Two Kinds of Naturalism in Ethics

NEIL SINCLAIR

neil.sinclair@nottingham.ac.uk

Penultimate draft. Final paper published in *Ethical Theory and Moral Practice* 2006, 9(4): 417-439

ABSTRACT: What are the conditions on a successful naturalistic account of moral properties? In this paper I discuss one such condition: the possibility of moral concepts playing a role in good empirical theories on a par with those of the natural and social sciences. I argue that Peter Railton's influential account of moral rightness fails to meet this condition, and thus is only viable in the hands of a naturalist who doesn't insist on it. This conclusion generalises to all versions of naturalism that give a significant role to a dispositional characterisation of moral properties. I also argue, however, that the epistemological and semantic motivations behind naturalism are consistent with a version of naturalism that abandons the condition.

KEY WORDS: Ethical realism; explanation; naturalism; Railton; response-dependence.

1. Introduction

Many recent discussions of naturalism in ethics tie the feasibility of the naturalist programme to the possibility of moral explanations. In particular, it is often assumed that a necessary condition on a successful defence of ethical naturalism is a role for moral properties in good empirical theories on a par with those of the natural

and social sciences (such as biology and psychology).[1] Such theories are taken to involve informative explanations of observed phenomena or patterns of observed phenomena that are not available at any other level of description. The assumption is seldom explicit, but once exposed can be questioned, and this questioning opens up the possibility of two kinds of naturalism in ethics.

According to the first, the characterisation of moral properties as natural properties will only be possible if moral properties feature in good explanations of certain observable non-moral events.[2] According to this type of naturalism, the question whether moral factors can feature in explanations of such things as agents' material success, or processes of social change, will be of necessary importance to those who wish to defend the existence of natural moral properties. [3]

According to the second, less demanding, form of naturalism the characterisation of moral properties as natural properties is not threatened by the possible absence of moral explanations of the same sort. For this type of naturalist, it may be an interesting question whether moral factors feature in explanations of agents' material success, or processes of social change, but the answer to this question will not affect one's position on the nature of moral properties.[4]

In this paper I argue against Railton's influential naturalistic account of the property of moral rightness, on the grounds that it fails to meet the condition imposed by the first type of naturalism. This rejection is reasonable since Railton himself espouses this view. I also argue, however, that Railton's account might be salvaged were he to adopt the second type of naturalism.

The rejection of Railton's account has wider implications for ethical naturalism, since many of the arguments used against Railton generalise. In particular,

they suggest that an ethical naturalist who accepts a dispositional account of moral properties is best not to be a naturalist of the first kind. My argument will initially focus on the case of Railton, with general lessons to be drawn after this case is made.

One qualification before we proceed. I shall assume that both types of ethical naturalism accept that, at least in favourable circumstances, our moral judgements are responsive to the actual distribution of moral properties. Accordingly, both will accept the possibility of explanations of our moral judgements that cite moral factors, in particular, the very factors judged to obtain. Since the making of a moral judgement is a non-moral event (just as the making of a judgement about the weather is not a meteorological event) both types of naturalism will accept that, in this sense, there are moral explanations of non-moral events. The first type of naturalism demands, further, that in order for naturalistic moral properties to exist they must be involved in explanations of observable non-moral events other than our making of moral judgements – events such as an agent's material success, or processes of social change.[5] It is this further condition that I wish to question. In §7 I shall return to the issue of how the weaker form of naturalism can account for moral explanations of moral judgements.

2. Argument Summary

According to Railton, facts concerning moral rightness are constituted by natural facts.[6] Further, Railton holds that for any such account to be vindicated it must show how moral concepts can "participate in their own right in genuinely empirical theories" (205). These theories must also be "good theories, that is, theories for which

we have substantial evidence and that provide plausible explanations" (205). These and similar remarks (for example at 171-2) identify Railton as a naturalist of the first, more demanding, sort.

I will argue, however, that Railton's account of moral rightness provides no reason to think that this concept will appear in its own right in good empirical theories.

My argument for this claim is as follows. Railton's account of moral rightness can be considered as making one of two property identity claims. According to the first, moral rightness is a dispositional property. According to the second, it is the categorical ground of such a disposition. If the former, then to hold that moral rightness can play a role in good empirical theories is to hold that highly idealised counterfactual circumstances can be causally efficacious, which is implausible. If the latter, then Railton has provided no grounds for optimism that moral rightness appears *in its own right* in good empirical theories (given plausible empirical assumptions). Either way, Railton has not shown how the concept of moral rightness can appear in its own right in good empirical theories. (My argument actually focuses on Railton's account of moral wrongness, but it easily transfers to the case of moral rightness). In the penultimate section, I suggest that this result need not be fatal were Railton to adopt the second, less demanding, form of naturalism.

3. Railton's Methodology

In his paper "Moral Realism", Railton's aim is to provide a 'reforming naturalistic definition' (204) of moral rightness. According to such a definition, the property of moral rightness is defined as being (identical with) some naturalistically respectable

property. Such definitions are put forward, not as analytic claims about the meanings of the terms involved, but as synthetic claims about the nature of the putative properties those terms refer to. They are to be judged, not by a priori means, but through a posteriori consideration or whether or not they provide good explanatory accounts of the nature of the practices involving the term. Such methodology is required by Railton's 'methodological naturalism' according to which philosophy possesses no distinctive a priori method able to yield substantive truths.[7]

According to Railton, the a posteriori assessment of reforming naturalistic definitions takes place across two dimensions (204-7):

3.1. *Constraints of function*

First, the defining property must capture most or all of the intuitive force of the definiendum (203-4). Every meaningful term of our language plays distinctive roles in our understanding and discourse. It is these roles that are reflected in those pre-reflective truisms that surround the term. So, for example, it is a truism about water that it is the stuff that makes up the majority of the oceans. Likewise, it is a truism about moral rightness that judgements involving it are typically connected to agents' motivations (168). Any definition in terms of a property that doesn't fill these central roles of the definiendum will to that extent be defective. (It is possible, of course, that our intuitions concerning functional role are confused, so that no single property fills (or could fill) the roles that those intuitions demand – perhaps the case of *ether* is like this. In that case, any definition will be revisionary, but may still be justified so long as it can be shown how the function taken as central affords the best understanding of

the term and the discourse within which it is embedded. Railton's sees his own account of moral rightness as 'tolerably revisionist' in this way (205).)

3.2. Constraints of naturalistic respectability

Successful naturalistic definitions are also governed by the criterion of naturalistic respectability. That is, the defining property must be naturalistically respectable. Any theory that hopes to offer a definition of a term whilst remaining a version of naturalism is committed to this condition.

Railton provides a peculiar interpretation of what it is for a defining property to be naturalistically respectable that marks him out as an ethical naturalist of the first kind. For Railton, naturalistic respectability derives from the ability of the putative property to feature in its own right in empirical theories. He writes:

> What might be called the 'generic stratagem of naturalistic realism' is to postulate a realm of facts in virtue of the contribution they would make to the *a posteriori* explanation of certain features of our experience. For example, an external world is posited to explain the coherence, stability, and intersubjectivity of sense-experience. (171-2)

Later, having offered definitions of non-moral goodness and moral rightness, Railton reminds the reader of this condition:

> …[I]t remains to show that the empirical theories constructed with the help of these definitions are reasonably good theories, that is, theories for which we have substantial evidence and which provide plausible explanations. I have tried in the most preliminary way imaginable to suggest this. If I have been wholly unpersuasive on empirical matters, then I can expect that the definitions I have offered will be equally unpersuasive. (205).

Thus Railton considers it a necessary condition for his defence of moral naturalism that moral properties feature in good empirical theories.

Good empirical theories, for Railton, are those which

> …contain generalisations that may not be strict or exceptionless, but that do illuminate functional connections, causal dependencies and other relations at a particular…level of description of the phenomena…[and that] afford explanatory insights that would not be evident at [any other] level of…description of events.[8]

Such theories thus "insert explananda into a distinctive and well-articulated nomic nexus, in an obvious way increasing our understanding of them" (184).

There are two relevant points to note about Railton's version of the criterion of naturalistic respectability.

First, a definition will meet Railton's criterion only if the defining property can participate *in its own right* in genuine empirical theories (205). This is to say that those theories must not be formulable except in terms of the defining property.

Second, the availability of empirical theories that Railton's criterion demands is an a posteriori matter, to be determined by the actual process of theory-construction (204). He notes that the normativity of the notions he is attempting to define should not be thought to rule out all such theories a priori.[9] He admits, however, that were empirical investigation to show that no theory in terms of his proposed definition could be constructed, we would have reason to reject that definition (205). Hence Railton intends us to judge his reforming naturalistic definition of moral rightness at least partly on the basis of whether, a posteriori, it allows for that concept to feature in its own right in an informative explanatory nexus.[10]

The two criteria for assessing reforming naturalistic definitions come together in the moral case as follows. According to constraints of naturalistic respectability, postulation of a realm of facts is justified when such postulation brings explanatory gain. If the realm of facts thus postulated satisfies constraints of function for moral notions then they can also be labelled distinctively *moral* facts and we would have what Railton labels a "plausible synthesis of the empirical and the normative" (163).

Railton argues that his definition of moral rightness meets both sets of constraints. The question, therefore, is whether Railton's definition specifies a distinct realm of facts that both captures the pre-reflective functions of our notion of moral rightness and plays the requisite causal-explanatory role. I argue that the second condition – the condition imposed by the first kind of naturalism – remains unsatisfied.

4. Railton on Moral Rightness

Railton introduces his definition of moral rightness by first considering the distinctive nature of moral norms. He notes that moral norms, including the norm of moral rightness, are distinguished from other criteria of assessment by being *interpersonal* – in that they are concerned with the "assessment of conduct or character where the interests of more than one individual are at stake" – and *impartial* – in that the "interests of the strongest or most prestigious party do not always prevail, purely prudential reasons may be subordinated, and so on" (189). These two features are captured for Railton in the claim that "moral norms reflect a certain kind of

rationality, rationality not from the point of view of any particular individual, but from what might be called a social point of view" (190). He continues:

> By itself, the equation of moral rightness with rationality from the social point of view is not terribly restrictive, for depending on what one takes rationality to be, this equation could be made by a utilitarian, a Kantian, or even a non-cognitivist…Here I have adopted an instrumentalist conception of rationality, and this…means that the argument for moral realism given below is an argument that presupposes and purports to defend a particular substantive moral theory.
>
> What is this theory? Let me introduce an idealization of the notion of social rationality by considering what would be rationally approved of were the interests of all potentially affected individuals counted equally under circumstances of full and vivid information. (190)

So Railton is aware that the equation of moral rightness with rationality from the social point of view is seriously incomplete, for an account of rationality needs to be given. By adopting an instrumentalist account of rationality (188) Railton takes the equation of moral rightness with social rationality to amount to the equation of moral rightness with what would be approved of by instrumentally rational agents when counting equally the interests of all potentially affected individuals and when fully and vividly informed. Thus Railton's reforming naturalistic definition of moral rightness can be represented by the following biconditional:

(1) ϕ is morally right iff ϕ would be approved of by instrumentally rational agents were the interests of all potentially affected individuals counted equally under conditions of full and vivid information.[11]

There are two further points to note about this definition.

First, the notion of *interests* requires clarification. Railton draws a three-way distinction between subjective, objectified subjective and objective interests (173-5). An agent's subjective interests are his current "wants or desires, conscious or unconscious" (173). An agent's objectified subjective interests are those desires or wants that an idealised counterpart of the agent would want his non-idealised self to want were he to find himself in the actual condition and circumstances of the non-idealised agent (174). The idealised counterpart is an agent possessing "unqualified cognitive and imaginative powers, and full factual and nomological information about [the actual agent's] physical and psychological constitution, capacities, circumstances, history, and so on" (173-4). Finally, an agent's objective interests are "…those facts about [the actual agent] and his circumstances that [the idealised agent] would combine with his general knowledge in arriving at his views about what he would want to want were to step into [the actual agent's] shoes" (174).[12] Thus objective interests explain the presence of objectified subjective interests, not vice versa (175). Railton is clear that the interests involved in the account of moral rightness are objective interests (190-1). Given his earlier definition of an agent's non-moral goodness in terms of that agent's objective interests (176), this entails that, for Railton, moral rightness is equivalent to "what is rational from the social point of view with regard to the realization of…non-moral goodness" (191).

Second, the notion of full and vivid information requires clarification. Railton assumes that the idealisation involved here is the same as that involved in the move from an agent's subjective to his objectified subjective interests (190-1). Thus an individual is fully and vividly informed when he has "…unqualified cognitive and imaginative powers, and full and factual information about [the] physical and

psychological constitution, capacities, circumstances, history and so on [of the potentially affected individuals]" (174).

5. Response Dependence

Railton's account of moral rightness can be considered an example of a response-dependent account of moral facts, according to which moral facts obtain in virtue of acts, objects, situations or features thereof being disposed to elicit a certain reaction from a certain group of people in certain circumstances.[13] Railton's view can be presented schematically thus:

(2) ϕ is morally right iff ϕ is disposed to elicit [R_1] from [P] in [C].

Where ϕ is any putative bearer of moral rightness, R_1 is approval, C is when considering equally the objective interests of all those potentially affected by ϕ under conditions of full and vivid information and P are people that are ideally instrumentally rational.[14,15]

Though he doesn't explicitly mention it, Railton would presumably accept a similar schema for moral *wrongness,* that is:

(3) ϕ is morally wrong iff ϕ is disposed to elicit [R_2] from [P] in [C].

The difference being that for moral wrongness the reaction involved – R_2 – is *dis*approval, so that where moral rightness involves a positive attitude towards ϕ, moral wrongness involves a negative attitude towards it. (3) is preferable to an

alternative view according to which an act is wrong when it is disposed *not* to elicit approval from P in C, since it accommodates the intuition (constraint of function) that *morally right* and *morally wrong* are contrary but not contradictory.

Notice a crucial feature of these schemas: though they provide necessary and sufficient conditions for an action to be morally right and morally wrong respectively, they do not tell us about the nature of the moral properties themselves. There appear to be two possible identifications for each.

In the first case, the moral properties may be identified with the relevant dispositional properties. In the case of moral rightness this would be the claim that:

(2a) The property of moral rightness is identical with the property {being disposed to elicit approval from instrumentally rational people were the objective interests of all potentially affected individuals counted equally under conditions of full and vivid information}.

And for moral wrongness:

(3a) The property of moral wrongness is identical with the property {being disposed to elicit disapproval from instrumentally rational people were the objective interests of all potentially affected individuals counted equally under conditions of full and vivid information}.

In the second case the moral properties may be identified with the categorical grounds of these dispositional properties, that is, with whatever it is about a certain

class of actions or situations that makes it the case that they are disposed to elicit a certain reaction from certain people in certain circumstances.[16] In the case of moral rightness this would be the claim that:

(2b) The property of moral rightness is identical with {that property or properties of actions that make it the case that: such actions are disposed to elicit approval from instrumentally rational people were the objective interests of all potentially affected individuals counted equally under conditions of full and vivid information}.

And, again, for moral wrongness:

(3b) The property of moral wrongness is identical with {that property or properties of actions that make it the case that: such actions are disposed to elicit approval from instrumentally rational people were the objective interests of all potentially affected individuals counted equally under conditions of full and vivid information}.[17]

Note that, on the second set of views there is no a priori guarantee that the set of properties with which moral rightness and wrongness are identified are unified by anything other than the fact that their instantiations elicit a certain reaction from certain people in certain circumstances.

6. Moral Explanations

Given the above account, Railton aims to show how "moral rightness could participate in explanations of behaviour or in a process of moral learning" (191). So what sort of thing might the notions of moral rightness and wrongness be called upon to explain? Railton's favourite example is of social instability.[18] He claims:

> Just as an individual who significantly discounts some of his interests will be liable to certain sorts of dissatisfaction, so will a social arrangement – for example, a form of production, a social or political hierarchy, etc. – that departs from social rationality by significantly discounting the interests of a particular group have a potential for dissatisfaction and unrest. (191)

By 'social rationality' Railton means what would be approved of by instrumentally rational agents when counting equally the interests of all potentially affected individuals under conditions of full and vivid information. Thus, a departure from social rationality is something that would be actively disapproved of in such conditions, that is, something which – according to Railton's definition (3) – is morally wrong. But since such a departure would seem to explain a certain potential for unrest, it seems as if moral wrongness has an informative explanatory role.

To simplify somewhat, the sort of explanatory role for moral wrongness that Railton is suggesting is a role in explanations such as:

(A) Arcadian society is unstable because its institutional arrangements are morally wrong.[19]

The question is whether Railton's reforming definition of moral wrongness can be substituted into such explanations to provide the good empirical theories that his methodology – and the first kind of naturalism – requires. Given that there are two

possible property-identifications that Railton may be making, there are two possible ways in which these explanations can be understood.

6.1. *Moral properties as dispositional properties*

Take first – (3a) – the view that the property of moral wrongness is identical with the dispositional property: being disposed to elicit approval from instrumentally rational people were the interests of all potentially affected individuals counted equally under conditions of full and vivid information. The moral explanation offered in (A) would then be equivalent to:

(B) Arcadian society is unstable because its institutional arrangements would be disapproved of by instrumentally rational agents were the objective interests of all those potentially affected counted equally under conditions of full and vivid information.

The problem with this understanding of explanations such as (A) is that they cannot be understood on either a straightforward causal or dispositional model.

In the first case, the explanans in (B) cannot be directly causally efficacious in bringing about the explanandum, for this would be for a highly idealised hypothetical situation to bring about an actual situation.[20]

Perhaps the explanans in (B) is *indirectly* causally efficacious in bringing about the explanandum. For this to be the case, the explanans would have to be directly causally efficacious in bringing about some intermediary which is itself directly causally efficacious in bringing about instability in Arcadian society. What

might this intermediary be? A dispositional model of explanation provides one answer. In dispositional explanations we can explain why a particular object undergoes a particular change by citing a relevant disposition to undergo just that change in certain conditions, given the assumption that those conditions are realised; the change is a particular manifestation of the disposition. This is the model of explanation at work cases such as: “The glass broke because it was fragile”. Adopting this model for the present case, we can construe the explanation in (B) as follows. First, we cite the fact that Arcadian institutional arrangements are disposed to be disapproved of by certain ideal agents in certain ideal circumstances in explaining why some agents – specifically, agents who have realised these conditions – disapprove of those arrangements. Second, we cite this disapproval in explaining why Arcadian is unstable. Thus the instantiation of the dispositional property – which on the present account just is wrongness – explains the disapproval, which in turn explains the instability. Hence explanation (B) is restored.

There are two problems with this suggestion. The first is that, unlike the case of fragility, for the dispositional property of wrongness the conditions under which the manifestation of the disposition – in this case disapproval – occurs are idealised and seldom, if ever, realised. For the fragility of a glass to explain its breaking we must assume that the glass has been dropped onto a hard surface or hit with a hard instrument – that is, been placed in the conditions which help characterise fragility. Given the present account, for the wrongness of an institutional arrangement to explain disapproval directed at it amongst some group of agents we must assume that those agents are instrumentally rational and have reflected on the role of the institution whilst considering equally the objective interests of all those potentially affected by it. But such agents and such reflections are extremely rare, if not

impossible. Accordingly the number of instances of disapproval that could be explained this way is likely to be negligible.

The second problem with this reading of (B) builds on the first. Given that the number of agents whose disapproval of social institutions is to be explained by their wrongness is likely to be small, it is highly implausible to suppose that this disapproval will explain any social unrest. As Railton himself seems to admit (191), and as I discuss in more detail below (§6.2), it is the non-satisfaction of a significant number of the objective interests of a particular group of individuals that is the likely explanation of instability, not the disapproval of a small number of ideal individuals. Thus even if we accept that instances of wrongness might explain some attitudes of disapproval among a privileged few, there is no reason to think that these attitudes will explain anything else. In particular, there is no reason to suppose that these attitudes of disapproval might explain unrest. Thus no reason to think that wrongness itself explains anything other than these attitudes. Thus, again, we should reject explanation (B).

In sum, if Railton takes moral wrongness to be a dispositional property, explanations citing moral wrongness are either highly counterintuitive (in that they involve attributing causal efficacy to idealised counterfactual situations) or severely limited (in that the range of explananda is restricted to the reactions of certain idealised agents). In neither case are they able to be part of the good empirical theories that Railton's methodology – and the first kind of naturalism in ethics – requires.

6.2. *Moral properties as the categorical grounds of dispositions*

Perhaps Railton would be wiser to identify moral rightness and moral wrongness not with idealised dispositional properties, but with the categorical grounds of such dispositions. Objects are disposed to behave in various ways in virtue of other 'lower-level' properties. So, for example, a glass is such as to be disposed to break when dropped in virtue of its microphysical structure (plus the physical laws). Similarly, a society is such as to be disposed to be disapproved of by instrumentally rational agents equally considering all objective interests in virtue of some 'lower-level' property it has. For example, it may be so disposed in virtue of it having an unequal distribution of resources. The underlying property is the categorical ground for the dispositional property (173).

Suppose, therefore, that we identify the property of moral wrongness with this categorical ground, that is, we accept (3b). On this view, the moral explanation offered in (A) would be equivalent to:

(C) Arcadian society is unstable because it has certain properties that make it the case that: its institutional arrangements are disposed to elicit disapproval from instrumentally rational people were the objective interests of all potentially affected individuals counted equally under conditions of full and vivid information.

Explanation (C) can be understood on a straightforwardly causal model: the possession of the categorical ground underlying the dispositional property can be taken as causally productive of the instability of Arcadian society. Given that the categorical ground is, on the present view, an instantiation of moral wrongness, such

explanations go some way to placing moral wrongness in the empirical theory that Railton's strategy demands.

There is, however, a general problem with this move. As previously noted, there is no a priori guarantee that the categorical grounds of the disposition to elicit disapproval in the conditions relevant to moral wrongness will share any similarity other than the fact that they all ground such a disposition. If there is no such similarity, then, under the present suggestion, the property of moral wrongness will be identified with a disjunctive set of properties having nothing else in common than that their instantiations are all apt to ground the appropriate disposition. Whether or not this is the case will be an a posteriori matter, confirmed by testing instrumentally rational agents' responses under the relevant idealised conditions. Railton, however, is committed to the view that such a posteriori testing would show that the property of moral wrongness is not disjunctive in this way, for if it were, it could not figure in its own right in the empirical theories that his strategy – and the first from of naturalism – demands (205).

To see this, suppose for the moment that a posteriori testing would confirm that the property of moral wrongness is irrevocably disjunctive, with the disjuncts having nothing more in common than that their instantiations in a situation will elicit disapproval from certain idealised agents in certain idealised circumstances. If the various disjuncts have no more than this is common, then we couldn't know, just from knowing that one of the disjuncts is instantiated, the likely causal effects of this instantiation (other than that the situation would elicit disapproval from certain people in certain circumstances). Since, on the view presently under consideration, predication of the property of moral wrongness would tell us no more than that one of the disjuncts of the set with which that property is identified is instantiated, then

predication of such a property couldn't tell us anything of the likely causal effects of its instantiation (other than that it will elicit disapproval from certain people in certain circumstances). Hence moral wrongness could not participate in its own right in any informative explanations (other than that its instantiation will elicit disapproval from certain people in certain circumstances). Of course, each of the various disjuncts may participate *in its own right* in informative explanations, but this would not be for the concept of *moral wrongness* to so participate.

If Railton is to maintain, therefore, that moral wrongness has, in its own right, a role to play in good empirical theory, he is committed to the view that, a posteriori there is something in common between all those categorical grounds of the disposition involved in the definition of moral wrongness (something in common that is, over and above the fact that they are all grounds for the disposition).

Railton would certainly not object to the above line of argument, and goes as far as to welcome the fact that his account of moral rightness is constrained by a posteriori testing in this way (205). He is optimistic, however, that such testing will vindicate and not undermine his position. He is optimistic, in other words, that the various categorical grounds which are the instantiations of moral wrongness will fall into an explanatorily useful category. What might be the cause of such optimism?

A clue was given earlier. When introducing his preferred example of moral explanation, Railton claims that:

> …a social arrangement…that departs from social rationality *by significantly discounting the [objective] interests of a particular group* [will] have a potential for dissatisfaction and unrest. (191, emphasis added.)

Situations that depart from social rationality, remember, are situations would be disapproved of by instrumentally rational agents when counting equally the interests of all potentially affected individuals under conditions of full and vivid information, in other words, situations that are morally wrong. So Railton is here giving one way in which situations might be morally wrong: by significantly discounting the objective interests of a particular group. Assuming that situations that involve the significant discounting of the objective interests of a particular group have a potential for dissatisfaction and unrest, it follows that situations that are morally wrong in this way will be prone to dissatisfaction and unrest.

Unfortunately this will not suffice to show that moral wrongness is itself an explanatorily useful category, since for all that has been said, there may be situations that are wrong in ways other than significantly discounting of the objective interests of an group. Such situations, though morally wrong, will not on this account have any potential for dissatisfaction and unrest. To avoid this difficulty, Railton might claim that *every* situation which is morally wrong involves the significant discounting of the objective interests of some group. If all situations of moral wrongness involve the discounting of the objective interests of some (groups of) people, and if in all situations in which some peoples' objective interests are discounted those people have a tendency for dissatisfaction, then it follows that whenever there is moral wrongness there will be a tendency for social unrest. Hence explanation (A) would be vindicated and the property of moral wrongness could play a part in its own right in a good empirical theory.

It is debatable, however, how far this would go to placing the property of moral wrongness in an informative explanatory nexus such as Railton's strategy demands. For clearly, in such a case, the causal work is being done by the property of

being such as to discount the objective interests of some (group of) people, and not by the property of moral wrongness. For comparison, consider that every situation in which an object is coloured is a situation in which the object has a mass, but the colour of the object will not help explain why the object is subject to a constant gravitational force.

Railton's position might be saved if he were to make a still stronger claim – not only that *every* situation which is morally wrong involves the significant discounting of the objective interests of some affected group of people, but that the property of moral wrongness is simply *identical with* the property of being such as to discount the objective interests of some (groups of) people. Given the present interpretation of Railton's view – (3b) – this amounts to the claim that the property that (alone and always) grounds the disposition of situations to elicit disapproval from instrumentally rational people when considering equally the objective interests of all potentially affected individuals under conditions of full and vivid information is the property of being such as to significantly discount the interests of some affected group of people. In which case, the causal work that being done by the latter property is *ex hypothesi* the same work that is done by the moral property.[21] Since the property of being such as to discount the objective interests of some (groups of) people is causally explanatory, then, given such an identification, so is the property of moral wrongness.

Unfortunately for Railton, this approach fails. To be successful, it would need to defend the following two claims:

(4) The property of moral wrongness is identical with the property of being such as to discount the objective interests of some groups of people.

(5) The property of being such as to discount the objective interests of some groups of people is an informative explanatory property.

To assess these claims, it is necessary to enquire into what it is for interests to be 'discounted'. There appear to be three reasonable candidates. In the first case, we may say that interests are discounted when they are frustrated, or go unsatisfied. A food distribution system that fails to ensure that every member of a society is properly nourished would be discounting interests in this sense. In the second case, we may say that interests are discounted when they are not considered in processes of deciding what to approve or disapprove of, or more broadly, in processes of decision-making. A law-making body that failed to consider the interests of the elderly, or of the young, in deciding which laws to enact would be discounting interests in this sense. In the third case, we may say that interests are discounted when they both go unsatisfied and fail to be considered in processes of decision-making. A ruling body that failed to consider the interests of the elderly in forming seasonal policy, resulting in many of that group dying due to under-heated homes during winter, would be discounting interests in this third sense. Unfortunately, on none of these three interpretations of 'discounting interests' has Railton done enough to support both claims (4) and (5).

First, suppose we mean by 'discounting interests' simply not satisfying them. Following this interpretation, together with the definition of moral wrongness (3b), the claim (4) amounts to:

(4a) The property that makes it the case that: a situation would be disapproved of by instrumentally rational agents when considering equally the objective interests of all those potentially affected under conditions of full and vivid information *is identical with* the property of being such as not satisfy the objective interests of some groups of affected people.

This property identity claim is implausible. It is implausible because there may be situations in which the objective interests of some groups of people would not be satisfied, but that would *not* be disapproved of by instrumentally rational people when considering equally the objective interests of all those potentially affected under conditions of full and vivid information (indeed, that might even be approved of). Consider that in any social situation, it seems likely that peoples' objective interests would diverge considerably, and that, furthermore, there may be no way in which that society *could* be organised that guarantees that the objective interests of all its citizens were satisfied. A ruling body, therefore, might quite possibly be in a situation where, whatever it does, the objective interests of some subset of its citizens would remain unsatisfied. Nevertheless, there may still be some decisions of such a body that would not be disapproved of by instrumentally rational people considering equally the objective interests of all potentially affected individuals under conditions of full and vivid information (indeed, some decisions may even be approved by such people – perhaps because they ensure the highest possible number of objective interests are satisfied). It follows that the property of being such as to not satisfy the objective interests of some groups of affected people is not identical with the property that makes it that case that a situation would be disapproved of by instrumentally rational agents when considering equally the objective interests of all those potentially affected under conditions of full and vivid information. Thus, on this interpretation, the property identity fails and claim (4) is false.

Suppose, alternatively, we mean by 'discounting interests' not counting them equally. In that case the property identity is more plausible: it seems probable that any social arrangement which fails to count equally the interests of all those potentially affected would be disapproved of by instrumentally rational agents when considering

equally the objective interests of all those potentially affected. What such agents would disapprove of is, of course, an a posteriori matter, but let's grant Railton this claim for the sake of argument. Unfortunately, on this view of 'discounting', the second claim is no longer plausible. On this interpretation, (5) amounts to:

(5a) The property of being such as to not count equally the objective interests of some groups of people is explanatorily informative.

The problem is that it is the category of not *satisfying* objective interests, not the category of simply not counting them equally, that is the plausible explanatory category. Consider a situation that does not count equally the objective interests of some groups of people yet still satisfies those interests. For example, a particularly affluent society may have a law-making body that considers only the interests of high-earners in forming its decisions, yet that on this basis enacts laws that result in the satisfaction of the objective interests of all members of society. Such a situation would not be prone to satisfaction or unrest. Thus the property of being such as to not count equally the objective interests of some groups of people cannot be involved in the explanation of any social dissatisfaction or unrest. Thus on this interpretation the explanatory claim fails, and claim (5) is false.

Finally, suppose we mean by 'discounting interests' both not satisfying them and failing to consider them in processes of decision-making. Once again, that would make the property identity plausible: it seems probable any social arrangement which fails to count equally the interests of all those potentially affected and thereby leaves many interests unsatisfied would be disapproved of by instrumentally rational agents when considering equally the objective interests of all those potentially affected. This identity claim is, of course, incompatible with the one accepted for the sake of argument when discussing the second sense of 'discounting interests', so Railton can

only accept one of them. Unfortunately, accepting either leaves Railton's claim (5) unsupported. On this final interpretation, (5) amounts to:

(5b) The property of being such as to both not count equally the objective interests of some group of people and leave those interests unsatisfied is explanatorily informative.

The problem here is the similar to that facing (5a): it is the simple category of not *satisfying* objective interests, not any more complex category of both not satisfying and not counting equally objective interests, that is the plausible explanatory category. Consider two situations, both of which leave the objective interests of a significant group of people unsatisfied, but only one of which includes a consideration of those interests in its decision-making procedures. In this latter case, that their names are mentioned in the Halls of Power will be scarcely much consolation to those whose objective interests continue to be frustrated. Thus the tendency for instability will be just as strong in both cases. Thus, it is the simple category of not satisfying objective interests, and not any more complex conjunctive category, that is explanatorily informative. Hence (5b) is false. On no interpretation of 'discounting interests', therefore, has Railton provided sufficient a posteriori grounds to believe that the property of moral wrongness – when identified with the categorical grounds of the disposition he specifies – is an explanatorily informative property. Once again, therefore, Railton has not provided sufficient reason to think that moral properties play a part in the good empirical theories which his methodology – and the first kind of naturalism in ethics – requires.

It is important to note that the arguments of the last two sections against Railton are not of mere parochial interest. They contain general arguments against any

naturalist who hopes to account for moral properties in dispositional terms and who accepts the condition on naturalistic respectability imposed by the first kind of naturalism in ethics.[22] Such views face the same choice I posed for Railton: either identify moral properties with the dispositional properties themselves, or with the categorical grounds of those dispositions. In the first case, it is a general point that should the conditions specified in the disposition be idealised or otherwise uncommon, little explanatory force will accrue to moral properties (§6.1). In the second case, it is likewise a general truth that there can be no a priori guarantee that those grounds themselves fall into an explanatorily informative category, and hence no a priori guarantee that moral properties, thus identified, are themselves explanatorily informative (§6.2). Railton recognises this deficit, and provides the beginning of some a posteriori considerations to support his belief in the explanatory potency of moral properties. I have argued that these considerations are not persuasive. But the need to provide them is incumbent on any naturalist who takes this path.

These general points amount to a presumptive case against the possibility of naturalists providing both a dispositional account of moral properties and a defence of the claim that those same properties play an informative explanatory role of the type demanded by the first kind of naturalism in ethics. It is worthwhile considering, therefore, how the naturalist might escape this inconsistency. In the next section, I will consider the prospects for ethical naturalism were it to drop the condition imposed by the first kind of naturalism in ethics.[23]

7. Other Moral Explanations

I have argued that Railton's reforming definition of moral wrongness – as captured by (3) – can be understood as making one of two property identifications. On neither understanding, however, has Railton shown how such a property can participate in the good empirical theories that his methodology – and the first form of naturalism – requires. A closely parallel argument also counts against Railton's definition of moral rightness.

One response to these arguments is to hold fast to Railton's methodology – and hence the constraint imposed by the first type of naturalism in ethics – and reject his definitions. An alternative response is to reject Railton's methodology and move to the second kind of naturalism in ethics.[24] Below, I suggest that this response is consistent with the underlying motives for Railton's naturalism.

What are the advantages of defining a concept in terms of a naturalistically respectable property? Besides simplifying ontology, the key benefit, as recognised by Railton, is that it makes available naturalistic accounts of our semantic and epistemological access to the property thus defined (205).[25] For any domain of putative properties which we claim to discourse meaningfully about, we must be able to show both how our terms can get to refer to such properties and how, in favourable cases, we might have knowledge of them. If the properties of the domain are naturalistically respectable, then the possibility arises that our access to such properties is a result of being related to them in some naturalistically respectable way – causally, perhaps. For example, there appears to be an explanatory constraint on epistemic access to a domain of facts according to which one can only be said to know that *p* if that very fact can play some part in the explanation of one's belief that

p.[26] If *p* is a natural fact then this explanation can be part of a causal empirical theory, and our epistemic access to the realm of facts is explained naturalistically.

If this is the motivation behind defining moral properties in naturalistically respectable terms, however, it doesn't entail that those properties should participate in their own right in well-articulated nomic nexus that afford explanatory insight into non-moral phenomena. All that is required is that we can offer particular explanations in instances whenever we are semantically or epistemologically in contact with those properties. An example of the epistemological sort will help make this clear.

Suppose that Donnie believes that ¬ϕ is morally right, and does so because he has been taught this at his mother's knee. Suppose that in actual (moral) fact, it is not the case that ¬ϕ is morally right, rather, ϕ is morally right. According to Railton's account of moral rightness (1), what makes ϕ right is that it would be approved of by instrumentally rational people when considering equally the objective interests of all potentially affected individuals under conditions of full and vivid information. Suppose that Donnie comes to realise this, that is, comes to realise that ϕ would be approved of by such people in such conditions. This realisation may the result of Donnie himself coming to meet such conditions and sharing the approval, or indirectly by realising that someone else has come to satisfy them and share the approval. There is no reason to suppose that this realisation will cause Donnie to alter his moral evaluation of ϕ (and ¬ϕ), but if it does, that is, if it causes Donnie to stop believing that ¬ϕ is morally right and start believing that ϕ is morally right, then we might offer the following explanation of Donnie's resultant belief:

(D) Donnie believes that ϕ is morally right because ϕ is such as to elicit approval from instrumentally rational people when considering equally

> the objective interests of all potentially affected individuals under conditions of full and vivid information and Donnie has come to realise this.

Given Railton's naturalistic definition of moral rightness (1), this explanation is equivalent to:

> (E) Donnie believes that ϕ is morally right because ϕ is morally right and Donnie has come to realise this.

Notice that in (E) the property of being morally right has a genuine explanatory role to play in the evolution of Donnie's belief – Donnie's belief has successfully 'tracked' the responses of instrumentally rational people when in the relevant idealised situation.[27] Since this is what, on Railton's view, defines moral rightness, Donnie's belief has successfully tracked the property of moral rightness. This property therefore plays a role in the explanation of his belief.

However, though such explanations allow properties such as moral rightness explain things such as Donnie's belief, this is far from the type of explanation required by the first type of naturalism in ethics. For that type of naturalism – embraced by Railton – requires that moral properties are causally explanatorily *independent* of their effects on agents' moral judgements. Therefore, even if Donnie's belief were to be causally explanatory of some non-moral event, this would not be sufficient for moral rightness to satisfy the condition on naturalistic respectability imposed by the first kind of naturalism in ethics.

Nevertheless explanations such as (E) do satisfy one of Railton's desiderata, namely they allow the moral realist to employ naturalistic accounts of how we come to have knowledge of moral truths, since they can be interpreted as providing naturalistic explanations that meet the explanatory constraint on epistemic access. The moral realist who accepts Railton's account of moral rightness, therefore, can accommodate the insights of a naturalistic epistemology (and naturalistic semantics) without necessarily meeting Railton's more stringent criterion of naturalistic respectability. In other words, he can accept Railton's definitions so long as he becomes a naturalist of the second, less demanding, sort.

8. Conclusion

I have argued that Railton's definitions of moral rightness and moral wrongness fail on his own terms. They fail because they do not show how either property can play a role in good empirical theories on a par with those theories of the natural and social sciences. They thus fail to meet a condition on naturalistic respectability imposed by the first kind of naturalism in ethics. I have also argued, however, that naturalistic definitions such as Railton's may be acceptable were naturalists to drop this condition and espouse the second, less demanding, form of naturalism. Finally, I have suggested that this methodological revision may be consistent with the underlying motivations for naturalism, as given by Railton. These conclusions do not entail that the first type of naturalism in ethics is misguided, but they do entail that those who find Railton's account of moral properties appealing would do better to become naturalists of the second kind.[28]

References

Blackburn, S., Just Causes, Philosophical Studies 61(1) (1991), pp.3-17.

Blackburn, S., Circles, Finks, Smells and Biconditionals in Tomberlin, J. (ed.) Philosophical Perspectives 7: Language and Logic. California: Ridgeview, 1993, pp.259-279.

Boyd, R., How to be a Moral Realist, in Sayre-McCord, G. (ed.) Essays on Moral Realism. Ithaca, N.Y.: Cornell University Press, 1988, pp.181-228.

Brink, D.O., Moral Realism and the Foundations of Ethics. Cambridge: Cambridge University Press, 1989.

Majors, B., Moral Explanation and the Special Sciences, Philosophical Studies 113(2) (2003), pp.121-152.

McDowell, J., Values and Secondary Qualities, in Honderich, T. (ed.) Morality and Objectivity, Boston: Routledge & Kegan Paul, 1985, pp.110-129.

Nagel, T., The Limits of Objectivity, in McMurrin, S. (ed.) The Tanner Lectures on Human Values I. Salt Lake City: University of Utah Press, 1980, pp. 75-139.

Nozick, R., Philosophical Explanations. Oxford: Oxford University Press, 1981.

Railton, P., Moral Realism, Philosophical Review 95(2) (1986), pp.163-207.

Railton, P., Naturalism and Prescriptivity, Social Philosophy and Policy 7(1) (1989), pp. 151-174.

Railton, P., Moral Explanation and Moral Objectivity, Philosophy and Phenomenological Research 58(1) (1998), pp. 175-82.

Sturgeon, N., Moral Explanations, in Copp, D. and Zimmerman, D. (eds.) Morality, Reason and Truth. Totowa, N.J.: Rowman & Allanheld, 1985, pp.49-78.

Sturgeon, N. What Difference Does it Make if Moral Realism is True?, Southern Journal of Philosophy, supp. vol. 24,1986, pp. 115-142.

Wiggins, D., Moral Cognitivism, Moral Relativism and Motivating Moral Beliefs, Proceedings of the Aristotelian Society 91 (1990), pp. 61-86.

NOTES

[1] See, for example, Sturgeon (1985), Railton (1986), Brink (1989) and Majors (2003). A distinct issue is whether the availability of moral explanations of this sort suffices to show that moral properties exist. For this latter debate, see Sturgeon (1986) and Blackburn (1991).

[2] For ease of reading I assume an events-based ontology, though my arguments do not depend on it.

[3] The explanations I have in mind are common in the literature: "Children thrive when treated with decency and humanity" and "The revolution was a result of the injustices suffered by the working classes". See Sturgeon (1986).

[4] The rejection of the explanatory condition is more commonly associated with non-naturalistic moral realists: see Nagel (1980) and McDowell (1985).

[5] One further caveat: since explanation is transitive, and since the making of moral judgements can often explain other non-moral events (such as agents' actions), even the second type of naturalist will concede that moral properties can participate, indirectly, in explanations of non-moral events other than the making of moral judgements. But the first type of naturalist demands further that moral explanations of such events sometimes be *direct*, in the sense that they do not transmit through the making of moral judgements. This is why I choose as my examples moral explanations that are not easily considered elliptical for explanations involving the making of moral judgements. See Blackburn (1991).

[6] Railton (1986). Subsequent numbers in brackets are page references to this paper.

[7] Railton (1989, pp.155-6).

[8] Railton (1998, p.179).

[9] Railton (1998, p.180).

[10] Sturgeon (1986) shares this deference to empirical discovery and, like Railton, is optimistic that such explanations will be forthcoming.

[11] See Miller (2003, p.197).

[12] Note that subjective and objectified subjective interests of agents are (actual or hypothetical) desires of the agent whereas objective interests are features of the agents' situation. Railton notes (175, n.16) that in the latter case 'interest' is not a happy term and suggests that 'positive-valence-making characteristic' may be a more accurate expression.

[13] See, for example, Blackburn (1993). Note that this characterisation of response-dependence accounts entails no particular view about the status of proposed biconditional equivalence. Following Blackburn (*ibid*.), we may distinguish three uses to which such a proposal may be put: an *a priori* analysis of the concept appearing on the left-hand side of the biconditional; an elucidation of the logic of the same concept and an *a posteriori* identification of the property specified on the left-hand side with the property specified on the right-hand side. Railton's position is most similar to the third approach, although by treating his *a posteriori* identification as a reforming *definition* he avoids the possible problems raised by seeming to offer a contingent identity statement.

[14] As Blackburn points out (1993) the distinction between those conditions that are part of P and those that are part of C is somewhat arbitrary.

[15] Strictly speaking, the reference to ideally instrumentally rational and fully informed *agents* is unnecessary, since Railton's view can be expressed in terms of the deliverances of the decision procedure that such agents would employ. The agent-based interpretation of Railton is common (see, for example Miller 2003 p.196), but the use of it here is not necessary for the argument, since the distinctions I draw below between dispositional and categorical interpretations of Railton's view would apply whether the dispositions concerned are those of instrumentally rational agents or of their decision procedures.

[16] I use the term 'categorical ground' to refer to those properties in virtue of which objects possess specified dispositional properties. The term is not intended to prejudge any metaphysical issues as to the nature (dispositional or otherwise) of these underlying properties.

[17] Note a parallel here with objectified subjective and objective interests. Claims about objectified subjective interests are made true by dispositional facts (facts about what idealised counterparts would want the non-idealised agent to want) whereas claims about objective interests are made true by categorical facts (facts that make it the case that these dispositional facts obtain). Railton identifies facts about *non*-moral goodness with the latter (176), but gives no indication that a similar view is intended for the case of *moral* facts. In any case, my argument is that neither identification satisfies Railton's criterion of naturalistic respectability.

[18] Other naturalists fond of this example include Brink (1989) and Sturgeon (1985, 1986).

[19] Railton goes on to claim that "the discontent produced by departures from social rationality may produce feedback that, at a social level, promotes the development of norms that better approximate social rationality" (193). In other words, if explanations such as (A) are acceptable, the instantiation of moral wrongness may play a wider role in the explanation of processes of social change. However, since I reject explanations such as (A) I also reject any such wider explanatory role.

[20] By 'directly' I simply mean not acting through some causal intermediary.

[21] Railton (1989 p.161).

[22] Dispositional accounts of moral properties are not uncommon. See for example, [[]]. These authors are less clear in whether they would accept the condition imposed by the first kind of naturalism.

[23] Due to lack of space, the other option – that of maintaining the explanatory condition whilst abandoning a dispositional account of moral properties – will not be discussed here. See Sturgeon (1985, 1986).

[24] It is an interesting question – not addressed here – which of these options Railton would himself prefer (assuming, of course, he accepts the arguments of §6). My hunch is that he is more attached to the explanatory condition on naturalistic respectability than the particular definitions he offers; but this is only a hunch. In any case, my arguments have shown that Railton cannot have both.

[25] See also Railton (1989 p.161, 1998 p.175) and Boyd (1988).

[26] Wiggins (1990).

[27] Nozick (1981).

[28] I would like to thank an anonymous referee at the British Society for Ethical Theory.