9 The Evolutionary Debunker Meets Sentimental Realism[1]

Mauro Rossi and Christine Tappolet

1 Introduction

Value Realism holds that there are evaluative facts that are objective, or stance-independent, in the sense that they exist independently of what we think or feel about them. On this view, facts such as pain being bad, knowledge being desirable, or helping others being admirable are as objective as facts postulated by the natural as well as the social and human sciences. But what would evaluative facts be? What, if anything, distinguishes evaluative facts from other kinds of facts? Is the value realist just adding more grim facts to our ontology or are evaluative facts distinct? If evaluative facts are distinct, what is the difference with natural facts? A useful way to frame the problem is to ask what values are, if one were to follow Hilary Putnam (2002) and accept that the fact/value dichotomy is a failure. Clearly, this question needs to be answered if we want to assess Value Realism.

Value Realism is opposed to Value Anti-Realism, the thesis that there are no objective evaluative facts. Both Value Realism and Value Anti-Realism come in very different versions, which can each be paired with a variety of stances regarding the epistemology of values, the semantic of evaluative sentences, and the moral psychology of evaluative judgments.[2] In spite of the large number of possibilities that results, some combinations are standard. Thus, Value Realism is generally associated with Rationalism, the view that the evaluative is grounded in reason, and which often takes the form of the epistemological claim that knowledge of what are conceived of as *a priori* evaluative truths can be gained through the exercise of reason. Quite generally, Rationalism is opposed here to Sentimentalism, the view that the ground of evaluative judgments lies in the sentiments or more generally in affective-*cum*-motivational states. Because they take affective and motivational states to be non-cognitive, purely motivational states, sentimentalists most often reject Value Realism (Nichols 2004, Prinz 2007).

In this paper, we propose a defense of Value Realism that relies on the unusual combination of Value Realism with Sentimentalism. What

this account, which we call "Sentimental Realism," holds, in a nutshell, is that what makes evaluative facts special is their relationship to emotions. More precisely, Sentimental Realism claims that evaluative facts are fully objective facts, but that such facts are picked out by concepts that are response-dependent, in the sense that they are essentially tied to emotions. Our plan is as follows. We shall start with a presentation of Sentimental Realism and a discussion of its main virtues. On the basis of this, we shall discuss an objection to Value Realism that draws on evolutionary considerations, the Evolutionary Debunking Argument. We shall argue that Sentimental Realism safely escapes from this dilemma.

Before we start, let us clarify the scope of our discussion. The normative domain can be divided into what belongs to the evaluative, which concerns the good and the bad as well as the different ways of being good and bad, such as *admirable* or *courageous*, on the one hand, and the deontic, which is exemplified by concepts such as *ought*, *obligatory, forbidden*, and *permissible*, on the other. As should be clear from the terminology we used, our discussion focuses only on one part of the normative, namely, the evaluative. We have little to say in this paper about deontic facts, such as the facts that we ought to reduce pain or that it is forbidden to kill innocents.

2 Sentimental Realism[3]

What makes Sentimental Realism a form of Sentimentalism is its commitment to the thesis that evaluative concepts are essentially tied to affective states, and more specifically to emotions, such that such concepts count as being response-dependent. This claim is particularly plausible in the case of evaluative concepts such as *admirable, shameful*¸ or *amusing*, which wear their response-dependence on their sleeves. As many have highlighted, a plausible way to spell out the response-dependence of such concepts is as follows:

(S) By conceptual necessity, *x* is admirable/shameful/amusing/etc., if and only if admiration/shame/amusement/etc., is appropriate in response to *x*, if one were to contemplate x.[4]

There are very different ways to understand (S). A first question that arises is how one should understand the equivalence. According to one reading, (S) consists in a conceptual analysis, in which evaluative concepts are broken down to what is taken to be simpler conceptual elements, namely, affective concepts and the concept of appropriateness (e.g., Chisholm 1986, 58–60). The problem with this interpretation is that it is far from clear that affective concepts can be understood without appealing to evaluative concepts. How, for instance, could one explain the concept of admiration without invoking admirableness? In our view, it is better

to understand (S) as offering a conceptual elucidation—as opposed to an analysis—in which none of the two sides of the biconditional has conceptual priority (Wiggins 1987). On such a no-priority view, the grasp of evaluative and affective concepts would be interdependent, such that none would be considered more fundamental. A second question concerns the notion of appropriateness, which can be taken to mean quite different things. According to some, such as Roderick Chisholm, appropriateness has to be explicated in terms of the deontic notion of requirement (Chisholm 1986, 58–60). Others, such as Thomas Scanlon spell out the equivalence in terms of reasons for the responses at stake (Scanlon 1998, 95–100). We believe that a better way to understand such equivalences is to claim that being appropriate in this context is to be correct from an epistemic point of view.[5] The idea is that someone is admirable, for instance, just if feeling admiration towards her is correct, or accurate, from an epistemological point of view. Put differently, an appropriate emotion is one that represents the evaluative facts as they are.

The worry that will immediately arise here is that understood this way, (S) is much too circular to be of interest. By proposing to elucidate evaluative concepts in terms of responses that are appropriate just if their objects fall under the concepts in question, the equivalence draws nothing but an uninterestingly little circle. The circularity cannot be denied, but we believe that (S) is nonetheless of interest. As David Wiggins stressed when defending what he called "sensible subjectivism," the aim is to "elucidate the concept of value by displaying its actual involvement with the sentiments. One would not [. . .] have sufficiently elucidated what value is *without* that detour" (1987, 189). The point underscored by (S) concerns our epistemic "access" to values. To paraphrase Wiggins, the important point is that when we try to find out whether something is admirable or shameful, for instance, there is nothing more fundamental to appeal to than our responses of admiration and shame. What this means is that (S) makes a point that concerns the epistemology of values. Clearly, the expertise in affective concepts involves the ability to feel the relevant emotions. Thus, (S) points toward the epistemic indispensability of our emotional responses.

Given this, it is easy to see that this account of evaluative concepts is compatible with the claim that evaluative facts are objective, in the sense of being stance- or response-independent. The emotion of admiration toward someone can be appropriate only if in fact this person is admirable; and nothing prevents us from saying that the evaluative fact in question is objective. Indeed, on a natural understanding, it is because someone is objectively admirable that feeling admiration towards her is appropriate. Thus, what Sentimental Realism adds to (S) is the following claim:

(R) There are objective evaluative facts.

Now, here too, there are different ways to go. In particular, (R) allows for both naturalist and non-naturalist versions, depending on how the relationship between evaluative facts and natural facts is thought of. According to radically non-naturalist versions, evaluative facts are *sui generis* normative facts, which are distinct from natural facts. What is denied, more precisely, is that evaluative facts are identical or reducible to natural facts. What both non-naturalists and naturalists accept in general, however, is that evaluative facts supervene on natural facts, in the sense, roughly, that there is no evaluative difference without a difference at the level of natural facts. The view we find most plausible is that evaluative facts not only supervene on the natural, but are constituted by natural facts that involve disjunctive natural properties. Whether this view should count as a form of non-naturalism or not is better left open here.

Sentimental Realism, the conjunction of (S) and (R), is clearly a possible account. But why would one want to embrace it? Independently of the reasons to embrace both (S) and (R), which we will briefly discuss, we believe that the combination of the two claims makes for an attractive conception of values, which is in a position to answer important objections to Value Realism.

Let us start with the second conjunct, (R). Quite generally, as most would agree, (R) can be considered to be the default meta-normative position (Brink 1989). Indeed, this claim is even stronger in the case of values, where there is little motivation for the view that what we are concerned with are rules that are supposed to govern human interactions. Evaluative predicates, such as 'good' or 'admirable', behave like ordinary predicates. Thus, the structure of 'Sarah is admirable' appears in no way different from that of 'The ball is round.' Both types of sentences can be evaluated in terms of truth, for we can ask 'Is it true that Sarah is admirable?' just as we can wonder whether the ball is round. The two types of sentences appear to have cognitive contents that are genuinely truth-assessable. Similarly, when making evaluative judgments, we seem to express cognitive states that appear to be ordinary, truth-assessable beliefs. Indeed, both our evaluative sentences and our evaluative judgments seem to be about an objective evaluative reality, which exists independently of our own attitudes. This is evidenced by the fact that disagreement about value presents itself as genuine disagreement, in which at least one of the disagreeing parties makes a mistake.[6] According to the value realist, these considerations should be taken at face value. This means that, unless we have strong reason to think otherwise, we should admit that there really are objective evaluative facts and properties, which our evaluative judgments attempt to capture, and about which we disagree.

Let us turn to the other conjunct, (S). What supports (S) is that it captures the close relationship between the evaluative and the affective. Given the lexical proximity between terms such as 'admirable' and 'admiration,'

it is difficult to deny that such concepts are closely connected to affective concepts. This observation might suggest that being admirable is nothing but to be admired by someone. But in contrast to this subjectivist thesis, (S) is in a position to explain why what we *de facto* admire is not always admirable, while what is admirable is not always what *de facto* causes admiration. What (S) holds, is that it is only when admiration towards someone is appropriate that this person is admirable. Correlatively, given that it does not aim at reducing normative concepts to natural concepts, (S) does not fall prey to the natural fallacy, i.e., the error that consists in analyzing normative concepts in natural terms (Moore 1903).

In our view, an important reason to adopt Sentimental Realism is that this approach is grounded in an independently plausible account of emotions, namely, the Perceptual Theory. According to the Perceptual Theory, emotions are a kind of perceptual experiences, namely, perceptual experiences of evaluative facts.[7] The idea is that the admiration you feel towards someone represents that person as admirable, in the same way as color or shape experiences represent things as having color or shape properties.

The Perceptual Theory gains to be compared with its main competitors, Feeling Theories (James 1884, Lange 1885, Whiting 2011), Conative Theories (Frijda 1986, Scarantino 2014) and Judgmental Theories (Solomon 1976, Nussbaum 2001).[8] By contrast to Feeling Theories, which hold that emotions are mere feelings, and Conative Theories, according to which emotions are constituted by motivational states such as desires, the Perceptual Theory holds that emotions are intentional states that have representational contents, in the sense that they represents things as having certain evaluative properties. It can thus easily account for the fact that emotions not only have intentional objects, but are naturally taken to be assessable in terms of the nature of these objects, fear being appropriate when it is directed at fearsome and dangerous objects. In contrast to Judgmental Theories, according to which emotions are normative or evaluative judgments, the Perceptual Theory holds that it is not necessary to have the ability to make evaluative judgments and to possess evaluative concepts to feel emotions. Thus, the Perceptual Theory has no difficulty to account for the fact that non-human animals and infants can feel a number of emotion kinds, such as fear or anger.

This is not the place to spell out the full argument for the Perceptual Theory, let alone to consider the different objections that have recently been raised against it.[9] Let us simply note that the main argument for the Perceptual Theory is an argument by analogy. Emotions and sensory perceptual experiences share a number of striking features, the main of which are the following. Emotions and sensory perceptual experience both have phenomenal qualities, in the sense that there is a way it is like to feel an emotion, just as there is a way it is like to have sensory perceptions; they both are characterized by a lack of direct control by the agent,

in the sense that you cannot feel an emotion or have a sensory experience merely because you decide to do so; they both manifest "recalcitrance," such as when you fear something which you judge to be harmless, or you see something as bent while you judge that it is straight.

Sentimental Realism is a promising account of evaluative facts concerning the admirable, the fearsome, the amusing, the shameful, etc. But what about other kinds of evaluative facts, such as the fact that something is good or bad? What about facts involving properties that correspond to so-called thick concepts, such as the fact that some action is courageous or some person generous? It is quite easy to see how the proposed account can be extended to the more general evaluative properties and their concepts. The basic idea is that one can use the general notions of positive and negative emotions to handle such cases. Let us say that to favor something is to have a positive emotion towards it, while to disfavor it is to have a negative emotion towards it. Given this, it can be claimed that something is good (*pro tanto*) just if it is appropriate to favor it, and something is bad (*pro tanto*) just if it is appropriate to disfavor it. Similarly, it seems plausible to tie thick concepts, such as *courageous* or *generous*, to affective responses. For instance, it seems plausible to claim that an action is courageous just if it is done in spite of danger to oneself, and in virtue of this, it is admirable, so that admiring it would be appropriate.

Let us accept, then, that Sentimental Realism is a plausible account of the evaluative. One virtue of Sentimental Realism is that it allows for straightforward answers to two traditional objections to Value Realism, both of which can be traced back to John Mackie (1977, ch. 1). Mackie's so-called Argument from Queerness has two parts: one metaphysical (or ontological) and one epistemological. He claims that "[i]f there were objective values, they would be entities or qualities or relations of a very strange sort, utterly different from anything else in the universe. Correspondingly, if we were aware of them, it would have to be by some special faculty of moral perception or intuition, utterly different from our ordinary ways of knowing everything else" (Mackie 1977, 38). According to Mackie, objective evaluative facts provide "the knower with both a direction and an overriding motive; something's being good both tells the person who knows this to pursue it and makes him pursue it" (Mackie 1977, 40). Since ordinary facts appear to lack the same action-guidingness and motivational force, objective values look "queer." So, we do best without objective evaluative facts in our ontology.[10]

The good news is that Sentimental Realism is well-placed to answer both parts of the argument from queerness. In reply to the epistemological challenge, Sentimental Realism can rely on the claim that emotions are perceptual experiences of values in order to develop an account of the justification of evaluative beliefs, and more generally of evaluative knowledge, which grounds such beliefs on our emotional responses.

There are different ways to go here, but what seems plausible is that our emotions be considered to constitute defeasible reasons for our evaluative beliefs, so that when we form a belief on the basis of an emotional reaction, our belief is *prima facie* justified (see Tappolet 2000, 2016, chap. 5; and Kauppinen 2013, *inter alia*). In our view, the justification of evaluative beliefs turns on the absence of any reason to believe that the emotion on which the belief is based be inappropriate. So, your belief that someone is admirable would be justified on condition that it is based on the admiration you feel for that person, and you have no reason to believe that what you feel is biased or more generally has to be distrusted. One question that arises is how we go about determining what conditions count as unfavorable. As David Hume stressed, finding out what conditions are favorable or unfavorable is something that we undertake with others. What Hume argues is that the very possibility of understanding each other presupposes that we adopt what he calls a "common" or "general point of view," that corrects for perspectival effects, such as biases due to spatial or temporal distance (1739–1740, III, 3, i). The idea is that in order to determine what counts as an unfavorable condition—a defeater, in epistemological terms—we need to take into account not only our own experiences at different times and in different conditions, but also the experiences of other persons.[11] In Philip Pettit's terms (1991, 600–1), when we want to determine whether something should count as a defeater, we have to look to our shared practice of discounting certain conditions as likely to interfere with our responses, a practice that aims at making sense of intra-personal, but also of interpersonal, discrepancies. In any case, evaluative knowledge can then be conceived of in terms of justified evaluative beliefs that are based on emotions that are not defeated. Suppose you form the belief that a friend is admirable on the basis of your feeling of admiration. Your belief will be justified in case you have no reason to believe that your admiration is interfered with. But of course, your admiration might be nonetheless misguided. So, to know that the action is admirable, your belief need not only be justified, it also needs to be such that *de facto* no defeater interferes with your admiration.

Let us now address the metaphysical part of Mackie's argument. Given that emotions normally come with related motivations, Sentimentalist Realism is also in a position to handle the challenges regarding the action-guidingness and motivational force of values. Value Realism is generally associated with a defense of the externalist thesis according to which evaluative judgments are not essentially, or necessarily, tied to motivation. Now, even though it can be agreed that making an evaluative judgment does not necessarily involve a motivation to act, Sentimental Realism can hold that evaluative judgments will normally be accompanied by such a motivation, given that such judgments are grounded in the corresponding emotional reactions. This simply follows from the fact

that emotions, or at least most of them, at least normally involve motivations. So, insofar as evaluative judgments are concerned with facts that are relevant to what we ought to do, so that they are in this sense action-guiding, they will normally involve a motivation to act accordingly. Given this, it is not necessary to postulate objective entities possessing magical motivational properties, in order to account for the motivational force of evaluative judgments.

If these arguments are on the right track, Sentimental Realism is an attractive account, which might well be better placed than other Value Realist accounts. But this does entail that Sentimental Realism is immune to all possible objections against Value Realism.

3 The Evolutionary Debunking Argument against Value Realism

In this part, we want to discuss one important objection, in particular, which has attracted considerable interest in recent years and which is known as the so-called Evolutionary Debunking Argument against Value Realism.[12] We shall argue that Sentimental Realism can successfully address this objection. We will proceed as follows. We will start by presenting the general version of the Evolutionary Debunking Argument. We will then show how the argument applies to the sentimental realist account proposed in the first part of the paper. After presenting some replies offered in the literature, we will put forward a new version of the "third factor argument" against the Evolutionary Debunking Argument.[13]

The debunking argument starts from the empirical claim that evolutionary forces have greatly shaped our evaluative faculties and that they have thereby disposed us to form specific evaluative beliefs. The argument proceeds by noticing that evolutionary forces aim at fitness, not at detecting mind-independent evaluative facts. If so, we have reason to think that evolutionary forces have been selected for evaluative faculties that yield fitness-enhancing beliefs, not beliefs tracking mind-independent evaluative facts. Indeed, it would be a massive coincidence—something akin to a miracle—if faculties shaped by forces that aim at survival and reproductive success also managed to deliver beliefs that track a stance-independent evaluative reality. Thus, absent some other reason to think the contrary, we have decisive reason to conclude that our evaluative faculties are unreliable, i.e., that they tend to generate mistaken evaluative beliefs. Insofar as Value Realism is an ontological and not an epistemological stance, this is not yet a conclusion that shows Value Realism to be wrong. However, the stark skepticism that would follow would put Value Realism in an uncomfortable position.

It is clear that Sentimental Realism is a target of the Evolutionary Debunking Argument. As we have seen, Sentimental Realism is

committed both to the existence of objective evaluative facts and to the possibility of having knowledge of such facts. More specifically, Sentimental Realism emphasizes the role of emotions, which are conceived of as perceptual experiences of evaluative properties, for the acquisition of evaluative knowledge. This is where the Evolutionary Debunking Argument kicks in. According to it, sentimental realists must recognize that evolutionary pressures have exercised a large influence on our emotions. Indeed, from an evolutionary perspective, emotions (at least basic emotions, that is, innate and pan-culturally shared emotions[14]) can be seen as adaptations which gave our ancestors an evolutionary edge by allowing them to deal more effectively with the specific challenges that they typically encountered. If this is true, then it is hard to maintain that emotions reliably track genuine evaluative facts. For this to be true, it would have to be the case that reliably detecting objective evaluative facts somehow increased our ancestors' fitness. According to the debunker, however, this is scientifically questionable. There are two more plausible options. The first consists in maintaining that emotions have nothing to do with evaluative properties and that, insofar as they are perceptual experiences of some sort, they are perceptual experiences of non-evaluative properties whose detection was significant for our ancestors' survival, but which are uncorrelated to evaluative properties. The second consists in maintaining that, while emotions are indeed perceptual experiences of evaluative properties, they are unreliable ones. Accordingly, when we experience an emotion towards a particular object, we perceive such an object as possessing an evaluative property that it does not really possess. Either way, the debunker's line of thought implies that the evaluative beliefs that we regularly form on the basis of our emotions are likely to be systematically mistaken.

In what follows, we will understand the debunker's challenge in the sense of the second option just mentioned. The reason is the following. For the sake of the argument, the debunker grants the assumption that there exists an objective evaluative reality. If evaluative properties exist, then, according to most emotion theorists,[15] they constitute the formal objects of emotions, where the formal object of an emotion type is what determines the correctness conditions of the tokens belonging to that particular type. For instance, the formal object of fear is the fearsome, while the formal object of admiration is the admirable. Moreover, if emotions are conceived of as perceptual experiences, then, insofar as their formal object determines their correctness conditions, it does so by determining whether the content of such emotions is correct or not. But if the formal object of a given emotion is constituted by an evaluative property, then it seems that the formal object determines whether the content of a given emotion is correct in the sense that such an emotion is assessed as correct if and only if, and because, it represents its object as possessing the relevant evaluative property and its object does indeed possess such a

property. From this account, it follows that emotions are unreliable only if they systematically misrepresent their objects as possessing evaluative properties that they do not really possess. Framed in these terms, the Evolutionary Debunking Argument claims that the objects of emotions possess properties that are of evolutionary significance, but it adds that these properties do not correlate with the perceived evaluative properties.

Suppose we grant the evolutionary story about the origin of emotions. One may be tempted to insist that this does not necessarily spell trouble for Sentimental Realism. In fact, emotions admit some degree of plasticity. So, even if we assume that emotions were selected for to track fitness-enhancing properties and that such properties are not correlated with evaluative properties, emotions can nevertheless be "corrected," e.g., by submitting them to rational scrutiny, in such a way to ensure that they reliably track genuine evaluative properties. Not only that, but if we think that emotions provide us only with the basis for particular evaluative judgments, which are then used as raw data to construct more general evaluative principles through a method of interpersonal reflective equilibrium, then, to the extent that such a method leads us to abandon the particular evaluative judgments that are clearly mistaken, then we may also think that some degree of error in the initial data is not necessarily an obstacle to the obtainment of evaluative knowledge.

However, things are not that easy. The debunker may reply, first, that we have good reason to think that most of our emotions are incorrect and, second, that we have no independent way of identifying which of them are incorrect. Together, these points can be used to reject the sentimental realist's initial reply. For a start, it follows from the debunker's claims that we cannot simply assume that the standard defeaters (e.g., drugs, fatigue, cultural biases, etc.) constitute the only distorting influences on otherwise correct emotions, since this is precisely what the debunker disputes. In fact, the debunker argues that we cannot make any substantive assumptions as to which emotions are correct, which we could then use to assess and "reprogram" our incorrect emotions, since such assumptions may simply be the product of our evolutionary history. If this is the case, and if we do have good reason to think that most of our emotions are unreliable, then no method that relies on our emotional responses as a starting point can lead us to evaluative knowledge. The method is as good as the data on which it operates and if the data are inadequate, then the method will deliver inadequate results. As it is often put: "Garbage in, garbage out."

The situation may appear bleak for Sentimental Realism. However, appearances are often deceitful. Value realists of different brands have proposed various replies to the Evolutionary Debunking Argument. The sentimental realist can co-opt at least some of them for her own purpose, as we will now show.[16] The starting point of the sentimental realist's counterattack is the observation that the debunker cannot merely assert

that, if emotions have been shaped by evolutionary forces, then most of them are incorrect. In other words, the debunker must give us some reason to accept the claim that it would be a massive coincidence if mental states selected for fitness were also reliable detectors of objective evaluative facts. What does she have to say in this regard?

The natural option for the debunker consists in showing that, more often than not, our emotions lead us to form evaluative beliefs that we know, or can confidently hold, to be incorrect. This would provide compelling evidence that evolutionary forces aimed at survival do not track objective evaluative facts. However, as it should be clear by now, this strategy is not open to the debunker, for it would contradict her claim that we have no independent way of identifying which emotions are correct and which are incorrect and, most importantly, it would provide an easy way out to the value realist.[17]

An alternative for the debunker consists in formulating her argument in terms of probabilities. If we cannot make any substantive assumptions about evaluative facts, then we are left with an essentially unlimited set of *possible* evaluative facts. But then, it is extremely unlikely that evolution has endowed us with mental states that track precisely the *actual* evaluative facts. This gives us reason to endorse the massive coincidence thesis against Value Realism. However, as several authors have noticed,[18] this argument can be challenged. For one, we can resist its starting point and deny that, if we forbid all substantive assumptions, then we must take the space of possible evaluative facts to be infinite. The reason is that there may be conceptual constraints on what counts as an evaluative truth, which significantly narrow down the set of possible evaluative facts.[19] Furthermore, the probabilistic argument seems to overgeneralize. Indeed, if we assume that evolutionary forces have exercised a great influence on the development of our sensory perceptual capacities and if we avoid making any substantive assumptions about perceptual facts, then we must accept that the space of possible perceptual facts is infinite. But then, we should conclude that it is extremely unlikely that capacities selected for survival ended up tracking precisely the actual perceptual facts. However, this does not seem to be a conclusion that debunkers would readily accept.[20]

The debunker may explore other possibilities. One of these consists in claiming that, insofar as evolutionary forces have shaped our emotions, the latter are *insensitive* to the evaluative facts, in the sense that our emotions would not have been different, had we lived in a different possible world.[21] Indeed, had the evaluative facts been different, evolutionary forces would have still pushed us to experience the same fitness-enhancing emotions (at least the same basic emotions), thereby getting us to form the same fitness-enhancing evaluative beliefs. Given this, we should conclude that any match between our actual emotions and the evaluative facts would be massively coincidental. A similar option consists in

claiming that, insofar as evolutionary forces have shaped our emotions and insofar as our evaluative beliefs are formed on the basis of such emotions, then our evaluative beliefs are not *safely* formed.[22] Indeed, had we evolved differently, we would have experienced different emotions, which, in turn, would have led us to forming different evaluative beliefs. By our current light, however, such beliefs would have been false. This shows that our evaluative beliefs are simply the result of an accident of our evolutionary history. If so, their correspondence to a mind-independent evaluative reality would be massively coincidental.

However, both arguments can be challenged on general grounds. In fact, as Tomas Bogardus (2016, 644) has stressed, recent work in epistemology gives reason to believe that neither sensitivity nor safety is required for justified belief and knowledge. So, even if we grant that our emotions might be insensitive or unsafe in the way suggested above, this has little implications on their epistemic status. In fact, both the argument from sensitivity and the argument from safety are part of a broader argument that aims at showing that emotions cannot lead us to evaluative knowledge. However, if it is shown that neither sensitivity nor safety, in the technical sense at stake, matters for knowledge, then, the alleged failure of emotions to satisfy these conditions does nothing to support the claim that emotions are not reliable grounds for the acquisition of evaluative knowledge.

One last possibility that we wish to consider is for the debunker to argue that, if evolutionary forces have shaped our emotions, then the best explanation of how we typically form our emotions does not presuppose their reliability.[23] For instance, one may argue that, to the extent that mind-independent evaluative facts are causally inert, then the best explanation of how we form our emotions is simply that the latter are caused by those non-evaluative properties of their objects that were evolutionarily salient for our ancestors' survival. Thus, even if we grant that our emotions are perceptual experiences of evaluative properties, we have no explanatory good reason to hold that they are reliable perceptual experiences. The upshot is that it would be a massive coincidence, if they were actually reliable.

The most natural reply against this argument is for the sentimental realist simply to deny its starting point, namely, that evaluative properties are causally inert. After all, Sentimental Realism holds that evaluative knowledge is largely a kind of synthetic *a posteriori* knowledge. This fits comfortably with the thesis that the evaluative is causally efficacious. But the sentimental realist may also consistently opt for a weaker thesis, according to which, if evaluative properties supervene on causally efficacious non-evaluative properties, they possess causal power in a derivative way. Either way, the sentimental realist has an explanation of how we can have perceptual experiences of genuine evaluative properties that is at least as good as that proposed by her opponent.

Let us take stock. So far, we have examined various arguments that the debunker may give in defense of her claim that the correlation between our actual emotions and mind-independent evaluative facts would be massively coincidental. We have seen that the sentimental realist has a reply to all of these arguments. However, a die-hard skeptic may not be satisfied by this essentially defensive strategy. She may insist that the alleged correlation between fitness-relevant properties and evaluative properties remains fishy and ask the sentimental realist for a more positive defense. It is unclear whether the debunker is entitled to such a demand, without having herself provided more convincing evidence about the unreliability of our emotions.[24] Be that as it may, in what follows we want to put forward precisely an argument of this sort. Our argument is a version of an argument that has been variously explored by value realists of different sorts: the "third factor" argument. In its general form, third factor arguments consist in showing that there is (at least) an indirect correlation between fitness-enhancing facts and objective evaluative facts, in the sense that, there exists a "third factor" that is responsible for both and explains their correlation. Our goal in the rest of this section is to see how an argument of this sort can be constructed in defense of Sentimental Realism.

Following David Enoch (2010, 428), our starting point is the observation that, in order to reject the debunker's challenge, it is not necessary to show that there exists a perfect correlation between our emotions and the objective evaluative facts. In other words, it is not necessary to show that all emotions are reliable. Rather, it is sufficient to show that our emotions are reliable more often than not. As we have seen, emotions provide the starting points for the acquisition of evaluative knowledge. However, reasoning mechanisms play an important role as well, and provided that they operate on mostly correct data, they can help us getting closer and closer to the evaluative truth, e.g., by eliminating inconsistencies and logical mistakes, and by taking into account defeaters. Notice that this may be possible even if we do not know exactly which particular emotions are correct, provided that we know that they are *mostly* correct.

How then can we proceed in order to show that our emotions are reliable more often than not? As we have seen, the debunker holds that evolutionary forces have selected for fitness-enhancing emotions, that is, emotions providing evolutionarily relevant information for our ancestors and leading to behaviors that increased their chances of survival and reproductive success. The first substantive step of our argument consists in claiming that, more often than not, whatever contributes to an individual's fitness contributes also to the individual's well-being. If this is true, then, by selecting for fitness-enhancing emotions, evolution may have also selected for emotions that are, more often than not, well-being-enhancing, i.e., emotions that either (a) favor, or lead to the promotion of, things that are good for the individual, or (b) disfavor, or lead to the

avoidance of, things that are bad for the individual.[25] Consider, as an example, a fairly universal emotion like the fear of snakes. Such an emotion was obviously adaptive for our ancestors, in the sense that those who experienced such an emotion had more chances to survive that those who did not. By helping them survive, such an emotion had an impact also on their well-being, at the very least by ensuring that a preliminary condition for well-being be satisfied, namely, remaining alive.

The debunker may immediately raise an objection against this first step. Isn't the claim that, more often than not, whatever contributes to an individual's fitness contributes also to the individual's well-being a *substantive* claim? And isn't it precisely the sort of claim that is presumptively infected by evolutionary influences, so that it cannot be presupposed without begging the question against the debunker? In response, the sentimental realist may appeal to the idea, mentioned above, that there are some conceptual constraints on what counts as well-being and that denying the existence of even a weak correlation between individual fitness and individual well-being would violate such constraints. In fact, in order to deny the first step of our argument, the debunker would have to deny that the adoption of effective survival strategies contributes, more often than not, to the individual's well-being, in the sense that it is, more often than not, beneficial for the individual or in the individual's interest. However, it is not clear whether this move can be defended without threatening to make the resulting view unintelligible as a view of well-being.

Suppose, then, that the debunker's objection can be resisted. The argument might be still thought to be very limited, since it establishes only a correlation between emotions and goodness for. In fact, if emotions are conceived of as perceptual experiences of non-relational evaluative facts, what we need in order to reject the Evolutionary Debunking Argument is a correlation between emotions and goodness *tout court*. This is the goal of the second step of our argument. Once again, the debunker will be skeptical. After all, it seems that this requires defending another of those substantive claims that the debunker will take to be question-begging. In order to defuse the debunker's worry, the sentimental realist can pursue a similar strategy as the one pursued with respect to the first step.

As we have seen, emotions are conceived of as perceptual experiences of evaluative properties. More specifically, positive emotions are directed at things that supposedly possess positive evaluative properties, e.g., the property of being pleasurable, joyful, etc.; while negative emotions are directed at things that supposedly possess negative evaluative properties, e.g., the property of being fearsome, disgusting, etc. Our claim is that things that *do*, in fact, possess properties of the former kind are also things that *tend* to be good for the individual. Likewise, things that *do* possess properties of the latter kind are also things that *tend* to be bad for the individual. In order to deny these claims, one would have to

maintain that pleasurable, joyful, etc., things do not tend to be well-being-enhancing, and that fearsome, disgusting things do not tend to be well-being-diminishing. Arguably, this would once again render the resulting view unintelligible as a view of well-being.

These considerations have some important implications. They imply that, *if* emotions were *reliable* perceptual experiences of evaluative properties, *then* they would typically be well-being-enhancing, since they would favor (or disfavor) things that tend to be good (or bad) for the individual. In fact, as we have seen above, fitness-enhancing emotions *are*, more often than not, well-being-enhancing. Of course, it would be fallacious to infer from this that fitness-enhancing emotions are, more often than not, reliable perceptual experiences of evaluative properties. However, if we combine the previous considerations with the claim that evaluative properties are causally efficacious, then the thesis that fitness-enhancing emotions are, more often than not, reliable perceptual experiences of evaluative properties appears to be *a plausible explanation* (perhaps, *the* most plausible explanation) of why fitness-enhancing emotions are also, more often than not, well-being-enhancing. If this is the case, then the second step of our argument can be vindicated.

We want to conclude by considering one possible objection that the debunker may raise at this stage. According to it, our argument simply presupposes that our evaluative concepts are not subject to distorting evolutionary influences. However, this presupposition is illegitimate in the context of a discussion on the Evolutionary Debunking Argument. We wish to make two points in response. The first is that conceptual assumptions about evaluative concepts seem to stand on firmer grounds than substantive assumptions about evaluative facts. Indeed, while the claims about the latter may reflect the distorting effects of evolution on our emotions, the former seem to be relatively immune to such effects, since conceptual assumptions are more the product of rational reflection than of our emotions, and we have no evolutionary grounds for thinking that rational reflection is unreliable. The second point is that, as our previous discussion shows, the debunker has not given us any reason to think that our substantive assumptions are systematically mistaken. In fact, we have rejected all her arguments aimed at showing that our emotions are unreliable. All in all, the sentimental realist seems thus to have the edge over her opponent.

4 Conclusion

Let us draw together the treads. We started out by asking what evaluative facts, as compared to natural facts, would be if Value Realism were true. In reply, we argued that what is specific of the evaluative domain lies in its ties to the affective. Thus, we proposed a version of Value Realism, Sentimental Realism, which combines the claim that evaluative facts

are fully objective, or stance-independent, with the thesis that evaluative concepts are characterized by response-dependence. We argued that Sentimental Realism is in a good position with respect to the Argument from Queerness as well as the Evolutionary Debunking Argument. In both cases, the emotions we feel did the heavy lifting in the counter-arguments.

Notes

1 We gratefully acknowledge the FRQ-SC and the SSHRC for their support, and wish to thank Samuel Dishaw for his useful comments.

2 Indeed, moral realism and anti-realism are often considered to involve more than an ontological stance. See Sturgeon (1984); Sayre-McCord (1988); and Railton (1996); and Tappolet and Rossi (2015a) for a general discussion.

3 This section draws partly on Tappolet and Rossi (2015a) and Tappolet (2016).

4 See Brentano (1889: 11); Ewing (1947, 1959); Wiggins (1976, 1987); Blackburn (1984, 1998); McDowell (1985); Chisholm (1986); Gibbard (1990); Anderson (1993); Mulligan (1998); Scanlon (1998); D'Arms and Jacobson (2000a, 2000b); Oddie (2009), *inter alia*.

5 The terminology is inspired by Brentano, who claimed that something is good if and only if loving it is correct (*richtig*) (1889: 11). See also Oddie (2014: 66).

6 See Moore (1922); Enoch (2011: chap. 2).

7 See Meinong (1917); de Sousa (1987); Tappolet (1995, 2000, 2016); Prinz (2004); Deonna (2006); Döring (2007); Tye (2008).

8 See Deonna and Teroni (2012) for a useful overview of emotion theories.

9 See Tappolet (2016: ch. 1); and for objections, see Deonna and Teroni (2012); and Brady (2013).

10 As this formulation makes manifest, Mackie's argument cannot concern normativity in general, including the epistemological normativity (see Putnam 2002; Cuneo 2007).

11 This is an insight that has often been highlighted in recent discussions. See Wiggins (1987: 196); Davidson (1991); Pettit (1991, 600–601).

12 Amongst the authors that have put forward versions of the evolutionary debunking argument, see Joyce (2006); Kitcher (2006, 2011); Ruse and Wilson (1986); Ruse (1998); and Street (2006).

13 Amongst the authors that have defended this kind of argument, see Brosnan (2011); Enoch (2010, 2011); and Skarsaune (2010).

14 See Ekman (1972); Griffiths (1997); and Cosmides and Tooby (2000).

15 See for instance, Kenny (1963); de Sousa (1987); Teroni (2007); and Deonna and Teroni (2012).

16 Amongst the value realists that offered some replies to the evolutionary debunking argument, see Berker (2014); Bogardus (2016); Brosnan (2011); Copp (2008); Enoch (2010 and 2011); FitzPatrick (2015); Kahane (2011); Huemer (2016); Shafer-Landau (2012); Skarsaune (2010); Vavova (2014 and 2015); and Wielenberg (2010).

17 On this point, see also Shafer-Landau (2012) and Vavova (2014).

18 See Shafer-Landau (2012); and Vavova (2014, 2015), *inter alia*.

19 We will come back to this line of thought below, when developing a "third factor" argument against the evolutionary debunking argument.

20 On this reply to the evolutionary debunking argument, see Shafer-Landau (2012, 12–13).

21 For a discussion of the argument from sensitivity, see also Wielenberg (2010); Clarke-Doane (2012); Shafer-Landau (2012); and Bogardus (2016).
22 For a discussion of the argument from safety, see Bogardus (2016).
23 See Shafer-Landau (2012), especially 24–6.
24 For a similar point, see Vavova (2015, 105–6).
25 In Tappolet and Rossi (2015b), we argued that the claim that emotions have been selected for as fitness-enhancing mechanisms does not entail the claim that emotions have been selected for as well-being-enhancing mechanisms. Furthermore, we have also argued that there exists no necessary relation between emotions and well-being. Both of these claims are compatible with the thesis defended here, according to which *some* fitness-enhancing emotions are also well-being-enhancing emotions.

References

Anderson, Elizabeth S. 1993. *Values in Ethics and Economics*. Cambridge, MA: Harvard University Press.

Berker, Selim. 2014. "Does Evolutionary Psychology Show that Normativity Is Mind-Dependent?" In *Moral Psychology and Human Agency: Philosophical Essays on the Science of Ethics*, edited by Justin D'Arms, and Daniel Jacobson, 215–52. Oxford: Oxford University Press.

Blackburn, Simon. 1984. *Spreading the Word*. Oxford: Oxford University Press.

Blackburn, Simon. 1998. *Ruling Passions*. Oxford: Clarendon Press.

Bogardus, Tomas. 2016. "Only All Naturalists Should Worry about Only One Evolutionary Debunking Argument", *Ethics*, 126: 636–61.

Brady, Michael. 2013. *Emotional Insight*. Oxford: Oxford University Press.

Brentano, Franz. 1889. *The Origin of Our Knowledge of Right and Wrong*. Translated by Roderick M. Chisholm, and Elizabeth H. Schneewind. London: Routledge and Kegan Paul, 2009.

Brink, David Owen. 1989. *Moral Realism and the Foundations of Ethics*. Cambridge: Cambridge University Press.

Brosnan, Kevin. 2011. "Do the Evolutionary Origins of Our Moral Beliefs Undermine Moral Knowledge?" *Biology and Philosophy*, 26: 51–64.

Chisholm, Roderick M. 1986. *Brentano and Intrinsic Value*. Cambridge: Cambridge University Press.

Clarke-Doane, Justin. 2012. "Morality and Mathematics: The Evolutionary Challenge", *Ethics*, 122: 313–40.

Copp, David. 2008. "Darwinian Skepticism", *Philosophical Issues*, 18: 186–206.

Cosmides Leda, and Tooby, John. 2000. "Evolutionary Psychology and the Emotions." In *Handbook of Emotions*, edited by Michael Lewis and Jeannette M. Haviland-Jones, 91–115. New York: The Guilford Press.

Cuneo, Terence. 2007. *The Normative Web: An Argument for Moral Realism*. Oxford: Oxford University Press.

D'Arms, Justin, and Jacobson, Daniel. 2000a. "Sentiment and Value", *Ethics*, 110: 722–48.

D'Arms, Justin, and Jacobson, Daniel. 2000b. "The Moralistic Fallacy: On the 'Appropriateness' of Emotions", *Philosophy and Phenomenological Research* 61: 65–90.

Davidson, Donald. 1991. "Three Varieties of Knowledge." Reprinted in Davidson, D. 2001, *Subjective, Intersubjective, Objective*, 205–20. New York: Oxford University Press.

de Sousa, Ronald. 1987. *The Rationality of Emotion*. Cambridge, MA: MIT Press.
Deonna, Julien. 2006. "Emotion, Perception and Perspective", *Dialectica*, 60: 29–46.
Deonna, Julien, and Teroni, Fabrice. 2012. *The Emotions: A Philosophical Introduction*. New York: Routledge.
Döring, Sabine. 2007. "Seeing What to Do: Affective Perception and Rational Motivation", *Dialectica*, 61: 361–94.
Ekman, Paul. 1972. *Emotions in the Human Face*. New York: Pergamon Press.
Enoch, David. 2010. "The Epistemological Challenge to Meta-Normative Realism: How Best to Understand It, and How Best to Cope with It", *Philosophical Studies*, 148: 413–38.
Enoch, David. 2011. *Taking Morality Seriously: A Defense of Robust Realism*. Oxford: Oxford University.
Ewing, Alfred C. 1947. *The Definition of Good*. London: Hyperion Press.
Ewing, Alfred C. 1959. *Second Thoughts in Moral Philosophy*. London: Routledge and Kegan Paul.
Fitzpatrick, William. 2015. "Debunking Evolutionary Debunking of Ethical Realism", *Philosophical Studies*, 172: 883–904.
Frijda, Nico. 1986. *The Emotions*. Cambridge: Cambridge University Press.
Gibbard, Allan. 1990. *Wise Choices, Apt Feelings*. Cambridge, MA: Harvard University Press.
Griffiths, Paul E. 1997. *What Emotions Really Are: The Problem of Psychological Categories*. Chicago: University of Chicago Press.
Huemer, Michael. 2016. *Approaching Infinity*. London and New York: Palgrave Macmillan.
Hume, David. 1739–1741. *A Treatise of Human Nature*, edited by David F. Norton and Mary J. Norton. Oxford: Oxford University Press, 2000.
James, William. 1884. "What is an Emotion?" *Mind*, 9: 188–204.
Joyce, Richard. 2006. *The Evolution of Morality*. Cambridge, MA: MIT Press.
Kahane, Guy. 2011. "Evolutionary Debunking Arguments", *Noûs*, 45: 103–25.
Kauppinen, Antti. 2013. "A Humean Theory of Moral Intuition", *Canadian Journal of Philosophy*, 43: 360–81.
Kenny, Anthony. 1963. *Action, Emotion and Will*. London: Routledge and Kegan Paul.
Kitcher, Philip. 2006. "Biology and Ethics." In *The Oxford Handbook of Ethical Theory*, edited by David Copp, 163–85. Oxford: Oxford University Press.
Kitcher, Philip. 2011. *The Ethical Project*. Cambridge, MA: Harvard University Press.
Lange, Carl G. 1885. "The Emotions: A Psychological Study." In *The Emotions*, edited by Carl G. Lange, and William James, 33–90. Baltimore: Williams and Wilkins, 1922.
Mackie, John L. 1977. *Ethics: Inventing Right and Wrong*. London: Penguin.
McDowell, John. 1985. "Values and Secondary Qualities." In *Morality and Objectivity: A Tribute to John Mackie*, edited by Ted Honderich, 110–29. London: Routledge and Kegan Paul.
Meinong, Alexius. 1917. *On Emotional Presentation*. Translated by Marie-Luise Schubert Kalsi. Evanston Ill: Northwestern University Press.
Moore, George E. 1903. *Principia Ethica*. Cambridge: Cambridge University Press, revised edition 1993.

Moore, George E. 1922. "The Conception of Intrinsic Value." In *Philosophical Studies*, edited by Moore, George E. (253–275). London: Routledge and Kegan Paul, 1965.

Mulligan, Kevin. 1998. "From Appropriate Emotions to Values", *The Monist*, 81: 161–88.

Nichols, Shaun. 2004. *Sentimental Rules: On the Natural Foundations of Moral Judgment*. New York: Oxford University Press.

Nussbaum, Martha C. 2001. *Upheavals of Thought*. Cambridge: Cambridge University Press.

Oddie, Graham. 2009. *Value, Reality, and Desire*. New York: Oxford University Press.

Oddie, Graham. 2014. "Thinking Globally, Acting Locally: Partiality, Preferences and Perspective", *Les Ateliers de l'éthique/The Ethics Forum*, 9: 57–81.

Pettit, Philip. 1991. "Realism and Response-Dependence", *Mind*, 100: 587–626.

Prinz, Jesse. 2004. *Gut Reactions: A Perceptual Theory of Emotion*. New York: Oxford University Press.

Prinz, Jesse. 2007. *The Emotional Construction of Morals*. New York: Oxford University Press.

Putnam, Hilary. 2002. *The Collapse of the Fact/Value Dichotomy and Other Essays*. Cambridge, MA: Harvard University Press.

Railton, Peter. 1996. "Moral Realism: Problems and Prospects." In *Moral Knowledge? New Readings in Moral Epistemology*, edited by Walter Sinnott-Armstrong and Mark Timmons, 49–81. New York: Oxford University Press.

Ruse, Michael. 1998. *Taking Darwin Seriously*. New York: Blackwell.

Ruse, Michael, and Wilson, Edward O. 1986. "Moral Philosophy as Applied Science", *Philosophy*, 61: 173–92.

Sayre-McCord, Geoffrey. 1988. "Moral Theory and Explanatory Power." In *Midwest Studies in Philosophy*, edited by P. A. French, T. E. Uehling, and H. Wettstein, 433–56. Minneapolis: Minneapolis University Press.

Scanlon, Thomas M. 1998. *What We Owe to Each Other*. Cambridge, MA: The Belknap Press.

Scarantino, Andrea. 2014. "The Motivational Theory of Emotions." In *Moral Psychology and Human Agency*, edited by Daniel Jacobson, and Justin D'Arms, 156–85. New York: Oxford University Press.

Shafer-Landau, Russ. 2012. "Evolutionary Debunking: Moral Realism and Moral Knowledge", *Journal of Ethics and Social Philosophy*, 7: 1–37.

Skarsaune, Knut. 2010. "Darwin and Moral Realism: Survival of the Iffiest", *Philosophical Studies*, 152: 229–43.

Solomon, Robert C. 1976. *The Passions*. Indianapolis, Ill: Hackett Publishing Company.

Street, Sharon. 2006. "A Darwinian Dilemma for Realist Theories of Value", *Philosophical Studies*, 127: 109–66.

Sturgeon, Nicholas. 1984. "Moral Explanation." In *Morality, Reason, and Truth*, edited by David Copp, and David Zimmerman, 49–78.Totowa, NJ: Rowman and Allenheld.

Tappolet, Christine. 1995. "Les émotions et les concepts axiologiques." In *La Couleur des Pensées*, edited by Patricia Paperman and Ruwen Ogien, *Raisons Pratiques*, 6: 237–257.

Tappolet, Christine. 2000. *Emotions et valeurs*. Paris: Presses Universitaires de France.

Tappolet, Christine. 2016. *Emotions, Values, and Agency*. Oxford: Oxford University Press.

Tappolet, Christine, and Rossi, Mauro. 2015a. "What Is Value? Where Does It Come From? A Philosophical Perspective." In *The Handbook of Value: The Affective Sciences of Values and Valuation*, edited by Tobias Brosch and David Sander, 3–22. Oxford: Oxford University Press.

Tappolet, Christine, and Rossi, Mauro. 2015b. "Emotions and Wellbeing", *Topoi* 34: 461–74.

Teroni, Fabrice. 2007. "Emotions and Formal Objects", *Dialectica*, 61: 395–415.

Tye, Michael. 2008. "The Experience of Emotion: An Intentionalist Theory", *Revue Internationale de Philosophie*, 243: 25–50.

Vavova, Katia. 2014. "Debunking Evolutionary Debunking." In *Oxford Studies in Metaethics*, Vol. 9, edited by Russ Shafer-Landau, 76–101. Oxford: Oxford University Press.

Vavova, Katia. 2015. "Evolutionary Debunking of Moral Realism", *Philosophy Compass*, 10: 104–16.

Whiting, Demian. 2011. "The Feeling Theory of Emotion and the Object-Directed Emotions", *European Journal of Philosophy*, 19: 281–303.

Wielenberg, Erik J. 2010. "On the Evolutionary Debunking of Morality", *Ethics*, 120: 441–64.

Wiggins, David. 1976. "Truth, Invention and the Meaning of Life." In his 1998, *Needs, Values, Truth: Essays in the Philosophy of Value*, 87–127. Oxford: Basil Blackwell.

Wiggins, David. 1987. "A Sensible Subjectivism." In his *Needs, Values, Truth: Essays in the Philosophy of Value*, 185–214. Oxford: Basil Blackwell Press.