

Moral Value, Response-Dependence, and Rigid Designation

BRAD THOMPSON
Southern Methodist University
Dallas, TX 75275-0142
USA

I Introduction

It is part of our notion of moral properties (certain forms of relativism to the contrary) that they are in some sense independent of our moral beliefs. A murderer cannot make his action moral simply by believing that it is so. Slavery was immoral even if a large number of people once believed that it was permissible, and it would remain so in the future even if every person came to believe that it was morally acceptable.¹ But views that take moral properties to be objective and thoroughly mind-independent constituents of reality face familiar metaphysical and epistemological obstacles.²

Furthermore, moral facts do seem to bear an intimate relationship to our moral attitudes and capacities. It is perhaps inconceivable that, at the end of moral deliberation and inquiry, fully rational human beings invested with our moral concepts could be radically incorrect in their moral beliefs. Moral properties seem to be *essentially* knowable. We hope that the fundamental truths of physics are epistemically available to us, but our conception of the physical world certainly does not *guarantee* it.

1 I shall simply assume without argument that relativism is false.

2 Such as those raised by J.L. Mackie (1977). Whether such obstacles are insurmountable is an open question. But one of the attractions of the view discussed here is that it helps avoid them.

However implausible, it is certainly conceivable that the cognitive capacities that are our evolutionary inheritance, adapted for survival in the terrestrial niche in which our ancestors found themselves, are not well suited for doing cosmology. Is a similar pessimism about our ability to know the moral facts even coherent? I don't think so. No matter how difficult it is to arrive at correct moral judgments, barring ignorance of any relevant non-moral empirical facts, the moral facts are by their very nature within our reach. And it is difficult to see how this could be so unless moral properties are in some way dependent on us.

A rigidified response-dependent account of moral properties promises to provide us with a thoroughly naturalistic moral realism that explains these senses in which moral properties are both mind-dependent and mind-independent. The response-dependent aspect of the analysis makes moral properties to a certain extent mind-dependent. But by rigidification, objectivity is secured. However, Peter Railton (1996) has argued against rigidifying on actual human responses as a way to avoid moral relativism. After an exposition of response-dependent accounts of value and their virtues, I will turn to Railton's thought experiment and show that it fails to pose a problem for rigidified response-dependence.

II Response-Dependence

In recent years, the analogy between moral properties and so-called 'secondary qualities,' such as colors, has generated considerable philosophical interest.³ One way this analogy is frequently articulated is in terms of our concepts of the corresponding properties. In particular, concepts for moral properties and concepts for colors are both said to be *response-dependent*. *C* is a response-dependent concept if a biconditional of the following form holds a priori:

x is *C* iff *x* is disposed to produce a response of type *R* in subjects *S* under conditions *K*.⁴

For example, a plausible analysis of 'red' is that something is red if and only if it is such as to produce a red sensation in normal human observers under standard conditions. Specifying what a normal human observer is, and what conditions are to count as standard, is certainly not a simple task. But presumably a normal human observer is one with a

3 See McDowell (1985), Wright (1988), Johnston (1989), Wiggins (1991).

4 This way of formulating the notion of response-dependence is from Johnston (1989). One must also add the condition that *R*, *S*, and *K* are given non-trivial specifications.

particular physiology, and standard conditions are particular types of environments and particular lighting conditions. Response-dependent accounts of color can be offered as either reductive or non-reductive analyses of color properties. If the term 'red' in the above analysis has the same meaning on both sides of the biconditional, the analysis is clearly non-reductive. But those who are realists about phenomenal properties and hold that we can secure reference to those properties without the use of color terms (as applied to physical objects) need not confine themselves to a merely elucidatory version of the response-dependent analysis. We might, for example, use Peacocke's (1983) term *red'* to rigidly denote the phenomenal quality that is typically caused by red things (fire engines, ripe tomatoes, etc.). We then have a promising reductive analysis of 'red':

x is red iff *x* is disposed to produce *red'* sensations in normal human observers under standard conditions.

Sometimes the notion of response-dependence is applied to properties rather than to concepts (e.g. Wedgwood 1998). When a property is understood as being response-dependent, it is necessarily a mind-dependent property — such as the disposition to cause certain responses in certain subjects. It should be noted that a *concept's* being response-dependent in the manner above, by contrast, leaves open whether or not the property in its extension is the corresponding disposition or the categorical grounds of the disposition (Pettit 1989).

III 'Good' as a Response-Dependent Concept

Response-dependent accounts of value can take many different forms, depending on how the responders, responses, and conditions are specified. Some have thought of value by way of analogy with secondary qualities.⁵ These views often identify moral properties as those which tend to elicit an emotion or moral sentiment. Sensibility theories, as they are sometimes called, resemble expressivism but with a cognitivist spin. One such account, which I will use as my working analysis of 'morally good' is:

x is morally good iff *x* is disposed to produce a sentiment of moral approbation in normal human beings under conditions of complete descriptive knowledge.

5 Such as McDowell (1985) and Wiggins (1991).

David Lewis (1989) offers a response-dispositional account of value that is quite different.⁶ According to Lewis' analysis:

x is valuable iff normal human beings are disposed to desire to desire x under conditions of full imaginative acquaintance with x.

Ideal observer theories, like the one offered by Firth (1952), can also be viewed as a species of response-dependent accounts. Such views are similar to Lewis', but where the relevant responder or observer is not a 'normal human being' but an agent idealized in some way, such as being fully rational or impartial.

My discussion and defense of response-dependent accounts of value will be general, intended to cover all such approaches. I will leave it open what the relevant response should be, who the relevant responders are, and how one should specify the relevant conditions. The virtues of a response-dependent analysis, however, will vary depending on how these elements of the response-dependent analysis are characterized.⁷

Many response-dependent accounts of value offer the promise of giving a purely naturalistic analysis of value. Typically, this will be a virtue of theories that provide non-circular specifications of the relevant responders, standard conditions, and responses. Naturalism might further require that the analysis not contain any additional normative terms, unless perhaps those terms themselves possess naturalistic analyses. The phrase 'normal human observers,' for example, should be read in a statistical rather than normative sense if it is to be used in a

6 However, as an anonymous referee points out, Lewis did not think that color provided a good analogy for value.

7 For instance, Elijah Millgram (1999) argues that there is a disanalogy between color experiences and moral sentiments, in that our moral sentiments in response to a type of situation can fade over time. For example, a person might react quite differently the first time she has encountered a street person versus the hundredth time. This poses a potential problem for response-dependent accounts of value. Either we must say that moral value itself fades over time, or we must find some principled basis for discounting responses that have dulled due to habituation. One solution would be to idealize away the dulled responses, either by appealing to an ideal responder who is not subject to habituation or by making only early encounters with a situation part of the relevant conditions. A better response, I think, is to make the relevant responses the ones that are arrived at under conditions of rational reflection and full descriptive knowledge. These are not the conditions typically present during one's daily walk along city streets. So although one's overt behavior and reflexive reactions to the plight of the homeless on that walk might change, this cannot be taken as evidence that our moral sentiments when reflecting on the problem of homelessness under the relevant conditions are subject to fading.

naturalistic analysis. Such an analysis can be viewed as reductive, assuming that the way of filling out the biconditional does not introduce any entities or properties which we didn't antecedently have reason to believe in. Nothing new is added to our ontology.

Many response-dependent accounts of value are also able to accommodate internalist intuitions, although this is not a necessary feature of such accounts in general. A *modest* internalism can be secured if the relevant response is in some way connected with motivation. It is plausible that a sentiment of moral approbation, for example, might have motivational force. Insofar as such sentiments are a kind of emotional response, it is interesting to note that emotion and motivation are intimately related both psychologically and neuroanatomically.⁸ 'Action tendencies' are in fact considered by most emotion researchers to be an essential component of emotion.

One feature of internalism which a sensibility variant of response-dependence might be able to defend is the claim that if a subject believes or judges that *x* is morally good, she will be motivated (*ceteris paribus*) to bring it about that *x*. According to a sensibility theory, many of our moral beliefs are formed by our having a moral sentiment or feeling. We might say that such beliefs are formed on the basis of moral *perception*. Assuming that a positive moral sentiment toward *x* tends to motivate an action to bring it about that *x*, the above internalist conditional is satisfied by beliefs formed in this way.⁹

David Lewis' account also establishes a connection between motivation and value, by way of including desires as part of the relevant response. As Lewis succinctly puts it:

8 See for example Lazarus (1991), Izard and Ackerman (2000), and Tucker, Derryberry, and Luu (2000).

9 I'm imagining a view that takes moral sentiments to be causes of moral beliefs. Moral beliefs caused by moral sentiments will satisfy the internalist conditional if moral sentiments tend to motivate action. The moral belief, by itself, does not have motivational force — which is why I call this a 'modest' internalism. Since the moral sentiment and the moral belief are taken to be distinct psychological states, it would appear possible that a person could form a belief with the same propositional content on the basis of something other than a moral sentiment. Beliefs formed in this way would tend not to satisfy the internalist conditional.

There is another interesting possible view in the vicinity. A sensibility theorist might hold that moral sentiments somehow constitute moral beliefs, such that a moral belief cannot be formed without the having of a moral sentiment. Perhaps moral sentiments partially or wholly constitute the belief token, or perhaps moral sentiments enter into the content of moral beliefs. Either way, on such a view it would be impossible to have a moral belief without having the related moral sentiment. Consequently, all moral beliefs will satisfy the internalist conditional.

If something is a value, and if someone is one of the appropriate "we," and if he is in ideal conditions, then it follows that he will value it. And if he values it, and if he desires as he desires to desire, then he will desire it. And if he desires it, and if this desire is not outweighed by other conflicting desires, and if he has the instrumental rationality to do what serves his desires according to his beliefs, then he will pursue it. (1989, 116)

As Lewis acknowledges, the form of internalism that is secured by his theory is again very modest. But like Lewis, I think that only a weak form of internalism is acceptable.

Response-dependent accounts of value also provide us with an interesting and reasonable model of how knowledge of moral properties is possible. Those who are members of the relevant class of observers and are in the appropriate conditions are in an epistemically privileged relation to moral properties. For such observers in such conditions, the enjoying of the relevant response as specified by the response-dependent analysis is criterial for the presence or absence of the property in question. Of course, most plausible versions of response-dependence theories will have it that reaching this special epistemic location is quite difficult. Standard conditions on Lewis' theory, for example, are conditions of 'full imaginative acquaintance.' And it's attractive to hold, following ideal observer theories, that the relevant responders are fully rational. Much of the resulting moral epistemology is one that we antecedently have good reasons to accept. Moral knowledge is hard work, but not impossible. It takes a great deal of rational reflection and requires a careful consideration of the multifarious empirical circumstances. The philosophical enterprise of exchanging arguments and counterexamples can be seen in this light as a mechanism for, collectively, attempting to get closer to the epistemically privileged position from which moral knowledge is secured with the greatest certainty.

IV Response-Dependence and Objectivity

One problem with the above analyses of 'red' and 'good' is that they are relativistic. As the relevant subjective responses of human beings vary across possible worlds, so does the extension of the response-dependent concept. In order to make moral properties non-relative the approach must be able to rigidify the analysis on the actual world. This strategy for converting a relativistic analysis into an objectivist analysis can be understood by analogy with color. Colors, as some have argued, are mind-independent properties that would persist even if there were no human observers. The grass would remain green even if there were no one present to see it. Furthermore, objects would retain their colors even if we all became red-green inverted. If I were ignorant of the fact that

physiological changes had occurred and that I had become spectrum-inverted, I would believe that the grass in my immediate environment which was formerly green had become red. But this would be an error due to my ignorance. I have changed, not the grass. Green things now *look red*, so to speak.¹⁰

In order for the response-dependent analysis of color to depict color properties as suitably objective, it must hold fixed the extension of red as determined by the biconditional in the actual world when considering counterfactual worlds. This can be done by means of the rigidifying device 'actually':

x is red iff x is disposed to produce 'red' sensations in normal human observers as they actually are under standard conditions as they actually are.¹¹

Redness on this analysis is an objective, mind-independent property. What is mind-dependent is simply the matter of why or whether something, to use David Lewis' locution, *deserves the name* 'red.' It is a matter of meaning that something is red if and only if it produces a certain type of sensation in normal subjects. Whether something deserves the name is determined a priori by the nature of our concept 'red,' independently of how the world actually is. But the property of objects that we succeed in picking out via this concept is thoroughly mind-independent.

Under the rigidified analysis, the dispositions of an object to produce sensations in human beings in a counterfactual world *w* are completely irrelevant to the matter of that objects' color in *w*. It is this feature of rigidified response-dependence which promises to deliver a similarly objectivist account of moral properties. The working analysis above, after rigidification, becomes:

x is morally good iff x is disposed to produce a sentiment of moral approbation in normal human beings as they actually are under conditions of complete descriptive knowledge.

Without rigidification, the response dependent analysis delivers troubling results when considering certain counterfactual scenarios. What if

10 Though it is likely that a widespread spectrum-inversion would cause us to shift the meanings of our color terms. But this again would not entail that grass has changed color. Rather, the word 'green' would denote the property that was once denoted by our word 'red.'

11 The rigidifying device 'actually,' but not the analysis of 'red,' is due to Davies and Humberstone (1980).

human beings had been disposed differently? There is a possible way the world could have been, in which normal human beings in that world have a *positive* moral sentiment in response to torturing babies. We want to say that this is a world in which most human beings are immoral. But without rigidification, we look to what satisfies the analysis in each world in order to determine the distribution of moral properties in that world. Human beings in that world have a sentiment of moral approbation toward torturing babies, so in that world torturing babies is morally good or permissible.

By rigidifying on *actual* human responses, we don't look to our counterparts to determine the extension of our moral concepts in a possible world. The torturing of babies does not tend to elicit a sentiment of moral approbation in *us*, and so torturing babies is not morally good in any world. On the rigidified response-dependent account of value the moral sentiments of human beings in counterfactual worlds, perverse or otherwise, are entirely irrelevant to the matter of what moral properties are and what actions or events possess them.

It is precisely this feature, of completely divorcing the responses of our counterparts in a counterfactual world *w* from the determination of moral value in *w*, to which Peter Railton (1996) directs his criticism of rigidified response-dependent accounts of value.¹² By way of a thought experiment, he argues that rigidification cannot be the proper means for capturing the objectivity of moral properties. In what follows I will argue that Railton's critique is unsuccessful, and that it rests on a mistaken conception of the consequences of rigidification on the response-dependent analysis.

12 G.W. Fitch has suggested to me that Railton's criticism is of rigidified response-dependent accounts of *intrinsic* value rather than *moral* value. There are times when Railton speaks of 'intrinsic value' rather than 'moral value.' But it is clear that Railton intends to reach the conclusion that RRD for moral value fails. He specifically articulates the upshot of his critique as one that shows that there are 'serious obstacles to an approach to moral non-relativism through fixing reference by actual moral responses' (70). And the thought experiment to be described below is explicitly discussed in terms of moral value. An anonymous referee observes that Railton may think that the points about intrinsic value carry over to moral value, given his own views about the relationship between the two. See Railton (1986a, 1986b, 1989). This looks to be a plausible diagnosis for why Railton discusses both non-moral and moral value somewhat interchangeably. However, Railton offers his critique of rigidified response-dependence for moral value without invoking his views about their relationship. I will treat Railton's criticism primarily as a criticism of RRD for moral goodness. At the end of the paper I consider Railton's argument concerning intrinsic non-moral value, and whether this has any consequences for RRD for moral value.

V Railton's Criticism of Rigidified Response-Dependence

Railton is not unsympathetic toward the analogy between moral value and secondary qualities, nor does he direct his criticism toward the notion of response-dependence as a means of articulating that analogy. His own view about non-moral goodness in fact might be seen as a kind of response-dependent account.¹³ But Railton does not think that response-dependent theories of moral value can deliver an appropriate account of the objectivity of moral properties through rigidification. Here he finds a disanalogy with the case of color, and suggests that a better analogy might be between moral value and taste qualities. Whereas a population-wide spectrum-inversion would not change the colors of things, such a change in the way things taste would seem to change the facts about what is bitter and what is sweet. Railton attempts to show that moral value is more like bitterness than redness in this regard, and thus that rigidifying on the relevant responses of human beings as they actually are would cause a response-dependent account of value to deliver the wrong verdict about the extension of 'morally good' in some counterfactual scenarios.¹⁴

Railton (1996) criticizes rigidified response-dependence by way of a thought experiment. He asks us to imagine a scenario in the future in which human beings have come to reproduce via a form of cloning. The cloning process more strongly resembles that of science fiction than that of Dolly the sheep. Railton's clones, it appears, are essentially born adults and quickly acquire verbal and motor skills. The 'replicas' created by the cloning process strongly resemble their 'parents,' though they are not perfect replicas. A replica looks quite a bit different from her parent, and she lacks the parent's episodic memories. Replicas go through a brief period of recuperation after being created, which is almost like a kind of childhood. They are often but are not always 'raised' by the person of whom they are a replica.

13 Railton (1986) gives a theory of intrinsic non-moral goodness upon which, roughly, *X* is good for *A* if and only if *A*'s ideally rational and fully factually and nomologically informed self *A+* would desire that *A* desire that *X*. Here the relevant subject is *A+*, the response is a particular kind of desire, and the appropriate conditions are those of having full empirical information.

14 This should not be seen as implying that Railton endorses subjectivism about moral goodness. Rather, the conclusion of Railton's argument is that rigidification is not the correct mechanism for accounting for the objectivity of moral goodness.

The clones, which Railton calls 'new humans,' are like traditional human beings in virtually every respect. However, one important difference is that they place no value on kinship relations. Railton claims that ordinary human beings intrinsically value in a way that is sensitive to biological relatedness. But new humans do not care about biological relatedness. They are like us in their moral sensibilities in every way except this one. They form meaningful bonds with other human beings, and often with those who 'raise' them. But they place no special value on genetic kinship.

Eventually, the cloning process becomes so popular that all human beings are created in this way. All humans are new humans. In this context, Railton has us imagine that a new human named 'Ethan' was, as is common, raised apart from his biological father Ed. One day, Ethan has a chance opportunity to meet his father in a taxi. Neither Ed nor Ethan knows that he is kin to the other; they are strangers who happen to be sharing a taxi. The father and son are also sharing a taxi with a third man who is unrelated to them. If Ed and Ethan were to chat, it would be pleasant. But the stranger would enjoy a conversation with Ethan slightly more, we might suppose, than would Ed. Ethan would enjoy talking to either of them to about the same degree—he pretty much just likes to talk about himself. And as new humans, Ed and Ethan place no intrinsic value on becoming acquainted with their own kin.

Railton claims that our moral intuition in this case is that there is no greater good in Ethan speaking to his father than to the other unrelated man in the taxi. In fact, it seems that it would be slightly better overall if Ethan spoke to the stranger. But for normal human beings as they actually are, there is a special intrinsic value in a son becoming acquainted with his father, even if neither of them know that they are kin (Railton claims). Railton concludes from this that what is good is not to be rigidly designated by what actually normal humans are disposed to find valuable. We 'normal' humans do in fact find being acquainted with one's kin to be intrinsically valuable. So according to Railton, if 'good' acted as a rigid designator on the response-dependent analysis, we would be forced to conclude that the new human should speak to his father rather than to the other stranger. But this is not the correct result; hence, the rigidified response-dependent analysis is mistaken.

VI In Defense of Rigidification

Railton's critique of rigidified response-dependence is unsatisfactory in two respects, one which is relatively superficial (but no less troubling for his conclusion) and one which is more philosophically interesting. The former problem concerns his choice of example. The latter is that Rail-

ton's very method of argumentation against rigidification cannot be successful, even after granting him the features of his thought experiment and the intuitions which he hopes to evoke.

Problems with the Thought Experiment

Is knowing one's kin really a moral good for us actual humans?

We should be suspicious of Railton's chosen example. It isn't entirely clear that knowing one's kin is a moral good, at least independently of a person's wanting to know his or her kin. That is, it isn't clear that knowing one's biological kin is intrinsically valuable for actual human beings. Railton's thought experiment isn't very different from ordinary cases of adoption in which a child has not met one of his biological parents. And though many adopted children come to want to know about their biological parents or to become acquainted with them, it isn't obvious that this is a moral matter. I don't share the intuition that if an adopted child has no desire to meet her biological parents, and no other desires that she or others have would be satisfied by such a meeting, a moral good would nonetheless be realized by her meeting them. But if knowing one's biological kin isn't an intrinsic moral good in the actual world for actual humans, then showing that it isn't an intrinsic good in counterfactual circumstances does nothing to argue against rigidity.

Perhaps Railton should just change examples. But notice that alternative examples to make his point are difficult to entertain. A clear case of moral vice such as torturing other people for sport seems to be a vice in all possible worlds for all human beings. Or imagine that new-humans are just like us except that they don't have a sentiment of moral approbation with regard to helping those in need. Let's suppose that new-humans are Social Darwinists. Are we the slightest bit compelled to hold that helping others is not a moral good for new-humans? No. Instead, we think that new-humans are morally deficient and would perhaps be inclined to discontinue the cloning procedure. This accords with the rigidified response-dependent account — we rigidify on our moral responses rather than indexing what counts as morally good to the responses of our counterparts in counterfactual worlds.

One *can* imagine Railton-like cases if the wrong-making feature in the actual world is absent in the counterfactual scenario. For example, suppose that people's feelings were not hurt as ours typically are.¹⁵

15 An anonymous referee presented me with this example.

Then many actions that we think are rude or wrong in the actual world, such as insulting a person's grandmother, might not be wrong in the counterfactual scenario. But this kind of an example does not count against rigidification, since it merely shows that the wrongness of certain actions is derivative on their effects on others (such as hurting their feelings). I will discuss this kind of example in more detail in section VII below.

Are new-human clones kin?

Suppose we grant Railton that for actual human beings, it is a moral good to be acquainted with one's parent independently of any desires that may be satisfied by that state of affairs. We can also grant that the same is not true for new-humans. This is still not sufficient for Railton's thought experiment to be effective against rigidification, unless we can also assume that new-human clones are kin to each other in the relevant respect. That is, the property that we actual humans value or to which we are sensitive in our value judgements must be a property which is instantiated in the counterfactual scenario.

The moral intuitions that Railton attempts to pump are intuitions about the moral relevance of the biological parent and child relationship. But in Railton's counterfactual scenario, the factual conditions are different from the actual world. It isn't clear that the new-humans even have biological parents. They certainly don't have parents in the ordinary sense, so why think that the peculiar relationship between replicants and 'originals' falls under the scope of what is morally good about knowing one's parents?

To argue against the rigidified analysis, Railton has to give us a counterfactual possible world in which the correct application of the concept 'morally good' diverges from what is disposed to produce a sentiment of moral approbation in normal human beings as they actually are under the ideal conditions. Following Railton in claiming that we actual humans have a relevant response toward becoming acquainted with or communicating with one's kin, the thought experiment must show that a situation of an identical type is not morally good in some counterfactual scenario. Another comparison with the case of color helps make this clear. Consider the rigidified response-dependent analysis of 'red' given above. In arguing against rigidification, one would have to present a counterfactual possible world that contains an instance of the property that elicits a red sensation in human beings in the actual world. And then one would have to argue that the object which bears that property in the counterfactual world is not in fact red.

For Railton's thought experiment to threaten rigidification in the case of moral goodness, it is necessary that the very same property which

elicits the relevant moral sentiment in actual human beings be present in Railton's counterfactual scenario. Whether this is the case is difficult to assess, given that there is no philosophical consensus about the proper account of parenthood.¹⁶ Genetic relatedness appears to be one property that is shared by both new-human and actual human 'parents' and 'children.' But it isn't clear that our own responses are simply toward genetic relatedness.

First of all, we can imagine that by some fluke accident a person on the other side of the globe is my genetic twin. Perhaps radiation from a failed nuclear power plant causes genetic mutations in a developing fetus. The result is the birth of a child who is genetically identical to me. I certainly would like to meet this person — it would be quite interesting to meet my genetic twin. But should I have a greater intrinsic *moral* interest in this child as opposed to a child in relevantly similar circumstances who is less genetically similar to me? I think not.

A more plausible candidate for moral significance might be a historical property such as derivation from the same genetic source. This is an interesting suggestion, though it too faces counterexamples. For example, suppose that some of my genetic material comes into contact with an alien slug which then incorporates some of my genetic material into its own in order to create a new offspring. This technique of incorporating the genetic information of other species might be, we can imagine, the primary mechanism by which the slugs ensure the diversity of their gene pool. Suppose the offspring survives as a generally 'slug-like' creature. I'm not inclined to think that the relationship between me and the slug is one which has special moral significance.

As we imagine cases of kinship that are further and further from the paradigmatic situation of ordinary conception and birth without medical intervention, and in which the resulting child is raised by those parents, I suspect that our intuitions about the moral relevance of those relationships becomes weaker. Should a man who anonymously donates his sperm have a special intrinsic moral interest in the resulting offspring? Perhaps, but this is less clear than in ordinary cases of fatherhood. The point is that even if we place an intrinsic moral interest in kinship relations, it doesn't look to be a simple valuing of genetic similarity or being causally derived from the same genetic source. And this raises the question, do new-human clones have the relationships among each other to which we actual humans are supposedly sensitive?

16 For a nice overview of different accounts of parenthood, see Bayne and Kolers (2003).

Railton assumes that they do, and that their relationships are similar enough to actual human kinship relations to elicit an intrinsic preference in us. Railton then argues that that preference conflicts with our judgment about what is best in the Ethan and Ed scenario. But Railton's scenario seems to be a border-line case with regard to our actual preferences. Given the complexities regarding what our responses regarding kinship are sensitive to in the actual world, it is just as likely that the reason why it does not seem better for Ed to communicate with Ethan is that the relationship between them is too dissimilar to 'normal' kinship relations to elicit the relevant response in us.

VII How To Argue Against Rigidification

Leaving the above considerations aside, is Railton's thought experiment effective in showing that rigidification is the wrong route for the response-dependence theorist to secure objectivity for moral properties? No, and the way in which Railton's argument fails is instructive for understanding both rigidification and the nature of moral response-dependence.

Rigidification requires that we take the responses of actual relevant responders as authoritative about the extension of 'good.' Railton's intuition pump is designed to show us that the extension of 'good' in his counterfactual scenario diverges from what would be determined by rigidification. So, the argument goes, rigidification gives us the wrong result.

But something peculiar is going on with this line of argumentation. We actual and presumably normal human beings judge that there is no intrinsic value due to kinship in Ethan becoming acquainted with Ed.¹⁷ This is the intuitive judgment that Railton hopes to elicit in us with his thought experiment. But this suggests that our own moral responses are sensitive to the facts that differ in the thought experiment from the conditions in the actual world. It doesn't show that moral properties cannot be picked out rigidly as those properties that actually cause our moral sentiments. It is on the basis of our own moral sentiments in response to this hypothetical scenario that Railton is trying to argue against rigidifying on our own moral sentiments. Something has gone wrong here; this strategy of arguing against rigidification is incoherent.

17 As before, I'm assuming here with Railton that the relevant responders in the response-dependent analysis are actually normal human beings.

For Railton's argument against rigidification to avoid this problem, it needs to be the case that we correctly judge that no moral good is realized by Ethan communicating with Ed. But it must also be the case that we (qua relevant responders) have in response to that scenario moral sentiments which conflict with that judgment. From the perspective of a response-dependent theory, this is an unusual state of affairs. Response-dependent theories, including rigidified versions, lead naturally to a certain epistemology of the relevant domain. This is in fact one of the features of moral response-dependence which makes such theories attractive. Consider first the case of response-dependence about color properties. Blind persons can know what color something is. But they come to know color facts in a relatively roundabout way — by being told the colors of things. Blind persons are not relevant subjects in the response-dependent theory of color, since they do not have color sensations in response to external objects. How do relevant subjects typically come to know the colors of things? They simply look at the objects in question. More precisely, it would be natural for the response-dependent theorist to say that we come to know the colors of things by having the relevant response. If we are normal subjects, and if we are in the appropriate conditions, our having the relevant response puts us in an epistemically privileged position with regard to the colors of things.

An analogous epistemology for moral properties will be attractive to the response-dependent theorist. Our best judgments about whether something is morally good will be arrived at through something like 'moral perception.' Our judgment that something is morally good is normally the result of putting ourselves in the relevant conditions and having the appropriate moral sentiment. It will be tempting to think, on this view, that our judging with Railton that no moral good is realized by Ethan communicating with Ed is formed on the basis of our not having a sentiment of moral approbation toward that state of affairs. But Railton cannot argue against rigidification on our actual moral sentiments on the basis of moral judgments that are in accordance with our actual moral sentiments — as noted above such a strategy would be incoherent. Railton's argument requires that we do judge that no moral good is realized by Ethan meeting Ed, but that we nonetheless have a sentiment of moral approbation toward Ethan meeting Ed. But Railton does not argue explicitly that we lack the relevant sentiments in response to the scenario in his thought experiment. Rather, Railton appears to make certain assumptions about how rigidification would work in the moral case that would obviate the need for such an argument.

Railton appears to assume that we actual normal human beings have the relevant sentiment toward new-humans meeting their kin on the basis of the fact (which I'm granting now) that we actual normal humans have a sentiment of moral approbation toward humans in the actual

world meeting their kin.¹⁸ But this assumption can only be justified if it is known in advance that the differences between being a human and being a new-human are not relevant in the production of a sentiment of moral approbation in the relevant responders (actual normal humans). In the face of the considerations Railton raises, the advocate of RRD will be inclined to conclude that the property in the actual world that grounds the disposition to cause a positive moral sentiment is a property that is missing in the counterfactual scenario. New-humans meeting their kin simply fail to instantiate the relevant property. But this is compatible with rigidification.

The fact that all x 's are p in the actual world, where ' p ' is a rigid designator, does not entail that all x 's are p in any possible world. But this seems to be what Railton is assuming. He infers from the fact that in the actual world a human being's communicating with her kin instantiates a moral good (is p), that in the counterfactual scenario a human being's communicating with her kin instantiates a moral good (if 'morally good' is rigid). But this would be a fallacious inference. Even though all comedians might be mammals in this world, that wouldn't entail that all comedians in a nearby possible world are mammals.

There may be a confusion here about the proper sense in which rigidification on a response-dependent account requires that we take actual relevant responses as authoritative about the application of the concept in counterfactual worlds. On a rigidified response-dependent account of some predicate p , a particular individual or state of affairs x causing in the actual world the relevant response in the relevant subjects under the appropriate conditions entails that x is p in the actual world. This of course does not entail that ' x is p ' is true in all possible worlds. Suppose x is a particular tomato and p is 'red.' Assuming that objects do not have their colors essentially, there is a world in which that particular tomato is blue rather than red. And even if all tomatoes in the actual world are red and satisfy the RRD analysis of 'red,' this would not entail that in all possible worlds all tomatoes are red. On a rigidified response-dependent analysis of a predicate, actual responses are not *directly* authoritative about what individuals or states of affairs are in the extension of the concept. Rather, to say that actual responses are authoritative is to say that our actual responses determine what properties deserve the

18 This might in fact be why Railton goes to such effort to make new-humans *humans*. After all, he could have just described an alien species that differed from us in the relevant respect. But then I think it would have been even clearer that our having the relevant response toward humans meeting their kin does not entail that we have the relevant response toward the individuals in his scenario meeting their kin.

name '*p*.' This *in turn* determines what things in counterfactual worlds are *p*.

Consider again the possibility of spectrum inversion, an example which Railton appeals to as being analogous to his own thought experiment. Rigidified response-dependence for color predicts that objects would not change their colors simply in virtue of a species-wide spectrum inversion. The reason is that in such scenarios the properties of objects which are causally relevant in the production of the sensory responses in the relevant responders remain unchanged. A ripe tomato remains red because it is still such as to produce a red' sensation in normal human beings as they actually are. In the inverted spectrum scenario, the dispositions of objects to produce particular color sensations have changed while the causal grounds of the relevant dispositions remain unchanged. To the extent that consideration of inverted-spectrum worlds supports rigidified interpretations of color terms, it is because the extension of color terms is found to track the properties which causally ground the relevant response-dispositions rather than the dispositions themselves.

But in Railton's thought experiment, are the grounds of the relevant moral responses unchanged? If not, the analogy is misleading. In fact, Railton's thought experiment will have at best given us a reason to think that our moral responses are caused by properties which vary between the actual world and the world of his hypothetical scenario.

What Railton cannot assume without argument is that human beings in the actual world meeting their kin instantiates moral goodness simply in virtue of being an instance of the generic event type 'human being meeting her kin.' One way of determining if there is some other property that tokenings of that type in the actual world have which grounds their disposition to elicit in relevant responders a sentiment of moral approbation is to consider various counterfactual scenarios in which the tokenings of that type have different properties than they actually do. I suggest that, as far as has been shown by Railton, the new-human thought experiment can be seen as providing evidence that there is such a property. The intuitions that Railton attempts to evoke with his thought experiment would demonstrate that the situation type described as 'a person meeting his father' has a neutral moral value (assuming that a new-human meeting his father instantiates that type). Following Railton in assuming that it is intrinsically good for human beings like us to become acquainted with our kin, this would show that such situations inherit their positive moral value in virtue of being a token of some other type. Given that the factual conditions in the thought experiment differ in many respects from those in the actual world, there will be many properties that an actual human's knowing his father will exemplify that Ethan knowing his father will not exemplify. One signifi-

cant type which Ethan's meeting his father would not instantiate is 'a person who desires to know his father meeting his father.' There are many others.

VIII Objections and Replies

It might be suggested in Railton's defense that he depends only on our 'reflective judgements' about moral goodness, which need not be formed under the conditions of any particular response-dependent account. Rather, they might be influenced by various background theories in a search for reflective equilibrium. If this were right, then there would be no reason to think that our actually judging that no intrinsic moral good due to kinship is realized by Ethan communicating with Ed indicates that we fail to have the relevant sentiment.

But Railton does not make appeals to anything beyond our initial intuitive judgement. It doesn't appear that Railton intends the moral judgement concerning his thought experiment to be one which we arrive at through more complex reasoning, given that he does not invoke any background theoretical considerations. More importantly, Railton's task in arguing against rigidification is in several respects more difficult than this objection makes it seem.

This defense of Railton only undercuts my suggestion that our actual moral judgements concerning the thought experiment provide *prima facie* support for the claim that we fail to have the relevant response. It remains true that Railton needs to establish that we do indeed fail to have the relevant response. I examined one possible reason that might be given and found it lacking. Also, for Railton's general criticism of rigidified response-dependence to succeed, it must be the case that the considerations he raises apply to any version of rigidified response-dependence. That is, it must apply regardless of what particular responses or conditions are specified. It is no defense of Railton to point out that the responses or conditions of reflection upon which Railton depends for the formation of our moral judgement may not be the same as those specified by any particular response-dependent analysis. It is Railton's burden to argue against all possible versions, so a defense of rigidified response-dependence depends only on there being at least one version which survives his critique. That there is any way of specifying the relevant conditions or responses which survives Railton's critique would demonstrate that Railton's critique does not directly challenge the rigidified element of rigidified response-dependence.

This is especially clear when one considers a popular version of response-dependence for moral value. Many advocates of response-dependence, including McDowell (1985), formulate their analysis not in

terms of what produces a sentiment of approval but instead in terms of what *merits* such approval. On such a view, it will be our actual considered judgments that are relevant, presumably the very judgments that Railton appeals to in the case of Ed and Ethan. Such a view is certainly non-reductive, employing an evaluative term in the right-hand side of the biconditional. But it still promises to elucidate the nature of the concept of moral goodness.

A different worry might be that my defense of RRD proves too much.¹⁹ Consider a possible response-dependent concept which is clearly non-rigid, such as ‘nauseating.’ If my line of reasoning generalizes, then it might seem to imply that our merely judging that something is nauseating indicates that we are having the relevant response. This in turn could be used to defend a rigid reading of the response-dependent analysis of ‘nauseating’ in the same way that I defend a rigid reading of ‘moral goodness.’ Since it is always our actual judgements about counterfactual cases which we go by in determining the extension of a concept in those cases, my argument against Railton could be used to establish that any concept is rigid. This result is clearly too strong.

Whether something is nauseating is relative to an individual. Let’s suppose that X is nauseating for S iff x is disposed to produce a ‘nauseous experience’ in subject S under conditions C (such as, when x is in the immediate vicinity of S, S is not ill, etc.). ‘Nauseating’ is shown to be non-rigid precisely because what we judge to be nauseating in a counterfactual world is determined not by whether actual subjects have the relevant response but whether denizens of the world in question have the relevant response. This becomes clear by considering a case of something which does not cause the response in me here in the actual world but which does cause that response in a counterpart of me in another world. That thing is nauseating for me in the counterfactual world even though it doesn’t actually cause the response in me in the actual world.

My criticism of Railton in the final section has simply been that he has not clearly given a case in which actual relevant subjects lack the response. He needs his thought experiment to be a case in which a moral good is not exemplified — which in practice amounts to eliciting in us a judgement that no moral good is exemplified in the scenario. But he also needs for the relevant responders (us actual normal subjects in the sample analysis) to *have* the relevant response (a sentiment of moral approbation). Otherwise the rigidified response-dependent analysis

19 David Chalmers raised this objection to me.

does not deliver a verdict in conflict with the facts about the distribution of moral goodness across possible worlds.

Why does my reasoning against Railton not apply to the 'nauseating' case (which would constitute a *reductio* of the reasoning)? In the case of 'nauseating' it is clear that our judgements track the responses of the individuals in question rather than our own. But Railton has not established that the same is true for 'moral goodness.' While it is true that our judgement concerning the new-humans is in line with the responses of the counterfactual individuals, this doesn't threaten rigidification unless *we* respond differently (by having a sentiment of moral approbation toward the counterfactual scenario). And it is our having of the relevant response which needs establishing. Railton doesn't argue explicitly that we have the response toward the counterfactual scenario. Rather he seems to assume as much on the basis of our having the response toward similar actual scenarios. But the actual scenarios are only similar, given that actual humans differ from the new-humans. As I've pointed out, it is compatible with Railton's considerations that his thought experiment merely establishes that our own actual responses are sensitive to the differences between the actual human and the new-human case.

I also claimed that our judgments in the moral case give a *prima facie* reason for thinking the relevant response is present. Perhaps it will be worried that the same could be said about the case of 'nauseating,' which would again seem to defeat the claim. But here there is a crucial difference. The conditions under which 'nauseating' judgements are epistemically privileged and caused by our having the relevant response are conditions under which we are in the sensory presence of the substance in question. My armchair judgements about counterfactual scenarios and whether something is nauseating in that scenario are clearly not formed on this basis. And so it is clear that my having a judgment about whether something is nauseating in such instances does not give a *prima facie* reason for thinking that I am having a nauseous experience. Even more importantly, as noted above, 'nauseating' looks to be a relative term. My best judgments about whether something is nauseating will be formed on the basis of my having a certain response only when the judgement concerns what is nauseating *for me*. I am not a relevant subject as per the response-dependent analysis for determining what is nauseating for other individuals. By contrast, at least on many versions of moral response-dependence the judgment that Railton needs to elicit is plausibly formed under conditions which are the relevant ones as specified by the analysis, such as conditions of full descriptive information or Lewis' condition of 'imaginative acquaintance' (1989). This is of course a substantive issue and there will be other forms of response-dependence for which this point will not apply. But it is not my burden to defend every possible version of rigidified response-dependence. Rather, Rail-

ton has set himself the task of arguing against those views and this task fails if there is a form of rigidified response-dependence that survives his critique.

Finally, Railton's paper has two targets, rigidification for non-moral intrinsic value and for moral value. I have focused only on the latter. But does Railton's thought experiment successfully apply for the case of non-moral intrinsic value? And if so, might not this undermine rigidification for moral value as well?

To take the last question first, the answer is no. It is perfectly coherent to endorse rigidified response-dependence about moral value while holding a non-rigid view about intrinsic value. Even if facts about what things are non-morally intrinsically valuable are morally relevant, a non-rigid account of non-moral value would not entail a non-rigid account of moral value. Suppose, for example, that we conclude with Railton that his thought experiment shows that intrinsic value is non-rigid. A rigidified response-dependent account of moral goodness like my working analysis would be in no way impugned by this. For it is quite plausible that our actual moral sentiments, under conditions of full knowledge, are sensitive to what is intrinsically valuable to creatures affected by an action. Thus, to the extent that what is morally valuable varies in part with what is intrinsically valuable, a rigidified response-dependent account of moral value can reflect this.

Rigidification in the case of moral value is more plausible than it is for non-moral value, for the simple reason that objectivity about the former is more plausible than for the latter. And as argued above, if the goal of Railton's thought experiment is more modest than I have assumed (targeting rigidification for intrinsic non-moral value), then even if he is successful, he falls far short of his conclusion that '[r]igidification, then, seems to this extent inappropriate as a way of capturing the objectivity of moral assessment' (81).

Is Railton's thought experiment successful against rigidified response-dependence about intrinsic non-moral value? Certainly it is possible for our judgment about what is best (either in a moral or non-moral sense) to diverge from our intrinsic preferences. And this very well could be the case with regard to our reactions toward Railton's thought experiment. Railton claims that we have an intrinsic preference that Ethan and Ed communicate, but that upon reflection we judge that it is not intrinsically better (either in a moral or non-moral sense) that they do so communicate. This would count against the following rigidified response-dependent theory of intrinsic non-moral goodness:

x is intrinsically valuable iff actual human beings have an intrinsic preference for x under conditions *k*.

Even if Railton's thought experiment is successful in challenging such a view (and similar ones) about intrinsic non-moral value, it is far from obvious that it successfully challenges rigidification. For there is an important difference between being non-rigid and being relational. The distinction between rigidity and non-rigidity concerns modality — the application conditions for a concept across possible worlds. By contrast, if a property p is relational, then there is no such thing as an object or event being p simpliciter.²⁰ An object or event will be p only relative to some group or individual S . We might then say that any true statement of the form ' x is p ' or ' x is not p ' is shorthand for a statement of the form ' x is p -for- S ' or ' x is not p -for- S .'

With regard to intrinsic non-moral value, even if Railton's thought experiment is successful, it supports some form of relationalism about intrinsic value rather than non-rigidity. Consider a relational version of response-dependence:

x is intrinsically good-for- S iff human beings of kind S have an intrinsic preference for x under conditions k .

We might be led to a view like this by thought experiments like Railton's. But this is not the same as concluding that 'intrinsically good' is non-rigid. To distinguish between the two, one needs to consider cases in which the relevant responders are the same, and the circumstances are the same, but in which those responders have different responses in different possible worlds. Railton's thought experiment fails to do this.

If intrinsic value is relational, then presumably there is principled reason that grounds the fact that what is good for creatures of one sort may not be identical to what is good for creatures of a different sort. On Railton's view, what is good for a type of creature S depends on the nature of such creatures. If we accept his conclusion that non-moral value is relational, then the resulting response-dependent account is such that we actual human beings are not members of the class of relevant responders for evaluating the case of Ed and Ethan. This is not because we are *actual*, but because we are a relevantly different kind of human being from the new-humans. Good-for-humans is a different relational property from good-for-new-humans. The thought experiment has no bearing on rigidification, because the irrelevance of actual

20 I am using Railton's term 'relational' for the idea that, in the case of intrinsic goodness, 'what is intrinsically good for an individual I of kind K depends upon the nature of I and K ' (77).

human responses is due to the relational character of non-moral value rather than the failure of 'good' to function as a rigid designator.²¹

It may be that, normally, if a property is relational in Railton's sense then rigidification for the corresponding concept will be implausible. Such is the case with 'nauseating.' This is because the kind of objectivity that is secured by rigidification is stronger than warranted for such relational properties. But rigidification is compatible with a property's being relational, and Railton's argument against rigidified response-dependence for non-moral value is most charitably interpreted as an argument against non-relationalism, which only derivatively lends credence to a non-rigid view. By contrast, moral goodness has a greater form of objectivity, for which rigidification is most suitable for securing on a response-dependent theory.²²

Received July 2004

Revised December 2004

Revised June 2005

Revised August 2005

References

- Bayne, T., and A. Kolers. 2003. 'Toward a Pluralist Account of Personhood.' *Bioethics* 17 221-42.
- Davies, M., and L. Humberstone. 1980. 'Two Notions of Necessity.' *Philosophical Studies* 38 1-30.
- Izard, C. and B. Ackerman. 2000. 'Motivational, Organizational, and Regulatory Functions of Discrete Emotions.' In *Handbook of Emotions*, 2nd edition, M. Lewis and J. Haviland-Jones, eds. New York: Guilford Press.

21 It thus remains open whether or not, even if we accept Railton's thought experiment, non-moral 'good' is rigid. I am inclined to think that one still should accept rigidification if one shares Railton's view that non-moral value is objective. For it seems plausible that if a view like this is to be considered objective, then for any counterfactual possibility in which there is a change in what is intrinsically valuable for a particular creature, it will be because the creature in that world is of a different relevant kind *S* (one with a different nature). If this view is non-rigid, then a change in preferences without an accompanying change in nature would result in a change in what is intrinsically good-for-*S*. Value on such a view does not seem to be objective in any sense.

22 Many thanks to David Chalmers, Mark Lebar, Dan López de Sa, and Mark van Roojen. An earlier version of the present paper was presented at the 2002 meeting of the Pacific APA. Thanks to the audience there, and especially my commentator, G.W. Fitch.

- Johnston, M. 1989. 'Dispositional Theories of Value.' *Proceedings of the Aristotelian Society*, suppl. vol. 63 139-74.
- Lazarus, R. 1991. *Emotion and Adaptation*. New York: Oxford University Press.
- Lewis, D. 1989. 'Dispositional Theories of Value.' *Proceedings of the Aristotelian Society*, suppl. vol. 63 113-37.
- Mackie, J. 1977. *Ethics: Inventing Right and Wrong*. Harmondsworth: Penguin.
- McDowell, J. 1985. 'Values and Secondary Qualities.' In *Morality and Objectivity: A Tribute to J.L. Mackie*, T. Honderich, ed. London: Routledge and Kegan Paul.
- Millgram, E. 1999. 'Moral Values and Secondary Qualities.' *American Philosophical Quarterly* 36 253-55.
- Peacocke, C. 1983. *Sense and Content*. Oxford: Oxford University Press.
- Pettit, P. 1989. 'Terms, Things, and Response-Dependence.' *European Review of Philosophy* 3 61-72.
- Railton, P. 1986a. 'Facts and Values.' *Philosophical Topics* 24 5-31.
- _____. 1986b. 'Moral Realism.' *Philosophical Review* 95 163-207.
- _____. 1989. 'Naturalism and Prescriptivity.' *Social Philosophy and Policy* 7 151-74.
- _____. 1996. 'Red, Bitter, Good.' *European Review of Philosophy* 3 67-84.
- Shoemaker, S. 1994. 'Self-Knowledge and "Inner Sense."' *Philosophy and Phenomenological Research* 54 249-314.
- Tucker, D., D. Derryberry, and P. Luu. 2000. 'Anatomy and Physiology of Human Emotion: Vertical Integration of Brain Stem, Limbic, and Cortical Systems.' In *The Cognitive Neuroscience of Emotion*, R. Lane, L. Nadel, and G. Ahern, eds. New York: Oxford University Press.
- Wedgwood, R. 1998. 'The Essence of Response-Dependence.' *European Review of Philosophy* 3 37-60.
- Wiggins, D. 1991. 'A Sensible Subjectivism?' In *Needs, Values, Truth: Essays in the Philosophy of Value*. Oxford: Blackwell.
- Wright, C. 1988. 'Moral Values, Projection and Secondary Qualities.' *Proceedings of the Aristotelian Society*, suppl. vol. 62 1-26.