

## Expressivist Explanations

Neil Sinclair  
Department of Philosophy  
University of Nottingham  
University Park, Trent Building  
Nottingham, NG7 2RD, UK  
[neil.sinclair@nottingham.ac.uk](mailto:neil.sinclair@nottingham.ac.uk)

Note: This is a final draft of the paper which appears in *Journal of Moral Philosophy* 9 (2012) pp.1-32. The final version can be found here: <http://www.brill.nl/journal-moral-philosophy>

*Abstract: In this paper I argue that the common practice of employing moral predicates as explaining phrases can be accommodated on an expressivist account of moral practice. This account does not treat moral explanations as in any way second-rate or derivative, since it subsumes moral explanations under the general theory of programme explanations (as defended by Jackson and Pettit). It follows that the phenomenon of moral explanations cannot be used to adjudicate the debate between expressivism and its rivals.*

*Key Words: Expressivism; Moral explanation; Programme explanation.*

Moral predication can intelligibly occur in the explaining phrase of explanations. At least, this is an assumption of ordinary engagement of moral practice, exhibited in our tendency to understand (if not agree with) explanations such as:

- (1) The revolution of 1917 was a direct result of the injustices suffered by the working classes.

- (2) Hitler ordered the Holocaust because he was morally depraved.
- (3) Brian thrived because his parents treated him with decency.
- (4) Lord Earl judges that capital punishment is wrong because he is decent and fair.
- (5) I hold that testing cosmetics on animals is wrong because testing cosmetics on animals *is* wrong.

Most generally, by a ‘moral explanation’ I shall mean a putative explanation of a particular (or type of) non-moral event (or act or state of affairs) that features moral predication in the explanans. Since moral explanation appears to be an everyday part of our moral practice, it is desirable that our meta-theory show it to be neither mistaken nor dispensable. Some moral realists have argued that, if expressivism is the correct meta-theory of moral practice, the practice of offering moral explanations is expendable and therefore that the persistence and apparent usefulness of such explanations provides an argument against expressivism.<sup>1</sup> In this paper I assess this line of argument. I will argue that not only can expressivists accommodate our linguistic practice of offering moral explanations, but the accommodation doesn’t necessitate viewing moral explanations as in any way less serious than non-moral explanations. This last point emerges from consideration of the theory of programme explanations put forward by Jackson and Pettit.

## **1. Expressivism and realism**

Expressivist accounts of moral practice hold that moral judgements such as ‘Torture is never permissible’ do not express mental states that represent the obtaining of moral

---

<sup>1</sup> Moral explanations of moral features, such as saying that an action is wrong because it is unjust, are not my concern here. Such explanations merely offer one moral verdict in support of another and so present no particular problem for expressivism.

states of affairs, but instead express practical, non-representational states such as approval and disapproval, the purpose of such expression being the mutual coordination of attitudes and hence action. One point of contrast is with moral realists, who hold that moral judgements express moral beliefs answerable to moral states of affairs. It has long been recognised that, in comparison with such views, expressivism lacks a natural explanation of many of the forms and assumptions of our engagement with ordinary moral discourse. For example, ordinary moralisers assume that some moral judgements are true, some knowable, that moral sentences can be intelligibly embedded in unasserted contexts (such as negations and conditionals) and that moral features can (sometimes) be explanatory. It is unclear how such assumptions might be preserved on an expressivist view, for if there is no moral reality there can be no true descriptions of it, no knowledge of its contours, no truth-conditional content on which to deploy the tools of logic and no explanatory role for moral properties. Indeed, the lack of a plausible expressivist vindication of such assumptions is often held to create a strong presumptive case in favour of moral realism.<sup>2</sup> It follows that insofar as the expressivist is able to show that the relevant phenomena *can* be accommodated, this presumptive case is undermined. The following argument can be seen in this light. I will show that expressivism is compatible with the practice of offering moral explanations (once that practice is properly understood) and hence this practice will not help determine the issue between realists and expressivists. This is to domesticate the phenomenon of moral explanation.<sup>3</sup>

---

<sup>2</sup> See D. Brink, *Moral Realism and the Foundations of Ethics* (Cambridge: Cambridge University Press, 1989), chapter 2 and R. Shafer-Landau, *Moral Realism* (Oxford: Oxford University Press, 2003), chapter 1.

<sup>3</sup> For the notion of 'domestication' see S. Blackburn, *Essays in Quasi-Realism* (Oxford: Oxford University Press, 1993), p. 3.

## **2. Two unsuccessful expressivist strategies for accommodating moral explanations**

To help understand the phenomenon to be accommodated it is useful to consider two expressivist strategies that fail. The first strategy draws on the thought that the practice of offering moral explanations is – unlike the assumption of truth-aptness, knowledge-aptness, intelligible embedding and the rest – peripheral to the core functions of moral practice. Given this minor role, this strategy recommends that the practice of using moral sentences as explaining phrases is one that we can happily do without. This revisionary approach leaves the expressivist with no phenomenon to accommodate.

The costs of offering a revisionary account of moral practice varies according to the nature of the practice that it is proposed we revise. Two desiderata for successful meta-ethical theories are (i) their ability to accommodate the forms of ordinary moral practice and the deeply embedded assumptions of those who engage in it and (ii) their consonance with our wider theories of the topics they make claims about, notably semantics, epistemology and metaphysics.<sup>4</sup> It is plausible, however, to think that no meta-ethical theory will be wholly successful in satisfying the first desiderata. One reason for this is that moral practice and its assumptions are neither homogenous nor static, and are influenced by prevailing intellectual culture, including philosophical culture. So, for example, some philosophers may include moral realism as a deeply embedded assumption of their ordinary moral practice, while others may include expressivism. Both cannot be vindicated. Another reason to doubt that any meta-ethical view will satisfy the first desiderata is the possibility that our actual

---

<sup>4</sup> M. Timmons, *Morality without Foundations* (Oxford: Oxford University Press, 1999), chapter 1.

moral practices contains residuals of older and now discredited practices. For example, Anscombe has argued that the notion of *obligation* makes no sense without belief in a (divine) law-giver and hence should be abandoned.<sup>5</sup> One needn't agree with Anscombe's particular example to agree that it is likely that our actual moral practices contain now misguided elements such as these, elements that no (non-misguided) meta-theory can accommodate. Hence it is likely that any meta-ethical theory will be to some extent revisionary, that is, propose that certain features or assumptions of moral practice cannot be accommodated and should be abandoned, perhaps to be replaced by theoretically less problematic surrogates. How then, are we to judge competing revisionary theories? The answer is that each feature or assumption that is revised carries an explanatory cost, but that this cost can be met so long as the resulting meta-ethical theory can show how all the important things that we want to say and do using moral discourse, and all those substantial debates moral discourse involves, are preserved in the recommended revised practice.<sup>6</sup>

The question arises, therefore, whether abandoning the practice of offering moral explanations carries an unacceptable revisionary cost. Unfortunately for the expressivist, there are three considerations which together show that the revisionary cost of abandoning moral explanations is too high to bear.

First, moral explanations are not uncommon and often appear informative, at least in the sense that it is possible to imagine apparently rational people being satisfied by them. Being told that the revolution of 1917 took place because of the injustices suffered by the working classes seems to be an explanatory gain. At the very least (and if true) it entails that it was the conditions suffered by the working

---

<sup>5</sup> G.E.M. Anscombe, 'Modern Moral Philosophy', *Philosophy* 33 (1958), pp. 1-19.

<sup>6</sup> See C.L. Stevenson, *Facts and Values: Studies in Ethical Analysis* (New Haven and London: Yale University Press, 1963), pp. 11 & 117; S. Darwall, A. Gibbard and P. Railton, 'Toward Fin de siècle Ethics: Some Trends', *Philosophical Review* 101 (1992), pp. 115-89, p.127; A.J. Ayer, 'The Analysis of Moral Judgements', in A.J Ayer, *Philosophical Essays* (London: Macmillan, 1954), pp. 231-49.

classes, and not some other factor, that started the revolution. There may be historical reasons to doubt this claim (perhaps it was the stirrings of political agitators or a clash of class ideologies that *really* brought about the revolution) but the proposed rejection of moral explanations that we are considering is more systematic than this. The claim is that *any* explanation featuring a moral predication in the explanans can be rejected, even before we do any further investigation into the actual antecedents of the explanandum. But the prevalence of such explanations and their apparent informativeness gives us cause to doubt this claim: *prima facie*, there seems nothing wrong with this type of moral explanation.

Second, moral explanations, at least in some instances, involve counterfactual dependence of the explanandum on the explanans. That is, in some cases it is true that if the moral verdict given in the explanans had been different then the event (act or state of affairs) described in the explanans would not have occurred (or been brought about). This provides a more concrete sense in which such explanations are informative.<sup>7</sup> To see this, consider the following explanation:

(6) Gilbert judges that setting fire to cats is wrong because setting fire to cats  
*is* wrong.

To test the counterfactual dependence of this we need to ask whether Gilbert would hold that setting fire to cats is wrong even if setting fire to cats wasn't wrong, that is, even if setting fire to cats was permissible. Now, it is commonly held that a criterion of competence for the use of moral predicates is that one's ascriptions of them supervene on one's ascriptions of natural predicates.<sup>8</sup> Accordingly, when asked to

---

<sup>7</sup> N. Sturgeon, 'Moral Explanations' in D. Copp and D. Zimmerman (eds.), *Morality, Reason and Truth* (Totowa, N.J.: Rowman and Allanheld, 1985), pp. 49-78, §III.

<sup>8</sup> S. Blackburn, 'Moral Realism' in J. Casey (ed.), *Morality and Moral Reasoning* (London: Methuen, 1973), pp. 101-24 and 'Supervenience Revisited' in I. Hacking (ed.), *Exercises in Analysis* (Cambridge: Cambridge University Press, 1985). This is the linguistic counterpart of the claim that moral properties supervene on natural properties: see S. Blackburn, *Spreading the Word* (Oxford: Clarendon Press,

imagine a situation where one is no longer warranted in asserting that ‘Setting fire to cats is wrong’, one is also asked to imagine a situation in which certain ascriptions of natural predicates are no longer warranted, that is, situations in which the natural facts are different. Suppose, for example, that the wrongness of setting fire to cats supervenes on its being an act that causes unnecessary pain. On this assumption being asked to imagine a situation in which setting fire to cats is permissible is also being asked to imagine a situation in which setting fire to cats causes them no unnecessary pain (perhaps, as in Greek myth, a world in which fire does not consume). If Gilbert takes the wrongness of setting fire to cats to be a result of it causing unnecessary pain, then there will indeed be counterfactual dependence of this judgement on the wrongness of setting fire to cats: if setting fire to cats was permissible, it wouldn’t cause cats unnecessary pain, and accordingly Gilbert wouldn’t judge it wrong. Hence explanation (6) would exhibit counterfactual dependence of the explanandum on the explanans. Of course, Gilbert may not be like this: he may take the wrongness of setting fire to cats to be a result of it being contrary to divine scripture, even though in fact it is a result of it causing unnecessary pain. In that case, there would be no counterfactual dependence of Gilbert’s judgement on the wrongness of setting fire to cats: even if setting fire to cats was permissible, and hence wouldn’t cause the cats unnecessary pain, Gilbert would still hold that it was wrong, because it would still be contrary to divine scripture. In such a case there would be no counterfactual dependence of the explanandum on the explanans. But this is just because the explanation is a bad one. The general point is that there is in principle no bar on (good) moral explanations revealing relations of counterfactual dependence and hence being informative. To test whether the explanandum counterfactually depends on the

---

1984), p. 184. I use the linguistic claim since at this point I wish to beg no questions on the issue of whether a distinct realm of moral properties exists.

explanans in a moral explanation of the type we are considering we need to consider the closest possible case where the predication of the moral feature (the explanans) is different. Since moral features (or our ascriptions of them) supervene on natural features, this will also be a case where some of the natural features are different. If these natural features were in any way productive of the original explanandum (say, Gilbert's judgement), then changing them will mean that the explanandum will no longer come to be, and there will indeed be a counterfactual dependence of the explanandum on the explanans.

Third, abandoning moral explanations would require abandoning talk and thought of moral knowledge, since it is plausible that a certain type of moral explanation is a necessary condition for moral knowledge. To explain. Gibbard – an expressivist who hopes to vindicate many of the assumptions of ordinary moral discourse – recognises the requirement that judgements that aspire to knowledge must be able to be explained by the facts that feature in them, or else we should abandon them. To illustrate the point Gibbard asks us to imagine that our belief in the evolution of humanity by natural selection is the result of that doctrine being made up by scientists who 'dimly realised that if Darwinian explanations were accepted, they would attain the cultural role of priests' and that the traces supposedly left by evolution 'had nothing to do with it'.<sup>9</sup> If this explanation is correct, then we would give up not only our former view that the facts of evolution caused us to believe in it, but also our belief in evolution itself.

It seems that this point is naturally extended to all judgements. In order to think of judgements as aiming at truth, and as possible candidates for knowledge, it must at least be possible to explain those judgements in terms of the very facts that

---

<sup>9</sup> A. Gibbard, *Wise Choices, Apt Feelings: A Theory of Normative Judgement* (Oxford: Oxford University Press, 1990), p. 124.

feature in them. In other words, the following is a necessary condition for a subject S knowing a proposition  $p$ : S knows that  $p$  only if S's judgement that  $p$  can be explained by  $p$ .<sup>10</sup> This condition seems plausible even in those cases where we are uncomfortable talking about *causal* explanation, for example with one's knowledge of mathematics. No one (to my knowledge) suggests that one's judgement that  $7+5=12$  is actually caused by the fact that  $7+5=12$ , yet we are comfortable with offering the (non-causal) explanation: 'Jane believes that  $7+5=12$  because  $7+5$  is equal to 12'. The effect of doing so is to confer epistemic credibility on Jane's judgement. Other cases confirm the condition. For example, we wouldn't be comfortable saying that John *knows* that the Earth is round, unless we could provide some sort of explanation (however long or complex) linking John's belief with the Earth's actually being round. All knowledge, therefore, seems to require the possibility of such an explanation. This simple point is potentially damaging for the expressivist who hopes to abandon moral explanations. For it raises the theoretical stakes. If a certain type of moral explanation is necessary for moral knowledge, then abandoning all moral explanations entails abandoning talk and thought of moral knowledge. And thought and talk of moral knowledge – unlike perhaps moral explanations – is an integral part of moral practice. Furthermore, the expressivist cannot simply accept one particular type of moral explanation (where the explanandum is a moral judgement) and reject the others, both because of a lack of any principled criterion to do so (all such explanations involve a moral explanans and a non-moral explanandum) and because of the transitivity of explanation (if moral predication can explain my judgement and this judgement can explain future events then moral predication can explain those

---

<sup>10</sup> See D. Wiggins, 'Moral Cognitivism, Moral Relativism and Motivating Beliefs', *Proceedings of the Aristotelian Society* 91 (1990), pp. 61-86; S. Rieber, 'Scepticism and Contrastative Explanation', *Noûs* 32 (1998), pp. 189-204; C. Jenkins, 'Knowledge and Explanation', *Canadian Journal of Philosophy* 46 (2006), pp. 137-64.

events). Thus the revisionary cost of abandoning explanations cannot be insulated from the (higher) cost of abandoning moral knowledge. (This is not to say, of course, that expressivists can give a plausible account of moral knowledge, only that their acceptance of moral explanations is a necessary condition of doing so.)

For these reasons, therefore, the expressivist is best to look for an approach that can preserve the practice of making moral explanations.

One unsuccessful way of doing this is to treat moral explanations as elliptical for longer explanations that replace apparent reference to moral properties (rightness, wrongness and so on) to reference of perceptions of, beliefs about, or judgements regarding such features. So, for example, citing injustices suffered by the working classes as the cause of the revolution might be adverting to the working classes' *perception* that they were the victims of injustice, just as when explaining Johnny's success with women by saying 'He's very funny' one is adverting to the fact that most women perceive Johnny to be funny. In these cases it is this conglomeration of psychological states, and not the injustice or funniness itself, which is the true explanans.

There are, however, two problems with this treatment of moral explanations.

First, the defender of this view would have to rule out the possibility that any moral predication is involved in explaining the psychological states themselves, since, if he doesn't, the transitivity of explanation would ensure that the original moral explanation was reinstated.<sup>11</sup>

Second, several of the examples cannot be elliptical in this way. Consider (3). The decency of Brian's upbringing would explain his thriving even if neither he nor his parents perceived it as decent; it is enough that it *was* decent. Further, any

---

<sup>11</sup> N. Sturgeon, 'Contents and Causes: A Reply to Blackburn', *Philosophical Studies* 61 (1991), pp. 19-37, fn. 16.

paraphrase of (5) that replaced the second ‘wrong’ with ‘is held by me to be wrong’ would render the explanation trivial, when *prima facie* at least it is not. More generally, it is premature to suppose that all moral explanations are elliptical in this way. Whether it is more accurate to say that it was injustice, or perceptions of injustice, that caused the revolution is a question for social historians and one that should not be foreclosed by philosophical reflection.

### **3. The possibility of an expressivist account of moral explanations**

Fortunately, there are ways for the expressivist to accommodate moral explanations without reducing them to explanations of other kinds. Before coming to the details of one such account, however, it is first necessary to defend its very possibility. For it can seem that the expressivist is committed to rejecting moral explanation, that is, that expressivism entails that there are no correct moral explanations. The argument is as follows: In moral explanations moral features appear as explanatory items. But according to expressivism there are no moral features, that is, no properties of rightness, wrongness, justice and so on. Therefore expressivism cannot accommodate moral explanations, and must look either to abandon or paraphrase them.<sup>12</sup>

This argument is unsound because it trades on an ambiguity in the phrase ‘moral features’. When one claims that according to expressivism ‘there are no...moral features’ one might have one of two thoughts in mind. First, that according to expressivism, it is never correct to sincerely employ moral predicates. This would be to accept that no act or state of affairs is ever good, wrong, permissible, cruel, just, unjust, deserved and so on. It would be to deny that there are any standards

---

<sup>12</sup> N. Sturgeon, ‘What Difference Does it Make if Moral Realism is True?’, *Southern Journal of Philosophy, Supplementary Volume 24* (1986), pp. 115-42.

appropriate to the application of moral predicates and thus to be a moral sceptic. Second, one might think that, according to expressivism, there is no distinct realm of moral facts and instantiated moral properties to which our moral judgements are somehow responsive. This doesn't entail that all possible moral judgements are false or unwarranted, just that they cannot be understood in descriptive terms.

It should be clear that expressivism only denies that there are 'moral features' in the second sense. Expressivists deny that moral judgements can be understood to express representational states that are answerable to the actual distribution of a distinct realm of moral properties (such as are believed to exist by the moral realist), but they insist that there is an alternative way of understanding moral predication and the standards governing it: the judgements are expressive of attitudes for the purposes of mutual co-ordination and the standards governing such expression are those appropriate to such a co-ordinating practice. Expressivists do not suggest that we should desist from making (genuine, full-blooded) moral proclamations – they just have a different view of what it was to do so. So expressivists can in good faith hold onto their moral convictions (and so accept that there are moral features in the first sense) whilst denying that those convictions express states of mind that describe a distinct realm of facts (and so deny that there are moral features in the second sense).

If expressivism is not committed to giving up all moral predication then the above argument against expressivism can be resisted. The first premise points to the fact that moral predications are employed in the explanans of non-moral phenomena. For the argument to be valid, the second premise must therefore be read as saying that expressivism rules out the use of such predication in such contexts. If expressivism does indeed rule out such predication, then the conclusion follows. But, of course, the expressivist is not in the business of ruling out all moral predication as false or

unwarranted – he is not a moral sceptic.<sup>13</sup> Once expressivists have accounted for sincere moral predication in general then there is no reason why he should rule out the use of moral predicates in explanatory contexts such as (1)-(6).

#### 4. The ontic model

It may seem, therefore, that moral explanations like (1)-(6) present no problem for the expressivist. Such explanations involve moral predication in the explanans and so long as expressivism can account for the phenomenon of moral predication in general, there is no reason to think that the explanatory context presents him with any special problem. In other words, the use of moral predication in explanatory contexts no more refutes expressivism than the use of moral predication *tout court*.

But this is not yet to give an account what we are doing when we use moral sentences as explaining phrases. As in the case of unasserted contexts, an account of the semantics of simple indicative moral sentences doesn't suffice for an account of the role of such sentences in other contexts. To defend moral predication is not yet to understand its role in moral explanations.

Indeed a further problem arises for the expressivist. For a standard and plausible way of understanding explanatory contexts might be labelled the 'ontic model'. According to the ontic model we understand an explanation such as 'The plant died because it was over watered' in two stages. First, both the explanandum ('the plant died') and the explanans ('it was over-watered') are taken to refer to

---

<sup>13</sup> Pace Sturgeon, 'Moral Explanations' and 'What Difference Does it Make...'. A similar damaging ambiguity persists in some quarters over the claim that expressivists deny that there are moral 'facts'. See, for example, Shafer-Landau, *Moral Realism*, pp. 27 & 113. If the claim is that expressivists deny it is ever appropriate to sincerely utter indicative moral sentences, then it is simply false. If, on the other hand the claim is that expressivists deny the existence of distinctly realist realm of moral properties to which our moral predicates refer, it is true but undamaging.

aspects of the world. So, for example, on an events-based ontology, ‘the plant died’ is taken to refer to the event that we can characterise as *the death of the plant* and ‘it was over-watered’ is taken to refer to the event we can characterise as *the over-watering of the plant* (other ontologies are available). Second, the explanation as a whole is then taken to assert that the two aspects of the world thus picked out stand in some worldly relation – in this example the likely relation seems to be that of causation.<sup>14</sup> The explanation as a whole, therefore, is understood as saying that the occurrence of one event (the over-watering of the plant) was a causal factor in the occurrence of a later event (the death of the plant).<sup>15</sup> The problem for expressivism is that the ontic model cannot be applied to moral explanations, because if expressivism is true moral explanans cannot be understood as referring to worldly (moral) properties. And if the explanans fails to refer, the explanation as a whole cannot be understood as attributing a worldly relation to two distinct referents, and thus, given the ontic model, must be rejected as confused. The challenge thus posed to the expressivist is whether he can provide an account of what we are doing when we offer explanations such as (1)-(6) if it isn’t referring to ontologically entwined relata. Failure to meet this challenge will lead to the charge that the expressivist forecloses empirical questions regarding the veracity of moral explanations by *a priori* reflection.<sup>16</sup>

## 5. An expressivist account of moral explanations

---

<sup>14</sup> Other worldly relations referred to in explanations might be: constitution, the relation between a whole and its parts.

<sup>15</sup> This ontic account generalises the causal account given in D. Lewis, ‘Causal Explanations’ in D. Lewis *Philosophical Papers II*. Oxford: Oxford University Press, 1986) and W.C. Salmon, *Scientific Explanation and the Causal Structure of the World* (Princeton, N.J.: Princeton University Press, 1984).

<sup>16</sup> Sturgeon, ‘What Difference Does it Make...’, p. 122.

The expressivist can reject a direct application of the ontic model to moral explanations whilst still relying on some notion of reference to ontologically entwined relata. To see how this can be done, in this section I consider and develop suggestions made by Blackburn and Gibbard. In subsequent sections I defend this account against attacks from Sturgeon and Brink.

The basic point is that the use of moral predication in explanans ‘points downwards to’ or ‘picks-out’ the properties on which the moral verdicts depend. Blackburn explains:

According to me, an upbringing is decent and humane *in virtue of* other features – meeting the child’s needs, engaging with attempts at action and communication, and so on – and I may simply point towards those other, causally powerful properties by using the moral predicate. An education will have to be decent and humane in virtue of other properties, and those other properties will bring it about that children thrive (indeed, that’s largely why they get the honorific titles).<sup>17</sup>

To elucidate this theory, we can define a *moral standard* as a function from perceived non-moral features of the situation to moral verdicts. To adopt a moral standard is just to possess a certain way of forming moral attitudes in response to perceived features of the situation. For example, if expressivism is true, my use of the moral predicate ‘is unjust’ serves to express disapproval. But this disapproval is not randomly generated; rather, it is governed by my own moral standard. That is, there are certain *non-moral* features of situations which prompt me to respond to them with

---

<sup>17</sup> S. Blackburn, ‘Just Causes’, *Philosophical Studies* 61 (1991), pp. 3-17, p.11. See also Gibbard, *Wise Choices, Apt Feelings*, p. 186.

the condemnation signified by use the predicate ‘is unjust’. These features are those non-moral features *in virtue of which* I take the situations to warrant the predicate ‘is unjust’. One might also say that these features are what I take injustice to *consist in*. Note that the relation *in virtue of* is not that of entailment, for no description of an object of evaluation entails a moral verdict. On the other hand, there is reason to think that any moral standard must obey supervenience, which is to say that there can be no warranted change in the moral predication without their being a change in non-moral features of the situation.<sup>18</sup>

With a notion of moral standards in hand, we can grasp the import of moral explanations in terms of them. Consider explanation (1):

- (1) The revolution of 1917 was a direct result of the injustices suffered by the working classes.

(Suppose that this is a genuine moral explanation, not elliptical for an explanation citing only agents’ moral attitudes.) Suppose my moral standard is that situations are unjust in virtue of having an unequal distribution of resources. In other words, it is because certain situations exhibit an unequal distribution of resources that I disapprove of them in the way expressed by the predicate ‘is unjust’. In that case, my use of the predicate ‘is unjust’ in the explanans can be thought to ‘point down to’ or ‘pick out’ the presence of this very property. The explanation is then understood as saying that the property ‘picked out’ by my use of moral predication is ontologically entwined (in this case, most likely causally related) with the event referred to in the explanandum. In other words, by offering the explanation in (1) I allude, via my moral standard, to the contention that the revolution of 1917 was caused by an

---

<sup>18</sup> See Blackburn, *Spreading the Word*, pp. 182-84. This is a non-metaphysical version of the supervenience claim: see Darwall, Gibbard and Railton, ‘Toward Fin de siècle Ethics...’, fn. 125.

unequal distribution of resources in the immediately preceding times. Since this explanation is perfectly understandable, so is the moral explanation that alludes to it.

Of course, the relation of ‘picking out’ is not the same as directly referring. Predicates refer to properties when it is part of their meaning that they signify that property. Moral predicates function to ‘pick out’ properties because their competent usage requires their use to be in response to *some* feature of the situation judged, though their meaning doesn’t determine which. (Indeed, this non-moral property need not even be ‘sparse’; it may, for example, be disjunctive – the only thing bringing the various disjuncts together being that they all generate the attitude expressed by use of the predicate.) This entails that which property is ‘picked out’ by the particular use of a moral predicate can vary whilst that predicate preserves its meaning.<sup>19</sup> More particularly, the property ‘picked out’ will depend on one’s own idiosyncratic moral standard. For example, you may disagree with my standard of injustice. You may hold instead that situations are unjust in virtue of an unequal distribution of *opportunities*. In that case, the predicate ‘is just’ in your mouth would ‘pick out’ a different property than the same predicate in my mouth. But nevertheless it would still be the case that, in all cases, the use of moral predication ‘picks out’ some property, in virtue of which the agent takes that predicate to apply.

It follows that when assessing moral explanations we must distinguish our assessment of the non-moral explanation alluded to from the assessment of the moral standard via which that allusion takes place. In the simplest case, agents may cite the same moral explanans, employing the same moral standard, thereby alluding to the

---

<sup>19</sup> This is the residue of Moore’s Open Question argument. It explains why this account falls short of a naturalistic reduction of moral predicates, for the meaning of moral predicates is preserved across idiosyncratic moral standards. According to expressivism, the common meaning is given by the predicate’s role in the persuasive expressions of attitude.

same non-moral explanation. In such cases there is agreement in moral standard and in the non-moral explanation thereby alluded to.

In a more complicated case, agents may cite the same moral explanans, employing a different moral standard, thereby alluding to different non-moral explanations. For example, I claim that the revolution was caused by injustice, and take injustice to consist in unequal distribution of resources. So I am committed to the view that unequal distribution of resources caused the revolution. You claim that the revolution was caused by injustice, but take injustice to consist in unequal distribution of opportunities, so are committed to taking unequal distribution of opportunities to be causally productive in bringing about the revolution. We both accept the claim “The revolution was the result of injustice” yet in our different mouths it alludes to different non-moral explanations. We therefore exhibit *superficial* agreement – both accepting the moral explanation – but a deeper disagreement, since we accept different non-moral explanations. (Depending our knowledge of the other’s moral standard, this deeper disagreement may or may not be recognised.)

In another complex case, agents may cite incompatible moral explanans, but because of different moral standards, allude to the same non-moral explanation. Cain, for example, takes the global financial meltdown of 2008 to be caused by the immorality of Wall Street bankers, and takes that immorality to consist in greed. So Cain holds that greed caused the meltdown. Ayn also holds that the greed of the Wall Street bankers caused the meltdown, but considers greed a virtue, so claims that the meltdown was caused by the *virtue* of the Wall Street bankers.<sup>20</sup> In this case, Cain and Ayn agree on the underlying non-moral explanation at issue – it was greed that caused the meltdown – but they disagree over whether that cause was immoral or virtuous. In

---

<sup>20</sup> My thanks to an anonymous referee for this example.

a reversal of the previous case, here superficial *disagreement* concerning the moral explanation masks deeper *agreement* concerning the underlying non-moral explanation. (Again, depending on Cain and Ayn's knowledge of each other's moral standards, this deeper agreement may or may not be recognised).

Note that in such cases agents are not obliged to agree with their opponents' *moral* explanation just because they agree with the non-moral explanation that that explanation (in the other's mouth) alludes to. This is because explanation is a factive context, in the sense that to accept an explanation of the form '*e* because *c*' is also to accept '*e*' and '*c*'. For example, to agree with Ayn's claim that "The meltdown occurred because of the virtue of the bankers" Cain must agree that the bankers were virtuous. Since Cain does not agree this, he will not agree with Ayn's moral explanation. Generalising, to agree with a moral explanation involves accepting both the moral verdict given in the explanans and the non-moral explanation one's use of that explanans alludes to. Another's moral explanations will only be agreed to when both conditions are met. (Note that, even then, the agreement may still be superficial, as in the case of Greg and Chris.)

In a final case, agents may cite different moral explanans, employing different moral standards, thereby alluding to different non-moral explanations. In this case there is no agreement in moral explanations, superficial or otherwise.

The above expressivist account of moral explanations also provides an interesting window onto explanations such as (5) and (6) where the explanans is the very fact that is judged to obtain in the explanandum. Consider:

- (5) I hold that testing cosmetics on animals is wrong because testing cosmetics on animals *is* wrong.

Understood on the above model, the explanans ‘picks out’ the non-moral property or properties on which I take my moral verdict to depend. Supposing I am a competent moral reasoner, I will have a moral standard of wrongness, that is, I will take things to be wrong in virtue of some (perhaps complex disjunctive) non-moral property they are taken to possess. This is the property ‘picked out’ by the explanans. The non-moral explanation I thus allude to says that my commitment that testing cosmetics on animals is wrong is caused by that very thing that I take things to be wrong in virtue of. By assenting to such an explanation, therefore, I am claiming that, by my own lights, my commitment that testing cosmetics on animals is wrong is ‘true-to-form’, that is, the result of the very things that I think should prompt such commitments.<sup>21</sup> This is why such an explanation can confer epistemic credibility on my commitment: in adopting it I recognise that the things which (according to me) should give rise to that commitment are in fact doing so in this case. If what actually gave rise to this commitment doesn’t concur with my endorsements of moral standards, then both the explanation and the commitment are undermined.

Of course, there is a limit to the amount of epistemic credibility that such explanations can confer on one’s moral commitments. The import of such an explanation is simply that, at the very least, my moral judgement is not contrary to my own endorsements of standards of moral evaluation. Of course, however, those standards may themselves be questionable. This is why explanations of this form sometimes seem like mere table-thumping (one might imagine Hitler declaring ‘I hold that the Final Solution is our moral duty because it *is* our moral duty!’). If, those standards are agreed, then the effect of explanations like (5) will be to claim that one’s moral judgement of say, wrongness, is indeed the result of those very things in virtue

---

<sup>21</sup> Blackburn, ‘Supervenience Revisited’, p. 162. Note that if a third-party doesn’t know the agent’s moral standards, such an explanation will be completely vacuous since there will be no way of knowing whether the actual belief diverges from the endorsed standard.

of which (it is agreed) an act *is* wrong. The epistemic advantage of such an explanation is clear.<sup>22</sup>

Finally, the above account preserves the counterfactual dependence present in good moral explanations. For an explanandum will be counterfactually dependent on a moral explanans just in case the explanandum is counterfactually dependent on the non-moral property that the agents' use of the moral predicate 'picks out'. Whether this is the case will be an empirical matter, so the expressivist is not committed to any *a priori* rulings on the acceptability of moral explanations.<sup>23</sup>

## **6. Objections to the expressivist account**

I shall discuss three objections to the above expressivist account of moral explanations.

### **6.1. First objection**

There seem to be some situations in which an agent is more sure of a particular moral explanation than she is of a moral standard that governs the use of the moral predicate used in that explanation.<sup>24</sup> So, for example, someone might accept that injustice tends to destabilize societies, yet have no particular idea about what destabilizes societies,

---

<sup>22</sup> Such explanations can be generalised to explain why agents converge in their moral opinions. If Brian holds that vivisection is wrong because it *is* wrong and Ben does likewise, the reason they converge in their moral opinions is because vivisection is wrong. Thus explanations of convergence in moral opinion in terms of the feature judged to obtain can also be accommodated. Wiggins (in 'Moral Cognitivism, Moral Relativism and Motivating Beliefs') takes such explanations to be a 'mark of truth' and hence a test for descriptivism in some area, but this point undermines such a test.

<sup>23</sup> Pace Sturgeon, 'What Difference Does it Make...', p. 122.

<sup>24</sup> N. Sturgeon, 'Thomson against Moral Explanations', *Philosophy and Phenomenological Research* 58 (1998). pp. 199-206, p. 201.

and hence no particular idea about what injustice consists in. This case is problematic because it seems that we can justifiably offer moral explanations without having any particular moral standards for the use of the relevant term, and hence, on the present account, without alluding to any unproblematic non-moral explanation.

But in fact, this case is benign. There are at least two possible understandings of it.

In the first case, an agent may have reason (perhaps from authority) to believe that some property which justifies her verdict of injustice is present without having reason to believe that anything about the nature of that property. In that case, the explanation alluded to in the example above would be of the following form: that there is some natural property of societies (though I do not know precisely which) that both justifies the moral verdict of injustice and tends to destabilize them.

In the second case, an agent can be understood as holding a *causal* moral standard. That is, she can be understood as having a moral standard according to which injustice consists in *whatever it is that destabilizes societies*. Thus, in her mouth, the predicate ‘is unjust’ programmes for whatever property it is that destabilizes societies. Thus, said by such an agent, the moral explanation ‘injustice tends to destabilize societies’ is uninformative, for it says no more than whatever it is that tends to destabilize societies tends to destabilize societies. But a causal moral standard is still a moral standard, and though the non-moral explanation alluded to in this case is quite uninformative, the allusion still takes place.

## **6.2. Second objection**

A second problem-case concerns moral explanations that appear to have a level of generality, and thus explanatory import, beyond any non-moral explanation that they might possibly allude to.<sup>25</sup> Suppose that injustice can be variably realised, and that it can be cited to explain the occurrence of a political protest. Suppose that, for the purposes of the explanation, it is not relevant how the injustice is realised – the protest would have occurred no matter how the injustice was realised – but that the effect would not have occurred if none of the various ways in which injustice can be realised had come about. In that case, it appears that only an explanation in terms of injustice is at a sufficiently general level to explain the occurrence of the protest. It therefore seems that such a moral explanation can be more explanatory than any single non-moral explanation that might be alluded to.

Such cases are also unproblematic, for the generality on display here is just what one would expect from what might be called a *pluralistic* moral standard. When one has a pluralistic moral standard, a moral term is taken to apply in virtue of any one of a number of possible non-moral features of a situation. An explanation in terms of a moral term governed by such a standard can therefore be understood as saying that there are a number of possible non-moral features, each of which, by itself, would have sufficed for the effect, and one of which actually did (and in addition, all these features, according to the speaker, suffice for the moral verdict). Accordingly, moral explanations offered when the moral verdict reflects a pluralistic standard will be potentially more informative than the non-moral explanation alluded to. Yet this does not show that such moral explanations are not explicable within an expressivist framework. They merely involve a pluralistic moral standard.

---

<sup>25</sup> Brink, *Moral Realism*, pp. 193-94.

### 6.3. Third objection

Sturgeon has objected that treating moral explanations in the suggested way necessitates treating them as somehow second-rate, deriving their importance only from the non-moral explanations they allude to. This in turn, it is claimed, makes the account revisionary, since ordinary engagement with moral practice does not see such explanations as in any way second-rate. The objection is that expressivism hasn't accommodated moral explanations at all, merely a cheap simulacrum of them.<sup>26</sup>

There are two possible replies. In the first case the expressivist can deny that the assumption of moral practice noted here needs to be accommodated. More precisely, he can argue as follows. Either the idea that moral explanations are first-rate explanations is an assumption of moral practice or it is not. If it is not, it is no objection to the expressivist that he cannot accommodate it. If it is, it can be dismissed as an errant piece of meta-theory whose revision is defensible on the grounds that there is no significant function of moral practice that is lost by abandoning it. After all, what does it mean to say that moral explanations are first rate if not they are to be understood on the ontic model? But, once again, it can be no objection to a meta-theory of a practice that it fails to justify false meta-theoretical claims embedded in that practice.

This reply can seem unconvincing. In answer to the rhetorical question the opponent of expressivism might reply: moral explanations are first-rate in the sense that ordinary moral practice perceives no distinction between them and non-moral explanations. The expressivist hasn't accommodated this assumption. This leads to the second reply. Assume for the moment that moral practice sees no distinction

---

<sup>26</sup> N. Sturgeon, 'Nonmoral Explanations', *Philosophical Perspectives* 6, *Ethics* (1992), pp. 97-117, p. 103.

between moral and non-moral explanations. Does the above account provide one? The simple answer is ‘no’: for the above account of moral explanations is structurally identical with a popular and plausible account of non-moral explanations. This is the theory of programme explanations, which I will briefly summarise.<sup>27</sup>

Some properties are directly causally relevant to the production of an effect. This occurs when they are properties in virtue of whose instantiation (at least in part) the effect occurs, that is, the instance of the property helps to produce the effect and does so because it is an instance of that property. Other properties may be indirectly causally relevant to the production of an effect. One way of being so is by being part of a chain of directly causally relevant properties. But a more interesting case is where a property is causally relevant to the production of an effect only in so far as the instantiation of that property ‘ensures’, ‘guarantees’ or ‘programmes for’ the instantiation of another property which is directly causally relevant to the production of the effect. One property will programme for another property when the former (‘higher-level’ property) holds *in virtue of* the latter (‘lower-level’ property). For example, a glass is fragile *in virtue of* having some microphysical structure (plus the physical laws). This means that being fragile will *programme* for such a structure.

Given this notion of programming, we can say that a property is indirectly<sub>p</sub> causally relevant to the production of an effect just in case it is a property whose instantiation programmes for the presence of a directly causally relevant property. To exclude this type of case from the first, ‘programming’ is taken to exclude directly causing (thus the ‘p’ subscript is intended to distinguish this type of indirect causation from that involved in chains of causes and effects). A programme explanation is one whose explanans refers to properties that are indirectly<sub>p</sub> causally relevant to the

---

<sup>27</sup> See F. Jackson and P. Pettit, ‘Functionalism and Broad Content’, *Mind* 97 (1988), pp. 381-400 and ‘Programme Explanation: A General Perspective’, *Analysis* 50 (1990), pp. 107-17.

production of an effect. Conversely, a process explanation is one whose explanans refers to properties that are directly causally relevant to the production of an effect.

Many (if not all) of our everyday explanations are programme explanations. For example consider: ‘The conductor was annoyed because someone coughed’.<sup>28</sup> In this case, it is plausible to say that the directly causally relevant property is not that ‘someone coughed’ but that some particular person (e.g. Brian) coughed. That someone coughed programmes for or ensures that some particular person (e.g. Brian) coughed. And it was this particular person’s coughing that was directly causally relevant in producing the effect. Furthermore, there is good reason, in certain contexts, to prefer the programme explanation to the corresponding process explanation. For the programme explanation can give us more information than the corresponding process explanation. Programme explanations can tell us, not merely what the causal history of the effect was, but what it *could have been*. In this case, if the programme explanation is correct, it would have been the case that the conductor was annoyed if Ben coughed, or Gena coughed (even though he was actually annoyed because Brian coughed). Thus the programme explanation is more informative than the corresponding process explanation.<sup>29</sup>

Similar accounts can be given for explanations in terms of relational properties, properties that gloss statistical data, disjunctive properties and dispositional properties. In general, a good proportion of our everyday explanations can be understood as programme explanations.<sup>30</sup>

If many of our actual explanations are programme explanations, and if the expressivist account of moral explanations allows us to see these as programme

---

<sup>28</sup> From Blackburn, ‘Just Causes’.

<sup>29</sup> This is a generalised version of the point in §6.2 that explanations involving predicates with plural standards can be more general than the explanation alluded to.

<sup>30</sup> See Jackson and Pettit, ‘Programme Explanation’, §III.

explanations, then moral explanations are not second-rate and the expressivist will have domesticated the notion of *explanation*. So can moral explanations be understood as programme explanations? At first blush, it may appear not. For programme explanations involve explanans understood as referring to genuinely worldly properties – such as temperature – whereas according to expressivism moral explanans cannot be understood this way.<sup>31</sup> But this difficulty is easily overcome. The important feature of programme explanations is that the use of the explanans phrase programmes for a property that is taken to be directly causally relevant in producing the effect. To have the ability to programme for a directly causally relevant property it is not necessary that the explanans phrase and its predicate be understood referentially. All that is necessary for programming is that the use of the explanans phrase ensures the presence of some causally relevant property. There is no constraint on how that phrase is understood. In short *it is the standard-governed use of predicates, not properties, that programme*. Furthermore, we have seen how the use of moral predicates can, via idiosyncratic moral standards, ‘pick out’ putative directly causally relevant properties. Thus is just to say that our use of moral predicates programmes, via those same standards, for the presence of genuine non-moral properties. Thus expressivist moral explanations are programme explanations.

Note that how exactly a predicate programmes for a property (and hence what the relation ‘in virtue of’ amounts to) will depend on the standards governing the application of the programming predicate. For example, the standard governing the application of ‘someone coughed is something like: apply ‘someone coughed’ only if some particular person coughed. In some cases, such as this one, the standard governing the application of the programming predicate (that is, the function from

---

<sup>31</sup> At least, the expressivist cannot understand moral explanans as referring to the sort of moral properties that moral realists believe in.

perceived features of a situation to application of that predicate) is relatively uncontroversial and endorsed by any competent user of the predicate. As we have seen, however, competent users of moral predicates can employ different standards for their application (with supervenience being a minimal constraint on such standards). This means that when the programming predicate is moral, the property (or properties) programmed for will depend on the speakers own moral standards, and these can vary from person to person without confusion. Thus in the moral case, the property programmed for will vary from speaker to speaker in more diverse ways than in other cases. But this is no reason to think moral programme explanations second-rate. Programming takes place via standards of application and some of these allow more slack than others.<sup>32</sup> But the notion of programming is generic to all cases.

## **7. Alternative accounts of moral explanations**

By showing that moral explanations are a species of programme explanations, and that a good proportion of our everyday explanations are programme explanations, the expressivist has succeeded both in domesticating the notion of explanation and showing how it can come in an expressivist flavour. It is informative to compare this theory of moral explanations with two further expressivist alternatives. First, a ‘traditional’ account based on the notion of secondary descriptive meaning and second, a recent account of ‘practical explanations’ given by Gibbard.

### **7.1. Secondary descriptive meaning**

---

<sup>32</sup> Sturgeon (‘What Difference Does it Make...’ p. 125) disagrees, suggesting that standards governing the predicate ‘is just’ should admit of only as much slack as the standards governing ‘is in poverty’. But no argument is offered for this claim.

According to a traditional view, moral terms can acquire secondary descriptive meaning and it is this that has a role to play in moral explanations.<sup>33</sup> On this view, moral predicates such as ‘is unjust’ can, in virtue of a constant association with non-moral properties of situations, begin functioning to describe. For example, suppose that it is commonly agreed that situations are unjust in virtue of an unequal distribution of resources. Then an agent’s utterance of ‘Theban society is unjust’ can be understood both as offering a description of Theban society as one where there is an unequal distribution of resources and expressing an attitude of moral disapproval towards Theban society (an attitude that is a response to the description offered). The descriptive content is then given a decisive role in moral explanations: to say that the revolution occurred because of injustices is to cite the prior situation of an unequal distribution of resources as a directly causally relevant factor in the revolution. Thus moral explanations may latch onto important truths, albeit ones better expressed in non-moral language.

This account is inferior to the account that treats moral explanations as programme explanations. The important issue is what determines the ‘descriptive meaning’ of a moral predicate. If it is the (possibly idiosyncratic) moral standards of the user of that predicate (if, that is, descriptive meaning is ‘in the head’) then this account amounts to the same thing as the programme explanations account, only expressed in terms of meaning. Yet this terminology is unfortunate, since there are significant problems with such individualistic theories of meaning.<sup>34</sup> Alternatively if (as was supposed above) descriptive meaning is determined by commonly agreed conventions governing the use of the term, then moral explanations will only be understandable when there are such conventions, which is only a partial subset of all

---

<sup>33</sup> Sturgeon ‘What Difference Does it Make...’, pp. 122-25.

<sup>34</sup> Sturgeon, ‘What Difference Does it Make...’, p. 124.

actual cases. For example, moral explanations can be offered using highly abstract moral predicates such as ‘is wrong’ which cannot plausibly be considered to have any conventionally determined descriptive meaning. The programme explanations account offers advantages over both theories since it denies a necessary role for conventionally determined descriptive meaning and rejects the implausible equation of an individual moral standards with descriptive meaning. It is thus an improvement on the traditional account.

## 7.2. Gibbard on practical explanations

In his 2003 book, *Thinking How to Live*, Gibbard offers an expressivist account of practical judgements such as “In the face of overwhelming enemy numbers, retreat is *the thing to do*”. According to Gibbard such judgements express states of contingency planning (roughly, decisions concerning what to do should a particular contingency arise).<sup>35</sup> Such states are not understood as tracking or attempting to track worldly properties, making Gibbard an expressivist about practical judgements.

Gibbard considers whether predicates such as ‘is the thing to do’ can play explanatory roles. In one of his examples, a reader of *War and Peace* accepts that:

(C) Retreat’s being the thing to do caused Kutuzov to retreat.<sup>36</sup>

This is a practical explanation. Gibbard’s treatment of it shares at least three features with the foregoing account of moral explanations.

---

<sup>35</sup> A. Gibbard, *Thinking How to Live* (Cambridge, Ma.: Harvard University Press, 2003), pp. 48-53.

<sup>36</sup> Gibbard, *Thinking How to Live*, p. 202.

First, in a paradigm case, accepting the explanation involves accepting a standard governing the explanans as well as accepting the underlying explanation that one alludes to via this standard.<sup>37</sup> In Gibbard's example, Hera holds that the thing to do is that which maximises one's own pleasure. Thus explanation (C) in Hera's mouth makes the claim that Kutuzov retreating would bring him most pleasure and it was this fact that caused him to retreat.

Second, agents can agree with each others' explanations without agreeing either on the standard governing the explanans or on the underlying explanation alluded to. In Gibbard's example, Diana has a different standard to Hera: she holds that the thing to do is to do one's duty and that a general's duty is to maximise chances of victory.<sup>38</sup> Diana also holds that Kutuzov's retreating maximised his chances of victory, and that it was this that caused him to retreat. Thus Diana accepts (C), although in her mouth it alludes to a different underlying explanation than it does in Hera's. (Hera and Diana manifest superficial agreement in explanation, as in the example of you and me in §5).

Third and finally, in some cases, agents can accept an explanation of the relevant type without a clear view on the correct explanans-governing standard, therefore without a clear idea of the underlying explanation alluded to. In Gibbard's example, Joe accepts (C), although he is unclear precisely what it is that constitutes being the thing to do in this case.<sup>39</sup> Joe may accept the explanation on authority or because he thinks that that Kutuzov reliably tends to do the thing to do.<sup>40</sup>

---

<sup>37</sup> Gibbard, *Thinking How to Live*, pp. 202-5.

<sup>38</sup> Gibbard, *Thinking How to Live*, p. 204.

<sup>39</sup> Gibbard, *Thinking How to Live*, p. 201.

<sup>40</sup> Gibbard, *Thinking How to Live*, pp. 206-7.

These points of agreement aside, however, there are several ways in which Gibbard's account limits both the scope and usefulness of *practical* explanations, in ways in which the foregoing account of *moral* explanations does not.

First, Gibbard's practical explanations always proceed via a practical judgement on behalf of an involved agent. In the case of (C), Kutzov retreated because he judged that retreat was the thing to do.<sup>41</sup> More generally, on Gibbard's view, deciding to  $\phi$  just is to judge that  $\phi$  is the thing to do.<sup>42</sup> And since practical explanations will always involve an agent coming to a decision to  $\phi$ , such explanations will always proceed by first explaining practical judgements. The foregoing account of moral explanations is more general than this. It can allow for cases where the explanation does not proceed via a judgement concerning the explanans. For example, explanation (1) can be a good one even when the working-class revolutionaries do not judge themselves to be victims of injustice (perhaps because they think, falsely, that 'justice' is a bourgeois construct).

The second difference concerns the usefulness of the explanations. Gibbard holds that in the case of practical explanations, the 'real' explanatory work is done entirely by the non-practical explanation one alludes to.<sup>43</sup> For an agent who offers a practical explanation, "Being the thing to do, on her view, explains nothing beyond what is explained by the [underlying] facts alone".<sup>44</sup> But on the account of moral explanations just given, the analogous claim is false. As we saw in §6.2 (and again in the general context of programme explanations in §6.3) when an agent has a pluralistic moral standard, a moral explanation can explain more than any particular underlying non-moral explanation. For example, where a particular injustice caused

---

<sup>41</sup> Gibbard, *Thinking How to Live*, p. 215.

<sup>42</sup> Gibbard, *Thinking How to Live*, p. 41.

<sup>43</sup> Gibbard, *Thinking How to Live*, pp. 210-14.

<sup>44</sup> Gibbard, *Thinking How to Live*, p. 211.

the revolution but the revolution would also have occurred if injustice was realised in another way, the moral explanation provides information concerning not just what the causal history of the revolution was, but what it could have been. An explanation purely in terms the actual realiser of injustice doesn't convey this information and is thus less explanatory.

A final difference concerns generality. Gibbard considers whether his account of practical explanations merely assimilates them to the sort of higher-level explanations familiar to philosophers of science.<sup>45</sup> For example, structural damage can be explained in terms of the occurrence of a storm, or, at a lower-level, in terms of the motion of the storm's constituent molecules. Gibbard ultimately rejects the proposed assimilation on the basis of the following disanalogy. In the case of higher- and lower-level explanations, where both explanations are in play an agent who understands the explanations at both levels must see that if the lower-level explanation is correct, then so is the higher-level explanation. In the case of practical explanations, however, an agent can accept the underlying explanation that their opponent's practical explanation alludes to, but still disagree with the practical explanation. This will happen where the agents' disagree over the standard governing the application of the practical explanans (as in Gibbard's case of Hera and Poseidon, which has the same structure as the case of Cain and Ayn of §5). Gibbard expresses the point in terms of conceptual bridge principles: those linking the practical to the non-practical allow more conceptual slack (i.e. disagreement without conceptual confusion) than those linking the molecular and the meteorological. However, while Gibbard is correct to point out this disanalogy, it does not follow that there is no general account of explanation available that can accommodate both the slack and

---

<sup>45</sup> Gibbard, *Thinking How to Live*, pp. 212-13.

non-slack cases. As I argued in §6.3, the theory of programme explanations is one possible account. On this view, the generic notion is one of predicates that programme, and the difference between the cases lies in the more or less restricted range of standards via which that programming can take place. Furthermore, the programming account helps explain, in a way that Gibbard's account does not, both how moral explanations needn't always work through explanations of moral judgements and how they can be more informative than any underlying non-moral explanation. For these reasons, therefore, the programming account given can be seen as in part an extension of, and in part an improvement upon, Gibbard's 2003 account.

## **8. Programme explanations and moral realism**

I have been arguing that given an adaption of the theory of programme explanations (that speaks of predicates rather than genuinely realist properties) moral expressivists can rebut a possible objection to their position. It is worthwhile relating this discussion to a recent debate between Miller, Nelson and Bloomfield, concerning the connections between programme explanations and moral *realism*.<sup>46</sup>

The Miller/Nelson/Bloomfield debate has two distinct branches. The first concerns two related issues:

---

<sup>46</sup> See A. Miller, *An Introduction to Contemporary Metaethics* (Cambridge: Polity Press. 2003), pp. 150-55 & 168-74; M. Nelson, 'Moral Realism and Programme Explanation', *Australasian Journal of Philosophy* 84 (2006), pp. 417-428; A. Miller, 'Moral Realism and Program Explanation: A Very Short Symposium 1: Reply to Nelson', *Australasian Journal of Philosophy* 87 (2009), pp. 337-41; P. Bloomfield, 'Moral Realism and Program Explanation: A Very Short Symposium 1: Reply to Miller', *Australasian Journal of Philosophy* 87 (2009), pp. 343-44.

(a) What are the *necessary* conditions for a set of properties to be ontologically respectable (or ‘real’)?

(b) Supposing moral properties exist, would they meet these conditions?

Regarding (a), Miller and Nelson agree that a necessary condition for ontological respectability is featuring in explanations that are best from an epistemically unlimited perspective or ‘God’s-eye view’, that is, a perspective with acquaintance with all the relevant facts.<sup>47</sup> Regarding (b) Miller and Nelson disagree over whether moral explanations (construed as property-involving programme explanations) meet this test. Miller argues that they do not; Nelson demurs.<sup>48</sup>

The expressivist position elucidated above can agree with the suggested necessary condition, but remain agnostic about (b). This is because in the present context, the expressivist is concerned not with whether moral explanations are best from a God’s-eye view, but only with whether they can be best for ordinary moralisers. Recall that the current defence of expressivist explanations arose from a concern to rebut an objection which holds that ordinary participants in moral practice – epistemically limited beings like you and me – find moral explanations informative and that expressivism cannot accommodate this. To rebut this objection the expressivist need only show that moral explanations can be informative for ordinary moralisers and can thus remain neutral on the question of whether they would be informative for an epistemically unlimited being.

---

<sup>47</sup> Miller, *An Introduction to Contemporary Metaethics*, p. 173, Nelson, ‘Moral Realism and Programme Explanation’, pp. 423-24. Miller and Nelson also agree that “an explanation is ‘best’ if its unavailability would lead to genuine explanatory loss”, see Miller, *An Introduction to Contemporary Metaethics*, p. 296.

<sup>48</sup> Miller, *An Introduction to Contemporary Metaethics*, pp. 173-74; Nelson, ‘Moral Realism and Programme Explanation’, pp. 423-28.

The second branch of the Miller/Nelson/Bloomfield debate also concerns two related issues:

(c) What are the *sufficient* conditions for a set of properties to be ontologically respectable (or ‘real’)?

(d) Do moral properties meet these conditions?

Here the expressivist cannot afford to remain neutral. The worry is that any expressivist defence of moral explanations threatens to undermine expressivism by supporting a premise in an argument for moral realism. That argument proceeds as follows: A sufficient condition for the ontological respectability of a set of properties is that they feature in best explanations; Some moral explanations (construed as programme explanations) are best explanations; Hence moral properties (or at least those of them that feature in best explanations) are ontologically respectable. The expressivist could resist this argument by agreeing with Miller’s answer to question (b), that is, by denying that moral explanations are ever best. He would then be required to explain where the counter-argument of Nelson goes wrong.<sup>49</sup> There is, however, an alternative way to resist the realist argument, that preserves neutrality on question (b). This is to question the sufficient condition given in the first premise.

To explain. If the ‘explanationist’ argument for realism is to be persuasive, it must start from premises that are acceptable to all sides. The expressivist position I am defending will not accept that moral *properties* (understood in the realist way) feature in best explanations, though it will accept that moral *predicates* do. Therefore, the only non-question-begging formulation of the first premise is as follows: A

---

<sup>49</sup> See Miller, ‘Reply to Nelson’, for one such attempt and Bloomfield, ‘Reply to Miller’, for criticism.

sufficient condition for the existence of a set of properties, referred to by a set of predicates, is that those predicates feature in best explanations. However, while no longer questioning-begging, this premise is now false. For the preceding defence of moral explanations demonstrates how a set of predicates can feature in best explanations without referring to a distinct realm of properties. Thus, far from undermining the expressivist position, the defence of moral explanations offered here has a further consequence of highlighting a new way to resist the explanationist argument for moral realism.

In sum, the theory of explanations outlined here is concerned primarily to rebut an objection to expressivism by showing how moral explanations, understood expressively, can be informative to ordinary, epistemically limited, moralisers. Further, that account demonstrates that even *if* moral explanations were informative from an epistemically unlimited perspective, it would not follow that explanatory moral predicates refer to moral properties. The expressivist can therefore remain neutral on the question of whether whether moral explanations are informative from a God's-eye view.

## **9. Moral explanations and explanatory contexts**

There is of course one context where the expressivist will deny that moral features play any explanatory role. That is the context of explaining why we moralise at all. As Blackburn notes:

...when our project is to place ethical...discourse, that is, understand why we go in for it and its role in our activities and lives, a simple explanation of our judgements as responsive to the facts is markedly inferior to the longer story explaining their functional role, the reason we have them and give them the importance we do.<sup>50</sup>

In §3 I distinguished two senses of the claim that ‘there are moral features’. In the first sense to say there are moral features is just to say that there are warranted moral predications. In this sense, as we have seen, the expressivist can accept both that there are moral features and that they can feature in good explanations. In the second sense – the sense answering to the explanatory context mentioned in the quote above – to say that there are moral features is to say that there is a realm of moral properties to which our moral judgements are (in the right circumstances) responsive. In this sense, the expressivist denies that there are moral features and therefore also denies that they feature in good explanations.

This dual perspective creates a problem. As Sturgeon notes, one of the common assumptions of moral practice is ‘...that we take [moral judgements] to be undermined, exactly as scientific judgements are, by explanations showing them to be unresponsive to the corresponding facts’.<sup>51</sup> Recall Gibbard’s example of our belief in evolution: if the idea was shown to be invented by scientists in a ploy to attain a higher social position, then our commitment to it would be undermined. The assumption is that moral judgements are vulnerable to similar debunking explanations. The problem for expressivism is that by denying any explanation of

---

<sup>50</sup> Blackburn, *Essays in Quasi-Realism*, p. 8. See also Gibbard, *Wise Choices, Apt Feelings*, pp. 107-8 and *Thinking How to Live*, p. 181.

<sup>51</sup> N. Sturgeon, ‘Critical Study of Allan Gibbard, *Wise Choices, Apt Feelings*’, *Noûs* 29 (1995), pp. 402-24, p.411.

moral practice in terms of responsiveness to moral facts he is committed to a debunking explanation.

The problem is illusory, and can be dispelled by distinguishing explanatory contexts.<sup>52</sup> The sort of explanations by which particular moral judgements are undermined are not the same sort of explanations which the expressivist meta-ethic denies. When we are concerned with whether or not a particular moral judgement is undermined, we are concerned with its *justification*. As we have seen, one way in which the justification for a moral judgement (and hence quite likely the judgement itself) will easily be undermined is if we come to believe that it is formed via a mechanism which we ourselves do not endorse (prejudice, for example). In such cases we can reject explanations of the form of (5) and, in precisely this sense, we cannot view that particular moral judgement as responsive to the moral facts. But expressivism isn't committed to the view that *all* explanations of the form of (5) are to be rejected. For such explanations simply involve moral predication in the explanans, and to repeat, expressivism is not in the business of denying that such a thing can be understood (and acceptable). What the expressivist is committed to denying is that reference to worldly moral *properties* can occur in the explanans of good explanations, but this is not the same as moral predication (at least, not on the expressivist account). When it is reference to moral properties that occurs in explanans, the explanatory context shifts from one of *justification* of a moral attitude to one of *explaining* why that attitude is present, and this is to shift from a discussion *within* ethics to a discussion *about* it (that is, to a meta-theoretical discussion). Of course, a meta-theoretical explanation of a particular moral judgement *can* be undermining, since it can show that judgement to be the product of a mechanism that

---

<sup>52</sup> Blackburn, 'Supervenience Revisited', pp. 153-58.

one doesn't endorse. But it can equally show that judgement to be the product of a mechanism that one does endorse, and in that case there will be no undermining.

In sum, the expressivist can accept that moral judgements are unjustified if an explanation of the form of (5) is not available. This is the acceptable content of the common assumption given by the quote from Sturgeon. What the expressivist cannot accept is that moral judgements are unjustified if an explanation of the realist sort (in terms of sensitivity to worldly properties) is not available. But neither, he will insist, is ordinary moral discourse committed to this. Or, if it is, it is another piece of errant meta-theory that can be expunged without any loss in the functioning of the practice.

## 10. Conclusion

If the above arguments are correct, our everyday practice of offering explanations with moral explanans is no threat to an expressivist account of moral practice. Not only can expressivists provide an account of what we are doing when we offer moral explanations, but this account dovetails with a plausible *general* account of explanation: the theory of programme explanations. This is not, of course, to say that there are no other parts of our moral practice that cannot be brought within the expressivist fold. To provide an expressivist account of moral explanation is not yet to provide an expressivist account of moral belief, knowledge, logic or truth. But, at least as far as the issue of explanation is concerned, expressivism seems in better condition than is often supposed.<sup>53</sup>

---

<sup>53</sup> My special thanks to Hallvard Lillehammer, three anonymous referees for this journal and to seminar audiences at Cambridge, Stirling and Bristol for helping to develop this paper.