

Moral Attitudes for Non-Cognitivists: Solving the Specification Problem

GUNNAR BJÖRNSSON

TRISTRAM MCPHERSON

1. Introduction

According to *moral non-cognitivism*, moral judgements fundamentally involve desire-like ('non-cognitive') attitudes. Moral non-cognitivism is attractive in part because it promises to solve at least two central metaethical problems. First, judging that one ought morally to perform an action seems intimately related to being motivated to perform it. On the influential 'Humean' account of moral psychology, however, only non-cognitive attitudes (and not beliefs) can be intrinsically motivating. By claiming that moral judgements involve non-cognitive attitudes, the moral non-cognitivist can explain the motivating character of moral judgements while respecting Humean moral psychology.

Second, metaethicists need to explain how people with systematically different moral views can nevertheless be in substantive disagreement, rather than simply talking past each other. Call this the issue of explaining *univocality*. Non-cognitivists have traditionally sought to explain univocality by understanding substantive moral disagreement as disagreement in non-cognitive attitude, thus avoiding the need to characterize a univocal cognitive content.¹

This strategy faces a deep challenge, however. To make it plausible that univocality requires disagreement in attitude, the non-cognitivist must identify a distinctively *moral* non-cognitive attitude. To see why, consider a sports fan booing her favoured team for its poor performance. She is thereby expressing a negative attitude towards the team, but she need not thereby disagree with someone who thinks that the team is morally unimpeachable. Because of this, the non-cognitivist needs to offer a specification of the moral attitude that permits her to distinguish real moral disagreement from cases like this one.

¹ See e.g. Hare 1952 (Ch. 9), Horgan and Timmons 1991, and Tersman 2006 (Ch. 5–6).

Alexander Miller (2003, pp. 43–51) dubs this the ‘moral attitude problem.’ As Nicholas Sturgeon (1986) and Miller have shown, this problem is extremely difficult for the non-cognitivist to solve. Worse, David Merli (2008) has recently argued that this problem undermines the dialectical force of the canonical ‘moral twin earth’ arguments that non-cognitivists often deploy against their rivals. Solving this problem is thus one of the most pressing tasks facing the non-cognitivist program in metaethics.

This paper aims to show how the non-cognitivist can meet these challenges. We begin in section 2 by building on Merli’s work to develop a simple recipe for generating apparent counterexamples to *any* informative specification of the moral attitudes. This may appear to be a lethal objection to non-cognitivism, but we argue that it is not. Rather, Merli’s challenge is an instance of a more general problem that also faces most of the non-cognitivist’s competitors. This *specification problem* is the task of characterizing the conditions which underwrite the contrast between genuine and merely apparent moral disagreement. In section 3, we argue that the generality of the specification problem requires a systematic response, which we offer in sections 4 and 5. In section 4 we speculate that, in paradigmatic cases, wrongness-judgements constitute a certain complex but functionally unified state. Then, in section 5, we explain how the functional story that underpins this proposal explains and defuses the intuitive judgements required by the counterexample-generating recipe introduced in section 2.

We conclude by taking stock. One of the lessons of our discussion is that a central assumption that is traditionally associated with non-cognitivism is untenable: it is not the case that every party to a genuine moral disagreement must share the same type of non-cognitive attitude. We argue that despite abandoning this assumption, our account answers to traditional arguments and motivations for non-cognitivism. We also argue that the sort of solution that we propose to the specification problem is difficult for at least non-relativistic cognitivists to adopt. Surprisingly, then, despite appearing to be an intractable problem for non-cognitivism, the specification problem turns out on examination to support rather than to undermine it.

2. The non-cognitivist's specification problem explained

Non-cognitivism is a thesis about the nature of a certain class of mental states: the moral judgements. To begin, we will offer a more precise characterization of this thesis, as we will understand it:

Non-cognitivism Moral judgements have non-cognitive aspects, and these aspects are fundamental to explaining the nature of moral agreement and disagreement.²

As suggested in the Introduction, the idea of the *non-cognitive* should be understood in broadly Humean, 'direction of fit' terms. Roughly, the functional role of non-cognitive states is to have the world conform to their contents; for cognitive states, it is for their contents to conform to the world. The basic idea can be illustrated by G. E. M. Anscombe's helpful analogy (1957, Sect. 32). A non-cognitive attitude is like a shopper's grocery list: the latter performs its function if it guides the shopper to fill the cart in accordance with the list. By contrast, a belief is like the list kept by a detective tracking the shopper: it performs its function if it accurately represents whatever the shopper has put in the cart.³

Notice that understood in this way, non-cognitivism does not in itself rule out that moral judgements also have cognitive content that might (for example) explain various aspects of moral reasoning. What it rules out is merely that moral agreement and disagreement are well understood as disagreement about

² Non-cognitivists attracted to the quasi-realist program (e.g. Blackburn 1993, Gibbard 2003) want to claim that they 'earn the right' to talk of moral truth and moral properties. However, such properties are supposed to be in some sense explanatorily non-fundamental. In the remainder of this paper, we ignore complications stemming from quasi-realist developments of non-cognitivism or expressivism.

³ See Smith 1994 (Ch.4) for the classic contemporary defence of the Humean theory of motivation. Other helpful discussions of the idea of direction of fit include Humberstone 1992, Velleman 1992, and Ridge 2006a. Sometimes heirs to the non-cognitivist tradition have rejected the non-cognitivist label. They have typically not done so because they think that moral judgments have the direction of fit of beliefs, but because they have preferred a different characterization of the cognitive/non-cognitive distinction (see e.g. Horgan and Timmons 2006).

the truth of such content. It is thus compatible with many forms of what Mark Schroeder (2009) calls ‘hybrid expressivism’.⁴

The thought that an act is right is different from the thought that it is morally good, virtuous, etc. Because of this, the non-cognitivist will ultimately need to specify the relevant attitude for each type of moral judgement. However, both the deep challenge posed by the specification problem and our reply can be helpfully discussed by focusing on an exemplary case. In this paper, we thus focus on offering a non-cognitivist account of what we will call ‘wrongness-judgements’: the sort of judgement characteristically expressed by sincere utterances of the form ‘ Φ ing is morally wrong’. Our domain is thus (part of) morality, narrowly construed. We intend our focus to contrast with accounts of judgements of moral value or virtue, but also with non-cognitivist accounts of aesthetic judgements (for example), and with accounts of the sort of ‘all in’ normative judgements that Allan Gibbard (2003) discusses under the label ‘the thing to do’.

Non-cognitivists are sometimes tempted to respond to the moral attitude problem by saying that what *matters*, fundamentally, is not whether an action is morally wrong, but whether it is the thing to do, in roughly Gibbard’s (2003) sense. (Compare Simon Blackburn’s 1991a, pp. 7–8). However, in so far as the non-cognitivist aims to give an account of narrowly moral thought and talk, this suggestion appears to simply change the subject (Nicholas Sturgeon makes a related point in reply to Blackburn: 1991, p. 35 n. 12). The significance of this change is illustrated by the fact that important naturalistic realist replies to the moral twin earth challenge by David Copp (2000, p. 123) and David Merli (2002, p. 232) each suggest that one way to defend moral realism is by combining it with non-cognitivism about the thing to do. Non-cognitivism about moral judgements thus requires an account of moral attitudes, not merely an account of practical attitudes.

⁴ See e.g. Copp 2000 (pp. 120–4), Ridge 2006b, and Björnsson and Finlay 2010. Non-cognitivism is of course often associated with expressivism, which analyses moral speech-acts in terms of the non-cognitive states of mind that they express. But the specification problem primarily concerns such states of mind, not their linguistic expression, and not all expressivists understand disagreement as to be explained in terms of non-cognitive features (see e.g. Boisvert 2008).

According to the non-cognitivist, then, wrongness-judgements at least centrally involve *some* non-cognitive attitude or other. This idea faces a long history of attempted counterexamples. These include amoralists like Plato's Thrasymachus, as well as listless or disillusioned former moralists. All of these characters are said to make moral judgements that are unaccompanied by the relevant motivation or emotion.⁵ By themselves, however, such cases are dialectically weak. In the extreme cases, it can seem unclear whether the judge is really making a moral judgement. In the clearest cases, the judges are embedded within a community of persons normally moved by such judgements. The non-cognitivist can potentially use this fact to explain away the clear cases in at least two ways.⁶ First, if the judges suffer from cynicism or general motivational disorders, it can be argued that these conditions mask, rather than remove, the relevant motivational or emotional dispositions. Second, such judges might be treated as making moral judgements in virtue of possessing an intention to be concerned with the same issue as normal members of their community.⁷ These explanations suggest hope for the possibility of a non-cognitivist account of the normal cases, and a treatment of deviant cases as parasitic upon these.

This hope can seemingly be bolstered by our intuitive interpretation of whole linguistic communities. Consider a whole community where people apply a certain label to much the same actions that we take to be morally wrong but are and have always been entirely unmoved by these classifications, engaging in them out of mere curiosity. Here it seems much less clear that they think that the actions they label are wrong.⁸

⁵ See, e.g. Stocker 1979, Brink 1989 (Ch. 3), Svavarsdóttir 1999 (pp. 176ff), and Merli 2008 (p. 33).

⁶ For defences of a constitutive connection between moral judgement and motivational states in light of both extreme and less extreme cases, see Bedke 2009, Björnsson 2002, Blackburn 1998 (pp. 59–68), Dancy 1993 (pp. 4–6), Gibbard 1993 (pp. 318–19), and Smith 1994 (pp. 68–71).

⁷ Blackburn briefly suggests that the non-cognitivist needs to 'locate the attitude that gives the words their function publicly (like Fregean senses) leaving subjective differences to one side' (1991b, p. 40, cf. 1991a, pp. 5–7). For criticism, see Sturgeon 1991 (p. 35 n. 14) and Merli 2008 (pp. 45–7).

⁸ See Dreier 1990, Lenman 1999, Bedke 2009, Tresan 2009.

David Merli (2008) has recently shown how to strengthen the challenge facing the non-cognitivist. To begin, notice that in order to address the specification problem, the non-cognitivist's account of what attitude is expressed by wrongness-judgements must manoeuvre between two constraints. On the one hand, the solution must be *sufficiently discriminating*. A solution fails this constraint if it cannot distinguish moral from aesthetic disapproval, for example. On the other, the solution must not *over-discriminate*. Failure here would be evidenced by the inconsistency of the solution with a class of substantive normative beliefs. For example, consequentialists and Kantians are generally taken to disagree about which acts are wrong. It would therefore count against a solution to the specification problem if it entailed that one party to this dispute was simply failing to make wrongness-judgements. Merli argues that in trying to respect these constraints, the non-cognitivist faces a dilemma. On the one hand, the need to distinguish wrongness-judgements from aesthetic judgements, all-in disapproval, etc., creates pressure to characterize the relevant attitude involved in increasingly specific terms. On the other, once characterizations become specific enough, they inevitably seem to over-discriminate.

This dilemma is especially dialectically powerful for two reasons. First, it applies convincingly to assessment of whole linguistic communities, making it immune to the strategies mentioned above for managing apparent intra-community counterexamples. Second, it closely parallels a central non-cognitivist argument from disagreement, introduced by R. M. Hare and developed by Terence Horgan and Mark Timmons. Hare famously asked us to imagine that missionaries come upon a cannibal society that uses the word 'good' as a term of commendation much as we do, but applies that term to acts of violence and cruelty. Hare suggests that we should translate their word 'good' homophonically, and treat the cannibals and missionaries as having a substantive disagreement about what sorts of acts are good in spite of a difference in descriptive content (1952, pp. 148–50).

In a series of papers, Horgan and Timmons have updated Hare's thought experiment, appealing to broadly similar scenarios in which the differences between the acts that the two linguistic communities count as good are much

less radical than in Hare's example.⁹ After Horgan and Timmons, we will call this class of cases *moral twin earth cases*. The strategy of such cases is to keep constant the normal functioning of our moral judgements, in particular their connection to motivation and role in social criticism and coordination, while varying the factors that cognitivists take to determine their truth-conditions. Where cognitivists have offered a reasonably informative and specific account of such factors, these variations can be quite subtle, and people are still willing to treat the varying parties as disagreeing or agreeing about moral questions. Such cases are used to suggest that moral judgements cannot be understood in terms of any particular cognitive content, but should be primarily understood in terms of their practical role.

Merli shows that this strategy of subtle variation can also be deployed against the non-cognitivist. For illustration, consider Allan Gibbard's proposed specification of the attitude that constitutes wrongness-judgement. Gibbard's approach to moral attitudes begins with an account of what it is to think that something is 'rational' (1990) or 'warranted' (2006). Such thoughts are *planning states*, on Gibbard's view: very roughly, states of being decided on what course to take under certain contingencies. Narrowly moral judgements are plans to have certain emotions: to think that John's act was wrong is to plan to resent him for doing it, and to plan, given the contingency of being John, to feel guilty for doing it. More precisely, moral thoughts concern what one plans to feel from *a special standpoint*—in the case of resentment, the standpoint of full but impartial engagement (1990, p. 127).¹⁰

⁹ See Horgan and Timmons 1991, 1992a, 1992b, 1996, 2000, and 2009; see also Holland 2001.

¹⁰ We ignore two important aspects of Gibbard's view: the subtle differences between his 1990 and 2003 accounts, and his distinction between wrongness and blameworthiness:

An act is wrong if and only if it violates standards for ruling out actions, such that if an agent in a normal frame of mind violated those standards because he was not substantially motivated to conform to them, he would be to blame. To say that he would be *to blame* is to say that it would be rational for him to feel guilt and for other to resent him (1990, p. 45).

We take these to be innocent simplifications. Regarding the first: although Gibbard did change his basic terminology from 1990 to 2003, he takes his latter account to be a development of the former (Gibbard 2006, p. 198). Regarding the second, the problems for Gibbard's view discussed here do not rest on cases where wrongness and blameworthiness come apart.

In tying wrongness-judgements to guilt and resentment, Gibbard's analysis achieves considerable discriminatory power, letting us distinguish between questions of what ought not to be done, all told, and questions of what is morally wrong. It promises to make sense of general scepticism about morality (as the rejection of guilt and resentment), and could explain why it seems that someone who wants you not to Φ —yet insists that he wouldn't hold it against you if you did—does not seem to think that Φ ing would be morally wrong. Moreover, among emotional reactions, guilt and resentment seem to have a very close tie to judgements of moral wrongness (indeed, close enough to raise the worry that these emotions themselves involve judgements of moral wrongness).¹¹

However, while the connection to guilt and resentment provides discriminatory power, it results in over-discrimination. This can be seen by considering an 'attitudinal' moral twin earth. This features an alien society that is much like our own. Their word 'wrong' functions much as our word 'wrong' does: as a term of criticism that plays a distinctive role in deliberation and social coordination. The term is used to characterize a very similar range of actions, and there are very similar controversies about its proper extension. The crucial difference is that the twins' use of 'wrong' seems to be robustly linked to emotions that we recognize to be contempt and shame rather than to resentment and guilt.¹² Merli's point is that when we isolate the moral attitude

¹¹ It is doubtful whether the account has *enough* discriminatory power, in virtue of judgements that involve the 'wrong kind' of reasons for resentment or guilt. For example, suppose that one planned to resent someone for a perfectly innocent act, and planned (for the contingency of being him) to feel guilt for that act, *simply* because of the foreseeably good consequences of having those feelings. It does not seem that, in doing so, one would thereby judge that the innocent person has done something wrong. See especially D'Arms and Jacobson 1994, Rabinowicz and Rønnow-Rasmussen 2004, Kauppinen 2010 and Schroeder 2010. Gibbard (1992, pp. 211–13) considers and rejects one sort of worry about judgements based on the wrong kinds of reasons. However, the case considered there—a person who bizarrely takes bean-counting to warrant guilt and resentment—is different from the sort of cases appealed to in recent literature, where the reasons for thinking that guilt and resentment are warranted are quite intelligible. Moreover, Gibbard does not consider whether the person in his story might think that while bean-counting is intrinsically reprehensible, it is not *morally* wrong. It seems that she might.

¹² This characterization of the case is very quick. One might, for example, think that societies that characteristically regulated social behaviour by guilt/resentment attitudes and ones that

as the only salient locus of variation (as we have done in this case), we will be inclined to treat the twin word ‘wrong’ as translatable with our own, and to treat the twins as having substantive (dis-)agreements with us about moral questions (2008, p. 35). This strongly suggests that Gibbard’s account over-discriminates.¹³

This problem might seem to invite a natural amendment to Gibbard’s account: abstract away from details of his account and understand wrongness-judgements as plans to sanction those who perform the associated actions. Such sanctions can then take different forms in different societies. By hypothesis, we have plans for resentment and guilt in our society, and for contempt and shame, on this twin earth. Less emotional and more calculating plans to threaten or punish are also possible.

This amendment is *still* not enough to avoid over-discrimination. Consider a society whose use of the term ‘wrong’ is similar to ours in all respects, except that their judgements lack any essential connection to social sanctions. Instead, such judgements are characteristically accompanied by plans to prevent the acts in question, if feasible. In first-person cases, such plans typically ensue in decisions not to perform the acts. In second and third-person cases, attempts at prevention sometimes do involve threats of social

did so via contempt/shame attitudes would tend to show significant differences in other respects. However, note two things. First, the focal difference concerns only links between judgement and emotion that hold for clear or paradigmatic cases of wrongness judgement. Nothing prevents guilt and resentment from playing *some* role in the moral life on twin earth, just as shame and contempt play some role in our society; it is just that the connection is too weak to be seen as a constitutive link. Second, since (as we take it) this twin earth scenario is not *obviously* impossible, any a priori link between wrongness-judgement and guilt and resentment would be indirect at best. That is enough to undermine Gibbard’s analysis understood as a conceptual or a priori claim.

¹³ Other worries about over-discrimination operate on the individual rather than societal level. Nichols 2004 (Ch. 4) argues that children can make narrowly moral judgements before having concepts of moral emotions (Gibbard 2006 (pp. 198–203) suggests that such children still have ‘near-moral’ concepts). Similarly, some highly sophisticated thinkers of various stripes deny that guilt and resentment are appropriate responses to moral wrongdoing (e.g. Harman 2009, Scanlon 2008 (Ch.4), Glover 1975, and Parfit 1984 (pp. 31–5)). In both cases, a defender of Gibbard’s analysis might argue that their judgements qualify as wrongness-judgements because these thinkers intend to relate to normal cases (which, we are assuming, do involve plans for guilt and resentment).

sanctions, but very often people in this society think that there are better means. For example, making the agent aware of what he is about to do, pointing out the consequences of what he just did, asking him to consider the effects his actions would have on other people's expectations of him, and thus on their willingness to cooperate with him, etc. (Such unwillingness to cooperate with someone does not plausibly constitute a sanction unless it is done to punish or teach a lesson.) Intuitively, it seems perfectly possible that this society's word 'wrong' is translatable with ours and that we could have substantive (dis-)agreement with them about moral wrongness: they just have different ideas about how to cope with wrong-doers.

To avoid such over-discrimination, one might try to abstract away further, or even to abandon the idea that there is a determinate moral attitude. In *Ruling Passions*, Simon Blackburn recommends such a liberal approach:

An ethic may characteristically express itself in disdain of those who do not measure up, rather than anger at them, or in colourless administrative controls on conduct, rather than emotional public demonstrations. But this difficulty of definition arises not because the subject is mysterious, or especially 'sui generis', or resistant to understanding in any terms that enable us to understand the rest of our emotional and motivational natures. It arises because of the polymorphous nature of our emotional and motivational natures themselves. (1998, p. 14, compare also pp. 59–68).

By itself, however, this approach runs into the other horn of Merli's dilemma: it *under-discriminates*. For example, I can surely feel disdain towards your Φ ing or be disposed to prevent it without thereby judging it to be wrong. Because of this, Blackburn must insist that only some cases of disdain (e.g.) count as wrongness-judgements.

Liberality thus requires that we find another way of demarcating the moral attitudes. Blackburn gestures at an etiological criterion: we can count various states of mind as moral judgements because they arise from a history of modification of, or reaction to, a more unified, explanatorily basic attitude (1998, pp. 60–1). The problem with this proposal is that in moral twin earth cases, judgements of univocality do not seem to require similarity in etiology.

Because Blackburn's etiological suggestion is sketchy, this point can be made most clearly by considering the sophisticated etiological characterization of the moral attitude offered by Antti Kauppinen. According to Kauppinen, moral attitudes are attitudes of disapproval of a type that

... characteristically results from a process of simulating the non-moral (quasi-) reactive attitudes like anger, gratitude, and reactive disgust that any unbiased and informed participant whose advice on the topic the subject is disposed to take would have a) in the circumstances of those affected by the action or b) as a spectator or trustee, in case no subject is directly affected (2010, p. 241).¹⁴

Suppose that this account were extensionally adequate for moral judgements made in our society. Now consider a society very much like ours but where the counterparts of our wrongness-judgements are typically formed through a weighing of preferences of all involved through a process of identification (as in Hare 1981) rather than through simulation of emotional states (as Kauppinen's account requires). Normally, their judgements come with an aversion to the act in question and dispositions to anger (or guilt), much as our wrongness-judgements, and they are expressed using homophonic vocabulary in much the same fashion. Contrary to what one would expect on Kauppinen's account, it seems possible for us to agree and disagree with people in this society about whether actions are morally wrong. One might respond to this worry about over-discrimination by relaxing the etiological specification, but the account will then, predictably, threaten to under-discriminate.

Taken together, these considerations can seem to generate an insoluble dilemma for any non-cognitivist attempt to solve the specification problem. On the one hand, the need to distinguish moral judgements from sports-fan disapproval, aesthetic judgements, all-in disapproval, etc., creates pressure to characterize the moral attitude in increasingly specific terms. On the other, we noted that our twin earth intuitions concerning substantive moral disagreement simply do not track fine-grained differences in attitude, all else being held equal. These points seem to rule out any conception of the moral attitude fine-grained enough to solve the specification problem.¹⁵

A tempting (and damning) explanation of the force of this dilemma is that what distinguishes wrongness-judgements from other judgements are their

¹⁴ The added complexity of Kauppinen's full analysis (2010, p. 253) makes no difference here.

¹⁵ Notice that the dilemma generalizes fairly straightforwardly to relativistic forms of cognitivism that take wrongness-judgements to be beliefs about the relations of acts to certain of our attitudes.

cognitive contents, not some non-cognitive attitude.¹⁶ However, what makes this dilemma particularly serious is that it threatens to undermine the two central arguments for non-cognitivism. First, without an adequate specification of the relevant moral attitude, we lack a substantive non-cognitivist explanation of the motivational character of wrongness-judgements. Second, the moral attitude dilemma seems to directly undermine the non-cognitivist's claim to have the better explanation of univocality. If cognitivist realists should be troubled by intuitions about moral disagreement and moral twin earth cases, it seems that non-cognitivists should be worried too.

3. Towards a Solution

Merli's argument suggests that, while the non-cognitivist's task of specifying the attitude expressed by wrongness-judgements and her opponent's moral twin earth problem initially appear to be distinct, they are in fact instances of the same specification problem. Almost everyone, it seems, is vulnerable to moral twin earth-style cases.¹⁷ Judging from the full range of these cases, intuitions about agreement and disagreement in wrongness-judgements seemingly require neither specific substantive assumptions about cognitive content, nor specific substantive assumptions about non-cognitive attitudes.

One plausible diagnosis of the problem is that our intuitions of disagreement are sensitive to both cognitive and non-cognitive features of agents' psychologies. Retain enough similarity on a variety of cognitive and non-cognitive dimensions, and any specific dissimilarity will be tolerated. This suggests a kind of 'cluster theory': an attitude counts as a wrongness-judgement just in case its combined and balanced score on the various dimensions reaches over some (vague) threshold. However, taken on its own, such a cluster theory

¹⁶ Cf. Svavarsdóttir 1999 and Zangwill 2008, 2009.

¹⁷ Moral twin earth arguments have typically been deployed against views that wed cognitivism to non-relativistic naturalistic realism, with non-cognitivists offering other arguments against relativists and non-naturalists. We don't need twin earth to see that relativists have a problem explaining intuitions of disagreement: they already evidently need a solution to this problem. More interesting is non-naturalism. One hypothesis is that it has distinctive resources to escape this problem. Another is that twin earth thought experiments require an informative theory of reference for moral language to work against, and few non-naturalists have offered such a theory.

is superficial (Zangwill 2009, 421–22). It tells us nothing about *why* certain dimensions count to the extent they do, and what unifies the cluster of relevant features.

As we saw above, Blackburn's etiological proposal in *Ruling Passions* does not look like a promising response to this explanatory challenge. In an earlier exchange with Nicholas Sturgeon, Blackburn briefly suggests a more promising alternative. This is that the pattern of variation in the moral attitude is explained by the fact that it suits certain *practical needs* served by moral discourse. We count certain attitudinal deviants as having ethical views because we need to be able to engage conversationally with their attitudes, either to change them, or to campaign against them. Using ethical terms, Blackburn says, is simply the best means for doing so (1991a, pp. 5–9). This points to a broad kind of explanation for the weights given to various factors that underwrite our univocality judgements: they depend on our needs for practical interaction.¹⁸ However, Blackburn says almost nothing to develop this idea, leaving it considerably less developed than standard non-cognitivist accounts of disagreement in terms of disagreement in attitude (Sturgeon 1991, 25–27).¹⁹

A convincing solution to the specification problem should provide what the cluster theory, and Blackburn's brief suggestion here, do not. That is, it should offer an account of the nature of moral thinking and discourse that offers a satisfying *explanation* of our intuitions of agreement and disagreement. We take Blackburn to be right to suggest that the practical needs served by moral discourse should play a central role in this explanation, and that the non-cognitivist is uniquely placed to provide such an explanation. In what follows, we offer a solution to the specification problem that builds on these insights to provide a broad and explanatorily robust account of our intuitions of moral agreement and disagreement.

¹⁸ One might object that this account cannot work with twin earth cases, as we do not engage in conversation with our twin earth counterparts. However, our interpretation of them might plausibly be guided by an understanding of them as *potential* interlocutors.

¹⁹ Blackburn has not himself returned to this suggestion in later work: for example, his discussion in *Ruling Passions* countenances attitudinal variability primarily in terms of the etiological suggestion mentioned in section 2 above. We are thus unsure whether he still endorses this suggestion, or how he would want to develop it.

We develop our solution in two stages. First, we spell out our positive account of the psychology of what we call *paradigmatic wrongness-judgements* (section 4). Our encounters with judgements of this kind constitute nearly all of our experience with what we think of as wrongness-judgements and can be expected to have shaped our understanding of (and dispositions to react to) these judgements. Second, we provide an explanation of the troublesome classificatory intuitions that generate the non-cognitivist's specification problem (section 5). We show how our account of paradigmatic wrongness-judgements predicts that people who make judgements of this kind should have classificatory intuitions that are *accommodating*, counting psychological states as wrongness-judgements in the absence of some signature marks of paradigmatic wrongness-judgements. We also show that the functional nature of wrongness-judgements plays an important independent role in explaining some of these intuitions. We argue that together, these mechanisms provide a principled explanation of the intuitions that give rise to the specification problem.

4. A non-cognitivist account of paradigmatic wrongness-judgements

In this section, we first describe five features that are central to our understanding of wrongness-judgements. We then explain why it is plausible that states that share these features form a naturalistically respectable functional kind, held together by homeostatic mechanisms.²⁰ For simplicity, we will call wrongness-judgements that have all of these features to a significant degree *paradigmatic wrongness-judgements*.²¹

Before turning to two features that are distinctive of paradigmatic wrongness-judgements, we introduce three features that are shared with instances of many other types of judgements. These features are *aversion*, *personal-level acceptance*, and *engagement*. Consider these in turn.

²⁰ We take our proposal to fit Gibbard's promising *general* conception of moral attitudes as what he calls 'adaptive syndromes': '... syndromes of tendencies to action, expressive tendencies, and tendencies to be caused in certain ways...' that have a certain *function* (1990, p. 134).

²¹ Notice that something might be uncontroversially counted as a wrongness-judgement without being a 'paradigmatic' wrongness-judgement in our sense.

First, in paradigmatic wrongness-judgements, the judge is at least to some extent averse to the action being judged: that is, the judge has a *ceteris paribus* preference that the target action does not take place, or wishes that it had not taken place. (This feature is shared with many other states, including other types of negative evaluative judgement.)²²

Second, paradigmatic wrongness-judgements involve personal-level acceptance of the judgement. For example, my judging that an act is wrong is different from that act merely seeming or feeling wrong to me. In the former case, I accept or make the judgement, in a way that I do not in the latter case. The feature of personal-level acceptance is shared with other types of personal-level judgement: consider the difference between what our senses tell us ('the oar sticking into the water is bent') and what we believe ('it is straight'), and the difference between our desires and what we take to be desirable. In many cases, personal-level acceptance will be connected to taking there to be adequate reasons (of the right kind) for making the judgement, even if one does not have those reasons before one's mind.²³

Third, in paradigmatic wrongness-judgements, the judge is disposed to engage with the moral judgements of others. Engagement involves being disposed to treat arguments or grounds for one's own wrongness-judgements as arguments or grounds for others to share those judgements. It also involves being disposed to argue with someone who rejects one of one's wrongness-judgements. This feature is shared with other types of judgement that are not overtly subjective.²⁴

²² Aversion is absent from Gibbard's 1990 account of narrowly moral judgements. The reason, it seems, is that Gibbard is concerned to make sense of cases where one can think that something is morally wrong while asking whether doing it is nevertheless warranted (or 'rational', or 'the thing to do'). But even given this constraint, Gibbard's ambition to reduce normative thoughts to thoughts about what is warranted does not immediately rule out a role for aversion. To think that something is morally wrong could, in part, be to think that some degree of aversion is warranted on the basis of particularly moral considerations while leaving open whether it is trumped by other considerations. For discussion, see Gibbard 1992.

²³ Arguably, personal-level acceptance grounds various other belief-like properties of this and other attitudes (compare Björnsson 2001).

²⁴ Features like personal-level acceptance and engagement are part of Horgan and Timmons' 2006 case for claiming that theirs is a 'cognitivist' expressivism. Again, because we have adopted a broadly Humean, functionalist account of the contrast between cognitive and non-

The features just sketched are characteristic of paradigmatic wrongness-judgements, but do not distinguish them from paradigmatic instances of all other types of judgement. For example, certain paradigmatic negative aesthetic judgements, perhaps about a performer ruining a dance or play, might share these features. Paradigmatic moral judgements are distinctive in two connected ways: the *grounds* on which such judgement are accepted, and their association with *social hostility*. Consider these two features in turn.

First, paradigmatic wrongness-judgements have distinctive grounds. They are grounded in the perception or belief that someone has performed an action with one of two features: intentionally harming or risking harm to others or things that they care about,²⁵ and failing to respect certain boundaries that play a central role in sustaining social cooperation.²⁶

Second, paradigmatic wrongness-judgements involve a pattern of emotional and behavioural dispositions that we will call *social hostility*.²⁷ We will

cognitive states, we are unmoved by this suggestion. Together with some aspects of what we call ‘social hostility’, personal-level acceptance and engagement are also part of Gibbard’s (1990, pp. 71–5) account of norm-acceptance.

²⁵ The clearest cases of such wrongness-judgements concern acts that involve both intended and actual harm. Where these come apart, there are some important complications. For example, Cushman (2008) suggests that judgements of wrongness, in comparison to judgements of *blameworthiness*, are much more sensitive to the beliefs and desires of the agent than to actual outcomes.

²⁶ Lest our list seem too short, keep in mind the variety of boundaries that could be understood to play this role. For example, given widespread belief in an Old-Testament style deity, violations of purity norms would be seen to threaten to undo cooperative gains by inviting supernatural smiting. Similarly, a person’s impure act might itself be seen as an intrinsic harm to the community or what they care about.

Notice that the class of morally relevant others allows for borderline cases. One familiar example concerns the treatment of non-human animals. Some people take animals to be objects of direct moral concern in more or less the same way that people are (e.g. Singer 1977 and Regan 1983), others to deny that they are objects of direct moral concern at all (e.g. Carruthers 1992), and finally, others to suggest that they warrant a kind of moral concern that is different in kind from that owed to persons (e.g. Scanlon 1998, pp. 180–4).

²⁷ As we have seen, something akin to social hostility is part of accounts of wrongness-judgements proposed by Gibbard (1990) and Kauppinen (2010). Similarly, Blackburn says that ‘the vocabulary of right and duty, rights and obligations’ and thus presumably moral wrongness should be understood in terms of ‘states of mind ... that prime us to insistences and to hostility to others’ (1998, p. 68).

develop this idea in modest detail, considering first the judgement that *another's* action is morally wrong. Social hostility comes in degrees: for example, mild social hostility can take the form of a disposition to refrain from acts of kindness towards the target that one would otherwise be disposed to perform, or to refrain from treating the target as a candidate for mutually beneficial cooperation. More extreme social hostility can involve dispositions to active ostracism and violence. More subtly, social hostility might consist in a *conditional* disposition to downgrade one's relationship with the target unless certain responses are forthcoming, coupled with a disposition to communicate this disposition, thus constituting a kind of *demand* for such responses. Social hostility comes not only in degrees, but in patterned varieties of flavor. For example, aggressive or violent social hostility may be more common in response to acts that cause physical harm, while shunning or isolating forms of hostility may be more common in response to perceived disloyalty or sexual deviance.²⁸ Like aversion, social hostility is a disposition, and need not actually produce the characteristic emotional expressions and behaviours in any given case.²⁹

This basic picture of social hostility needs to be enriched along at least two dimensions. First, we think that there is a story to be told about guilt, understood as an emotion designed to placate social hostility. Second, we can think of the attitude expressed by a paradigmatic moral utterance as directed not only at potential wrongdoers, but also at potential evaluators (people in a position to respond to potential wrongdoers and their actions). It is notable that moral attitudes are characteristically *intolerant*: taking something to be wrong typically involves a disposition to some degree of social hostility towards potential evaluators who fail to share one's attitudes towards potential

²⁸ For a valuable discussion of some forms of social hostility, see Bennett 2002. Social hostility is found in a variety of social species, and has been given considerable attention. See e.g. Clutton-Brock and Parker 1995 and Aureli et al. 2002.

²⁹ Social hostility is more closely related to expressions of resentment or blame that Gibbard associates with narrowly moral judgements. One contrast with Gibbard (1990) is that for Gibbard, wrongness-judgements are higher-order judgements that such hostility is *warranted*. On our view, paradigmatic moral judgements involve dispositions to such hostile emotions, rather than higher-order judgements about their warrantedness. At least phenomenologically, we take this to be a strength of our proposal.

wrongdoers. We are just not happy with people who think that there is nothing wrong with others torturing babies.³⁰

We have just argued that paradigmatic wrongness-judgements involve five features: aversion, personal level acceptance, engagement with other judges, grounding in a distinctive class of beliefs or perceptions, and social hostility. We do not think that the five features just identified are merely accidentally conjoined in paradigmatic wrongness-judgements. Rather, we hypothesize that these features are core properties in a naturalistically respectable psychological kind, functionally unified by homeostatic mechanisms.³¹ By this, we mean that there are mechanisms that tend to reinforce the presence of each feature, given the presence of the others. In particular, presence of each of the two features that are *specific* to paradigmatic wrongness-judgements will tend to make presence of the other more likely. In what follows, we briefly sketch some of these mechanisms.

It is common for non-cognitivists to emphasize the way that moral norms function to enable and enforce social coordination and cooperation.³² We follow this tradition and suggest that it is because human psychology is adapted to the needs of cooperation in a certain way that the five features we have identified tend to go together. In particular, it seems plausible that our capacity for paradigmatic wrongness-judgements has been shaped so as to direct social hostility towards threats to social cooperation, because the disposition to such hostility tends to forestall such threats.

Consider first the paradigmatic grounds for wrongness-judgements. We identified two such grounds above: that an action intentionally harms or risks harm to others or things that they care about, and that it fails to respect certain boundaries that play a central role in sustaining social cooperation. We should expect moral judgements to be paradigmatically sensitive to the latter sorts of grounds, if their function is to enforce coordination and cooperation. We should also expect such judgements to be sensitive to harming. This sensitivity

³⁰ Recall the analogous (but friendlier) imperative element of Stevenson's gloss on calling something good: *I approve of Φ ing, do so as well* (1963, p. 25).

³¹ Compare Richard Boyd's (1988) suggestion that the morally good is constituted by a homeostatic property cluster.

³² See especially Blackburn 1988b and Gibbard 1990.

allows moral discourse to provide some protection against the threats to cooperative gains that are posed by inconsiderate or hostile parties.

It is notable that the class of paradigmatic grounds is broad enough to allow wrongness-judgements on both sides of deep moral disagreements to be paradigmatic wrongness-judgements. This too can be expected given that the function of moral judgement and moral discourse is to bring about coordinated attitudes and expectations, and given that different people can have interest in differing cooperative schemes, in light of their personalities, cultural backgrounds, and material circumstances.

Consider next why social hostility would tend to be an especially effective response to such threats, given this function. A number of points are relevant. First, it is dangerous to trigger hostility in others, especially given the possibility that such hostility can escalate to violence. Humans thus have a tendency to avoid eliciting hostility. However, a hostile response to threats to cooperation can itself be costly and risky. A *cooperative* mechanism for addressing such threats would mitigate costs and risks. Because social hostility is addressed partly to potential evaluators, it has the effect of encouraging coordination in the response to threats to existing patterns of cooperation. This has the effect of raising the potential costs of norm-violation, while simultaneously lowering the enforcement costs to each enforcer. Together, we might expect that these evaluator-mediated effects of the moral attitude will tend to reduce the numbers of wrongdoers, thus tending to sustain the pattern of cooperation.

Maintaining a cooperative arrangement also involves managing and minimizing the temptation to violation by the cooperators. Here, we would expect guilt to play a prominent role. From the perspective of the evaluating group, an agent's possession of a (properly-tuned) capacity for guilt would signal an internalization of the relevant cooperative norms, and an enhanced sensitivity to social hostility, thus making agents with such a capacity more reliable cooperators. From the perspective of the agent, display of guilt over wrongdoing would tend to mitigate social hostility towards the agent, because the guilty party would be signalling existing internal pressures to desist from

the offending behaviour. The protection provided by this mitigating role would in turn encourage the learning of guilt-reactions.³³

Given the advantages of cooperation, we suggest that the sorts of pressures identified will tend to help to create and maintain a type of psychological state with the two distinctive features identified above. Our hypothesis that wrongness-judgement is such a psychological state can also help to explain the three non-distinctive features of paradigmatic wrongness-judgements. Thus, like guilt reactions, an agent's *aversion* towards acts eliciting her social hostility would tend to make her a more reliable cooperator, and agents who rely on general expectations of cooperative behaviours will be disposed to prevent breaches of such expectations. Further, the complexity of conditions for cooperation calls for the ability to engage in non-strategic reasoning both personally and interpersonally, and attitudinal responsiveness to the personal-level judgements that such reasoning delivers.³⁴

It may be helpful to contrast the account just offered with the sort of account that might be sketched about paradigmatic negative aesthetic judgements. Such judgements have different typical grounds: they are triggered by the appearance of things; by how they strike us when we contemplate them in some paradigmatically aesthetic way. On our account, the grounds of paradigmatic wrongness-judgements, by contrast, involve

³³ Similar homeostatic mechanisms are at play in upholding quasi-moral 'wrongness-judgements' concerned with preventing harm and upholding cooperation within restricted groups, such as families, gangs, groups of professionals, or members or fans of a particular sports team. There are important differences between quasi-moral judgements and paradigmatic wrongness-judgements, however, both in the range of harms and cooperative preconditions that are relevant and in the scope of engagement and the degree of intolerance involved. Moreover, these differences are linked: when the grounds for the attitude are crucial to cooperation in general, not just within the group, social hostility can resonate outside the group and a wider scope of engagement is likely to yield better results.

³⁴ Haidt (2001) emphasizes two countervailing features of moral discourse: (1) the way in which people tend to act as 'intuitive lawyers' in defence of their views, and (2) the way in which people's intuitively held norms are sensitive to the emergence of coordinated social norms on a topic. The extent of these phenomena is controversial, but were Haidt correct about their central significance, this would not undermine our explanation of the role of reasons-sensitivity in allowing for discursive coordination. It would merely remind us that social and psychological mechanisms that deploy the rhetoric of reasons do not imply a *substantively rational* process of shared reflection.

beliefs about the actions and motivational structures of the agent, and their potential to harm others or contravene certain social boundaries. Moreover, rather than involving social hostility, paradigmatic aesthetic judgements dispose the agent to make certain choices and recommendations for the activity of aesthetic contemplation: what movies to go to, what museums to visit, what artists to support, and so forth.

This contrasting account of the distinctive grounds and practical upshot of paradigmatic aesthetic judgements helps to explain some notable contrasts with moral judgement. It is not surprising that aesthetic considerations tend to matter less than considerations with strong implications for the expectations that we rely on in cooperation, thus providing less support for social hostility. Moreover, since paradigmatic aesthetic judgements are perhaps primarily sustained by phenomenological reactions, whereas paradigmatic wrongness-judgements are sustained by beliefs about complex psychological, social and causal states, they tend to come with less of an appearance-reality distinction than moral judgements do.

The account that we have offered in this section could be considerably developed in each of three dimensions. First, it is natural to want a much more detailed account of the features of paradigmatic wrongness-judgements identified here. Second, we have barely sketched the potential homeostatic mechanisms that might serve to make paradigmatic wrongness-judgement a naturalistically respectable psychological kind. Third, in focusing only on paradigmatic wrongness-judgements, we abstract away from a host of interesting questions about how our account could be extended to a broader class of moral judgements (for example, about value or virtue).³⁵ It is also evident that the account centrally involves empirical conjectures about human psychology. Our largely ‘armchair’ motivation of these conjectures clearly

³⁵ We think that there is reason to be optimistic about this last point. In extending the account to other sorts of moral judgements, the expressivist can appeal not only to emotive and conative aspects of paradigmatic judgements of the relevant kind, but also to the kinds of grounds on which such judgements are made. Since any purely cognitivist account of such judgements would have recourse only to their cognitive content and various cognitive attitudes towards these, the expressivist would seem at least as well positioned to deliver accurate specifications of these states.

needs to be augmented by serious empirical enquiry.³⁶ For the time being, we take it to garner plausibility from its ability to explain certain (hopefully uncontroversial) features of our shared moral experience.

This section has sketched the first part of our non-cognitivist solution to the specification problem: a homeostatic functionalist explanation of the nature of paradigmatic wrongness-judgements. The next section builds on this account to provide the second part of our solution: an explanation of why people making paradigmatic wrongness-judgements would have the classificatory intuitions that generate the non-cognitivist's specification problem.

5. A non-cognitivist account of our classificatory intuitions

The account of paradigmatic wrongness-judgements just offered does not by itself solve the specification problem. This is because we will intuitively treat many possible interlocutors as making wrongness-judgements, despite their lacking one or more of the features identified in the previous section. In this section, we show that the functional account developed in the previous section can predict and explain the classificatory intuitions that otherwise threaten to make the specification problem appear intractable. We proceed in two stages. First, we argue that our account predicts that paradigmatic wrongness-judges would have robustly *accommodating* classificatory intuitions: intuitions that count psychological states as wrongness-judgements in the absence of some of the marks of paradigmatic wrongness-judgements sketched in the previous section. Second, we argue that the fact that paradigmatic wrongness-judgement is a *functional* state underwrites a complementary explanation of the remaining apparently divergent classificatory intuitions.

The traditional non-cognitivist proposes that moral disagreement is best understood as *practical* disagreement: disagreement about what to do or what

³⁶ The beginnings of the relevant body of empirical enquiry can be sampled in, e.g., Nichols 2004, and Sinnott-Armstrong (ed.) 2007, 2008a, 2008b. Two recent book length attempts to integrate empirical studies and functionalistic considerations into general accounts of moral thinking are Joyce 2006 and Prinz 2007. Significantly, Joyce defends an error-theoretic and Prinz a constructivist rather than non-cognitivist account of morality, but many of the considerations they discuss are straightforwardly relevant within a non-cognitivist framework.

attitudes to have. Our account of classificatory intuitions refines this suggestion. On our view, classificatory intuitions should themselves be understood as elements of the functional story told in the previous section. According to this story, moral discourse and moral thinking have the function of producing and enforcing cooperative convergence among agents. The functional significance of classification stems from the fact that classifying someone as making a wrongness-judgement involves deploying a characteristic set of defeasible expectations about her reasoning, emotional reactions, and behaviour. Among these are expectations concerning her aptness for discursive engagement, in particular the exchange of reasons and expressions of agreement or disagreement, and the corresponding expression of aversion and social hostility. Such engagement will play a key part in the central functional role of wrongness-judgements: the creation and preservation of patterns of social coordination and cooperation. We thus hypothesize that our classificatory intuitions have been shaped to contribute to this coordinating function.

Agents can differ not only in their socially significant goals, as non-cognitivists have traditionally stressed, but also in how closely they resemble paradigmatic moral judges on each of the dimensions sketched in the previous section. As our account emphasizes, the practice of moral discourse in part aims at *producing* (approximate) coordination or convergence in these respects. We should thus expect paradigmatic wrongness-judges to tend to treat judges whose attitudes are only a rough approximation of paradigmatic wrongness-judgements as potentially legitimate parties to moral discourse. Doing so makes it more likely that, by exchanging reasons and deploying social influence, these non-paradigmatic judges will be led to converge on the paradigm. We call this mechanism the *functional demand for accommodation*.

This demand for accommodation explains the existence of most of the intuitions that generate the moral attitude dilemma. These intuitions seem to show that *no* specific attitude is necessary for counting as making a wrongness-judgement, because we would intuitively categorize some individuals or groups who lacked that attitude as making wrongness-judgements and hence really (dis-)agreeing with us about which acts are wrong. This intuitive tendency to accommodation is predicted by our functional account of wrongness-judgement. The function of paradigmatic wrongness-judgements demands that paradigmatic judges treat a range of psychological states diverging from the

paradigm as potential targets of substantive moral agreement or disagreement. This demand is met by paradigmatic judges being disposed to treat such judgements as belonging to the same kind as their own judgements.

It is important to emphasize that our account of the functional demand for accommodation does not imply that attributions of wrongness-judgements track whether engaging with these judgements will *actually* lead to coordination in a particular case. (The death of a moral judge, or his waxing stubbornness in old age, undermines coordination but not attributions of wrongness-judgements.) Rather, on our view, these attributions track *types* of states that *tend to be suitable* for reinforcement or modification in moral discussion. Similarly, we are not suggesting that attributions of wrongness-judgements involve the explicit thought that the attributee is in a kind of state that tends to be open to reinforcement or modification. The idea, rather, is that the weights that we intuitively give to certain factors in determining whether someone is making a wrongness-judgement have been shaped by the importance of these factors for the function of paradigmatic wrongness-judgements.

How much accommodation does our account predict? It predicts that if a psychology is similar enough to ours with respect to the five dimensions mentioned in the previous section to permit us to successfully coordinate with the bearer of that psychology, then there will be pressure to treat that psychology as deploying wrongness-judgements. Perhaps some other basis for cooperation will be possible with agents whose psychologies diverge radically enough from our own on many of these dimensions, but moral discourse will tend to be inert in such cases. The functionalist account thus predicts failure to accommodate psychologies that diverge radically from the paradigm with respect to these five dimensions. Between the near-paradigmatic cases and the radically divergent ones, there is a range of possible psychologies for which the prospects for coordination via moral discourse are unclear. Our account predicts that we will have unstable or uncertain judgements about whether such psychologies include wrongness-judgements, and increased individual variation in judgements about such cases.³⁷ Happily, intuitions appear consilient with this prediction.³⁸

³⁷ If someone has a psychology that includes two imperfect candidates for wrongness-judgement, functional purposes favour counting only the better fit of the two as involving

Consider some ways in which degrees of divergence from paradigmatic cases might affect intuitions, starting with *grounds* for wrongness-judgements. On the near end of the spectrum, we seem to have no problem at all treating a society's otherwise similar judgements, that are sensitive to only deontological, or only consequentialist reasons, as making wrongness-judgements (these are the intuitions that drive Horgan and Timmons' canonical moral twin earth cases). At the far end of the spectrum, we find Philippa Foot's character who claims that clasping one's hands three times a day is uniquely fundamentally good (1958, 92–4). As Foot points out, we are uncomfortable attributing a moral judgement to this person if he takes this good to be independent of any conjecture about positive further effects or features of such an act. In between we might find groups who take harming to be wrong only if it harms some select genetic or religious elite. Here we might waver between treating this group as erroneously thinking only certain people matter morally, and treating them as lacking the concept of moral wrongness, and caring only about their elitist ideal. This variation is what we could expect given the coordinating function of wrongness-judgements. The less sensitive someone's judgements are to the typical grounds for moral judgement, the less likely they are to be affected by our moral arguments.

Consider next variations in *social hostility*. At the near end of the spectrum, we tend to accommodate variations in the form of this hostility: for example, as the attitudinal twin-earth scenarios discussed in section 2 suggest, we tend to

wrongness-judgement, even if we would count another agent, who had only the lesser fitting state, as making wrongness-judgements with that state. Cf. Dreier (2006, pp. 257–8).

³⁸ To say that there is a functional demand for accommodation is not to deny that there are worries about equivocation in this area. For example, people have certainly felt the need to make distinctions, between subjective, prospective and objective notions of 'ought' (e.g. Ross 1939, p. 146–7; Zimmerman 2008, p. 1), or between what ought to be the case and what one ought to do (Harman 1983, p. 38), or between maximizing or satisficing uses of 'right' (Jenkins and Nolan 2010). These distinctions (which correspond to distinctions between uses of 'wrong') seem to correspond directly to differences in naturally occurring attitudes towards those who do not perform the target action, and once they have been made it is sometimes unclear exactly what the core disagreement would be between for example consequentialists and proponents of agent-centred notions of moral wrongness. Here, as in the previous note, ascription of wrongness-judgement may be sensitive to the richness of alternative types of judgement available for ascription.

accept the possibility of a society where wrongness-judgements are typically accompanied by contempt and shame rather than resentment and guilt. At the far end of the spectrum, we tend to reject the idea that a twin earth word could express a wrongness-judgement if its use is not linked to any mechanisms for discouraging violation.

The three other features of wrongness-judgements—*aversion*, *personal-level acceptance* and *engagement*—all come in degrees and affect attributions of moral judgement in ways that one would expect given the functional demand for accommodation.

Aversion: We readily allow that people can agree that some act is wrong but be more or less averse to it. At the same time, a very weak preference that an act is not performed tends to undermine attributions of the judgement that it is *very* wrong, and the complete absence of such a preference suggests that the judge does not think that the act is wrong (unless there is a special explanation for the discrepancy—see below). Again, this is to be expected, given the demand for accommodation. As long as the judgements in question carry some corresponding motivational force, this can provide an adequate basis for coordination (provided that there is agreement in judgement), as well as an opportunity to strengthen the pre-existing motivational force by social support.

Personal-level acceptance: Some borderline cases arise because of lack of personal-level acceptance. For example, an agent might feel guilt and regret about having done something that hurt someone else, or subtly display social hostility towards someone else who has done so, but reject the understanding of the action that triggers these feelings. In such cases we can be tempted to say that the agent *feels* as if the target action was morally wrong though she *believes* that it was not. Alternatively, we might say that at heart, she thinks that it was morally wrong but fails to admit this to herself.³⁹ In these cases, the lack of personal-level acceptance complicates discursive engagement, but the existence of an appropriately tuned sensibility suggests substantial possibility of successful coordination nonetheless.

³⁹ Thus, in the case of Huckleberry Finn, who accepted the judgement that it was wrong to help fugitive slaves but nevertheless proceeded to do so, one might be tempted to say that he grasped, intuitively, that it would be wrong to tell on Jim. Cf. Prinz (2007, 113–14) and Blackburn (1991a, 7–8).

Engagement: We intuitively allow that people can be more or less disposed to engage with the moral judgements of others. Indeed, quite a few people engage only reluctantly, frequently relativizing moral judgements and arguments to individuals or cultures rather than confronting them with their own. The more extreme their relativistic tendency is, the less clear it will be that they really have moral views of their own. Again, this seems to be well explained on the assumption that our interpretation is shaped by the purpose of coordination: relativization disengages the cogs of moral argument from the wheels of moral emotions and motivation.

The degree to which we are sensitive to these five factors in our attributions of wrongness-judgements seems to support the non-cognitivist's functionalistic account. However, the account might seem incompatible with a familiar group of more radical characters who have neither the social hostility nor any degree of aversion, and perhaps also accept non-standard grounds for judgement, but who nevertheless seem capable of judging actions right or wrong. This group includes cynics, amoralists, sadists, the chronically listless, and children who are learning about what is right or wrong from their parents but are yet to acquire the corresponding moral motivation themselves.⁴⁰ Although the interpretation of these unmotivated judges and their unmotivating judgements is highly controversial, they seem to show that there is no *obvious* a priori necessary connection between moral judgements and any degree of social hostility or motivation to act. Moreover, it is unclear how the functional demand for accommodation can explain why people attribute moral judgements in these cases. Why would one try to coordinate with people who do not care about any sort of paradigmatic moral reasons by engaging them in moral discussion?

We think that the explanation of our (hesitant and varying) tendency to attribute wrongness-judgement to such agents can be explained by the fact that paradigmatic wrongness-judgements constitute a *functional* kind, as we argued in the previous section. Since many aspects of this function are obvious enough, and since everyday encounters with paradigmatic wrongness-judgements constitute nearly all of our experience with what is naturally

⁴⁰ For discussion of these and related examples, see Stocker 1979, Brink 1989 (Ch. 3), Dreier 1990, Smith 1994, Svavarsdóttir 1999, Gert 2002, and Zangwill 2008.

understood as wrongness-judgements, we should expect our ordinary understanding of wrongness-judgements to reflect this fact. In particular, we should expect our intuitive typing of such judgements to share features with our typing of other kinds that we are primarily interested in for their function, such as artefacts with a purpose and physiological organs.

Consider examples of such functional kinds. Something is normally typed as a heart or a car not only if it actually pumps blood or serves as a vehicle in normal ways, but also under any of the following conditions:

- (1) It is the upshot of a sufficiently completed process of a sort normally responsible for fully functioning items (a heart under development, a car still lacking its steering wheel).
- (2) It has deteriorated in ways that prevent it from performing its function while retaining enough structure (a wrecked car, a thoroughly clogged heart), or it is placed under conditions where it cannot function even though it is internally intact (a car with wheel clamps, a heart on the surgeon's table during transplant).
- (3) It is intended to be of a functional kind, and is similar enough to paradigmatic instances of that kind (an artificial heart). (The typing of instances is perhaps less clear here than for (1) and (2).)

It is an interesting question *why* functionally interesting kinds are individuated in this way, but whatever the answer is, our claim here is that cases of unmotivated wrongness-judgements fall into the same pattern.⁴¹ This is not the place to go through the full variety of cases that have been discussed in recent literature on the connection between moral judgements and moral motivation. However, there is a growing consensus that the more plausible cases of unmotivated moral judges fall under one of the following descriptions:⁴²

- (1') The judge is learning what is right or wrong from others who make moral judgements, but lacks the relevant preferences and moral emotions (children and people with developmental difficulties).

⁴¹ We thus follow Björnsson 2002 and Bedke 2009 in suggesting that the intuitive connection between moral judgements and moral motivation is best understood in functionalist terms.

⁴² Compare Dreier (1990, pp. 9–14), Blackburn (1998, pp. 59–68), Björnsson (2002, pp. 334–42), Tresan (2006, pp. 149–52), and Bedke 2009.

- (2') The judge had previously been motivated to act by her moral judgements under normal psychological conditions, but no longer is because of psychological abnormalities (listlessness, brain damage), an internal (and perhaps deeply flawed) critique of morality, or psychological reversal (sadists, Milton's Satan).
- (3') The judge engages with the moral judgements of others, taking herself to be in moral agreement or disagreement with them, and mimics enough features of paradigmatic moral judgements (Plato's Thrasymachus, psychopaths). (Again, the attribution of moral judgements may be less clear here than for (1') and (2').)

Given a functionalist understanding of moral judgements, judgements exemplifying (1') fall under (1) while those exemplifying (2') fall under (2). Instances of (3') fall under (3), finally, because the judge intends to be making wrongness-judgements. This correspondence between classifications of paradigmatic objects-with-a-function and classifications of moral judgements further supports the assumption that our attributions of moral judgements are driven by an interest in their practical function.

It should be noted that the functional explanation of these intuitive judgements explains an interesting asymmetry: we are more prepared to treat an amoralist in our own society as perhaps making moral judgements, than we are to treat an imagined twin earth society whose use of the term 'wrong' uniformly resembles the amoralist's here. The functional account suggests that this is because we treat the amoralist's judgements as a functionally degraded or not yet fully functioning member of our functional kind wrongness-judgements. However, the twin earthlings' judgements cannot be seen as similarly degraded or not yet fully developed, thus explaining why we are less tempted to treat them as wrongness-judgements proper.

It should be clear that the explanations offered in this section predict rough tendencies to treat various attitudes as wrongness-judgements, not sharp and uniform conceptual rules. We take such tendencies to be shaped by one's understanding of a multidimensional practice, and although some of the functionalist features are particularly salient, people might weigh the five features we have discussed here differently. Perhaps some features have played a more important role in one's experience, or better fit one's theoretical commitments (for example, we might expect those inclined to an absolutist cognitivist interpretation of moral judgements to stress sameness of cognitive

content and be less sensitive to differences in non-cognitive aspects). Again, such variation seems to be exactly what we see.⁴³

This section has argued that two mechanisms—the functional demand for accommodation, and sensitivity to the nature of wrongness-judgement as a functional kind—help to explain the complex set of intuitions that underwrite the non-cognitivist’s moral attitude problem. Together with the functional account of paradigmatic wrongness-judgements from section 4, this account sketches what we take to be a satisfying non-cognitivist answer to the specification problem. The answer denies the existence of a philosophically sharp boundary between wrongness-judgements and other attitudes. Our account can thus acknowledge Blackburn’s insistence that no definition of ‘the’ moral attitude is possible ‘...because of the polymorphous nature of our emotional and motivational natures themselves’ (1998, pp. 13–14). However, this is no admission of defeat in the face of intractable data, or retreat into under-discriminating liberalism. This is because we have provided what the simple cluster view and Blackburn’s liberal view about moral attitudes do not: a plausible explanation for why our intuitive judgements of moral univocality have the shape that they do. We have argued that the apparent complexity of these judgements is driven and maintained by the same pressures that link the features we identify with these judgements into a natural grouping. Those pressures *both* explain why the paradigm wrongness-judgements that share all five of the features that we identify form a sort of privileged class, *and* explain why we are prepared to accommodate judgements with strikingly divergent features as (non-paradigmatic) wrongness-judgements.

6. Conclusions: is this non-cognitivism, and is there a cognitivist alternative?

We conclude by briefly canvassing three issues: whether the view defended here is really best understood as a form of non-cognitivism, whether our view is compatible with other crucial elements of the non-cognitivist program, and the significance of our discussion for the broader dialectic in metaethics.

⁴³ See e.g. Francén 2010, for some of the variation and discussion of its implications for a priori methodologies in metaethics.

To begin, we have of course departed from the common assumption that a complete non-cognitivist analysis of moral judgements must provide a determinate attitude for each kind. For example, although attitudes of social hostility play a central role in our account, a wrongness-judgement need not involve any specific such attitude, and we have identified several other features that play a similarly central yet negotiable role. Further, and perhaps more surprisingly for a non-cognitivist theory, the account allows that some states count as wrongness-judgements not because they involve any distinctive non-cognitive attitude, but because they are grounded in paradigmatically moral considerations and sensitive to paradigmatic moral reasoning in thought and discourse.⁴⁴

Despite these features of our view, we think that it is clearly non-cognitivist in spirit. First, what holds the class of paradigmatic wrongness-judgements together is not that they have the same cognitive content but rather that they are closely enough related to judgements that display paradigmatic features of wrongness-judgements; features that together have a practical, non-cognitive function. Consequently, the wrongness-judgements expressed by sincere utterances of the form ‘ Φ ing is wrong’ in moral contexts are best understood in terms of their practical function. Second, our classificatory intuitions about wrongness-judgements reflect this underlying coordinating function. This coordinating function explains our intuitions about paradigmatic wrongness-judgements as well as our willingness to accommodate psychological diversity by classifying non-paradigmatic cases as wrongness-judgements. For the same reason, our view rejects the cognitivist idea that paradigmatic moral disagreements are best understood as concerning some fact that holds independently of the beliefs or attitudes of the disagreeing parties. All this reflects non-cognitivists’ long-standing insistence that morality is ‘a practice with a purpose’, and that this fact has to be accommodated

⁴⁴ The centrality of the assumption that each kind of moral judgment corresponds to a determinate attitude stems at least partly from its role in standard non-cognitivist accounts of disagreement in terms of disagreements in attitude. When Sturgeon (1991, 25) notes Blackburn’s (1991a) departure from the assumption, he takes the departure to be striking and costly. We have sought to explain how non-cognitivists can have a compelling account of moral disagreement even if they abandon the assumption.

within moral semantics and psychology in general, and in a plausible account of moral disagreement in particular.

In providing a non-cognitivist solution to the specification problem, we have simply assumed that non-cognitivism is an otherwise plausible metaethical position. However, our solution is adequate only if it fits well with central arguments for non-cognitivism, and with non-cognitivist answers to standard objections. We think that it does.

To begin with, it leaves intact plausible versions of the two central desiderata favouring non-cognitivism that were mentioned at the beginning of this paper. First, the core of our view is an account of how univocality and thus moral disagreement is possible despite wide psychological variance, including radical substantive disagreement. Second, our view takes paradigmatic wrongness-judgements to be intrinsically motivating, and gives this fact a central explanatory role. Further, it does this without committing itself to an implausibly strong version of the connection to motivation, which struggles to accommodate attitudinal variation.

One might worry that our view undermines plausible expressivist explanations of the logical and inferential properties of moral judgements and moral statements (the so-called ‘Frege-Geach problem’). Central to many such accounts is the idea that we take two moral judgements to be inconsistent because we take them to be constituted by practically inconsistent attitudes (see for example Blackburn 1988a; Schroeder 2008). Our view might seem to immediately rule out any such explanation, since classificatory intuitions do not require any particular attitude. However, it is compatible with an independently attractive way of thinking about the problem, as consisting in two tasks. The first is intra-personal: to explain how one person’s moral and non-moral judgements are logically or inferentially related. The second is inter-personal: to explain how to translate others’ wrongness-judgements into our own. The central Frege-Geach problem concerns the former task. Our view in effect proposes a sophisticated solution to the second task, explaining when another’s judgements are translatable to our own. Because of this, it is compatible with leading non-cognitivist strategies that address the first task.⁴⁵

⁴⁵ Obviously, purely non-cognitivist strategies can only be straightforwardly applied for wrongness judgments that are constituted by non-cognitive attitudes, but could be supplemented with ordinary cognitivist accounts. Hybrid accounts (see e.g. Ridge 2006b)

We wish to conclude by very briefly considering the implications of our view for moral cognitivism. While it is possible to develop certain kinds of relativistic cognitivist views out of the materials that we have provided, our discussion may suggest significant barriers to non-relativist versions of cognitivism.

At first blush, our account might seem to challenge all forms of cognitivism. This is because our functional account seems to violate the fundamental teleology of belief: our account of paradigmatic wrongness-judgements does not require that they track specific properties in the world, or be sensitive to determinate ‘moral facts’.

The *relativist* cognitivist has several potential ways of finessing this problem. First, our functionalist explanation rests heavily on the role that coordination on a social norm plays in explaining the various features of paradigmatic wrongness-judgements. A relativist cognitivist might seek to claim that such target norms can provide truth-conditions for wrongness-judgements. Since there are many possible social norms that we might have coordinated on, and since contingent facts will settle which ones various actual agents refer to, such an account suggests a deeply relativistic cognitivism. Alternatively, our psychological account might be adapted by a subjectivist theory, according to which the existence of certain wrongness-judgements as we have characterized them in an agent’s psychology play a central role in determining the truth-conditions for *her* statements of the form ‘ Φ ing is wrong’ (cf. Dreier 1990). Yet another way to incorporate cognitivist elements allows that moral utterances lack determinate cognitive contents—no such contents are part of *what is said* by these utterances—but insists that *acts of moral judgement* nevertheless aim to determine whether the object of judgement satisfies some standard or is conducive to some goal endorsed by the moral judge and thus that the resulting judgements can be correct or incorrect depending on whether this is indeed the case.

might apply to all cases, as would accounts that are neutral between cognitivism and non-cognitivism (see e.g. Björnsson 2001).

For the strategy of reducing the problem of logical relations between judges to the first-person case via translation affected by pragmatic considerations, see Wong (1984, p. 73) and Björnsson and Finlay 2010.

Such relativist cognitivist amendments to our view may or may not be successful. Perhaps wrongness-judgments are too diverse and some wrongness-judges too flexible in the grounds they accept to be subject to any plausible story of content determination. Assessing the coherence and defensibility of these proposals would require significant further discussion of relevant semantic issues.

We take the challenge posed to the non-relativist cognitivist to be more serious, however. We argued in section 3 that the specification problem is a problem for non-cognitivist and cognitivist accounts alike. In sections 4 and 5, we sketched and defended what we take to be the most plausible non-cognitivist solution. It is natural to wonder whether the non-relativist cognitivist can offer some equally promising solution. Our discussion suggests that this will not be easy. Our explanation of the accommodation of non-paradigmatic wrongness-judgements leaned heavily on the characteristic practical function of paradigmatic wrongness-judgements. By contrast, a non-relativistic cognitivist account of wrongness-judgements would seem to necessarily take our practices of attributing such judgements to be fundamentally regulated by epistemic goals. It is unclear how such an account could give us reason to ignore differences in cognitive content, as moral twin earth cases suggest that we robustly do.

It is worth concluding by reflecting on the broader dialectical significance of our discussion. If Merli's argument adumbrated in section 2 is correct, non-cognitivists' standard appeal to moral twin earth thought experiments is unstable: *no one* can offer a twin-earth-immune answer to the specification problem in terms of necessary and sufficient cognitive or non-cognitive properties of moral judgements. Our argument suggests that the twin-earth dialectic can be reframed: if we are right, (one version of) non-cognitivism has a plausible explanation of the complexity of moral twin-earth intuitions. Unless the realist can offer an equally attractive account, then, the specification problem seems, surprisingly, to support rather than undermine non-cognitivism.⁴⁶

⁴⁶ For helpful comments on earlier versions of the paper, we are indebted to David Plunkett, Wlodek Rabinowicz, audiences at Macquarie University, Macalester College, Lund University, the 2011 MMER workshop at University of Gothenburg, the Second Annual Dutch Conference on Practical Philosophy, Groningen, and three anonymous referees.

References

- Anscombe, G. E. M. 1957: *Intention*. Oxford: Blackwell.
- Aureli, Filippo, Marina Cords, and Carel P. van Schaik 2002: 'Conflict Resolution Following Aggression in Gregarious Animals: A Predictive Framework'. *Animal Behaviour*, 64, pp. 325–43.
- Bedke, Matt 2009: 'Moral Judgement Purposivism: Saving Internalism from Amoralism'. *Philosophical Studies*, 144, pp. 189–209
- Bennett, Christopher 2002: 'The Varieties of Retributive Experience'. *The Philosophical Quarterly*, 52, 145–63.
- Björnsson, Gunnar 2001: 'Why Emotivists Love Inconsistency'. *Philosophical Studies*, 104, pp. 81–108
- 2002: 'How Emotivism Survives Immoralists, Irrationality, and Depression'. *The Southern Journal of Philosophy*, 40, pp. 327–44.
- Björnsson, Gunnar and Stephen Finlay, Stephen 2010: 'Metaethical Contextualism Defended.'. *Ethics*, 121.1, pp.: 7–36.
- Blackburn, Simon 1988a: Attitudes and Contents. In his 1993, pp. 182–197. Originally published in *Ethics*, 98.
- 1988b: 'How To Be an Ethical Anti-Realist.' In his 1993, pp. 166–81. Originally published in *Midwest Studies in Philosophy*, 12.
- 1991a: 'Just Causes'. *Philosophical Studies*, 61, pp. 3–17.
- 1991b: 'Reply to Sturgeon'. *Philosophical Studies*, 61, pp. 39–42.
- 1993: *Essays in Quasi-Realism*. Oxford: Oxford University Press.
- 1998: *Ruling Passions*. Oxford: Oxford University Press. Boisvert, Daniel R. 2008: 'Expressive-Assertivism'. *Pacific Philosophical Quarterly*, 89, pp. 169–203.
- Brink, David O. 1989: *Moral Realism and the Foundations of Ethics*. Cambridge: Cambridge University Press.
- Boyd, Richard 1988: 'How to be a Moral Realist,' In Sayre-McCord 1988, pp. 181–228.

Björnsson's work on this paper was supported by the Swedish Research Council (grant number 2009–1517); McPherson's work was supported by a fellowship with the Many Moral Rationalisms research project, funded by the Australian Research Council.

- Carruthers, Peter 1992: *The Animals Issue*. Cambridge: Cambridge University Press.
- Clutton-Brock, T. H. and G. A. Parker 1995: 'Punishment in Animal Societies'. *Nature*, 373, pp. 209–16.
- Copp, David. 2000: 'Milk, Honey, and the Good Life on Moral Twin Earth'. *Synthese*, 124, pp. 113–137.
- (ed) 2006: *The Oxford Handbook of Ethical Theory*. Oxford: Oxford University Press.
- Cushman, Fiery 2008: 'Crime and Punishment: Distinguishing the Roles of Causal and Intentional Analyses in Moral Judgement'. *Cognition*, 108, pp. 353–80.
- D'Arms, Justin and Daniel Jacobson 1994: 'Expressivism, Morality, and the Emotions'. *Ethics*, 104, pp. 739–63.
- Dancy, Jonathan 1993: *Moral Reasons*. Oxford: Blackwell.
- Dreier, James 1990: 'Internalism and Speaker Relativism'. *Ethics*, 101, pp. 6–26.
- 2006: 'Moral Relativism and Moral Nihilism'. In Copp 2006, pp. 240–264.
- Foot, Philippa 1958: 'Moral Beliefs'. *Proceedings of the Aristotelian Society*, 59, pp. 83–104.
- Francén, Ragnar 2010: 'Moral Motivation Pluralism'. *The Journal of Ethics*, 14, pp. 117–48.
- Gert, Joshua 2002: 'Expressivism and Language Learning'. *Ethics*, 112, pp. 292–314.
- Gibbard, Allan 1990: *Wise Choices, Apt Feelings*. Cambridge, MA: Harvard University Press.
- 1992: 'Moral Concepts: Substance and Sentiment'. *Philosophical Perspectives*, 6, pp. 199–221.
- 1993: 'Reply to Sinnott-Armstrong'. *Philosophical Studies*, 69, pp. 315–27
- 2003: *Thinking how to Live*. Cambridge, MA: Harvard University Press.
- 2006: 'Moral Feelings and Moral Concepts.' In Shafer-Landau 2006, pp. 195–215.
- Glover, Jonathan 1975: 'It Makes no Difference Whether or Not I Do It'. *Proceedings of the Aristotelian Society, Supplementary Volumes*, 49, pp. 171–90.
- Haidt, Jonathan 2001: 'The emotional dog and its rational tail: A social intuitionist approach to moral judgement'. *Psychological Review*, 108, pp. 814–34.

- Hare, Richard. M. 1952: *Language of Morals*. Oxford: Oxford University Press.
- 1981: *Moral Thinking*. Oxford: Oxford University Press.
- Harman, Gilbert 1983: 'Human Flourishing, Ethics, and Liberty.' *Philosophy and Public Affairs* 12: 307–22.
- 2009: 'Guilt-Free Morality': In Shafer-Landau 2009, pp. 203–214.
- Heil, John (ed) 1991: *Rationality, Morality, and Self-Interest*. Lanman Md: Rowman and Littlefield.
- Holland, Sean 2001: 'Dispositional Theories of Value Meet Moral Twin Earth'. *American Philosophical Quarterly*, 38, 177–95.
- Horgan, Terence and Mark Timmons 1991 'New-Wave Moral Realism Meets Moral Twin Earth.' In Heil 1991, pp. 115–33.
- 1992a: 'Troubles for New Wave Moral Semantics: The Open Question Argument Revived'. *Philosophical Papers*, 21, pp. 153–75.
- 1992b: 'Troubles on Moral Twin Earth: Moral Queerness Revived'. *Synthese*, 92, pp. 221–60.
- 1996: 'Troubles for Michael Smith's Metaethical Rationalism'. *Philosophical Papers*, 25, 203–31.
- 2000: 'Copping Out on Moral Twin Earth'. *Synthese*, 124, pp. 129–152.
- (eds) 2006a: *Metaethics after Moore*. Oxford: Oxford University Press.
- 2006b: 'Cognitivist Expressivism'. In Horgan and Timmons (eds) 2006a, pp. 255–298.
- 2009: 'Analytical Moral Functionalism Meets Moral Twin Earth'. In Ravenscroft (ed) 2009, pp. 221–36.
- Humberstone, I. L. 1992: 'Direction of Fit'. *Mind*, 101, pp. 59–84.
- Jenkins, Carrie and Daniel Nolan 2010: 'Maximising, Satisficing and Context', *Noûs*, 44, pp. 451–68.
- Joyce, Richard 2006: *The Evolution of Morality*. Cambridge, MA: MIT Press.
- Kauppinen, Antti 2010: 'What Makes a Sentiment Moral?' In Shafer-Landau 2010, pp. 225–256.
- Lenman, James 1999: 'The externalist and the amoralist'. *Philosophia*, 27, pp. 441–57.
- Merli, David 2002: 'Return to Moral Twin Earth'. *Canadian Journal of Philosophy*, 32, pp. 207–40.
- 2008: 'Expressivism and the Limits of Moral Disagreement'. *Journal of Ethics*, 12, pp. 25–55.

- Miller, Alexander 2003: *Introduction to Contemporary Metaethics*. Cambridge: Polity.
- Nichols, Shaun 2004: *Sentimental Rules*. New York: Oxford University Press.
- Parfit, Derek 1984: *Reasons and Persons*. Oxford: Oxford University Press.
- Prinz, Jesse 2007: *The Emotional Construction of Morals*. Oxford: Oxford University Press.
- Rabinowicz, Wolodek and Toni Rønnow-Rasmussen 2004: 'The Strike of the Demon: on Fitting Pro-attitudes and Value'. *Ethics*, 114, pp. 391–423.
- Ravenscroft, Ian (ed) 2009: *Minds, Ethics, and Conditionals: Themes from the Philosophy of Frank Jackson*. Oxford: Oxford University Press.
- Regan, Tom 1983: *The Case for Animal Rights*. Berkeley: University of California Press.
- Ridge, Michael 2006a: 'Saving the Ethical Appearances'. *Mind*, 115, pp. 633–49.
- 2006b: 'Ecumenical Expressivism: Finessing Frege'. *Ethics*, 116, 302–36.
- Ross, W. David 1939: *Foundations of Ethics*. Oxford: Oxford University Press.
- Sandis, Constantine (ed) 2009: *New Essays on the Explanation of Action*. Houndsmills: Palgrave Macmillan.
- Sayre-McCord, Geoff 1988: *Essays on Moral Realism*. Cornell: Cornell University Press.
- Scanlon, Thomas 1998: *What We Owe to Each Other*. Cambridge, MA: Harvard University Press.
- 2008: *Moral Dimensions: Permissibility, Meaning, Blame*. Cambridge, MA: Harvard University Press.
- Schroeder, Mark 2008: *Being For*. Oxford: Clarendon.
- 2009: 'Hybrid Expressivism: Virtues and Vices'. *Ethics*, 119, pp. 257–309.
- 2010: 'Value and the Right Kind of Reason'. In Shafer-Landau 2010, pp. 25–56.
- Shafer-Landau, Russ (ed) 2006: *Oxford Studies in Metaethics* Vol. 1. New York: Oxford University Press.
- (ed) 2009: *Oxford Studies in Metaethics* Vol. 4. New York: Oxford University Press.
- (ed) 2010: *Oxford Studies in Metaethics* Vol. 5. New York: Oxford University Press.

- Singer, Peter 1977: *Animal Liberation*. New York: Avon.
- Sinnott–Armstrong, Walter (ed) 2007: *Moral Psychology, Volume 1: The Evolution of Morality: Adaptations and Innateness*. Cambridge, MA: MIT Press.
- (ed) 2008a: *Moral Psychology, Volume 2: The Cognitive Science of Morality: Intuition and Diversity*. Cambridge, MA: MIT Press.
- (ed) 2008b: (ed) *Moral Psychology, Volume 3: The Neuroscience of Morality: Emotion, Brain Disorders, and Development*. Cambridge, MA: MIT Press.
- Smith, Michael 1994: *The Moral Problem*. Oxford: Blackwell.
- Stevenson, Charles S. 1963: *Facts and Values*. New Haven: Yale University Press.
- Stocker, Michael 1979: ‘Desiring the bad: An essay in moral psychology’. *Journal of Philosophy*, 76, pp. 738–53.
- Sturgeon, Nicholas 1991: ‘Contents and Causes: A Reply to Blackburn’. *Philosophical Studies*, 61, pp. 19–37.
- Svavarsdottir, Sigrun 1999: ‘Moral Cognitivism and Motivation’. *The Philosophical Review*, 108, pp. 161–219.
- Tersman, Folke 2006: *Moral Disagreement*. Cambridge: Cambridge University Press.
- Tresan, Jon 2006: ‘De Dicto Internalist Cognitivism’. *Noûs*, 40, pp. 143–65.
- 2009: ‘The challenge of communal internalism’. *The Journal of Value Inquiry*, 43, pp. 179–99.
- Velleman, J. David 1992: ‘The Guise of the Good’. *Noûs*, 26, pp. 3–26.
- Wong, David 1984: *Moral Relativity*. Berkeley: University of California Press.
- Zangwill, Nick 2008: ‘The Indifference Argument’. *Philosophical Studies*, 138, pp. 91–124.
- 2009: ‘Non–Cognitivism and Motivation’. In Sandis 2009, pp. 416–24.
- Zimmerman, Michael J. 2008: *Living with Uncertainty*. Cambridge: Cambridge University Press.