

Moral Sentimentalism in Counterfactual Contexts: Moral Properties Are Response-Enabled

Abstract

According to moral sentimentalism, there are close connections between moral truths and moral emotions. Emotions largely form our moral attitudes. They contribute to our answerability to moral obligations. We take them as authoritative in guiding moral judgement. This role is difficult to understand if one accepts a full-blown moral realism, according to which moral truths are completely independent of our emotional response to them. Hence it is tempting to claim that moral truths depend on our emotional responses. I outline a problem for this view: we are adamant that, if our moral sentiments were different, things would be the same, morally speaking. Moral truth does not seem to counterfactually depend on moral sentiments. I show how this independence can be reconciled with the role of moral sentiments in guiding our moral outlook. I draw on Yablo's distinction between response-dependent and response-enabled properties. I propose that moral truths are response-enabled: their supervenience base does not include anything about our emotions. Hence they do not counterfactually depend on changes in our emotional response. However, their factual supervenience base being naturally ineligible, it is ultimately our response that enables them to play their role as an independent moral compass.

1. Sentimentalism and Moral Truth

Consider an exemplary moral claim:

(TORT): It is morally wrong to torture people for pleasure.

There is a strong intuition that TORT is not only truth-apt; it is true, just as it is true that Earth is a planet. However, there is something peculiar about moral truths like TORT. Philosophers in the tradition of *moral sentimentalism* have noted that our stance toward moral truths seems to differ in several respects from our stance towards morally neutral truths.¹ I present some general intuitions, which I shall take for granted as a background of my discussion. I aim at a more general characterisation of the relationship between moral properties and emotions rather than an overview of moral sentimentalism.

First, moral beliefs systematically covary with certain emotions. If one believes TORT, one will normally feel a certain way about torturing people for pleasure. If one considers a situation where people are tortured for pleasure, one feels abhorrence.

Second, one normally accepts a moral truth like TORT not purely intellectually but embraces it. Embracing a moral truth involves a whole pattern of emotions and motivations to act. Even if one is not motivated to act and consequently will not act morally, one may feel bound to do so. Emotions partially account for this feeling. For instance, one may feel the prick of conscience if one does not act in accordance with one's moral assessment of a situation. The phenomenon of conscience may involve detached deliberation. A Kantian may check her past actions by the categorical imperative and come to reproach herself. But when we talk of a 'prick of conscience', often emotions will be involved, for instance sadness and remorse about an action which one judges to be morally wrong.

Third, not only does moral belief go together with a certain emotional response, often this response plays a role in our acceptance of moral claims. Our moral outlook is largely formed by our emotions.² One may believe TORT at least in part because one reacts with

¹ On moral sentimentalism cf. Kauppinen (2014).

² There is empirical evidence that 'moral judgement depends essentially on tacit affective processes.'(Gerrans and Kennett 2010, 586; in their critical review of this hypothesis, Gerrans and Kennett also give a survey of the literature). This evidence is compatible with some of the relevant processes not being tacit.

abhorrence to scenarios where people are tortured, be they imaginary or real. The role of emotions here is not merely causal but also rationalising. Our emotional reaction makes us *see* that it is wrong to torture people for fun. We treat our emotions as authoritative with respect to moral truth.

I shall elaborate a bit on the third point: one may often be in a position to give reasons for particular moral beliefs, but these reasons do not simply replace one's emotional response as an independent source of our moral stance. One's emotional response may be modified by reasoning processes, but it continues to play an independent role. Emotions are not fully penetrable to reasoning, and the same goes for their influence on our moral attitudes. For instance, it seems implausible that one will fully embrace an action as morally right although one has a persistent feeling of abhorrence towards it. At least a feeling of cognitive dissonance is to be expected. It will not do to simply disqualify the emotion as inappropriate in light of one's reasons (as one may use background knowledge to correct one's visual impression that a stick in water is bent). This suggests that there is at least some sort of reflective equilibrium between emotions and moral reasoning. Emotions have an authority on moral questions which cannot simply be dismissed.

In the following, I shall assume that these sentimentalist tenets are correct. The question becomes what to make of them. I oppose two salient reactions. A moral realist may insist that moral truths are independent of our emotional response. She may try to explain the observed relationships to emotions as follows: independent moral truths set a standard to which our sentiment has to be attuned. If it is properly attuned, it can guide our attitude towards moral truth. However, this explanation needs to be elaborated. Perhaps emotions just like perception reveal certain independent facts, in particular moral ones. But as contrasted to sense perception, the function of emotions does not seem exhausted by detecting certain independent facts.

It is difficult to understand how realism could account for the features that distinguish our embracing moral truths from our attitude towards facts like 'Earth is a planet'. In order to

avoid positing spooky high-level properties which are not grounded in the more fundamental make-up of the world, one should expect independent moral facts to supervene on more basic facts, for instance physical ones. Moral facts are tied to this supervenience base, and it is part of identifying moral facts that one reliably tracks the supervenience base. But how can tracking this supervenience base give rise to an attitude that goes beyond mere cognition of facts?³ How can it explain our peculiar *answerability* to moral truth? How can it explain the latter's *motivational* force? An answer in the spirit of sentimentalism is that our emotions account for the remainder beyond cognition of facts. This answer is compatible with a certain kind of realism. Indeed I shall outline a proposal how to spell out the resulting division of labour between moral sentiment and independent moral truth. So far I only note a challenge to moral realism.

A realist position faces the challenge to account for the function of moral emotions. The outstanding alternative reaction to this function is to deny that there are independent moral truths and to claim that emotional reactions both make for moral truth and our attitude towards it.⁴ The view that moral truths are constituted by emotional reactions allows to explain why we take our emotional response to be authoritative in settling moral questions. Moral sentiment reveals moral truth because it makes things as it depicts them.

While the move towards a constitutive role of emotions allows to explain the special relationship between our moral stance and moral sentiments, there is a suspicion that it comes with an undue relativity of moral truths to changes in our reactive dispositions. I shall elaborate this suspicion by considering a specific class of counter-intuitions. Then I shall

³ Of course, often our attitude towards external facts will not be that of a neutral observer. But our moral attitudes are significantly different from evaluating a situation by one's personal preferences. In the latter case, independent truths do not put up a standard for our attitude.

⁴ One may even question that moral discourse is truth-apt, but this leads to Frege-Geach-style problems. I shall not pursue this option and assume that moral discourse is truth-apt.

outline an alternative way of accounting for the special role of moral sentiments. My alternative allows to reconcile the role of emotions in forming our attitude towards moral truths with a robust realism about their metaphysical supervenience base.

A paradigmatic way of spelling out a constitutive role of emotions for moral truth is to treat moral properties as *response-dependent*.⁵ I shall assume that moral truths can be interpreted as attributions of moral properties like (*morally*) *wrong*. As a first stab, a response-dependent property P can be characterised by the following biconditional, making use of normalcy conditions (conditions C):

X is P if and only if for any subject S: if conditions C obtain, then S judges that X is P. (DeClerq 2002, 160)

In section 4, I shall replace this characterisation by a modally qualified one, but it will do so far. For instance, sugar is sweet iff everyone judges it to be sweet under conditions C.⁶ A certain gustatory perception, resulting from the proper functioning of the subject's tastebuds in their proper environment, leads to the judgement that sugar is sweet. The analogous move for the moral property *wrong* is that an action is wrong iff any subject judges it to be wrong under conditions C. When conditions C apply, a certain emotional response occurs, which leads to the judgement that the action is wrong.

The biconditional binds response-dependent properties to judgements. Accordingly, C-conditions are not framed as conditions of experiencing the property but conditions of judging that the property applies. One may object to this focus on judgements. For instance, surely small infants experience cookies as sweet, but they may not be disposed to judge them to be sweet. In response, the biconditional does not limit the *experience* of response-dependent

⁵ On the view that moral properties are response-dependent cf. the supplement to Kauppinen (2014).

⁶ 'Under conditions C' is read as: 'if conditions C obtain'.

properties to judging subjects. Children may track sweetness as reliably as adults without being disposed to judge. All the biconditional does is to provide a standard which settles when the property applies to an object. The biconditional does not preclude subjects who are not disposed to judge from engaging with that property. The only restriction is that subjects who are not disposed to judge do not directly figure in the standard for ascribing the property. In the moral case, subjects who are not disposed to judge would not be precluded from being moral and even figuring in the constitution of moral truth. They are only precluded from figuring in the standard settling when a moral property applies.

It has turned out to be difficult to spell out the conditions C for standard response-dependent properties. Moral properties come with additional challenges. Moral standards are highly dependent on culture, education, reflectivity, psychological sensitivity and so on. In that they resemble high-level properties like aesthetic properties. However, many would insist that they are distinct from aesthetic properties regarding the possibility of error. We want to say that whole periods of history were wrong in accepting slavery, but we might be more reluctant to say the same about their aesthetic valuations.

I shall make do with the simplifying assumption that all these difficulties can be surmounted by sufficiently refined conditions of *apt* moral feelings, which are written into the conditions C (cf. Gibbard 1992, ch. 7, Salmela 2006). I shall give some hints how to characterise aptness. It is required that conditions C ‘...can be specified in a *substantial* way, i.e., in such a way that nothing about the extension of the predicate “P” is presupposed.’ (DeClerq 2002, 160-161) In this vein, I assume that aptness can be characterised in terms of proper functioning without writing substantial moral truths into the conditions of aptness by hand. I add a highly tentative list of further conditions subject to enlargement and revision. The list is to ensure that whole periods can go wrong in their moral judgements: First, the judging subject should be fully informed; no additional factual information about the scenario considered should alter the judgement.

Second, the judgement should be considerate: no additional level of reflection should alter the judgement.

Third, the judging subject should not suffer from psychiatric disorder.

Fourth, the judging subject should have normally functioning cognitive capacities.

Fifth, the judgement should not vary with the degree of personal involvement of the judging subject. For instance, the judgement should not be different depending on whether the persons tortured are personally known to the subject considering TORT.

Even if all these prerequisites for the conception of moral truth as response-dependent to succeed are granted, still one key problem remains. This problem will be the subject of the next section.

2. The Puzzle: Moral Intuitions in Counterfactual Contexts

I shall now present specific counter-intuitions to the view that moral properties are response-dependent. These intuitions allow to contrast moral properties with clearly response-dependent properties. A paradigm response-dependent property is *sweet*. Brownies are sweet iff any observer under conditions C judges brownies to be sweet. She will do so because she has a certain gustatory experience when tasting brownies under conditions C. *Fear* is an emotion that gives rise to a response-dependent property. An object is *frightening* iff any subject under conditions C will judge it to be so. If conditions C apply, the subject experiences fear, which leads her to judge that the object is frightening.

Now consider how these properties behave in counterfactual scenarios. It seems perfectly in order to say:

(1) If our taste were different, brownies would not be sweet.⁷

⁷ Things might be a bit more complicated. In a certain context, one may feel inclined to accept

If our taste were different, we would not have a certain gustatory experience under conditions C and hence not judge brownies to be sweet. It also seems perfectly in order to say:

(2) If we were not disposed to fear lions, lions would not be frightening.

If we were not disposed to fear lions, we would not experience fear from lions under conditions C and hence not judge lions to be frightening. The same holds for properties like *wrong by convention*:

(3) If our conventions were different, driving through a red traffic light would not be wrong.

However, moral judgements significantly differ from response-dependent judgements of the sort just considered. This can already be seen in the reactions of children as reported by Neil Levy:

‘With regard to conventional transgressions (e.g. a boy wearing a dress to school), children were likely to say that if the teacher said it was permitted, it was (even if they had initially said that the behavior was wrong). With regard to moral transgressions (e.g. one child hitting

If our taste were different, brownies would still be sweet in virtue of containing sugar.

If this is right, it shows that the status of *sweet* and the like as response-dependent is not always clear. Sometimes we may be driven towards what I call a response-enabled property, depending on which sort of constitutive relationship we focus on. Still the contrast to moral properties is striking. I shall largely ignore the subtleties and treat *sweet* as a response-dependent property.

another) children were likely to maintain that the behavior remained wrong, even if the teacher permitted it.’(Levy 2005, 232)

One may say that the teacher is treated as authoritative with regard to conventions but not morals. But Levy takes his evidence to support a stronger point: children treat moral norms as necessary and thus not susceptible to counterfactual variations.

In the same vein, we adults tend to reject claims like

(4) If our moral sentiments were substantially different, torturing people for pleasure might not be wrong.

I use a ‘might’-counterfactual because it is weaker than the ‘would’-counterfactual: (4) says not that any situation but only that some relevant situations where our sentiments are substantially different are such that torturing people for pleasure is not wrong. It is more significant to see us resist even this weaker claim. The findings from developmental psychology are evidence that our intuitions about (4) are deeply entrenched in basic habits of thinking.⁸

⁸ There is a striking contrast between our rejection of (4) and our acceptance of the following indicative conditional, which is intuitively true:

If our moral sentiments about torturing people for fun are not in line with moral truth, torturing people for fun might not be wrong.

How does this contrast arise? I draw on the widely shared assumption that indicative conditionals are epistemic. They invite to consider the consequences of a hypothesis about what is actually the case. They force us to play through a revision of our beliefs by learning that our moral sentiments actually fail to track moral truth. The result is that the moral

Our resistance to (4) does not square well with the assumption that moral properties are response-dependent. Under that assumption, one should expect the following: if our sentiments were substantially different, we might not feel the way we actually feel towards torturing people under conditions C and hence not judge it wrong. As a consequence, it might not be wrong. Hence we should accept (4) if moral properties are response-dependent.

One caveat about the antecedents of conditionals like (1)-(4): they may seem underspecified, but this impression can be dispelled. They just require some implicit contextual supplement, which is provided by our tendency to cooperate in philosophical inquiry. I could also formulate the antecedents in a more convoluted way to make the contextual supplement explicit: if our moral sentiments with respect to torturing people were under conditions C such as to find torturing people for pleasure permissible..., but processing such a demanding explicit antecedent may spoil intuitions. In my view, the less convoluted antecedents, read in light of a suitable contextual supplement, are perfectly in order.

Summarising, typical response-dependent properties and moral properties behave differently under counterfactual suppositions. One should expect response-dependent truths to counterfactually covary with our response. But our intuitions about moral judgements do not display this kind of variability. My aim is to account for these findings.

3. Solutions I Reject

statements based on these sentiments might be false, among them TORT. In contrast, counterfactual conditionals usually do not invite to consider epistemic hypotheses about what is actually the case. They invite to consider possible situations which are taken to differ from the actual one. One upholds one's belief that actually, moral sentiments track moral truth, and considers a different world where one's feelings are different.

Before drawing my own conclusion, I shall discuss some reactions to these findings which I find not entirely convincing.⁹

I have already indicated that (assuming moral properties are response-dependent), in order to be constitutive of moral standards, moral sentiments have to be apt. One quick reaction to the puzzle outlined in the last section is to write into the conditions of aptness that moral sentiments, at least apt ones, *could not be* such as to license torturing people for fun. I cannot rule out that an account along these lines might succeed, but it comes with problematic commitments. I doubt that any conditions of aptness of the sort outlined in my tentative list will entail these commitments. Rather one will have to write moral standards into the conditions of aptness by hand; as a consequence, one will have to violate the constraint that normalcy conditions C for a response-dependent property P can be specified in a *substantial* way, without presupposing anything about the set of scenarios P applies to.

I shall turn to two accounts in the literature. Neil Levy provides an account in terms of imaginative resistance:

‘Some concepts are such that they apply *simply* in virtue of some more basic facts obtaining. Call these basic facts the supervenience base of the concept. Imaginative resistance in us is provoked if the author tells us (or gives us reason to believe) that the supervenience base of concept C, which is such that it applies simply in virtue of more basic facts, obtains in the story, and then denies that C applies (or vice versa).’(Levy 2005, 236)

In Levy’s account, moral concepts are analytically tied to a certain supervenience base. For instance, the concept (*morally*) *wrong* by conceptual necessity applies whenever someone is tortured for fun (this is only one of many applications, of course). The latter fact is a sufficient condition for applying *wrong* and thus part of the concept’s supervenience base. *Imaginative*

⁹ I dismiss the simple solution that our intuitions about counterfactuals like (4) are misguided; it would lead to an unattractive error theory.

resistance arises when one is asked to imagine that the factual conditions for applying a moral concept are satisfied but the concept does not apply.

For Levy's explanation to succeed, moral concepts must be tied to a supervenience base which does not include facts about us, as contrasted to the supervenience base of *conventionally wrong*. The latter does include facts about us, namely our conventions. Again we are faced with the challenge to make sense of the special role of emotions in our moral outlook. At some point Levy seems to acknowledge this role. He grants that our attitudes may be included in the supervenience base:

‘... our affective responses can figure in the supervenience base of a concept...’(Levy 2005, 239)

The idea seems to be that *wrong* does not simply apply in virtue of a situation being one where someone is tortured for fun but only in virtue of a broader supervenience base, including our attitude towards people being tortured for fun. Yet the straightforward conclusion is that application conditions should counterfactually covary with apt moral sentiments, contrary to (4). Thus, for Levy's explanation of counterfactual stability to succeed, he has to exclude that our affective responses actually figure in the supervenience base of a moral concept.

In sum, Levy's account may explain the difference between *morally wrong* and *conventionally wrong* and the like, but it remains open how to account for the role of emotions. The only explanation of this role he offers is that emotions might figure in the supervenience base of a concept. But this approach is irreconcilable with his account of the difference between morals and conventions. In the next section, I will provide an account which reconciles Levy's view of the supervenience base of moral concepts or properties with the role of emotions in forming our moral stance.

I shall now turn to a second proposal. Drawing on his projectivist metaethics, Simon Blackburn explains the observation that we reject (4) as follows:

‘...suppose a projective theory must involve us in believing things like “If we had different attitudes it would not be wrong to kick dogs”... Then clearly it is refuted, because these things are absurd. Fortunately, however, the projective account of indirect contexts shows quite clearly how to avoid them. The counterfactual “If we had different attitudes it would not be wrong to kick dogs” expresses the moral view that the feature which makes it wrong to kick dogs is our reaction. But this is an absurd moral view, and not one to which a projectivist has the least inclination. Like anyone else he thinks that what makes it wrong to kick dogs is that it causes them pain. ... A projectivist is only tangled in these unlovely counterfactuals if he makes the mistake of thinking that after all there is a state of affairs making the projected commitment true, only one about us.’ (Blackburn 1981, 179)

To Blackburn, counterfactual dependence tracks some ‘in virtue of’-relation. ‘If p were not the case, q would not be’ is used to express that p makes it the case that q. For instance, (4) conveys that it is wrong to torture people for fun because our moral sentiments are the way they are. If moral sentiments were not as they are, it would not be wrong to torture people for fun. But it is not our reaction which makes torturing people for fun wrong, says Blackburn.

Again the problem with this view is how to account for the difference between response-dependent properties like *sweet* and moral properties like *wrong*. Blackburn emphasises that there is a difference in kind between two relationships: the relationship between a moral truth like ‘kicking dogs is wrong’ and our attitudes on the one hand and the relationship between the moral truth and the natural fact that kicking dogs causes them pain on the other hand. Only the second relationship is a relationship of making something the case.

However, it is not clear what the difference to clearly response-dependent properties is.

It seems true that

(1) If our taste were different, brownies would not be sweet.

Now it seems that in the case of *sweet*, a similar distinction applies as in the case of *wrong*. On the one hand, there is the relationship to our taste. On the other hand, there is the relationship to the natural fact that brownies contain sugar. There is a sense in which one may say that what makes brownies sweet is their containing sugar. Whatever the constitutive role of the relationship to our taste, it seems different from the way in which brownies containing sugar makes it the case that brownies are sweet. Hence there is an argument that is perfectly analogous to Blackburn's argument about *wrong*: On a certain natural understanding of *making something the case*, it is their containing sugar and not our reaction to them that makes brownies sweet.¹⁰ This natural understanding is the one that matters to the pastry-maker who chooses to add sugar to the dough in order to make the brownies sweet. We need additional reasons to treat the case of *wrong* and the case of *sweet* differently.

Notwithstanding my criticism, the solution to be presented in the next section can be read as a partial vindication of Blackburn's claim. It lays out one relevant metaphysical 'in virtue of'-relation. This relation underpins a reading of 'because' according to which kicking dogs is wrong because it causes them pain but brownies are sweet partially because our tastebuds are the way they are. Moreover, this relation seems relevant to judging counterfactual dependence.

4. My Solution: Sentiment-Based Moral Properties Are Response-Enabled

4.1 Response-Enabled Properties

¹⁰ There is a connection to the uncertainties about the status of properties like *sweet* outlined in footnote 7.

Having critically reviewed selected reactions, I shall now propose my own way of accounting for the modal stability of moral properties, based on a proposal of Stephen Yablo. Yablo identifies a class of properties that are not *response-dependent* but *response-enabled*.

Key to understanding response-enabled properties is the classification of properties as naturally more or less eligible. The distinction is by no means a matter of course, but I shall assume that something like it can be made to work. The distinction can be neatly illustrated by an example due to Ted Sider (2011, 1): consider a universe which is filled with liquid. The liquid is perfectly homogeneous, the only difference being that there are two hemispheres. The liquid in one hemisphere is red and the liquid in the other is blue. There are many two-dimensional planes which may be used to divide three-dimensional space into hemispheres. However, the only naturally eligible division, the only one that carves nature at the joints, is the one dividing the hemisphere with the red liquid from the hemisphere with the blue liquid.

Accordingly, naturally eligible properties E1, E2, E3... are joint-carving in that the things falling under them form naturally eligible classes, e.g. red liquid vs. blue liquid. A pure enquirer, 'writing the book of the world', will find only E1, E2, E3... fit for being included in her theory of the world. In contrast to the naturally eligible properties, there are less eligible ones, which can be defined by the more eligible ones. We may define a property F as follows: x is F iff: x is E1 and E2 but not E3, or x is E3 and either E1 or E2 but not both. This property will be perfectly objective but less naturally eligible than E1, E2, E3... However, it may from some perspective be more conspicuous and play a more important role than the perfectly natural properties.

For instance, if we take fundamental physics to tell us what is naturally eligible, *having positive charge* and *having upwards-directed spin* are candidates for naturally eligible properties, *having positive charge or upwards-directed spin* does not seem (as) natural. However, it may well be that, for whatever reason, *having positive charge or upwards-directed spin* seems more conspicuous to a being with certain sensibilities. For instance, the

being may feel an itch precisely when approaching a positively charged particle or a particle with upwards-directed spin. Being only interested in avoiding the itch and not in a position to discern the underlying perfectly eligible properties, the being only attends to the less natural property which precisely covaries with *having positive charge or upwards-directed spin* rather than to the naturally eligible ones.

The distinction between naturally eligible and ineligible properties looms in the background of Yablo's distinction. His example of a response-enabled property is *oval*:

'A thing in [any world] w is oval if it is of a shape that would strike *me* as egg-shaped were I (with my sensibilities undisturbed) given a chance to look at it.' (Yablo 2002, 465)

It is crucial that 'sensibilities undisturbed' here means that sensibilities and conditions of their proper functioning are not changed compared to the actual world. Our actual reactions settle the extension of 'oval' in any counterfactual situation.

On Yablo's construal, the property *oval* is characterised as follows: on the one hand, there is a fixed mathematical structure that underlies *oval*. It can be described by a highly complex condition exclusively in terms of more eligible mathematical properties. Someone who combines immense cognitive capacities with comprehensive knowledge of both mathematics and of our dispositions to respond to visual shapes by calling them oval may state the condition for a thing to be *oval* in purely mathematical terms. She does not even need vision to do so. On the other hand, this structure fails to be mathematically eligible.¹¹ From a purely mathematical standpoint, it is uninteresting and gerrymandered. The very point of selecting this structure among the overwhelmingly many other mathematical structures lies in our perceptual endowment. The boundaries of the property are identified by what a competent

¹¹ *Oval* partly overlaps with certain eligible mathematical concepts, but not perfectly. For instance, mathematically eligible *cassini ovals* comprise many (idealised) everyday oval shapes. But there are some cassini ovals no untutored person would call oval.

observer actually judges to be oval under conditions C because it seems shaped like an egg. The property seems relevant only against the background of our practices of grouping egg-shaped things together.

The role of *oval* in our world view is not owed to its eligibility from a mathematical point of view but entirely to our dispositions to react in a certain way to visual shapes. Our perceptual endowment and sense of salience make us lump these shapes together due to their distinctive phenomenology. So far response-enabled properties perfectly resemble response-dependent properties like *sweet*. One may also figure out the natural make-up of the heterogeneous things we find sweet. Yet the decisive difference is that the supervenience base of a response-dependent property includes our reactions under C-conditions. Hence the extension of *sweet* counterfactually covaries with our taste. In a situation where our taste is different, the natural features of sweet things are different. In contrast, the extension of *oval* does not counterfactually covary with our sensibilities but with the mathematical structure, however gerrymandered.

The supervenience base of response-enabled properties includes only a highly complex configuration of naturally eligible properties. It does not include the reaction to these natural features by which we usually identify the properties. Since the supervenience base is independent of our perceptual endowment, the following is true:

(5) If our perceptual endowment were arbitrarily different, things which would have the geometrical shape of actual eggs would still be oval.

Our actual response *enables* us to identify an independent property; this property does not depend on our response. Although it is independent of our response, it is only in light of that response that the mathematical structure of oval things as selected from myriads of other gerrymandered mathematical structures plays a unified role in our dealing with the world.

Response-enabled properties can be further illustrated by an analogy to proper names in a Kripkean (1980) view. The role of our perceptual endowment in forming expressions like ‘oval’ corresponds to the role of a reference fixer, say a description used to select the referent of a proper name. Since the proper name is a rigid designator, its bearer does not have to satisfy the description in any counterfactual situation. For instance, we may fix the reference of *Aristotle* by describing him as Alexander’s teacher. But in a certain counterfactual situation, Aristotle would not have been Alexander’s teacher. The extension of the name is tied in all counterfactual situations to the actual physical make-up and origin of the person to which we actually gave the name. In the same vein, we identify the actual extension of ‘oval’ by what we find egg-shaped. But the extension of ‘oval’ in all counterfactual situations is tied to a certain geometrical structure. In a counterfactual situation where our perceptual endowment is different such as to yield a visual phenomenology of eggs as square, eggs would still be oval due to their having the same geometrical form as actual eggs.

Yablo does not give a precise characterisation of response-enabled as contrasted to response-dependent properties. I shall try to supply one. I modify the initial characterisation of response-dependent properties by a modal clause. The initial characterisation taken from DeClerq was:

X is P if and only if for any subject S: if conditions C obtain, then S judges that X is P.

I replace the initial condition by a modally qualified one:

For any possible world w , any x , and human subject S , at w , x is P if and only if: if conditions C obtain, then S judges that x is P.¹²

¹² ‘at w ’ takes scope over the whole rest of the sentence.

The response-enabled property R is a bit more complicated to characterise, using a natural supervenience base B. B is a configuration of naturally eligible properties E1, E2, E3... R can be characterised as follows:

For any possible world w, any x, and any human subject S: at w, x is R if and only if x has some natural property B such that: *actually*, if conditions C obtain and x is B, then S judges that x is R.

Natural properties are real properties instantiated by the material world whether we know it or not. The naturally eligible properties are a subset of the natural ones. I find it most convenient to distinguish response-enabled properties from the natural properties on which they supervene. The difference is that, just as in the case of response-dependent properties, we have special epistemic access to them: if conditions C obtain and a response-enabled property is instantiated, we judge that it is. The prize of this privileged access is that the property is not ensured to be naturally eligible. For the condition of a response-enabled property is not that its supervenience base is naturally eligible but that we tend to identify it as one property under suitable conditions. If one does not share our sensibility, there is no reason to select one gerrymandered natural supervenience base rather than another. Without constraints in terms of natural eligibility, any natural supervenience basis can be selected, just as it fits the sensibilities of the observing subject.

The crucial difference to response-dependent properties lies in the actuality operator. The natural supervenience base of a response-dependent property P (and thus the extension of P) changes from one possible world to the other together with our (the evaluating subjects') response under conditions C. In contrast, the natural supervenience base B in virtue of which a response-enabled property R applies to objects does not change from one world to the other. R is tied to B as selected by our response in the actual world. Response-dependent properties counterfactually depend on our reaction, response-enabled ones do not.

4.2 Response-Enabled Moral Properties

I shall now transfer the lesson to morals. Response-enabled properties are precisely what we need to reconcile a sentiment-based view with modal stability and eventually with moral realism. If moral properties are response-enabled, we can explain the privileged role of emotions in identifying them. Just as *oval*, *wrong* seems an interesting property to us mainly in light of our apt emotional reactions to independent features of the world.

The role of our response in tracking moral properties can be illustrated by the contrast to someone who does not share our reactive dispositions. Someone whose perceptual endowment substantially diverges from ours can mimic our use of the word ‘oval’ by tracking the mathematical truth-conditions of statements about things being oval, but she won’t find the property significant purely from her own viewpoint. She will not be disposed by herself to group oval things together and act accordingly. Her recognition of *oval* will be parasitic on our reactive dispositions. In the same vein, take someone who simply does not share our moral sentiments but knows which worldly facts figure in the supervenience base of moral properties. She can figure out the moral truths (it is wrong to torture people for pleasure, it is wrong to kick dogs...) in virtue of tracking the supervenience base; she may even form the intention and be motivated to do what is right and shun what is wrong; but her perspective on moral truth will inevitably be parasitic on ours. She has to know what we call right and wrong to find the same moral properties salient as we do. Our moral sentiments form part of the original perspective that provides access to the moral properties which figure in our true moral judgements. This perspective makes the moral properties stand out among naturally equally ineligible features of reality.

This is not to say that reasons cannot play a role in figuring out moral truth, but their role will have to be negotiated with moral sentiment. From the viewpoint of a pure enquirer ‘writing the book of the world’, there are no interesting properties wrong actions share in

being wrong. They are made salient to us as sharing a common feature partly by our common emotional reaction. For instance, people being tortured for pleasure and dogs being kicked surely have in common that living beings suffer pain. One may derive a more general rule that one ought to prevent suffering, which may serve as a reason why one ought to act so as to prevent suffering in a specific case. But normally the starting point for figuring out such a reason is that we have the same kind of feeling towards a range of scenarios. In each case, our abhorrence tells us that these scenarios have a common property *wrong*, and it bestows a certain role on this property: it makes us answerable to right and wrong. A rule extending the range to new cases will have to be weighed against our sentiment.

Our emotions contribute to our becoming sensitive to moral properties in the first place. But these findings are reconcilable with a robust moral realism: given a plausible understanding of dependence as counterfactual covariance, moral properties like *wrong* do not *depend* on our apt reactions. They are ‘out there’ in the world, tied to a robust factual supervenience base, which only includes a configuration of eligible natural properties and not our response to it. The extension of a moral property is perfectly independent of our reaction. I can take on board Blackburn’s point that what makes it wrong to kick dogs is that it causes them pain. There is a metaphysical relationship which obtains between *wrong* and its natural supervenience base but not between *sweet*, construed as response-dependent, and the corresponding natural property (containing sugar etc.). The relationship can be spelled out in terms of counterfactual dependence. It is sufficient for kicking dogs being wrong that it causes them pain, whether we react to dogs being kicked with abhorrence or not. And things would not change if we were to react otherwise than we do.

The hypothesis that moral properties are response-enabled does not only fit into the big picture of reconciling moral realism with the function of moral sentiment, it also allows to deal with the concrete intuitions about counterfactuals observed in section 2. Since moral properties are response-enabled, the following is true:

(6) If moral sentiments were substantially different, it would still be wrong to torture people for pleasure.

(6) is inconsistent with (4), which is consequently rejected:

(4) If our moral sentiments were substantially different, torturing people for pleasure might not be wrong.

This accounts for the counterfactual stability viz. necessity of moral verdicts.

In sum, the hypothesis that moral properties are response-enabled allows to reconcile two very distinct features of these properties: on the one hand, there is the function of apt emotions in making us see what is morally right and wrong. On the other hand, there is the intuition that our emotions do not make morally right and wrong what we feel to be so. They are supposed to conform to moral truth as an independent standard. We feel an action to be wrong because it is wrong and not vice versa. The independence of moral truth as a standard is manifested by tying the standard to the way the world is independently of us: the natural supervenience base. The standard is the same throughout any potential variation of our emotional reactions.

S. Blackburn, 'Rule-following and Moral Realism', in S. Holtzmann, C. Leich (eds.), *Wittgenstein: To Follow a Rule* (London, Routledge, 1981), 163–190.

R. DeClerq, 'Two Conceptions of Response-Dependence', *Philosophical Studies* 107 (2002), 159-177.

P. Gerrans, J. Kennett, 'Neurosentimentalism and Mental Agency', *Mind* 119 (2010), 585-614.

A. Gibbard, *Wise Choices, Apt Feelings* (Cambridge/M, Harvard University Press 1992).

A. Kauppinen, 'Moral Sentimentalism', *The Stanford Encyclopedia of Philosophy* (Spring 2017 Edition [originally 2014]), Edward N. Zalta (ed.), URL =

<<https://plato.stanford.edu/archives/spr2017/entries/moral-sentimentalism/>>

S. Kripke, *Naming and Necessity*. (Cambridge/M., Harvard University Press 1980).

N. Levy, 'Imaginative Resistance and the Moral/Conventional Distinction', *Philosophical Psychology* 18 (2005), 231-241.

M. Salmela, 'True emotions', *The Philosophical Quarterly* 56 (2006), 382-405.

T. Sider, *Writing the Book of the World* (Oxford, OUP, 2011).

S. Yablo, 'Coulda, Woulda, Shoulda', in T.S. Gendler, J. Hawthorne (eds.), *Conceivability and Possibility* (Oxford, Clarendon Press, 2002), 441-492.