Priorities: Comments on *Rational Sentimentalism* by Justin D'Arms and Daniel Jacobson

Vida Yao[1]

1 Introduction

D'Arms and Jacobson (hereafter, D&J) present a view, Rational Sentimentalism (hereafter, RS), which seeks to vindicate what they call the "sentimental values", such as the amusing, the disgusting, and the shameful, as well as our appeal to reasons in correcting those values' paired emotions – our own, and those of others. RS has three main components. The first is an argument that when we describe something as, for example, disgusting, we can be understood as claiming that it would be "fitting" (that is, that there is sufficient reason) for some appropriately placed person to feel disgust toward it. The second is to show that this kind of ascription and our emotional responses are subject to rational standards: they can be corrected through considering whether they are "fitting". What counts as a reason in favor of "fit" will be determined by each emotion's overall "appraisal": e.g. that anger appraises something as an offense. If the emotion in question is a "natural kind", this appraisal is determined through an interpretation of its empirically observable goal, phenomenology, elicitors and action-tendencies. Thus, to think of something as "anger-worthy" is to attribute a reason to feel anger toward it that references this standard; we are neither merely expressing an attitude, nor (the authors think) ascribing to the world any "queer" metaphysical properties (78). Third, while the view is consistent with competing "first-order" ethical views (i.e. differing views on what counts as an offence), the view nonetheless sets limits on them: (i) such views must respect the emotion's appraisal which is settled prior to ethical reflection and (ii) any adequate ethical position must respect that the natural emotions yield reasons to feel and act for all (and only) normal human beings. The view is in this sense "anthropocentric."

RS is presented as a form of "neo-sentimentalism", an approach the authors believe comprises the views of Simon Blackburn, Alan Gibbard, John McDowell, and David Wiggins. In earlier work, D&J ask us to overlook the "metaphysical and

✉ Vida Yao
  vidayao@humnet.ucla.edu

1 University of California, Los Angeles, USA

semantic" differences between these positions, and to focus instead on their similarities.[1] According to D&J, these include the importance placed on the emotions for ethical life, an anthropocentric conception of value, and their reliance upon a notion of an emotion being "appropriately" ("fittingly") felt. These apparent similarities also include – as I'll focus on here – the idea that ethical or moral truth must be "earned" (D'Arms and Jacobson, 2000, p. 733), as well as a wariness of forms of "scientism".

Given this framing, one may gain the impression that RS makes philosophical progress by building upon what all "neo-sentimentalist" positions fundamentally agree to, while shedding the relatively less significant, more esoteric disagreements between them that many philosophers have by now simply grown weary of. But this impression is misleading. The crux of just one of their disagreements – about the meaning of concepts, pressed by McDowell – puts pressure on the very *ambition* of a theory such as RS.

I will begin by presenting four objections to RS. In their general forms, these objections may be familiar to those within contemporary meta-ethics, and can be raised against any project which seeks to "ground" or "vindicate" something ethical (or "normative") in or with something that is not. One might think that while RS is vulnerable to these objections, there is, nonetheless, *some* view which could avoid them. Indeed, the idea that there could be is the *raison d'être* of many strands of contemporary meta-ethics. So, my central goal here will be to provide a unified explanation of why – *in principle* – the meta-ethical ambition of RS cannot be met, given a plausible insight about the meaning of concepts. This will involve making explicit aspects of McDowell's position that stem from that insight, which is not unique to his views, and which the authors do not directly address here or in their earlier discussions.[2] If one shares the meta-ethical ambition of RS and does not address this claim one will, at the very least, directly, and *predictably*, encounter forms of the objections I will now press.

2 Four objections

Schematically, RS strives to vindicate our use of certain evaluative (or "normative") terms (the sentimental values) and their rational correction through providing a metaphysically respectable (i.e. "naturalistic") basis for our use of them (the "natural" emotions and their fittingness conditions). Given this aim, I will raise four

---

[1] "…we will deliberately downplay their metaphysical and semantic disputes so as to bring out the point that these theories are all heirs to the sentimentalist legacy" (D'Arms and Jacobson, 2000, p. 724).

[2] D&J discuss McDowell's views jointly in (D'Arms and Jacobson, 2000) and (2006), and Jacobson discusses them alone in (Jacobson, 2005), but they do not address this claim in those places or in *Rational Sentimentalism*. The contemporary "fitting attitude" literature that RS is part of locates McDowell (along with Wiggins) as holding a "no priority" view of attitude and value, but with little to no engagement with his arguments for this position. This dismissiveness of such views on the basis of their "circularity" is expressive of just how different the ambitions and philosophical methodology of the contemporary literature are from either McDowell's or Wiggins'.

objections. First, the view has not identified and characterized a recognizable aspect of our actual ethical practices. Second, it is not clear that the authors have disentangled themselves from first-order ethical commitments in characterizing the non-ethical resources meant to ground and explain those commitments. And, even if the view were able to meet these first two challenges, the authors do not provide an adequate argument or explanation for how or why the non-ethical foundation of their view restricts first-order ethical thought. So, third, RS seems to commit the naturalistic fallacy.[3] And fourth, we are not given an adequate argument for how or why the nearby phenomena that their view cannot vindicate are *thereby* unable to earn their keep in our ethical practices.

First objection. RS aims to vindicate certain aspects of our ethical discourse: that we seem to think that some things really are funny, disgusting, or shameful, and that, as they put it, it is "…routine for people to deliberate, endorse, and regulate [these emotions] according to standards of fit" (D'Arms and Jacobson 2023, p. 68). To illustrate this second aspect, D&J present a disagreement between two people, Ben and Beth, about whether it is appropriate for Ben to be ashamed of his parents, given that they were "impoverished immigrants" (D'Arms and Jacobson 2023, p. 68). Ben thinks yes; Beth thinks no. Beth reasons no because he "cannot help who his parents are, and what they are like" (D'Arms and Jacobson 2023, p. 68). This, the authors claim, is the "wrong kind of reason" – it is not a reason of "fit" because irrelevant to the question of whether his shame is accurately felt. Likewise, they claim that it would be the wrong kind of reason if Ben were to reflect about what a more virtuous (perhaps more grateful, more understanding, less shallow) person would feel – this is also the wrong kind of reason, just like that it would be more "advantageous" if he did not feel ashamed.

I agree with the authors that we often criticize how people feel, and that this sometimes involves appealing to reasons there are to feel differently. I agree, too, that sometimes these reasons have nothing to do with what the situation calls for, but instead speak to the consequences of feeling one way or another – often, the moral consequences. But I find their interpretation of this kind of conversation unconvincing.

Beth does not need to be understood as offering "the wrong kind of reason." She can instead be interpreted as presenting an overall (and familiar) ethical picture that emphasizes the heightened importance (to one's self-conception) of only the things that one is responsible for. As I have argued elsewhere, this kind of claim does not need to be understood as a matter of committing what D&J call "the moralistic fallacy": citing a reason that is irrelevant to fit because a matter of consequence, or the violation of a duty.[4] She can instead be understood as saying, roughly: a person with the right kind of values (i.e. whose self-esteem depends only on what they are

---

[3] I do not mean in the strict Moorean sense, but the wider sense that has come into usage and which includes attempts to "derive" evaluative facts from "natural" ones. For discussion see (Williams 1985).

[4] (Yao 2023). There, D&J interpret (I'd argue misinterpret) certain remarks by philosophers such as Gabrielle Taylor, Richard Moran, and Ronald de Sousa as evidence that they are committing "the moralistic fallacy".

responsible for) would see the world accurately and so, not feel ashamed of what they cannot help. She may be wrong about this, but she is not giving the wrong kind of reason.

But this, by the author's lights, is not a possible interpretation, and so we are unnaturally constricted to theirs. There are other places in *Rational Sentimentalism* where this constriction expresses itself: in, for example, their suggestion that the best explanation of why some members of the Donner party refrained from cannibalism or eating their dogs is simply that they experienced these things as "disgusting", where it is "not plausible" that this was a moral reaction (D'Arms and Jacobson 2023, p. 7).[5] Thus, it is not clear that the view in fact refers to an actual aspect of our practices, rather than an interpretation of those practices that rings false because described through the narrow range of distinctions provided by, and needed by, their theory.

Second Objection. Rational Sentimentalism must disentangle ethical claims from non-ethical fact, in order to use the latter to vindicate the former. D&J do so by distinguishing between three different kinds of claim about the fittingness conditions of the emotions: empirical, interpretive, and ethical. The first relies on empirical research to both identify a set of emotions that can be understood as "natural kinds" (thus, forming the "scientific", or naturalistic basis of the view), noting the common elicitors, phenomenology and motivations of the members of this set. We then interpret what the fittingness conditions (or "appraisals") of each natural emotion are – for example, that anger is appropriately responsive to *offences* – using and unifying this data with literary and philosophical accounts of the emotions. It is then up to first-order ethical thinkers to determine what really counts as an offence.

But to settle between competing views of what anger's appraisal is – that it is appropriately responsive to offences rather than, for example, *wrongdoing,* or *being obstructed by another person,* or *under their power* – one must take an ethical stance when ruling out other possible appraisals (if intelligible) as nonetheless unfitting.[6] And this reliance on ethical values seems the only plausible way to go, given their own proposed methods. Consider their suggestion that our best interpretations will take into consideration literary and philosophical accounts of the emotions. But presumably these observations are themselves conditioned by the evaluative beliefs and commitments of those philosophers and writers, and it is through at least some of those commitments that the empirical data will be unified and made sense of. While D&J are right that many moral philosophers offer hyper-moralized accounts of the emotions' fittingness conditions that strain credulity and should be rejected, this is because they are *bad interpretations;* they are bad partly because they rely on a

---

[5] The authors stipulate that what they mean by "moral" is what moral philosophers tend to think of as the "moral": which focuses on right and wrong and "things that agents can control and potentially correct" (D'Arms and Jacobson 2023, p. 6). They associate this understanding of the moral with Bernard Williams, but do not also address Williams' arguments that what moral philosophers tend to think of as the moral is a moralistic *distortion* of actual moral life, which includes sharply distinguishing between the moral and the non-moral (as D&J themselves encourage us to do).

[6] See also (Achs, forthcoming).

hyper-moralized ethical stance – not because they rely upon evaluative assessments and commitments at all.

Third Objection. The non-ethical foundation of RS is our "natural emotions". These are universal among "normal" human beings, difficult to change or extirpate, resist enculturation, are psychological "natural kinds", and are essentially motivational. It is the sentimental values paired with just these emotions that can be vindicated by RS. But what is the relationship between this non-ethical foundation, and ethics, such that the former has authority over the latter? The view is said to *constrain* moral theories. The argument for this is that the natural emotions and their concerns are deeply entrenched into our psychologies, and the claim that: "When a concern can only be eliminated through such extraordinary processes [such as totalitarian indoctrination], that supports the claim that it provides genuine reasons for action" (50). And so, any moral theory must recognize that all the natural emotions are sometimes fitting and, in those cases, give us reason to act. This move from what is natural for us to care about and do, to what we have reason to care about and do, seems to commit the naturalistic fallacy.[7]

Fourth Objection. What about emotions which do not make the provisional list of "natural emotions", but which seem to have a robust life within our actual practices, or some other actual cultural practice? Consider gratitude, love, and hope, which happen to not show up on the list given (which is "hostage to empirical fortune" (D'Arms and Jacobson 2023, p. 19)). Or consider emotions that are more culturally bound, such as the form of guilt, mixed with fear, that perhaps only those raised in the shadow of Judeo-Christianity might characteristically experience, or the form of shame, mixed with anger, that perhaps only those raised in the shadow of Confucianism might characteristically experience. Another might be grief: a species of sorrow, but with much more specific cognitive content and so, more specific fittingness conditions. And another might be nostalgia: an emotion that perhaps *cannot* make the list because it is not essentially motivational. What does the view conclude about all of these non- "natural" emotions (where the relevant contrast here is *cultural* and not "supernatural"), the properties they ascribe, and their fittingness conditions?

We can extrapolate from one of D&J's objections to cognitivism about the emotions: that it "encourages philosophers to subdivide emotions with philosophically interesting distinctions, especially morally significant ones. It thereby proliferates emotions with *fictitious* subdivisions and sham pretense of accuracy" (D'Arms and Jacobson 2023, p. 98, my emphasis). Moreover, it undermines the ambition to "carve emotions at their psychological joints" (D'Arms and Jacobson 2023, p. 98) – an ambition satisfied by understanding the emotions as natural kinds: "…whose members share properties that cluster together contingently but non-accidentally, in ways that enable inductive generalizations within some branch of science." (117). Thus, "non-fictitious", "real" joints are those that will emerge given the needs of some branch of science. Accepting the body of science D&J rely upon (that is, the work of empirical and evolutionary psychologists such as Nico Frijda, Robert Plutchik

---

[7] I'll consider an alternative interpretation of this argument in my final section.

and Robert Zajonc, and neuroscientists such as Joseph LeDoux (D'Arms 2005)), we should conclude, for example, that there is "really" only *one* emotion here: *anger* felt fittingly at *offences* – no indignation, frustration, resentment, *ressentiment,* etc.

But cognitivism will not proliferate types of emotion *arbitrarily:* it will do so in a way that relies on background values which give the emotion in question its evaluative point.[8] Given a set of values, there may be a world of ethical difference between being resentful of someone and being cross with them, or feeling guilty and feeling contrite – this is not a "sham pretense" of accuracy. These groupings and "sharpenings" have a point, and do not exist because of some need in a branch of science to group them together, but because of certain ethical needs, embodied within a culture that encourages its inhabitants to learn to group and split their emotions in these ways. Why should the explanatory needs of a science *displace* these ethical needs? How could they even be *relevant* as reasons to refrain from, for example, encouraging a child to feel grateful for the good weather on her birthday, or persuading a friend that the circumstances do not yet call for despair? We are not here given a satisfying answer. If it is because the sentimental values that are paired with emotions that are not "natural", are thereby not "real"*,* RS is relying on an underlying metaphysical picture that without explicit defense seems reductively scientist. And even if we conclude that such emotions are in *that* sense not "real", this does not seem to be a relevant consideration for us to dislodge them from our practices.

3 Why these objections?

The objections I've raised are versions of familiar responses to any meta-ethical theory that shares the basic ambition of RS to "vindicate" something ethical with resources that are not: first, the ethical concepts under investigation do not mean what the theorists claim they mean; second, the non-ethical foundation used by the view is in fact entangled with ethics; third, the view commits the naturalistic fallacy; fourth, in so far as the view is relevant to our practices and first-order ethical thought at all, it calls for a revision of our practices and thinking without offering an adequate justification.

But there is a principled reason why a view of this kind will encounter these objections. This is the plausible idea that one cannot grasp the meaning of a concept (such as "shameful") without being inside of, and having at least partly internalized, the values and practices that give the concept its evaluative point.[9]

---

[8] This form of cognitivism does not need to be, and would in fact be principally against, the possibility of characterizing the fittingness conditions of an emotion in a "response-independent" way, which D&J attribute to their "cognitivist" opponents. Their taxonomy (I think unhelpfully) renders McDowell, Wiggins and Williams as not "cognitivists", in this sense.

[9] Hence, McDowell's idea that only the virtuous agent who has *perfectly* internalized those values could really know what its central concepts truly mean. But the basic point can be understood as having more general implications than that, and in its most basic form is Wittgensteinian. There are different ways this basic claim can be understood and developed without McDowell's elaboration of it; I will only discuss McDowell's handling of the claim here.

One way to put the challenge that arises from this idea is that when one attempts to supply a standard or justification for our ethical practices or some part of them that is fully external to them, one will be unable to get an adequate fix on what they are actually like. The risk is then that one will not be able to successfully refer to, and so provide a vindication for, some part of our *actual practice*: hence the first objection. And because the meaning of a concept depends on these values and practices, one *cannot* disentangle its meaning from those values and practices. This is why in order to arrive at a plausible interpretation of the appraisals of the emotions and to argue for them in favor of alternative appraisals, D&J must rely upon their own background evaluative commitments – hence the second objection.

To show why the third and fourth objections are also inevitable if one grants this claim about meaning, I will have to develop it. As mentioned, D&J note that despite their differences, Blackburn and McDowell agree that truth in ethics must be "earned". That is, we cannot simply take for granted that ethical truth consists in how it simply appears to us, and intuitionism (though perhaps now the dominant and largely implicit method of moral philosophers) supplies us with a bogus epistemology.

But what cannot be overlooked is how they conceive of what it would mean to "earn ethical truth". McDowell, making use of the initial claim about meaning, contextualizes the question of "earning truth" by imagining an actual situation and problem that a person could face: she is within a culture where *all* seemingly ethical and moral claims are really nothing other than attempts at manipulation, where there is no other ethical language available. All she can be interpreted as doing in such circumstances (even by herself) is attempting to exert her will because the existing ethical concepts available are too "sparse and crude" to mark a distinction between doing that, and anything else (McDowell, 1998, p. 156, fn. 12). In this context, she is bereft of is an ability to *justify* a moral or ethical claim; what is not relevant is that she lacks an adequate *metaphysical* basis for them. Rather, she lacks a culture with living concepts and practices of moral reasoning and rational persuasion. What is needed to "earn ethical truth" in this sense – and which philosophers could not do on their own – is what Alasdair McIntyre calls "conceptual reform" (McDowell, 1998, p. 156, fn. 12).

Thus, there are at least two different ways in which we can understand the challenge that ethical or moral truth must be "earned": one sense relies on being able to justify oneself with a reason, relying and developing the actual ethical concepts available; the other relies on the possibility of a "metaphysical move" that attempt to answer a quasi-empiricist understanding of truth (McDowell, 1998, p. 156). If one is trying to answer the first kind of challenge, an answer to the second will be beside the point. Moreover, the idea that the second *could* be given the kind of answer it seeks is itself questionable: the answer would have to isolate a set of concepts from what gives them meaning and expect that one will still be able to successfully use them to refer to what was under investigation in the first place. But again, given the initial claim about meaning, this move is incoherent. Hence, McDowell's conclusion that when one searches for a justification for the use of an ethical concept, the only

place one *can* look is, if they are available, to more ethical concepts: not to something outside of them entirely.[10]

Even if one does not accept this conclusion, we should not run these two senses of "earning truth" together and expect a metaphysical "vindication" to be relevant to a question of how to "earn ethical truth" by being able to justify oneself. This distinction between two senses of "earning truth" explains why the principle RS uses to bridge the fact-value divide seems so implausible: if someone is sincerely wondering whether they are *justified* in caring about something, it is almost never a good reason that they already do and that they can't help it. And at the same time, it explains why it seems wholly irrelevant, if one is wondering whether her hope in the future is justified, whether hope is a natural emotion.

In fact, I think this distinction is implicitly recognized by D&J. After describing how they will use our natural emotions to vindicate their paired sentimental values, they write, "It is notoriously difficult to defend fundamental evaluative claims in the abstract, and our defense of the significance of psychological depth and width relies in part on its plausibility in various specific cases addressed over the course of this book" (D'Arms and Jacobson 2023, p. 50). These cases involve highlighting what we would give up, ethically speaking and in terms of our own understanding of ourselves, if we were to accept the positions of their more rationalistic, scientistic, and moralistic opponents. The fact that D&J rely upon commitments they think we currently have and are in good enough order to justify (hold up, illuminate, or defend) other ethical values and concepts is not itself suspicious. It is the pre-tense of doing something *else* we should be suspicious of.

4 "Anthropocentricity" and "scientism" in ethics

As mentioned, while McDowell relies on and expands upon this basic point about meaning, it is not unique to him. So, the disagreement I've characterized is not best understood as one between his form of neo-sentimentalism, and the form defended by D&J. It is better understood as a deeper one: between those who think we can unproblematically extract some set of concepts we use in our ethical practices from whatever else gives them meaning and still succesfully refer to them in our theorizing, and those who think we cannot. For those who think we cannot, there is a straightforward explanation for why RS, and any view with its basic meta-ethical ambition, will face forms of the four objections I raised. We also arrive at an explanation for why we cannot extract some set of McDowell's commitments from his "semantic and metaphysical" ones, and still retain *their* meaning. While it seems to D&J that McDowell agrees with them about certain basic points – such as the need to "earn truth" in ethics, the need for a concept of emotional "appropriateness", and

---

[10] (D'Arms and Jacobson 2006) address and (quickly) reject McDowell's *diagnosis* of the search for such an external foundation as a symptom of modernity, but not his reasoning for it.

an ambition to provide a "anthropocentric" conception of value that avoids forms of "scientism" – he does not, with these terms, mean what they mean.

Consider now these last two ambitions. Recall that RS seems to commit the naturalistic fallacy given its claim that, "When a concern can only be eliminated through such extraordinary processes, that supports the claim that it provides genuine reasons for action" (D'Arms and Jacobson 2023, p. 50). But on another reading of the view, there is no attempt to "derive" an "ought" from an "is", or reason from non-evaluative fact. Instead, the argument is that there are certain constraints on human nature that make it extremely difficult if not impossible for us to change our ways. If a moral theory strives to be something that it is possible for human beings at large to live up to and internalize (at least stably, and without severe repression), then the fact that we are limited by our natures in some way is something that such views should pay attention to. This is, I think, a perfectly good ambition to have, and an equally good constraint to observe.

But if this how to understand the argument, and to avoid the other objections I've raised, the investigation into human nature would have to be very different from the kind envisioned by the authors here, which relies on the emotions as understood by empirical and evolutionary psychologists, and neuroscientists. It would have to take seriously the details of actual human culture, and the ways that our emotions have expressed themselves within those cultures presently and historically.[11] How our emotions become – including what they come to be *about* – under conditions of "normal" enculturation, will be the evidence of what they are able to become. And it would have to involve investigating cultures where a norm against feeling a natural emotion or acting on it was enforced, and seeing whether it is true that this norm could not really be internalized by people (or enough people) living within those cultures. This endeavor would have no metaphysical ambitions and could be done fully in the spirit of the basic point about meaning that I've restated from McDowell. In other words, the "anthropocentricity" of this approach would be more thoroughly humanistic, rather than covertly scientistic.[12]

---

[11] Consider the alternative approach in characterizing human nature taken by Martha Nussbaum (Nussbaum 1995), which she explicitly describes as not a metaphysical account, and as thoroughly normative.

[12] I am grateful to Samuel Reis-Dennis, Selim Berker, Oded Na'aman, and Ram Neta for comments and discussion.

References

Achs, R. (forthcoming).

D'Arms, J. (2000). Sentiment and value. *Ethics, 110*(4), 722–748. https://doi.org/10.1086/233304

D'Arms, J. (2005). Two arguments for sentimentalism. *Philosophical. Issues, 15*, 128–150. https://doi.org/10.1111/j.1533-6077.2005.00012.x

D'Arms, J., & Jacobson, D. (2006). Sensibility theory and projectivism. In D. Copp (Ed.), *The Oxford handbook of ethical theory* (pp. 186–218). Oxford University Press.

D'Arms, J., & Jacobson, D. (2023). *Rational sentimentalism*. Oxford University Press.

Jacobson, D. (2005). Seeing by feeling: Virtues, skills, and moral perception. *Ethical Theory and Moral Practice, 8*(4), 387–409. https://doi.org/10.1007/s10677-005-8138-3

McDowell, J. (1998). *Mind, value, and reality*. Harvard University Press.

Nussbaum, M. (1995). Human capabilities, female human beings. In M. Nussbaum & J. Glover (Eds.), *Women, culture, and development: a study of human capabilities* (pp. 61–104). Oxford University Press.

Williams, B. (1985). *Ethics and the limits of philosophy*. Harvard University Press.

Yao, V. (2023). The good fit. *Philosophy and Phenomenological Research, 108*(2), 414–429. https://doi.org/10.1111/phpr.12901