




The alchemists: on Justin D'Arms and Daniel Jacobson's *rational sentimentalism*

Oded Na'aman¹ 

Accepted: 6 September 2024
© The Author(s) 2024

Abstract

D'Arms and Jacobson's *Rational Sentimentalism* promises an alchemy: a view that grounds certain values and reasons in facts about human sentiments but also treats the very same facts about values and reasons as fundamental. I examine how they attempt to deliver on the promise, doubt that they succeed, consider their motivations, and offer an alternative interpretation of what they might be doing.

Keywords Rational sentimentalism · Fittingness · Reasons · Values · Sentiments

What? Rational sentimentalism? How could that be? The name of the book and of the view it defends might seem obscurely technical to the uninitiated, but to the initiated it is a bold promise. A provocation even. How can a sentimentalist view, which grounds values and reasons in facts about human sentiments, also be a rationalist view, which treats facts about values and reasons as fundamental? Justin D'Arms and Daniel Jacobson believe they can have it both ways. I am not sure they can but I will not spend too much time arguing the point. Rather, I wish to speculate about why they aspire to perform this alchemy and to offer an alternative approach that requires no philosophical wizardry. *Rational Sentimentalism* promises an alchemy: a view that grounds certain values and reasons in facts about human sentiments but also treats the very same facts about values and reasons as fundamental. I examine how they attempt to deliver on the promise, doubt that they succeed, consider their motivations, and offer an alternative interpretation of what they might be doing.

I am grateful to Rachel Achs and Vida Yao for insightful written comments on previous drafts of this piece; to an anonymous referee with a sharp eye; and to Justin D'Arms and Dan Jacobson for their work, for their openness and generosity, and for multiple discussions of their powerful ideas.

✉ Oded Na'aman
Oded.naaman@mail.huji.ac.il

¹ Hebrew University of Jerusalem, Jerusalem, Israel

Even the initiated might gloss over the provocation. D'Arms and Jacobson have already introduced various parts of rational sentimentalism in several of their papers, so the view might seem familiar and unsurprising. But this is rational sentimentalism's first comprehensive presentation and as such it invites a more sustained engagement with its striking ambition. The book is also a reminder that D'Arms and Jacobson's philosophical concerns and commitments differ importantly from those of the many philosophers who have been influenced by their work (myself included.) Their joint papers since the early 2000s have had a significant and lasting impact on contemporary metaethics and normative theory. First, they convinced many that there are distinctive normative standards for emotions, not to be conflated with considerations of morality and prudence, and best captured by the normative notion of *fittingness* (for instance, in D'Arms & Jacobson, 2000a). Second, by explaining values in terms of fitting emotions, they gave powerful tools to fitting attitude accounts of value ('FA accounts'), according to which to be valuable is to be the fitting object of a certain type of attitude (for instance, in D'Arms & Jacobson, 2000b). But due to its focus and scope, their view, rational sentimentalism (RS, for short), is unlike any of the views they have influenced.

1 RS and its motivations

RS is given various articulations throughout the book, with no single definitive statement, but as I understand it, RS includes three general theses. First, that there exists a distinctive group of values, which D'Arms and Jacobson call *sentimental values*. These include: the shameful, the disgusting, the regrettable, the fearsome, the amusing, the pride-worthy, the anger-worthy, and many more. Second, that these sentimental values are to be analyzed in terms of facts about fitting emotions, such that for an object to have one of these values (e.g., for a joke to be amusing) is for it to *merit* or make *fitting* a certain type of emotional response (e.g., amusement). Third, that the existence of these sentimental values is explained by psychological facts about natural human emotions: shame, disgust, regret, fear, amusement, pride, anger, etc.

The view is unusual for a metaethical theory. Though it has implications for general discussions of morality and value, it is concerned, primarily, with establishing and "vindicating" (D'Arms and Jacobson 2023, 50) a certain *subset* of values. Another noteworthy aspect of the theory is that it makes claims at three different metaphysical levels. A first-order normative claim about the existence of sentimental values; a metaethical claim about their place in the normative domain; and a further metaethical claim about the grounding of these values in natural facts about human emotion.

It is worth considering what led D'Arms and Jacobson to this combination of commitments. Why does RS have the peculiar shape it has: narrow in scope but heavy in metaphysics? What do D'Arms and Jacobson fundamentally care about such that they are driven to argue not only that certain sentimental values exist and are explained by fitting emotions, but that they are further explained by human psychology?

As I noted, the combination of commitments RS takes on is, at least initially, puzzling. What makes the view a *sentimentalist* view, according to D'Arms and Jacobson, is that according to it "certain values are to be explained by way of the

emotions” (13). What makes the view a *rational* form of sentimentalism, is that it analyzes certain values in terms of *fitting* emotions, which entail reason for emotion. Thus, while a dispositionalist version of sentimentalism may hold that to be funny is *to cause amusement*, a rationalist version of sentimentalism holds that to be funny is *to merit amusement* (or make amusement *fitting*). The label ‘rational’ is earned by the normative content in the explanans in conjunction with the assumption that fit is related to reasons and rationality. However, D’Arms and Jacobson also claim that the sentimental values are grounded in the nature of human emotions. Human psychology explains why there are values such as the amusing, the shameful, etc., and therefore it also ultimately explains facts about fitting amusement, fitting shame, etc.

Given the grounding of sentimental values in natural emotions, it may now seem that the rational aspect of RS is superficial. While dispositional sentimentalism explains the amusing in terms of causing amusement, rational sentimentalism explains the amusing in terms of meriting amusement and then proceeds to explain both in terms of human tendencies and inclinations. Rational sentimentalism therefore appears to be *dispositional sentimentalism in rational disguise*. If RS is meant to distinguish itself from dispositional sentimentalism, why does it take the further step of giving non-normative explanations of normative facts? Why not rest content with the fitting emotion analysis of sentimental values?

The answer, I think, is that D’Arms and Jacobson believe only a sentimentalist view can explain sentimental values and worry that if they do not explain facts about fitting emotions their view would be too rationalist and insufficiently sentimentalist. For example, while considering a possible objection that points to divergences between fitting and actual emotional reactions, D’Arms and Jacobson write:

If neo-sentimentalism deploys the merit schema in ways that allow sentimental value judgments to become too disconnected from actual responses, it loses its connection to the core sentimentalist idea that values depend on emotions. Any adequate view about what merits a natural emotion must be constrained by the nature of that emotion and the range of sensibilities humans can internalize. (28)

In a moment we will look more closely at how, according to D’Arms and Jacobson, sentimental values depend on the nature of emotions and the range of sensibilities humans can internalize and how their account of these dependence relations is meant to leave room for a genuine rationalism alongside sentimentalism. Before we do, however, I want to consider why they are attracted to sentimentalism to begin with. Why are they committed to the “core sentimentalist idea that values depend on emotions,” the idea in virtue of which they must explain facts about fitting emotions by facts about actual emotions?

They never say so explicitly, but D’Arms and Jacobson seem to believe that since sentimental values are *anthropocentric* — i.e., they are values *for* humans and ascribe reasons to humans and not to all rational beings — these values must be explained by facts about human emotions. It is implausible that the amusing and the shameful, for instance, are values that entail reasons for non-human beings or for any rational agent. Why would anything be disgusting or fearsome for a rational being that has

none of our emotional machinery? On the other hand, if something is disgusting or fearsome *for us* this must be because these values are somehow shaped by our disgust and fear. A related, internalist thought is that such values can only ‘get a grip’ on our emotions if they are dependent on them. This line of thought seems to lead D’Arms and Jacobson to the conclusion that to defend sentimental values as genuine values they must show that these values are couched in “deep and wide” human concerns (50).

So far, I find myself in agreement with this reasoning. My disagreement begins with their claims about *how* sentimental values are grounded in deep and wide human concerns.

2 Into the deep and wide

According to D’Arms and Jacobson, in order to argue that a certain type of emotional response is sometimes fitting one must show that “humans have *deep* and *wide* concern” for the associated sentimental value (50). To show that anger can be fitting, for instance, it must be shown that people have deep and wide concern “for what people do and why they do it” (ibid.).

The “depth” of a concern is understood counterfactually: a deep concern could be changed only through extremely oppressive and manipulative means, if at all. The “width” of a concern is understood by its connection with other concerns and inclinations: changing it would involve a global change in our psychological profile. D’Arms and Jacobson claim that, “the conjunction of these two features can vindicate sentimental values as the source of anthropocentric reasons” (ibid.).

On their view, when emotional inclinations are couched in deep and wide human concerns, they “give rise to” genuine reasons. It is not clear how they understand this relation. They claim that the fact that changing a deep concern would be very oppressive “supports the claim that it provides genuine reasons for action” (ibid.) This suggests that the depth and width of a concern are *evidence* that the concern provides genuine reasons, so the relation between the psychological facts and the normative facts is epistemic: the former is evidence of the latter. But this is surely not what D’Arms and Jacobson intend, because the issue at hand is the metaphysical dependence of sentimental values and associated reasons on human psychology.

In other places (e.g., 14, 28) D’Arms and Jacobson write that sentimental values and facts about fit are “constrained” by human psychology. This makes it sound as if human psychology does not *determine* sentimental values but merely sets limits on any plausible theory of sentimental values. But D’Arms and Jacobson also speak of “grounding” sentimental values in human psychology (13), or of human psychology “giving rise” to sentimental values (50). And this suggests they believe deep and wide concerns *determine* sentimental values and associated reasons. It is because a concern is deep and wide that the significance it attributes to its object is genuine. Thus, the idea that deep and wide concerns ground values goes along with their sentimentalist commitment and the idea that such concerns merely constrain values leaves room for their rationalist commitment. But how can both be true at once? How

can deep and wide concerns *merely constrain* sentimental values but *also determine* them?

Before offering my interpretation of how D'Arms and Jacobson intend to resolve this problem, let me note an instance of it. After making the point that on their view human psychology gives rise to anthropocentric reasons (50–51), D'Arms and Jacobson go on to explain that their view does not succumb to the relativism of other sentimentalist views, which reduce reasons and moral imperatives to non-normative facts about emotional reactions, or approval and disapproval (51–55). On these relativists' views, what is right in a given culture or society is what is approved in that society. By contrast, RS denies that conformity justifies. Still, D'Arms and Jacobson acknowledge, one might worry that since they argue “that human patterns of concern can provide genuine reasons for creatures like us” (55) they rely on conformity to tendencies that are species-wide and therefore succumb to “relativism-to-humanity.” In response, they insist that relativism-to-humanity does not follow from their view: “not everything that matters universally to human beings is a source of reasons, because some human tendencies are undermined by critical reflection” (55). This response is puzzling: how can they earn the appeal to critical reflection given their commitment to grounding sentimental values in universal human concerns?

An answer becomes apparent later in the book, when they explain in more detail how we should extrapolate values from facts about human emotions (142–143). They distinguish between three kinds of claims. First, there are *empirical claims* about the characteristic features of natural emotion kinds. Such are claims about conditions in which an emotion of a certain type is elicited, about when it is satisfied, about the phenomenology of the type of emotion, and about its motivational role. These claims, they say, are based on common descriptions of emotions as well as the characterizations of psychologists, novelists, and philosophers.

Second, there are *interpretive claims*, which offer a gloss of how someone in the grip of a natural emotion is best understood as appraising its object. It is because of what elicits fear and what satisfies it, what fear feels like, as well as the fact that to be afraid of something issues in direct and urgent motivation to avoid it, that it makes sense to gloss fear's appraisal of its object in terms of danger. Moreover, D'Arms and Jacobson claim — and this is crucial — that interpretations of goal-directed states like emotions must be expressed partly in normative terms. An adequate gloss will reveal the specific way that the emotion evaluates its object and purports to illuminate reasons to act as it motivates.

For instance, D'Arms and Jacobson argue that pride appraises its object as *reflecting well on me*. This is an interpretation of empirical claims about pride that does not yet resolve the question of what merits pride because it leaves open the question: what reflects well on me? Nonetheless, D'Arms and Jacobson argue that “a crucial test of any gloss is whether people with disparate sensibilities can accept it as capturing in a rough and ready manner the way they take the object of an emotion—whether or not they deem it fitting” (142). A fan of a sports team and a philosopher might disagree about whether the fan's pride in his team is fitting, but “if *reflects well on me* is a good account of pride's appraisal, then ... [the] fan and philosopher should agree that the fan's pride appraises the triumph of his team as reflecting well on him. They understand pride's appraisal similarly despite differing about when it is correct”

(*ibid.*). So interpretive claims go *some* way toward answering the question of what merits pride but not all the way.

To fully answer the question of what is fitting one must resort to a further, third, kind of claims, namely, *evaluative claims*. These are claims about what satisfies the agreed upon emotional appraisal. This is where the philosopher and the sports fan disagree about the fittingness of pride. They agree that pride appraises an object as *reflects well one me* but disagree about whether the team's win reflects well on the fan.

This brings us closer to understanding how psychological facts both *ground* and *constrain* sentimental values according to RS, as well as a more detailed understanding of its alchemical combination of sentimentalism and rationalism. RS is sentimentalist in virtue of claiming that facts about human emotions determine the evaluative property an object must have if it is to merit them; RS is rationalist because it denies that facts about human emotions determine what instantiates the relevant evaluative property. So non-normative facts about pride determine that pride is fitting to what *reflects well on me*, but they do not determine *what* reflects well on me. What matters universally to humans (in a deep and wide way) is a source of reasons, but it does not fully determine the reasons it gives rise to. We might all share a misguided notion of what *reflects well on me*, and such a mistake might be discerned through "critical reflection."

This is my best explanation of how the view is meant to combine rationalism and sentimentalism. But I do not think it works. First, consider their proposed test for determining a good gloss of an emotion's appraisal. They claim that the sports fan and the philosopher, while disagreeing on whether the fan's pride is fitting, would nevertheless agree on how pride appraises its object. They claim such agreement indicates that this is pride's appraisal. But a few pages later, they appeal to their disagreement with an actual philosopher, Philippa Foot, who maintains that pride appraises its object as *splendid and mine* (146). Since they believe *reflects well on me* is a good gloss of how pride appraises its object, and they acknowledge that there is disagreement about whether this is a good gloss of pride's appraisal, it cannot be a test of a good gloss that people with disparate sensibilities about the fittingness of an emotion accept it.

Second, as some have already argued (Achs forthcoming, Yao forthcoming), there are reasons to doubt that plausible claims about the nature of an emotion and about its appraisal can be made independently of assumptions about the fittingness of the emotion. Our interpretation of what the emotion is and of what it is about are both shaped by our normative judgments about when the emotion is fitting. D'Arms and Jacobson's reasons for rejecting Foot's gloss of pride's appraisal draw on judgements about fitting pride, such as the judgment that a winning lottery ticket can be *splendid and mine* yet not merit my pride (93). Interpretation and evaluation cannot be clearly distinguished.

But even if we grant that claims about the nature of emotions and about their appraisal can be made independently of any normative judgments, I still find it implausible that such psychological facts can *ground*, or *give rise to*, or *determine*, sentimental values. All that non-normative facts about human emotions can demonstrate is that we tend to *treat* certain things as valuable in various way. Why should

the fact that we have a deep and wide tendency to find certain things valuable entail that these things are in fact valuable? This strikes me as an instance of the naturalistic fallacy.¹

However, there is another possibility. Human psychology can give rise to anthropocentric reasons and sentimental values given certain normative principles. For instance, there might be a normative principle according to which [if A has a deep and wide concern for X, then X is valuable for A and A has special reasons with regard to X]. This principle may explain why deep and wide human concerns give rise to sentimental values. However, it is unclear whether such a view counts as sentimentalist in the sense D'Arms and Jacobson intend. Its explanation of normative facts by non-normative facts presupposes further normative principles. But maybe that is not all that bad?

3 Why be fitting?

Compare D'Arms and Jacobson's project of vindicating sentimental values to the age-old project of vindicating morality. What explains the authority of the various principles we appeal to in order to tell right from wrong? Why do moral principles have the reason-giving force we tend to attribute to them? This question is sometimes put simply (and misleadingly) as: Why be moral?

One response, famously given by Prichard (1912), is that the question rests on a mistake. Prichard presents a dilemma. Either the answer to 'Why be moral?' appeals to a further moral reason, in which case it is uninformative; or the answer appeals to a non-moral reason, such as the reason that abiding by moral principles serves one's long term self-interest, in which case it gives a reason that a moral person would not be motivated by. Prichard concludes from this that the morality system cannot be explained. Instead, he endorses intuitionism, which maintains that the principles of morality are explanatorily fundamental.

Utilitarianism, on the other hand, explains moral principles not by a moral consideration but by the value they promote: moral principles of right and wrong are those that serve the greatest happiness. T. M. Scanlon argues that utilitarianism ends up falling onto the second horn of Prichard's dilemma (Scanlon, 1998, ch. 4). Scanlon's own view, contractualism, purports to explain moral principles by appealing to a non-moral reason that is nevertheless properly related to morality and unifies moral motivation. For Scanlon, this is the reason to stand in relations of mutual recognition to others, which means being justifiable to others, where to be justifiable means living according to principles for the general regulation of society that no one similarly motivated could reasonably reject (*ibid.*).

In response to Prichard's dilemma, Scanlon's view does not attempt to offer a naturalist vindication of morality, but a normative one. The normative vindication it

¹ I use 'naturalistic fallacy' in a broader sense than G. E. Moore's original explication of it. As an anonymous referee points out, according to Moore, the naturalistic fallacy is the (supposed) fallacy of trying to define goodness in terms of something natural. I mean it in the way that, I believe, it has come to be understood, as the (supposed) fallacy of trying to derive something normative from something natural.

offers aims to be external to the moral domain and yet recognizably related to morality. The unifying explanation of moral principles in terms of justifiability to others would also serve as a guide for revision of moral principles, in accordance with the method of reflective equilibrium (Scanlon, 2016, p. 19). Whether Scanlon succeeds in his mission is another question, but I am interested in his insightful strategy.

D'Arms and Jacobson have shown us that alongside our moral practice, which involves principles for telling right from wrong, we also engage in a practice of emotional evaluation and judgment, a practice with its own distinctive standards. They have convincingly demonstrated that norms of fitting emotions cannot be reduced to, or analyzed in terms of, other normative standards. So the question arises: what explains the authority of norms of fit for emotions? Why be fitting?

One response can be found in arguments for the fundamentality of fit (McHugh & Way, 2016, 2022; Howard, 2019). If fittingness is fundamental relative to the rest of the normative domain such that all other normative facts can be explained in terms of it (Howard, 2023), then there is no further normative explanation of fittingness norms. This view still leaves room for a unifying account according to which most fittingness norms are derivable from a small number of fittingness norms. However, it seems that normally those who subscribe to the view that fit is fundamental in the normative domain also maintain that there are multiple non-derivative fittingness norms. This approach is similar to intuitionism about the moral realm because it rejects the demand for further explanation of the normative domain in question. But just as intuitionism might seem unsatisfying since it posits various moral principles without any unifying structure, so might a view that accepts multiple norms of fit as normatively fundamental.

My conjecture is that D'Arms and Jacobson themselves find a disunified view of emotional norms unsatisfying. They therefore set out to offer a unifying explanation of fittingness norms for emotions. They hope to accomplish this by reaching outside the normative domain and tracing sentimental values and fittingness norms to human psychology. But, taking inspiration from Scanlon, there is, I think, a more promising strategy for the vindication of fittingness norms for emotions.

Scanlon seeks to explain the morality system, which is a sub-domain within the normative domain, by arguing that moral principles make possible the (non-moral) value of being justifiable to others. Since everyone has a strong and normally decisive reason to be justifiable to others, everyone has reason to be moral. Perhaps fittingness norms for emotions could also be explained by a compelling account of the value or values these norms make possible.

Take norms for fitting fear and norms for fitting anger. To assess and vindicate them we might reflect on *what they do for us*. When we are in danger, we are often at a loss; there are many plausible but incompatible reactions that must be adjudicated. We need guidance. Norms for fitting fear might be part of the guidance we need. A fitting response to danger includes, in addition to norms for fitting fear, norms for various other reactions, such as caution, resourcefulness, resoluteness, etc. So norms for fitting fear have authority for us because they are part of a broader response to danger that has authority for us and that addresses our concerns.

When we suffer an offence, we struggle to navigate the tension between recognition and resolution, between backward- and forward-looking concerns. Norms for

fitting anger might be part of the guidance we seek. A fitting response to offense includes, in addition to norms for fitting anger, norms for various other reactions, such as condemnation, blame, demand for restitution, etc. So norms for fitting anger have authority for us because they are part of a broader response to offense that has authority for us and that addresses our concerns.²

What is crucial is that the vindication of fittingness norms for emotions is not earned by appeal to human nature, but by appeal to concerns that we judge important and valid. The normative status of the concerns explains the authority of the fittingness norms that serve them. The normative status itself might be captured in terms of their value, or our reason for having them, or the fittingness of the concerns themselves. So the vindication of fittingness norms for emotions need not take a stand in the metaethical debate about the structure of the normative domain. Still, it offers a more unified account of the normative domain and a way to systematically reflect on the norms by which we evaluate, criticize, and endorse emotional reactions.

Finally, throughout the book D'Arms and Jacobson illuminatingly discuss various emotions and their standards of fit. They intend in these discussions to discern and demonstrate the deep and wide concerns human beings in fact have, but I believe the same discussions are better interpreted as discerning and demonstrating the deep and wide concerns we find important and compelling. Thus, large parts of the book may be read as a non-sentimentalist, normative vindication of sentimental values. The alchemists are, in fact, chemists.

Funding Open access funding provided by Hebrew University of Jerusalem.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Achs, R. (forthcoming). Review of Rational Sentimentalism. *Philosophical Review*.
- D'Arms, J., & Jacobson (2023). *Rational sentimentalism*. Oxford: Oxford University Press.
- D'Arms, J., & Jacobson, D. (2000a). The moralistic fallacy: On the 'appropriateness' of emotions. *Philosophy and Phenomenological Research*, 61, 65–90.
- D'Arms, J., & Jacobson, D. (2000b). Sentiment and value. *Ethics*, 110, 722–748.
- Howard, C. (2019). The fundamentality of fit. *Oxford Studies in Metaethics*, 14, 216–236.
- Howard, C. (2023). Fitting attitude theories of value. In Edward N. Zalta & Uri Nodelman (Ed.), *The Stanford Encyclopedia of Philosophy*.
- McHugh, C., & Way, J. (2016). Fittingness First. *Ethics*, 126, 575–606.

²More generally, the fittingness of an emotional reaction may be determined by the fittingness of the process of which it is a part (Na'aman, 2021). By looking at the fitting process and at the concerns it addresses, we can better assess fittingness norms for emotions that are part of the process.

-
- McHugh, C., & Way, J. (2022). *Getting things right: Reasons, value, and fittingness*. Oxford University Press.
- Na'aman, O. (2021). The rationality of emotional change: Toward a process view. *Nous*, 55, 245–269.
- Prichard, H. A. (1912). Does moral philosophy rest on a mistake? *Mind*, 21, 21–37.
- Scanlon, T. M. (1998). *What we owe to each other*. Harvard University Press.
- Scanlon, T. M. (2016). Contractualism and justification. *Reason, Justification, and Contractualism: Themes from Scanlon*. Markus Stepanians and Michael Frauchiger (Eds.). Berlin, Boston: De Gruyter, 2021, pp. 17–44.
- Yao, V. (forthcoming, this volume). Priorities: Comments on *Rational Sentimentalism* by Justin D'Arms and Daniel Jacobson. *Philosophical Studies*.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.