivational oomph* (or *push*), that is, they incline towards or avert away from certain acts, even if they are not ultimately successful in producing action. We typically presume, of course, that *affects* or *feelings* are characterized, in part, by their ability to produce *motivational oomph*, and, indeed, the fact that they do so is one of the main points thought to count in favor of views, like sentimentalism, that understand moral judgments to involve essentially the expression of feelings.

It is equally a part of ordinary usage to take feelings to have another crucial characteristic: namely, that they are phenomenologically distinctive. There is something *it feels like* to be inclined to stop the child from sticking his hand in the fire, and there is something *it feels like* to be inclined to avoid killing a child. A *non-cognitivist* view of affects claims that they can be fully individuated by their distinctive phenomenal feel; a *cognitivist* view denies that, claiming, instead, that to individuate the affect one also needs to consider some aspect of its *cognitive* (i.e., truth-evaluable) content, such as a belief. Cognitivist views of emotions, notoriously, are forced to deny that human infants can have emotions, since they lack the concepts necessary for having truth-evaluable judgments (cf. Deigh 1994). That infants cannot have emotions is, by my lights (as someone who has cared at length for three infants!), a *reductio* of the cognitivist position. But less anecdotally, cognitivism about the emotions runs up against a large body of psychological research showing that, as Jesse Prinz puts it, “emotions can arise without judgment, thoughts, or other cognitive mediators” (2007: 57). While it may be doubtful that qualitative feel can distinguish all emotional states (Prinz’s example is differentiating “anger” from “indignation” [2007: 52]), that is not relevant here: the question is, following Nietzsche’s

proposal, that the *affective* source of moral judgment, in the first instance, consists in fairly basic or simple mental states of *inclination* and *aversion*, and it seems quite plausible that there is something it feels like to be *inclined towards X* or *averted away from Y*, even if the relevant *qualia* might be thought inadequate to pick out all conceptual nuances we might want to apply to such cases. It is also possible, of course, that our conceptually nuanced distinctions between, say, *anger* and *indignation* are explanatorily otiose with respect to the *motivational oomph* associated with the feeling in question.

But what about the most common worry about non-cognitivist views of feelings or affects, namely, how can they account for our propensity to think emotional responses can be assessed as warranted or not (*cf.* Prinz 2007: 60ff.)? Even in infants, of course, emotional responses have causal triggers: the infant *sees* a scary “monster” and reacts by crying in fear. The infant’s *fear* has an *intentional content*: it is *about* something, namely, the scary monster. We might believe the infant is wrong to be afraid (i.e., what the infant sees is not dangerous, since it is only make-believe), but that does not change the fact that the infant experiences upset and fear. I believe that a child is about to stick his hand in the flame, and so I *feel inclined* to stop him. “Feel inclined” understates the character of the feeling, of course: *I feel I must* stop him. But the fact that there is a causal relationship between a mental state with cognitive content and an affective response, does not show that to *identify* the affect one must understand it to be constituted, in whole or even in part, by a *cognitive* component. We need to account for the *intentionality* of emotions, to be sure, but that does not mean we have to identify emotions with *judgments*. It might suffice to introduce a kind of explanatorily primitive mental state, as Peter Goldie (2002: 19) famously proposes, like “feeling towards,” one that capture both the *intentionality* of emotional states *and* their distinctive qualitative character.

So feeling inclined to do X or averse to doing Y can have a distinctive phenomenal character, even if the causal trigger for the feeling is a false belief of the agent with the distinctive feeling. When we criticize the resulting emotion, perhaps, then, we are making a different normative judgment,

something like: a correct belief about the facts ought not cause such an affective response. Consider: the infant falsely perceives (without any conceptual content) the family dog as dangerous, and so bursts into tears. We think the infant is *mistaken*, because we know that the family dog is not dangerous. But we don't criticize the infant for having an unwarranted affective response, rather we console the infant. Why? Perhaps because we don't think the infant is *epistemically responsible* for its false cognition of the family dog as dangerous. But when I react with anger to a perceived slight by the store clerk that, in reality, was nothing of the kind, I can be criticized not because anger is anything other than non-cognitive, and not because my "belief that I've been slighted" is a causal trigger for my anger, but because I am epistemically culpable for having a false belief, one that produces unhappy consequences (my anger, a fight with the clerk at the store, an unpleasant scene, etc.). (Note that anti-realism about epistemic judgment is a live option: cf. Field 2009.)

Yet it seems I can also be criticized for responding with anger even when my belief was correct, i.e., when the store clerk *really was* rude. How can that criticism make sense if the cognitive belief is just a causal trigger for an evaluative/affective response that has no cognitive content? The obvious answer is that such judgments are, themselves, the expression of other non-cognitive attitudes or feelings, such as an inclination towards non-aggressive responses to insults or an aversion to aggressive ones (we might call the latter the "Christian attitude"). Of course, if we view such criticism as making a cognitive, rather than non-cognitive, claim, then we will be forced towards an error-theoretic interpretation of such judgments (that is, we will have to interpret all such judgments as false) or some more complicated non-cognitive (in the semantic sense) interpretation of them. Thus, the non-cognitivist about basic affects of inclination and aversion can claim that our basic affects of inclination and aversion are marked by a distinctive conscious, qualitative feel; that these feels are the causal root of our moral judgments; and the social practice of assessing such feelings as warranted or unwarranted can be compatible with the non-cognitive character of the feeling because to deem them warranted or

unwarranted is either (1) to render an epistemic judgment (perhaps itself amenable to anti-realist treatment) about the cognitive judgment that is the causal trigger for the non-cognitive state, or (2) to express a meta-affect (i.e., an affect *about* an affect) about someone else's affective response to a causal trigger. In sum, non-cognitivism about basic affects offers a simple and conservative explanation of the crucial claims in the sentimentalist hypothesis about moral judgment.¹⁴

The third premise makes a claim about the logical structure of arguments for reasons for action that seems hard to dispute: such arguments depend on a premise that appeals to what is supposed to be intuitively plausible. Now the "Kantian" loosely construed—that is, someone who thinks rational action itself, either as a matter of logic (e.g., arguably Kant or Gewirth) or its constitutive character (e.g., Korsgaard or Velleman), supplies the crucial normative premise—will deny that intuition is at issue, as opposed to the demands of rationality or action. I shall simply record my view that these arguments are implausible, so much so that Gewirth's has no defenders anymore, and Korsgaard's and Velleman's constitutivism will, one hopes, go the same way (see Enoch 2006). With apologies to members of this particular Kantian sect, I will say nothing further about these views.¹⁵

The fourth premise notes the key explanandum we have been discussing throughout: philosophers do not agree about reasons for acting.

Now we are ready for the anti-realist inference to the best explanation. Since philosophers develop arguments that depend on key intuitive premises, that are non-cognitive, but explicable in

¹⁴Someone might worry that it is not at all "conservative," but radically revisionary of ordinary "folk" beliefs, to tell people that there is (as Nethanel Lipshitz aptly put it to me) "no such thing as good and evil," and that there is nothing they *objectively* have a reason to do or not do. But this kind of worry involves a misunderstanding of conservatism as a demand on choosing better and worse explanations. We should be conservative with our explanations vis-à-vis well-confirmed theories and beliefs about the world, not about folk beliefs and intuitions, which are (outside, say, judgments about midsize physical objects) a notorious hodgepodge of falsehood, delusion and wish-fulfillments.

¹⁵Weber's point a century ago about specialization as increasingly essential to *Wissenschaft* in the universities has wreaked particular havoc in philosophy, since, as a university discipline, it still aims to be a *Wissenschaft* (unlike certain other fields in the so-called "humanities" like English) yet is obviously hostage to the intuitions or "common sense" of its practitioners, an interesting sociological and anthropological artifact to be sure. But the combination of specialization, the need for appearing *wissenschaftlich*, and the dependence on epistemically feeble data points like "intuitions" explains some of the peculiarities of recent professional philosophy—these may even provide the "best explanation"!

terms of facts about the psychology and sociology of those who make the judgments, it should hardly be surprising that their philosophical systems are locked in intractable conflict, since their arguments depend, both logically and causally, on views that are psycho-social artifacts that vary with person, time and place.¹⁶ Bentham thinks it intuitively obvious that the experience of pain is the most morally salient fact; Kant (on certain readings) thinks it intuitively obvious that the motive for action is the most morally salient fact. What would an argument for either premise look like that did not already presuppose the intuition?

Recall the children at the birthday party. Child A says, “The cake is disgusting, you ought not eat it!” Child B says, “The cake is delicious, eat it!” Child A, with some help from a philosopher, explains that (1) chocolate and strawberry are a revolting combination of flavors; (2) any combination of flavors that is revolting ought not be eaten; therefore (3) the cake should not be eaten. Child B, with the help of another philosopher, retorts: (1) chocolate and strawberry are a delightful combination of flavors; (2) any combination of flavors that is delightful ought to be eaten; therefore (3) the cake should be eaten. Their debate is intractable because a crucial premise in their argument—the first premise—is determined by their gustatory sensibilities, an artifact of each child’s psychology (perhaps also biology) and upbringing. There is no objective fact of the matter about whether chocolate and strawberry combined are delicious or not.

The intractable debate among philosophers about reasons for action has the same explanatory structure.¹⁷ The disputants depend on a premise that is determined not, of course, by their gustatory

¹⁶Max Etchemendy points out to me that there is another IBE in the offing here, which accepts my first three premises, but adds one more, namely, that basic affective responses are not amenable to change through argumentation in seminar rooms or in the pages of philosophy journals. That, too, would explain the failure of convergence among contemporary philosophers, but without taking a position on realism vs. anti-realism about reasons. It would, of course, entail that moral philosophy is irrelevant, and so, in that regard, is a version of the first realist rejoinder considered earlier, i.e., affirming that the truth about reasons for acting transcends what we can know about them. But that raises the question why we should think *that* is the explanation for the failure of convergence.

¹⁷The realist may here object that a key difference is that while adult observers will agree that the children are wrongly disputing mere matters of taste in the cake case, they will not so view their moral disagreements.

sensibilities, but by facts about their psychology and sociology: their sensitivity, their stamina, their openness, their neuroticism, their upbringing, their peer environment, etc.. That premise is non-negotiable, but itself has no cognitive status, and so cannot be refuted by logic or evidence. Let us recall, for example, that Derek Parfit's parents were Christian missionaries; that John Rawls had planned to be an Episcopal priest; that Joseph Raz's father was a Rabbi; and, more recently, that Rae Langton's parents were Christian missionaries. Are these facts about some contemporary moral philosophers *really* irrelevant to their moral philosophies?

Presumably even the realist will admit these psycho-social facts are relevant, the question is whether these facts about the prosaic origins of philosophers' central moral intuitions are part of the best explanation of why they end up disagreeing with each other? Of course, the disagreements among products of Judeo-Christian indoctrination can look like the narcissism of small differences compared to the disagreement between recent Western moral philosophy and Nietzsche or Homer or the rest of the world, so perhaps we should pause for a moment to mention the more dramatic cases of disagreement.

The standard-issue academic Anglophone philosopher thinks it intuitively obvious that "suffering is bad" or that "humans are possessed of dignity which precludes harming them" or some other familiar formulation. My argument is that, however much affection readers have for one or both of these premises, they are nothing more than affectively-induced intuitions that tell us more about "socialization"¹⁸ than about reality.

Someone part of the Homeric world—or perhaps the Mafia more recently?—would find these purportedly intuitively obvious claims bizarre. Suffering is essential to glory and honor, it is not

Assuming that is true (the evidence is mixed—consider the economists' talk of "preferences," which is ubiquitous in the neoclassical tradition), that simply raises the question what best explains such a meta-view about the status of moral disagreements. It is not clear to me the explanation is really that different in either case, though the habit of academic philosophers of talking only to each other perhaps encourages a different judgment.

¹⁸I use the term "socialization" very loosely to capture all those psycho-social facts about persons that explain their moral judgments per the IBE sketched earlier.

objectionable but welcome.¹⁹ Loyalty and the ability to exact revenge are far more important than suffering, and individuals can be harmed when honor, glory or revenge demands it. It is not simply that all of human history reveals the powerful role of this competing set of norms, but that there is no refutation of commitment to them that does not presuppose a premise involving a different affective commitment. These norms can be coherently formulated and applied and they play, as much as we observers might regret it, an enormous role in human affairs.

Most academics have a profound affective identification with the intuitions at the foundation of both the Kantian and utilitarian traditions (I find myself moved by aspects of both).²⁰ Glory, honor, revenge and loyalty lack academic defenders, but that is not to be regretted, since those normative ideals thrive and structure lives without academic philosophers writing in support. But calling attention to them indicates the diversity of views about reasons for action that have gripped human minds.

To be sure, the realist might retort that the appeal of glory or revenge as a reason for acting is vulnerable to being explained away, perhaps along precisely the lines Enoch proposed originally. Recall Enoch:

Many moral matters are complex, and not at all straightforward; people are the victims of any number of cognitive shortcomings...and so that some may be more likely to make moral mistakes than others; many find it hard—or do not want—to sympathize and imagine what it is like to occupy a different position in the relevant interaction, and different people are sensitive to the feelings of others to different degrees; we let our interests influence our beliefs (moral

¹⁹Someone might object that even those committed to the view that suffering is bad and that inflicting suffering is wrong can allow that those are defeasible normative considerations, and so in fact the Homeric hero is not disagreeing. That could be true if there were evidence that the Homeric hero's "practical reasoning" worked this way: but, first, Homeric heroes don't engage in much practical reasoning (at least not as Homer presents them), and second, there is no indication that Homeric heroes are committed to the wrongness of suffering being a normative consideration that needs to be defeated.

²⁰I put to one side Aristotle and the neo-Aristotelians with their fanciful assumptions about what it means to be human: they have the distinct advantage of not regarding empirical evidence from the natural and human sciences of what humans are actually like as having any bearing on their ideas.

and otherwise); we are subject to the manipulation of others, and so to the distorting effects also of their self-interests.... (Enoch 2009: 25-26)

Remember that our topic is reasons for action, not simply “moral” reasons for action as conceived within a particular parochial tradition, say, the Kantian or utilitarian. But to evaluate Enoch’s alternative, let us take as an example Odysseus, who is celebrated by Homer for his talents at deception and his capacity to exact revenge. Odysseus, recall, returns after his long journey to find suitors in his home trying to seduce his faithful wife Penelope. Disguised as a beggar, Odysseus is the only one strong enough to string his old bow (the challenge put to the suitors by Penelope), at which point he slaughters the other suitors in revenge.

Has Odysseus made a mistake about his reasons for acting? Did he fail to “sympathize and imagine what it is like to occupy a different position in the interaction”? Let us grant that this is, in fact, a legitimate demand on reasons for acting, though it looks dangerously close to assuming, without argument, that possibly self-regarding reasons (like honor) are out of bounds. But it is easy enough to imagine that Odysseus would reason as follows: “Were I to be found out trying to seduce another man’s wife during his absence, he would be right to slaughter me.” Was Odysseus insufficiently “sensitive to the feelings of others”? Again, let us grant that this consideration is even relevant to his reasons for acting. Has he really failed to be “sensitive to the feelings of others”? The feelings of others are, presumably on any view about reasons for action, a defeasible consideration: it no doubt hurt Donald Trump’s feelings that residents of London flew a giant balloon of a “Baby Donald” over the city during his recent visit to the country, yet his feelings were not a reason for the protesters not to do that. Even if Odysseus does not give much weight at all to the no-doubt strong preference of the suitors not to be slaughtered, this does not mean he gave it no weight: Odysseus surely knows his suitors would prefer not to be victims of his revenge. Yet he could reason, without error, as follows: “to try to marry a man’s faithful wife while he is away engaged in an arduous life-and-death struggle for a worthy cause is

an offense to honor and decency, and anyone who embarks on such a path has forfeited his claim to have his wish to live respected.”

Of course, that reasoning would be in error *if* one granted a premise like “inflicting death on others for the sake of revenge is always wrong” or, alternatively, “inflicting death on others to avenge their wrongful actions is always wrong.” But Odysseus accepts neither premise, not because either is false, but because the psychology and sociology of his own upbringing leaves him with no affective inclination toward accepting these premises such that they should influence his practical reasoning.²¹ Odysseus disagrees with the reasons for acting that seem decisive to we moderns because his psychology and sociology are different, not because there is an objective reason for him not to slaughter the suitors of Penelope—or, for that matter, an objective reason for him to slaughter them.

A realist might protest that the *best* explanation of why we judge Odysseus’s reasons for acting mistaken is precisely that we are sensitive to the infliction of death on sentient beings for trivial reasons, and that any action with that attribute is a token instance of “moral wrongness”: yes, we were brought up to be sensitive to that, but in so being raised, we were raised to be sensitive to actual moral wrongdoing! Now in order not to beg the question we have to ask whether the thesis that “moral wrongdoing” is identical with or supervenient upon, among other things, inflicting death on sentient beings for revenge, would be part of the best explanation of the Odysseus/contemporary dispute? The anti-realist explanation, recall, appeals to psychological and sociological facts, and their historical, cultural and temporal variation; the realist explanation presumably does not deny any of those facts *but then adds a further fact*, namely, that killing for revenge is a case of objective moral wrongness. Does that additional metaphysical thesis, which certainly makes our ontology less simple, make the explanation more consilient? *What else does it help us explain?* I have argued elsewhere that such an

²¹As noted, above *supra* n. 17, I do not think Odysseus actually engages in “practical reasoning,” rather than acting instinctively. But this is a separate debate: for skepticism about practical reason, see my discussion of Nietzsche’s view (Leiter 2019: Ch. 5).

assumption has no consilience benefits at all.²² Is it a methodologically conservative assumption? It would be if our best-going theory of the world already required that there be objective facts about reasons, but there is no evidence I can see that it does.²³ Anti-realism about reasons for acting is rampant in all the social (as well as physical) sciences; it is only some philosophers (and theologians) who think there are objective reasons for acting.

Realists about reasons have developed many “defusing explanations” (Doris & Plakias 2007: 311, 320-321; cf. Loeb 1998 for a useful survey and rebuttal of various strategies) to block the abductive inference from apparently intractable disagreement about our reasons for acting to skepticism about objective reasons. Disagreement, as I have acknowledged throughout, is an epistemic phenomenon, from which we propose to draw a metaphysical conclusion. The ‘defusing’ explanations of such disagreement propose to exploit that fact, by suggesting alternate epistemic explanations for the disagreement, many of which I have already touched upon. Here I want to comment on one residual defusing explanation.

Perhaps, this defusing argument goes, disagreements about reasons for action are not *really* intractable, but resolvable in principle. This has been the standard optimistic refrain from philosophers ever since “moral realism” was revived as a serious philosophical position in Anglophone philosophy in the 1980s. With respect to very particularized moral disagreements—e.g., about questions of economic or social policy—which often trade on obvious factual ignorance or disagreement about complicated empirical questions, this seems a plausible retort. But for over two hundred years, Kantians and utilitarians have been developing increasingly systematic versions of their respective positions. The Aristotelian tradition in moral philosophy has an even longer history. Utilitarians have become

²²Leiter (2001) argues at length that the ontological complexity produces no explanatory payoffs, criticizing claims to the contrary by Brink, J. Cohen, and Sayre-McCord, among others.

²³Enoch (2009: 27) mischaracterizes the issue when he says (correctly) that “[c]ompeting explanations are evaluated holistically and against a background of prior beliefs,” but then says (falsely) that competing explanations of moral disagreement will “depend on whether we were metaethical realists to being with.” But holistic forms of justification, informed by conservatism, do not assume that just any “prior beliefs” count: it is the “prior beliefs” that we already have good IBE reasons to think are true! But our prior beliefs about realism about reasons for action are precisely what is at issue in the IBE argument, so we cannot presuppose such beliefs in evaluating competing explanations.

particularly adept at explaining how they can accommodate Kantian and Aristotelian intuitions about particular cases and issues, though in ways that are usually found to be systematically unpersuasive to the competing traditions and which, in any case, do nothing to dissolve the disagreement about the underlying moral criteria and categories. Philosophers in each tradition increasingly talk only to each other, without even trying to convince those in the other traditions.²⁴ And while there may well be ‘progress’ within traditions—e.g., most utilitarians regard Mill as an improvement on Bentham—there does not appear to be any progress in moral theory, in the sense of a consensus that particular fundamental theories of right action and the good life are deemed better than their predecessors. What we find now are simply the competing traditions—Kantian, Humean, Millian, Aristotelian, Thomist, perhaps now even Nietzschean—which often view their competitors as unintelligible or morally obtuse, but don’t have any actual arguments against the foundational premises of their competitors. There is, in short, no sign—I can think of *none*—that we are heading towards any epistemic *rapprochement* between these competing moral traditions. So why exactly are we supposed to be optimistic?

As grounds for optimism, many philosophers appeal to the thought (due to Derek Parfit) that secular “moral theory” is a young field, so of course it has not made much progress. This strikes me as implausible for several reasons. First, most fields with factual subject matters have usually managed to make progress, as measured by convergence among researchers, over the course of a century—and especially the last century, with the rise of research universities. Moral theory is, again, the odd man out, when compared to physics, chemistry, biology, or mathematics.²⁵ Even psychology, the most

²⁴The one partial exception is Parfit’s last work, though the extent to which it really engages the competing views is open for some doubt: cf. the reviews cited in n. ____.

²⁵It is natural to worry that the problem of intractable disagreement is not limited just to moral philosophy: is not the same the apparently intractable disagreement mirrored in many other parts of our discipline? Are not metaphysicians and epistemologists also not locked in intractable disagreements of their own? Think of debates between internalists and externalists in epistemology, or between presentists and four-dimensionalists in the philosophy of time. If disagreement among moral philosophers supports an abductive inference to denying the existence of moral facts, what, if anything, blocks that inference in all these other cases? Some writers think this kind of “companions in guilt” consideration counts in favor of moral realism, notwithstanding the disagreement among moral philosophers. It is not entirely clear why they rule out, however, the other natural conclusion. Of course, we would need to think carefully about individual cases of philosophical disagreement, since not all of them, in all branches of philosophy, are as intractable or as foundational as they are

epistemically robust of the ‘human’ sciences, managed to make progress: e.g., the repudiation of behaviorism, and the cognitive turn in psychology in just the last fifty years. Second, Spinoza, Hume, Mill and Sidgwick (among many others) may not have advertised their secularism, but the idea that their moral theories are for that reason discontinuous with the work of the past hundred years does obvious intellectual violence to the chains of influence of ideas and arguments. Third, and relatedly, so-called “secular” moral theory regularly conceives itself in relation to a history that stretches back in time (sometimes back to the Greeks)—contrast that with the relative youth of modern physics!--so that it becomes unclear why the bogeyman of the deity was supposed to have constituted the insuperable obstacle weighing down intellectual progress. Most contemporary deontologists may be atheists, for example, but it is not obvious that their atheism enabled them to make stunning intellectual progress beyond Kant.²⁶

Psychological and sociological facts are part of the ontology of the social sciences already; objective facts about reasons for acting are a significant addition. Unfortunately, it is unclear that adding such facts offers any explanatory advantages. Such an ontological add-on seems to fare

in moral philosophy. Some philosophical disagreements can, in fact, be defused fairly easily. Thus, to take an example from one of my other fields, the debate in legal philosophy between natural law theorists and legal positivists about the nature of law has both an element of tractability (natural law theorists like Finnis have, in fact, conceded most of the claims that actually matter to legal positivism as a theory of law) and admits, in the intractable parts, of defusing by reference to the transparent and dogmatic religious commitments of the natural law theorists on the remaining issues they refuse to cede. In sum, the skeptical argument from disagreement among philosophers may have implications beyond moral philosophy, but what precisely they are will have to be decided on a case-by-case basis. (I discuss this issue in a bit more detail in Leiter [2014].)

²⁶Alastair Norcross suggests to me that the real problem is that ethics requires reliance on “intuitions,” and our intuitions are still strongly tainted by our religious traditions. That seems a more plausible point, though it is unclear what criteria we are going to appeal to in order to sort the “tainted” from “untainted” intuitions. As Nietzsche would be the first to point out, the utilitarian obsession with sentience and suffering is, itself, indebted to Christianity—an ironic fact, given the centrality of the wrongness of suffering to Parfit’s own moral philosophy (e.g., 2011b, pp. 565 ff.).

especially poorly in helping us understand the history of philosophy, which is the history of intractable dispute about reasons for acting.²⁷

References

- Cushman, Fiery, Liane Young & Joshua D. Greene. 2010. "Multi-system Moral Psychology," in Doris (2010): 47-71.
- Doris, John M. and the Moral Psychology Research Group eds. 2010. *The Moral Psychology Handbook* (Oxford: Oxford University Press).
- Doris, John M. & Alexandra Plakias. 2008. "How to Argue about Disagreement: Evaluative Diversity and Moral Realism," in W. Sinnott-Armstrong (ed.), *Moral Psychology, Volume 2: The Cognitive Science of Morality: Intuition and Diversity*. Cambridge, Mass.: MIT Press.
- Enoch, David. 2006. "Agency, Schmagency: How Normativity Won't Come from What is Constitutive of Agency," *Philosophical Review*
- 2009. "How is Moral Disagreement a Problem for Realism?" *Journal of Ethics* 13: 15-50.
- 2011. *Taking Morality Seriously: A Defense of Robust Realism* (Oxford: Oxford University Press).
- Field, Hartry. 2009. "Epistemology without Metaphysics," *Philosophical Studies*
- Goldie, Peter. 2002. *The Emotions: A Philosophical Exploration* (Oxford: Oxford University Press).
- Knobe, Joshua & Brian Leiter. 2007. "The Case for Nietzschean Moral Psychology," in B. Leiter & N. Sinhababu (eds.), *Nietzsche and Morality* (Oxford: Oxford University Press).
- Leiter, Brian. 2001. "Moral Facts and Best Explanations," *Social Philosophy & Policy* 18: 79-101. Reprinted in Leiter (2007) (page citations are to the reprinted version).

²⁷I received helpful comments on an earlier draft of this paper from Thomas Adams, Max Etchemendy, and Nethanel Lipshitz.

- . 2007. *Naturalizing Jurisprudence: Essays on American Legal Realism and Naturalism in Legal Philosophy*. Oxford: Oxford University Press.
- . 2013. "Moralities are a Sign-Language of the Affects," *Social Philosophy & Policy*
- . 2014. "Moral Skepticism and Moral Disagreement in Nietzsche," *Oxford Studies in Metaethics* (ed. Shafer-Landau): ___-___
- . 2015. "Normativity for Naturalists," *Philosophical Issues: A Supplement to Nous* (ed. Neta): ___-___.
- . 2019. *Moral Psychology with Nietzsche* (Oxford: Oxford University Press).
- Loeb, Don. 1998. "Moral Realism and the Argument from Disagreement," *Philosophical Studies* 90: 281-303.
- Schroeder, Timothy, Adina L. Roskies & Shaun Nichols. 2010. "Moral Motivation," in *Doris* (2010): 72-109.
- Tersman, Folke. 2006. *Moral Disagreement* (Cambridge: Cambridge University Press).
- Wright, Crispin. 1992. *Truth and Objectivity* (Cambridge, Mass.: Harvard University Press).