Fittingness and Idealization

Antti Kauppinen (a.kauppinen@gmail.com)

Trinity College Dublin and University of Jyväskylä

[T]here are some modern systems, according to which virtue consists in propriety; or in the suitableness of the affection from which we act, to the cause or object which excites it … None of those systems either give, or even pretend to give, any precise or distinct measure by which this fitness or propriety of affection can be ascertained or judged of. That precise and distinct measure can be found nowhere but in the sympathetic feelings of the impartial and well-informed spectator.[1]

1. Two Debates?

On the pages of *Ethics*, there are two separate long-running debates that would, I believe, benefit from being considered together. The first concerns the prospects of a Fitting Attitude (FA) or neo-sentimentalist analysis of value.[2] The second is more narrowly focused on the role of idealization in subjectivist or response-dependence accounts of normative and

---

[1] Adam Smith, *The Theory of Moral Sentiments*, ed. D. D. Raphael and A. L. Macfie (Oxford: Oxford University Press, 1976), 346.

[2] Justin D'Arms and Daniel Jacobson, "Sentiment and Value," *Ethics* 110 (2000): 722–48; Wlodek Rabinowicz and Toni Rønnow-Rasmussen, "The Strike of the Demon: On Fitting Pro-Attitudes and Value," *Ethics* 114 (2004): 391–423; Joshua Gert, "Color Constancy and the Color/Value Analogy," *Ethics* 121 (2010): 58–87; Mark Schroeder, "The Ubiquity of State-Given Reasons," *Ethics* 122 (2012): 457–88; Jonathan Way, "Transmission and the Wrong Kind of Reason," *Ethics* 122 (2012): 489–515.

evaluative properties.[3] In this note, I will propose that a satisfactory response to the idealization challenge for subjectivism simultaneously offers a promising solution to two problems for FA analyses. Roughly, as Adam Smith suggests in my epigraph, fitting attitudes are attitudes endorsed by an ideal subject.

According to a Fitting Attitude analysis, for someone to be admirable or blameworthy, for example, is nothing other than for admiration or blame to be a somehow appropriate attitude towards her, or for there to be sufficient reason to admire or blame. Sometimes, the term 'neo-sentimentalism' is used for the corresponding account of evaluative concepts rather than properties. This distinction will not be central for my purposes. The most notorious challenge to such accounts is the *conflation* or *wrong kind of reason problem*.[4] In general terms, it is that it may be appropriate to have attitude Y towards x without x being Y-able, and x may be Y-able even if having Y towards x is not appropriate. For example, it may be that it is appropriate for me to admire someone for reasons that have nothing to do with their being admirable. Maybe, in a bizarre enough circumstance, my sincere admiration of Kim Jong-un will save the lives of a hundred dissidents. It might then be appropriate for me to admire Kim Jong-un, even though he is not admirable. Conversely, it may be that someone is enviable even if it is not appropriate to envy her.

There are many attempts to solve this problem. Perhaps the most promising are those that appeal to standards of fittingness derived from the nature of attitudes themselves to

---

[3] David Enoch, "Why Idealize?" *Ethics* 115 (2005): 759 – 87; David Sobel, "Subjectivism and Idealization," *Ethics* 119 (2009): 336-352.

[4] D'Arms and Jacobson, "Sentiment and Value"; Rabinowicz and Ronnow-Rasmussen, "Strike of the Demon".

distinguish the right and wrong kind of reasons for the attitude.[5] But these solutions lead to a further problem that has not received enough attention. I'll call it the *many kinds of fittingness problem*. As I'll argue below, someone may be morally admirable or aesthetically admirable or even prudentially admirable, where this is not the same thing as it being morally or aesthetically or prudentially appropriate to admire her. This problem cannot be solved by appealing to standards of fittingness derived from the nature of the attitude, since the *same* attitude may be fitting or correct in *different* ways. I claim that there is no such thing as being admirable or enviable *tout court*, although contextual cues allow us to use the terms without further qualification in ordinary use. (Belief, in contrast, is fitting in just one way, and merits separate discussion, which I will not attempt here.)

If fittingness is determined by the attitudes of a morally or aesthetically or prudentially ideal subject, we have a ready explanation for why it comes in different varieties. One important challenge for such proposals is providing a rationale for appealing to idealized rather than actual responses in determining the extension of evaluative properties. Recently, David Enoch has made a strong case that idealization is either *ad hoc* or tacitly presupposes that there's a response-independent fact of the matter.[6] I'll argue that this challenge is met by the old-fashioned sentimentalist approach of David Hume and Adam Smith. These *paleo-sentimentalists*, to give them a label, begin with the idea that acting on immediate, uncorrected attitudes causes different practical problems. To solve them, we

---

[5] Justin D'Arms and Daniel Jacobson, "The Moralistic Fallacy: On the 'Appropriateness' of Emotions," *Philosophy and Phenomenological Research* 61 (2000): 65–90; Mark Schroeder, "Value and the Right Kind of Reason," *Oxford Studies in Metaethics, vol. 5*, ed. Russ Shafer-Landau (Oxford: Oxford University Press, 2010), 25–55 ; Christine Tappolet, "Values and Emotions: Neo-Sentimentalism's Prospects," in Carla Bagnoli (ed.), *Moral Emotions* (Oxford: Oxford University Press, 2010); Schroeder, "Ubiquity". See also Rabinowicz and Ronnow-Rasmussen, "Strike of the Demon", 421.

[6] Enoch, "Why Idealize?"

have to subject our unreflective responses to various kinds of discipline: to discount responses made in certain circumstances and to count responses made in others. When we treat responses of certain kinds of subjects in certain circumstances as authoritative in practice, we implicitly endorse certain perspectives as ideal or optimal for solving practical problems without making assumptions about response-independent truths.

It is important that there are many kinds of problem to solve, and consequently many ways in which a subject can be ideal. The moral point of view, perhaps, is that of an impartial spectator, and the aesthetic that of an ideal critic. Then, as Smith notes, the measure for the moral "fitness or propriety of affection" can be found "in the sympathetic feelings of the impartial and well-informed spectator". Someone is morally admirable if and only if admiring her is endorsed by any impartial spectator, merely as such. Similarly, for an attitude to be aesthetically fitting is for it to be endorsed by anyone occupying an aesthetically optimal point of view. The general recipe is that an attitude is W-ly fitting iff endorsed by anyone occupying the optimal W perspective, or any W-ly ideal subject. For any attitude Y, x is W-ly Y-able iff Y is W-ly fitting towards x. This approach, then, enables us to distinguish between fittingness and mere appropriateness while also allowing attitudes to be fitting in many ways.

2. The Wrong Kind of Reason and Many Kinds of Fittingness Problems

The wrong kind of reason (WKR) problem, again, is that it is possible for it to be appropriate to admire a person or inappropriate to be amused by a joke without the person being admirable or the joke being unfunny. Only some of the considerations that can make admiration appropriate bear on the person actually being admirable. We can stipulate that these *right kind of reasons* make the attitude *fitting* or *correct* rather than merely appropriate

or justified, but in the absence of a theory of what makes reasons of the right kind for all the different attitudes, this is a merely terminological move. The question is: what is the difference between fittingness and mere appropriateness?

One popular proposal has been to distinguish between reasons that derive from the properties of the object of the attitude (object-given reasons) and reasons that derive from having the attitude itself (state-given reasons), and then claiming that only the former are of the right kind.[7] Since Wlodek Rabinowicz and Toni Ronnow-Rasmussen convincingly argue that the wrong kind of problem arises for object-given reasons as well[8], and Mark Schroeder calls into question the importance of the distinction for identifying right and wrong kind of reasons[9], I will leave this proposal aside here, though it will in part inform my positive view. A related suggestion, wrong kind of reason scepticism, is that alleged wrong kind of reasons for attitudes do not exist; there are in fact simply right kind of reasons for wanting to have or bring about attitudes.[10] This solution still needs to make sense of the appropriateness of attitudes for which there are putative WKRs, and turns out to be compatible with the view of right kind of reasons defended here.

The most promising proposals for distinguishing the right kind of reasons appeal to what I will call *attitude-based fittingness*. On these views, we can derive a standard for what makes a reason a reason of the right kind for holding an attitude by reflecting on nature of the attitude itself. As Schroeder says, using the term 'activity' broadly to cover the formation of attitudes, "the class of right-kind reasons with respect to any activity will need to depend

---

[7] See Derek Parfit, *On What Matters* vol. 1 (Oxford: Oxford University Press, 2011).

[8] Rabinowicz and Ronnow-Rasmussen, "Strike of the Demon", 407.

[9] Schroeder, "Ubiquity".

[10] Way, "Transmission".

on the nature of that activity, in some way".[11] Justin D'Arms and Daniel Jacobson defend one version of this kind of view. They distinguish the right kind of reasons for attitudes on the basis of the way that different sentiments present things as being. The argument is simple. Assuming that an emotion has a cognitive component that presents or construes things as being in a certain way, it can be either accurate or inaccurate, depending on whether things are that way. The right kind of reasons for or against the attitude are those that bear on its accuracy. As they put it, "An emotional episode presents its object as having certain evaluative features; it is unfitting … when its object lacks those features."[12]

D'Arms and Jacobson use envy as their example. According to them, it, broadly speaking, presents a rival as having a desirable possession in a negative light. The right kind of reasons bearing on whether a rival is enviable are thus those that have to do with what she possesses and its desirability relative to what you have. If a colleague gets a promotion you wanted, envy may be fitting in this sense, and consequently, given a fitting attitudes account of evaluative properties, the colleague will be enviable. This is the case even if envy is morally or prudentially inappropriate – the latter kind of reasons do not bear on the fittingness of envy.

This type of account assumes that each emotion that is the building block of an evaluative property and thought has presentational content. D'Arms and Jacobson grant that it may be difficult to articulate and that no general recipe can be given.[13] The content of admiration, for example, is not easy to spell out. Maybe admiration presents its object as having done great things. It is hard to say much more than that. It is also not clear whether the model generalizes. Consider being desirable. Desire does not seem to have presentational

---

[11] Schroeder, "Ubiquity", 483.

[12] D'Arms and Jacobson, "Moralistic Fallacy", 73.

[13] D'Arms and Jacobson, "Moralistic Fallacy", 74.

content, unless it is that its object is *good*.[14] This would be problematically circular. For something to be desirable is for desire to be fitting, and what makes desire fitting, if anything, is that the object is good. This may be true, but endorsing it amounts to giving up the whole project of understanding evaluative properties or concepts in terms of fitting attitudes – in this case, it seems to be precisely being good (or desirable) that makes the attitude fitting.[15]

Mark Schroeder's proposal to solve the wrong kind of reason problem is based on a different notion of attitude-based fittingness. Schroeder's idea is that the right kind of reasons to engage in any activity A are those that are necessarily shared by anyone engaging in A, just because they are engaging in A.[16] Schroeder believes this generalizes for attitudes, *mutatis mutandis*. For admiration, he sketches two stories about the kind of reasons that anyone has merely in virtue of having the attitude of admiration. The first is based on the idea that admiration motivates emulation. This suggests that the standard of correctness from admiration derives from who one has reason to emulate. Anyone who engages in admiring has reason to admire only those they have reason to emulate. The right kind of reasons to admire derive from reasons for emulation.

Schroeder's second suggestion builds on the D'Arms and Jacobson proposal. Assuming that admiration as such presents the target as having a certain property, and that we have a standing reason to avoid false representations of any kind, anyone engaged in

---

[14] For recent defences, see Graham Oddie, *Value, Reality, and Desire* (Oxford: Oxford University Press, 2005) and Sergio Tenenbaum, *Appearances of the Good* (Cambridge: Cambridge University Press, 2007).

[15] On this point, see also Rabinowicz and Ronnow-Rasmussen, "Strike of the Demon", 421–3.

[16] Schroeder, "Value". Schroeder now says he is "agnostic" about this account and seems to find it incomplete ("Ubiquity", 485). Nevertheless, he doesn't offer a new systematic proposal, so the view discussed in the text is his most developed one. I return to his latest arguments in note 34.

admiring has reason not to admire someone who lacks that property, just because they are engaged in admiring. Thus, the right kind of reasons to admire are *alethic*, those that bear on the truth of the presentational component.

Schroeder's account gives the right answer in our test case. It is not the case that anyone who engages in admiration has reason to admire Kim Jong-un, just because she engages in admiration. To focus on Schroeder's distinctive proposal, this is because not everyone engaged in admiration has reason to *emulate* Kim Jong-un. So we get the right result: he is not admirable. Even I don't have reason to admire him just because I engage in admiration in general, but only because I can save a hundred lives by doing so.

Yet as I suggested above, these proposals face the underappreciated *many kinds of fittingness problem*. Consider a popular entertainer somewhat resembling Lady Gaga, say Lady Giga. Let us say that people who are into fashion widely admire her for her dress sense and inventiveness. If they're right, she is aesthetically admirable. Yet it's easy to imagine she's mean to her assistants and a bad influence on the kids, in which case she is not morally admirable. The *same* kind of attitude can be correct or fitting in many *different* ways. And there is no way that we can derive all these different standards of correctness from the very nature of the attitude itself. No matter how hard we stare at admiration itself, it will not tell us what the different grounds for moral as opposed to aesthetic admiration are. In principle, this problem could be resolved by adopting a more fine-grained view of either attitudes or their objects. I'll explain in a moment why both these moves are problematic.

But first, consider what the views just discussed would say about Lady Giga. Since D'Arms and Jacobson neither present a generalizable recipe nor discuss admiration, the implications of their view must remain speculative. But if the suggestion I made above is along the right lines, on their account, the right kind of reasons for admiration bear on

having done something great. I stipulated that Giga has indeed done something great in the way of the creative and the beautiful. So it is fitting to admire her. Yet it is also the case that she is not admirable, morally speaking. This is *not* accounted for by the view, since having done *something* great suffices to make admiration fitting. To be sure, D'Arms and Jacobson might propose a more narrow concern embedded in admiration. They might say that admiration presents its target as having done nothing but great things. But that would rule out the way in which Giga *is* admirable. It may, after all, be appropriate to admire in some way someone who is less than perfect in every respect – think of Picasso or Bill Clinton or any number of other people of high accomplishment and low fidelity.

Take Schroeder's view, then. His distinctive proposal appeals to reasons shared by all engaged in attitudinizing. Does anyone who engages in admiring have a reason to admire Giga just because she engages in admiring? Evidently not, if reasons for admiration derive from reasons for emulation. Not everyone who admires something has reason to emulate Giga. Only fashionistas do. So the account implies that she is not admirable. But she is – aesthetically so. Again, Schroeder might broaden the account and argue that everyone does have some reason – perhaps a very weak one – to emulate Giga, given that it would add some spice to their lives. But then the account would imply that she is admirable, and we want the result that she is morally un-admirable, given the way she treats her employees.

The many kinds of fittingness problem suggests that it is better to give up trying to make sense of admiration *tout court*. There are, in principle, two ways out for attitude-based fittingness theorists. The first is adopting a more fine-grained view of *attitude individuation*. Maybe there is a distinct attitude of *moral* admiration and *aesthetic* admiration and *prudential* admiration, and so on, and maybe each of these distinct attitudes somehow gives rise to distinct standards. As long as we confine ourselves to one type of attitude, this may have some plausibility, but we should recognize that the issue is perfectly general. People or

actions can also be morally or aesthetically contemptible, enviable, or preferable, among many other things. We are led to a vast proliferation of attitudes, without any systematic way of deriving standards of correctness for them. Other things being equal, this is best avoided.

The second way of evading the problem is adopting a more fine-grained view of the *object* of attitudes. Perhaps instead of Giga herself, what is admirable is her dress sense, while her employee policies are not.[17] Then maybe everyone has reason to emulate one and not the other, to take Schroeder's view. I grant that this move is possible at least in some cases, though attitudes often seem to take a person rather than some particular aspect as their object. (It seems possible to blame or indeed admire a person rather than some feature of hers.) But first, the problem arguably recurs even if we focus on a single aspect. Can't a certain kind of brutality be aesthetically admirable? Can't the very thing that makes something morally undesirable make it prudentially desirable? Second, we will still want to say that Giga's dress sense is aesthetically rather than morally admirable. So I don't think that finer grain of objects of attitudes will ultimately solve the many kinds of fittingness problem. We will have to look beyond the nature of attitudes themselves to understand correctness.

3. Why We Idealize

Can we make sense of the notion of fit or correctness in a way that is not subject to either the wrong kind of reason or many kinds of fittingness problems? I believe so. In this section, I will sketch a paleo-sentimentalist account of fittingness, which appeals to the attitudes of suitably idealized subjects. As David Enoch says, accounts of this type "tie the relevant

---

[17] This point was made by a referee for this journal.

normative fact … to hypothetical, idealized responses, or to the responses of hypothetical, idealized agents or thinkers, or to the responses of actual agents after going through some hypothetical, idealizing procedure."[18] What Enoch argues is that any such account lacks an adequate rationale for the appeal to idealization. It will not do for a subjectivist or sentimentalist to give what he calls the 'natural answer', which is that only idealized responses track the fact of the matter, since the motivation for such views is that the responses determine what the normative facts are.[19] It is *ad hoc* to appeal to getting the extension of normative properties right. Nor will actual justificatory practices support idealizing, since those practices, according to Enoch, themselves involve a commitment to response-independent facts. His claim is that "What best explains our justificatory practice is rather our (perhaps implicit) belief, false though it may be, that, say, conditions of full imaginative acquaintance are conducive to the reliable tracking of an independent order of value-facts."[20] So, Enoch concludes, "idealizers cannot motivate the idealization they employ in a way consistent with the considerations underlying their response-dependence views".[21]

How might subjectivists respond? David Sobel has offered a partial response to Enoch on behalf of full information accounts of well-being. Roughly, he argues that only desires based on a perfect forecast of what it's like to realize them are genuinely desires for the particular options, which accounts for their authority in determining prudential value.[22] It

---

[18] Enoch, "Why Idealize?", 760.

[19] Enoch, "Why Idealize?", 761–5.

[20] Enoch, "Why Idealize?", 774.

[21] Enoch, "Why Idealize?", 780.

[22] According to Sobel, "The rationale for granting the idealized agent information and experience is to provide her with a more accurate understanding of what the option she is considering would really be like. […]

is clear, however, that this reply has a limited scope. It provides a rationale for, at best, one aspect of an idealized subjectivist account of prudential value. There is no prospect of generalizing his proposal to Fitting Attitude analyses of moral or aesthetic value.

The paleo-sentimentalist response maintains that there are many evaluative points of view, each of which is the outcome of a way of idealizing our pre-reflective perspective on the world. Responses from an optimal evaluative perspective determine a standard for correctness of attitudes. Crucially for the idealization challenge, the rationale for appealing to ideal perspectives derives from *practical problems* with uncorrected attitudes. The starting point is that we have affective and conative responses to people and things, from admiration and envy to anger and disgust. But if we just go with our immediate, uncorrected responses, we are led to various kinds of trouble, by our own lights – outcomes we do not desire and are motivated to avoid. We face mutual recrimination, boredom, dissatisfaction, and unhappiness. For example, acting on just any desire may result in ill health, garish wallpaper, a broken marriage, or a distrustful community.

So, we have to step back from the immediate responses and scrutinize our attitudes in different ways to avoid these different problems. We do, in practice, defer to responses made by people who have certain characteristics and lack others, and occasionally aspire to acquire or shed such features ourselves. Our doing so reveals that we regard certain evaluative perspectives as authoritative when it comes to solving practical problems. An evaluative perspective is *ideal* or *optimal* when adopting it most reliably avoids the practical problems caused by the uncorrected attitudes.

[D]esires that do not involve this perfect forecast are, in a sense, not actually for the option as it is but rather for the option as it is falsely imagined to be." ('Subjectivism', 343, 345)

It is bound to be controversial what the particular problems with uncorrected responses are and how they are best solved, but let us make do with rough suggestions here. I will say that the function of moral evaluation is to guide action so as to robustly avoid interpersonal conflict between agents who have competing goals, and so to increase the desire-satisfaction of interacting individuals regardless of the content of their desires.[23] By 'robust' I mean that achieving the outcome is not accidental or contingent on some unusual circumstance obtaining. This is roughly Hume's position. He notes in the *Treatise* that if we allow our positive and negative attitudes to be guided by self-love and our naturally biased sympathy towards the near and the dear, we run into "perpetual contradictions" in society and conversation with those who do not share our position and interests.[24] Similarly, Adam Smith notes that unless we step out of our self-centered perspectives and let our sentiments resonate with each other, "[w]e become intolerable to one another. I can neither support your company, nor you mine".[25]

Let us say, very tentatively, that given such a problem to solve, the optimal moral point of view is that of any informed, impartial, and sympathetic spectator with otherwise normal emotional tendencies. When we defer to or take up such a perspective before we feel anger or guilt, or for that matter admire or laugh, it is no accident that we can all get along. That is something we very much care about, so it is no wonder that, as Hume puts it, "experience soon teaches us this method of correcting our sentiments"[26]. Much would need to be said to fill in the description, but let us assume for the sake of argument that it can be

---

[23] I owe this formulation to an anonymous referee for this journal.

[24] David Hume, *A Treatise of Human Nature*. Ed. L. A. Selby-Bigge, 2nd rev. edn., P. H. Nidditch (Oxford: Clarendon Press, 1978), 602.

[25] Smith, *Moral Sentiments*, 34.

[26] Hume, *Treatise*, 582.

done. Then for someone to be morally admirable is to be such that any impartial spectator, merely as such, would endorse admiration. For Helen to be morally *more admirable than* John is for it to be the case that any impartial spectator would endorse admiring Helen more than John.[27]

Our immediate and unreflective attitudes do not merely lead to choices that result in conflict with others. They also result in boredom and melancholy, and absence of inspiration and elevation. I will say that the function of aesthetic evaluation is to guide choices so as to robustly maximize pleasures of taste and imagination. Plausibly, the optimal aesthetic point of view is that of a good Humean critic – roughly, an experienced, discriminating, disinterested, and unprejudiced observer with otherwise normal hedonic tendencies. In Hume's slogan, what makes for a true critic is "[s]trong sense, united to delicate sentiment, improved by practice, perfected by comparison, and cleared of all prejudice".[28] Then for someone to be aesthetically admirable is for admiring her to be endorsed by any disinterested critic, merely as such.

Here it is possible, to be sure, that the verdicts of ideal critics systematically diverge due to a "a diversity in the internal frame or external situation", as Hume puts it.[29] At the margins, the young woman may be excited by what an old man finds annoying, even when both occupy an aesthetically ideal perspective. This means that there may be, in effect, a plurality of aesthetically optimal evaluative perspectives. Aesthetic evaluative predicates

---

[27] Schroeder, "Value", argues that since 'admirable' is a gradable adjective, the binary *more admirable than –* relation is explanatorily primary. The paleo-sentimentalist account can accommodate this as suggested in the main text. It can then deploy the same kind of semantics as Schroeder does.

[28] David Hume, "Of the Standard of Taste". In *Selected Essays*. Ed. Stephen Copley and Andrew Edgar (Oxford: Oxford University Press, 1993), 147.

[29] Hume, "Taste", 149.

will be tacitly indexical. This relativization in no way vitiates the point of stepping back from unreflective responses. There will still be a standard of correctness for each internal frame. And while relativization creates the danger of speaking past each other, there will be plenty of overlap, given that we're all human. Without knowing much about you, I can recommend a classic novel like *Crime and Punishment* to you, since I am confident that if you approach it with an open mind and a little patience, you will find reading it rewarding.

Finally, some problems with acting on unreflective attitudes arise for my long-term self-interest. I shoot myself in the foot, contract a painful disease, or pick a drink that doesn't alleviate my thirst. In general, I don't get what I want for myself. It is to solve a problem of this type that I have to stop to gather information, consider the effects of an action on my future desire-satisfaction, and aim for coherence in my pro-attitudes. Taken to the limit, such idealizing moves define what we might call the prudential point of view. Since my situation and desires differ from those of others, what is endorsed from the prudentially optimal perspective will be relative to each subject. That is why it is natural to talk of an *ideal advisor*'s endorsement as defining which attitudes are fitting for a subject to adopt, rather than of an ideal observer or critic.[30] Relativization is often explicit when it comes to the corresponding evaluative predicates – we easily say that something is desirable *for me*, but not for you. Once again, there will be plenty of overlap in some respects: given our shared physiology, a claim like "Drinking clear liquids is desirable when dehydrated" will be true at most indices.

These quick sketches of paleo-sentimentalist accounts for different kinds of idealization are not meant to be stand-alone theories of moral, aesthetic, or prudential

---

[30] Ideal advisor theories are proposed by Peter Railton, "Facts and Values," *Philosophical Topics* 14 (1986), 296 – 325, and Michael Smith, *The Moral Problem* (Oxford: Blackwell, 1994), among others.

evaluation, but just to demonstrate how such views are to be constructed within this framework. They should suffice to make it clear that the rationale for idealization, in each case, in no way assumes that there is a response-independent standard for correctness to aim for. Nor is there any *ad hoc* appeal to matching our intuitive verdicts. The paleo-sentimentalist view thus offers an alternative explanation of our actual justificatory practice. According to it, we don't, as a matter of fact, treat desires formed under conditions of full imaginative acquaintance as authoritative because we think they track some independent truth, but because, say, desires formed due to poor acquaintance lead to choices that do not satisfy for very long, if at all. However, the main issue isn't about historical explanation, but about the availability of a rationale for idealization that is consistent with subjectivist motivations. Starting from practical problems resulting from un-idealized responses and the subsequent need for a robust method of solving them provides just such a rationale.

One worry that someone might have concerns the individuation of the practical problems and their solutions. What kind of problem would it be if taking up the moral perspective were predictably boring in some contexts, for example, and what would it mean for the characterization of the morally ideal subject?[31] The paleo-sentimentalist line is that the demarcation of problems and solutions derives from the challenges we face in virtue of different aspects of the human condition, as it were – among other things, we are social beings who have to live together without pre-established harmony, goal-seeking agents without perfect foresight or self-control, and subjects whose experience is potentially enriched through imagination. It is an unfortunate fact that engaging in one kind of reflection to solve one sort of problem may itself give rise to a practical problem of a different sort. Still, boredom remains a broadly aesthetic issue. It's not the sort of problem that taking up

---

[31] This concern was raised by a referee for this journal.

the moral perspective is meant to solve, but rather a problem in the way of taking up the moral perspective.

Solving the Problems

How does appealing to idealization required to solve practical problems help with the many kinds of fittingness and wrong kind of reason problems? The answer to the first is evident. Lady Giga may be simultaneously such that any disinterested critic (of a certain type) would endorse admiring her and such that no impartial spectator would endorse admiring her. It is, then, aesthetically fitting (relative to a frame) and morally unfitting to admire her. The second issue is more complex. Is Kim Jong-un in my scenario morally admirable? Broadly, the paleo-sentimentalist says that there isn't sufficient moral reason (of the right kind) to admire him (and hence it isn't morally fitting to admire him), since not every impartial spectator, merely as such, would endorse admiring him.

But why not, if my admiring would save many lives? The answer turns on what it is to *endorse* an attitude, as opposed to merely approving of someone having it. I believe the best way to capture the relationship between ideal subjects' attitudes and our (right kind of) reasons is to say that for r to be a *pro tanto* (right kind of) W reason for S to have Y is for any W-ly ideal subject to take r to favour S's manifesting Y with success. (To avoid unnecessary controversy, I leave it open here whether the notion of favouring is reducible, although I do believe it is.) By 'manifesting' I mean doing or feeling whatever the attitude *essentially motivates* or *disposes* one to do or feel – for example, fear essentially motivates one to flee, shame to hide, guilt to make amends, admiration to emulate, intention to act, and desire to attend to and try to bring about its object. Motivating or disposing one to perform such actions is essential to the attitudes, because it is part of what makes them the attitudes they are. In fully endorsing an attitude, an ideal subject endorses manifesting it with success.

It is nevertheless possible to have the attitude without the manifestation – we can fear without fleeing and intend without acting. I'll say that an ideal subject may *approve of* someone's having an attitude without *endorsing* it, if merely having the attitude has beneficial consequences. I'll reserve the term '(merely) appropriate' for attitudes that are approved of but not endorsed (and hence not fitting for their objects).

How does this help to make sense of the distinction between fittingness and mere appropriateness? Let's say, for example, that Nelson Mandela's integrity is a moral reason for Joan to admire Mandela. For this to be the case is for any sympathetic impartial spectator to take Mandela's integrity to count in favour of Joan's manifesting admiration by emulating Mandela. (This obviously owes much to Schroeder's insight.) This may be a sufficiently strong reason to make admiration fitting. In the moral case, it arguably makes it fitting for *anyone* to admire Mandela. In contrast, it is not the case that any impartial spectator would take the fact that my admiring Kim Jong-un would save a hundred lives to favour my manifesting admiration by emulating him. It would be awful if I did the sort of things he does. Hence, there isn't sufficient (right kind of) moral reason for me to admire him, and he isn't morally admirable. Similarly, it is not the case that my ideal advisor would take the reward for *intending* to drink the toxin to favour my manifesting the intention by *drinking* the toxin in Kavka's famous puzzle[32], so there is no (right kind of) prudential reason for me to intend to drink it.

If there are wrong kinds of reasons, they are the considerations on account of which an ideal subject would approve of someone having an attitude without endorsing it. If sufficient, they make an attitude merely appropriate, not fitting. But the paleo-sentimentalist account is compatible with, and indeed provides some support for, *wrong kind of reason*

---

[32] Gregory Kavka, "The Toxin Puzzle," *Analysis* 43 (1983): 33–36.

scepticism, the view that putative wrong kind of reasons for attitudes are in fact simply reasons to *want* or to bring about an attitude. This is because it provides a natural account of why there is indeed reason to (only) want to have attitudes in certain situations. As I briefly noted, desires essentially manifest themselves (at least) in attending to and trying to bring about the states of affairs that are their objects. Consider what this means for the dictator case. Plausibly, any impartial spectator would take the fact that my admiring Kim Jong-un would save a hundred lives to favour my focusing on admiring him, and my trying to bring about my admiring him. Since these are essential manifestations of desire, it seems any morally ideal subject would endorse my *desiring* to admire Kim. Thus, there is moral reason for me to desire to admire Kim. Similarly, an ideal advisor with my best interests at heart would take the fact that I would get rich to favour attending to the intention to drink the toxin and to try to bring it about that I so intend, and thus endorse manifesting my desire to intend (in spite of not endorsing my manifesting my intention). If these reasons for desiring are sufficient, it is fitting to desire to admire or to intend. And since being W-ly desirable is just being a W-ly fitting object of desire, the account entails that my admiring Kim is morally desirable and my intending to drink the toxin is prudentially desirable.

This is exactly what WKR sceptics argue for, in Jonathan Way's case by appealing to what best explains patterns of transmission of reasons.[33] It is a point in favour of the paleo-sentimentalist account that it is compatible with this approach. The only cost of WKR scepticism that I can see is that we'll still need to make sense of it being somehow *morally appropriate* for me to admire Kim, in spite of there being *no reason* for me to admire him. It

---

[33] Although Way doesn't put it quite this way, the basis of his argument is the claim that rightness and putative wrongness of reasons transmit in different ways. He argues that what best explains this asymmetry is that the putative WKRs are in fact (right kind of) reasons for desiring to have the attitudes in question, and thus transmit the same way as reasons for desires in general do.

seems best to simply say that it is (merely) morally appropriate for S to admire X if there is

sufficient moral reason for S to want to admire X. After all, how could it be morally

desirable to have an attitude if having that attitude were entirely morally *inappropriate*? If it

were in no way morally appropriate for me to admire Kim, it would mean that realizing my

morally fitting desire to admire him would result in moral impropriety. That seems absurd.

So the paleo-sentimentalist who adopts WKR scepticism will make the distinction between

fittingness and mere appropriateness as follows. It is W-ly *fitting* for S to have Y towards X

if and only if any W-ly ideal subject would take some consideration(s) to sufficiently favour

S's manifesting Y towards X.[34] It is W-ly *appropriate* for S to have Y towards X if and only

if any W-ly ideal subject would take some consideration(s) to sufficiently favour S's wanting

to have Y towards X (that is, if having Y is desirable). For many attitudes, we express our

endorsement of Y being a W-ly fitting attitude towards x by saying that x is W-ly Y-able.

On this account, fittingness and appropriateness of having attitude Y are

distinguished in terms of reasons for manifesting Y or only desiring to have Y. It may seem

---

[34] Schroeder ('Ubiquity': 466–470) argues that there are right kind of reasons for attitudes that come apart from reasons for manifesting them (in my terms). For example, expecting more information to come can be an RKR not to make up one's mind about travelling yet, and thus not to intend, though it doesn't bear on whether to travel. The natural response, which Schroeder (475–477) anticipates, is to say that expecting information is not a reason for not intending, but rather for *withholding* an intention (insofar as it is a reason for an attitude at all). The view defended here may offer hope for this move. Let's say that the attitude of withholding A essentially manifests itself in refraining from forming A. Consequently, RKRs for withholding are the considerations on the basis of which an ideal subject (say my ideal advisor) would favour my refraining from forming A (now), such as the fact that I'll later have more information bearing on whether to manifest A. The putative WKRs for withholding (or reasons to desire to withhold) are considerations on the basis of which my ideal advisor would favour my merely having the attitude of withholding without manifesting it, such as being paid to withhold making up my mind (which is not the same as paying for *not having* the intention).

that it is vulnerable to counterexamples that involve *forced descent* or *blocked ascent*.[35] In forced descent cases, the incentive for having the attitude is there only if one also manifests it, though the attitude is intuitively unfitting. Kim Jong-un will only save lives if I not only admire him, but *also* manifest the admiration by emulating him. Without further qualification, an impartial spectator will still not approve of emulating the dictator, since that would mean going around ordering underlings to kill and terrorize people, among other things. But suppose further that I'm powerless and unable to actually harm anyone. Wouldn't any impartial spectator then endorse admiring Kim, in view of the benefits? No. In this variant, an impartial spectator would only approve of my emulating Kim on condition that I *fail* at it – that I don't actually do the sort of bad things he does. In other words, she would at best approve of my *trying* to emulate him. It's still not true that any impartial spectator would approve of my successfully acting in every way like Kim does. And admiration is only fitting if any relevant ideal subject would take there to be sufficient reason to manifest admiration by emulating the object (in every admired respect) *with success*. An incentive for manifestation won't make the attitude counterintuitively fitting.[36]

In blocked ascent cases, there is a cost to *wanting* to have an intuitively appropriate attitude. The dictator will only spare lives if I admire him *without* desiring to admire him. Won't any impartial spectator then oppose desiring to admire, so that my account wrongly implies that admiring is morally inappropriate? No. As in the original case, any impartial spectator would take the fact that lives will be saved by admiring to (sufficiently) favour manifesting the desire to admire, so admiring is morally desirable. However, at the same

---

[35] An anonymous referee pressed both types of counterexample. The term 'blocked ascent' comes from Andrew Reisner, "The Possibility of Pragmatic Reasons for Belief and the Wrong Kind of Reasons Problem," *Philosophical Studies* 145 (2): 257–272.

[36] There are tricky variants of this case that I can't discuss here for reasons of space.

time, she would take the fact that lives would be spared by not desiring to admire to (sufficiently) favour bringing it about that one doesn't desire to admire, that is to say, manifesting a desire to not desire to admire. (This could be supported by considering how the reasons are transmitted.) Hence, desiring to admire is morally undesirable. By my definition of appropriateness, it is then morally appropriate to admire the dictator, and morally inappropriate to desire to admire the dictator. That's exactly what the objector wants to say. Nevertheless, appropriateness is here understood in terms of an ideal subject's endorsement of wanting to have an attitude. There may seem to be a tension between the moral inappropriateness of desiring to admire and the moral fittingness of desiring to admire. But there is no genuine contradiction: there's just sufficient reason to desire not desiring to admire and sufficient reason to desire to admire. (What is special here is that the latter reason can't survive being responded to.[37]) The structure is just that of a classic WKR case: it can be simultaneously inappropriate for me to envy my colleague and fitting for me to envy her.

4. Conclusion

Our unreflective attitudes, when manifest, predictably cause different practical problems, because we are biased, fallible, and short-sighted, among many other shortcomings. The resulting dissatisfaction motivates us to step back and seek the advice of someone better situated, or to become better situated ourselves with respect to some domain of problems. Better situated in what way? In the way of lacking those of our features that predictably cause trouble, and having, fully developed, those of our features that predictably help avoid the kind of problem at issue. These dispositions to defer and aspirations to improve show that we regard, in practice, certain perspectives as ideal for settling which attitudes to have. I

---

[37] Way, 'Transmission', 512.

have offered brief and no doubt naïve sketches of what the optimal moral, aesthetic, and prudential perspectives might look like, drawing on Hume's insights. Even so, they should suffice to indicate the general form of the approach. No appeal to independent truths that would be better accessible from such a perspective is needed.

According to this kind of old-fashioned sentimentalism, what makes a non-doxastic attitude like admiration fitting in some way is that it would be endorsed from some ideal perspective. Since there are many ideal perspectives, there are many ways for an attitude to be fitting. Insofar as *having* an attitude would not be endorsed by an ideal subject unless she took there to be enough to be said for *manifesting* it by doing what it essentially disposes one to do, the account allows for distinguishing between fitting and merely appropriate attitudes. The latter are those that an ideal subject would only endorse *desiring* to have. In this way, the debates concerning Fitting Attitudes analysis of value and idealization in subjectivist theories turn out to be linked to each other. This note has only outlined how this connection might be made and how the various ideal perspectives might be specified in terms of the problems they're designed to solve. Among other things, I haven't even tried to show that ideal subjects would converge on endorsing plausible attitudes, though I'm more hopeful of that than most. No doubt the devil is in the details, which must await future work.[38]