The Empirical Case for Folk Indexical Moral Relativism

James R. Beebe (University at Buffalo)

Forthcoming in *Oxford Studies in Experimental Philosophy*, vol. 4

Apr. 15, 2020

Abstract

Recent empirical work on folk moral objectivism has attempted to examine the extent to which folk morality presumes that moral judgments are objectively true or false. Some researchers report findings that they take to indicate folk commitment to objectivism (Nichols & Folds-Bennett, 2003; Wainryb et al., 2004; Goodwin & Darley, 2008, 2010, 2012), while others report findings that may reveal a more variable commitment to objectivism (Sarkissian, et al., 2011; Wright, Grandjean, & McWhite, 2013; Wright, McWhite, & Grandjean, 2014; Beebe, 2014; Beebe et al., 2015; Beebe & Sackris, 2016; Wright, 2018). However, the various probes that have been used to examine folk moral objectivism almost always fail to be good direct measures of objectivism. Some critics (Beebe, 2015; Pölzler, 2017, 2018) have suggested that the problems with existing probes are serious enough that they should be viewed as largely incapable of shedding any light on folk metaethical commitments. Building upon the work of Justin Khoo and Joshua Knobe (2018), I argue that many of the existing probes can be seen as good measures of the extent to which people think that the truth of one moral judgment excludes the possibility that a judgment made by a disagreeing party is also true and that the best explanation of the findings obtained using these measures is significant folk support for indexical moral relativism—the view that the content of moral judgments is context-sensitive. If my thesis is correct, most contemporary moral philosophers are deeply mistaken about the metaethical contours of folk morality in one very important respect.

Keywords: folk morality, metaethics, moral realism, moral objectivism, moral relativism, moral psychology

As is well known, most contemporary Anglophone moral philosophers think that folk morality is committed to moral objectivism, the view that moral judgments are objectively true or false—i.e., that they are truth-apt ascriptions of moral properties whose truth values do not depend upon the attitudes or opinions of any person or set of persons (cf. Mackie, 1977; Dancy, 1986; Brink, 1989; Smith, 1994a; Darwall, 1998; Shafer-Landau, 2003; Huemer, 2005). Empirical research conducted during the last decade and a half has attempted to examine folk commitment to moral objectivism using different variations on what I will call 'the disagreement paradigm,' in which participants are asked to reflect on situations in which two people disagree in the moral judgments they make regarding some action and then to make metaethical judgments about the judgments of the disagreeing parties. Following Khoo and Knobe (2018, p. 111), I use 'disagree' in the activity sense of the term, where two speakers disagree when they are "engaging in the activity of disagreeing with one another" and where "they can appropriately reject each other's claims by saying 'No.'"

The central kind of judgment that researchers who employ the disagreement paradigm have wanted participants to make is whether there is an objective fact of the matter about which (if any) of the moral judgments expressed in a particular moral disagreement are correct. However, as I explain below, the vast majority of the probes used in this area of research never ask participants whether the moral judgments in question are objectively or mind-independently true or false. Attempts to measure folk moral objectivism have a significant tendency to fall short when no attempt is made to elicit judgments about the objectivity of moral judgments. Nevertheless, extending an interpretation of the disagreement paradigm articulated by Khoo and Knobe (2018), I argue that the findings uncovered with the use of these probes can be seen as shedding light on

folk moral objectivism after all when they are understood as revealing patterns in folk judgments about the extent to which the truth of one moral judgment excludes the possibility that a moral judgment made by a disagreeing party is also true. My thesis is that the observed findings reveal that ordinary individuals treat the truth conditions of moral judgments as having much more variability than philosophers have assumed and that this variability is best explained by a broad folk commitment to indexical moral relativism—the view that terms of moral evaluation such as 'right,' 'wrong,' 'good,' 'bad,' 'permissible,' and 'impermissible' express different contents on different occasions of use (Beebe, 2010).

To illustrate the kinds of judgment variability that will be important in what follows, consider the following pairs of sentences (due in part to Sarkissian et al., 2011, p. 482):

(1.1)　The number 2,377 is prime.

(1.2)　The number 2,377 is not prime.

(2.1)　I am an epistemologist.

(2.2)　I am not an epistemologist.

(3.1)　January is a winter month.

(3.2)　January is not a winter month.

It is a matter of necessity that if (1.1) is true, then (1.2) is false; and if (1.2) is true, then (1.1) is false. These sentences have what Khoo and Knobe (2018) call 'exclusionary content'—their contents exclude the possibility of both of them being true.

Things are more complicated with the other two pairs of sentences. Because sentences (2.1) and (2.2) contain the first-person indexical 'I,' the content expressed by utterances of these sentences will differ depending upon who the speaker is. When (2.1) is assertively uttered by me, it means that I am an epistemologist, but when it is assertively uttered by you, it means that you

are. In my case, the content expressed by this assertion is true, whereas the content expressed by your assertion might not be. Thus, if I assert (2.1) and you assert (2.2), we might both succeed in saying something true. And even if we both say something true by assertively uttering (2.1), your words will express a completely different proposition from the one mine express. The contents of (2.1) and (2.2) thus vary across contexts of use. For simplicity, in what follows I will call this phenomenon 'context sensitivity.' Utterances of (2.1) and (2.2) will have exclusionary content only when the context-sensitive terms they contain express the same contents. That is, it will be the case that if (2.1) is true, then (2.2) is false, and vice versa, only when the 'I' in each of them denotes the same person.

There is also a kind of variability involved in sentences (3.1) and (3.2), but it does not concern variation in content. Rather, it involves variation in the set of circumstances at which the truth values of these assertions should be evaluated (cf. Kaplan, 1989). If (3.1) and (3.2) are both uttered in the same climate in the northern or southern hemisphere, at most one of them will be true. However, if they are each uttered in a different hemisphere, they might both be true (or both false). Importantly, the predicate 'is a winter month' expresses the same content in every hemisphere (denoting the coldest season of the year in a polar or temperate zone), and the same is obviously true for 'January' as well. So, what varies is not the semantic values of these terms. What varies is the set of local circumstances that determine whether the proposition in question is true. Following John MacFarlane (2005a, 2005b, 2007, 2009, 2014), I will call this phenomenon 'assessment sensitivity.'[1]

---

[1] It is widely accepted that the truth value of a proposition should be evaluated at a world and a time (Kaplan, 1989). Thus, the proposition that I am an epistemologist is true at the present time in the actual world, but it was not true when I was a much younger person, and it is not true in possible worlds where I (perhaps wisely) chose a musical career instead of an academic career. The proposition that the number 2,377 is prime is true at all times in all possible worlds. Statements that are assessment-sensitive in MacFarlane's sense will thus be those that require the consideration of factors in addition to world and time for the determination of their truth values.

A generalization of Khoo and Knobe's (2018) notion of exclusionary content is needed to capture the ways in which assessment sensitivity either permits or rules out the possibility that the judgments of two disagreeing parties can both be true. In their original formulation, Khoo and Knobe (2018, p. 109) defined the notion as follows:

(4)   Two claims are *exclusionary* (or have exclusionary content) iff it has to be the case that at least one of them is false.

The definiens suggests a wider scope for 'exclusionary' than what is implied by subsequent uses Khoo and Knobe make of the term. Throughout their paper, they claim that the judgments of two disagreeing parties are exclusionary when there is something about the content of these judgments that prevents them from both being true. This shows that they are primarily thinking about the question of context sensitivity. However, the phenomenon of assessment sensitivity reveals that whether or not it is possible for the judgments of two disagreeing parties to both be true may have nothing to do with the contents of these judgments but may instead concern the set of circumstances at which the truth values of these assertions are to be evaluated—what MacFarlane (2005a, 2005b, 2007, 2009, 2014) calls the 'context of assessment.' Thus, in what follows I will be operating with a conception of 'exclusionary' that is neutral as to why the judgments of two disagreeing parties cannot both be true. Whenever someone judges a pair of moral judgments to be exclusionary in this sense, let us say that they are making an exclusionary judgment. From this perspective, uses of (1.1) and (1.2) will always be exclusionary, uses of (2.1) and (2.2) will be exclusionary when the context-sensitive terms they contain express the same contents, and uses of (3.1) and (3.2) will be exclusionary whenever they are evaluated from the same context of assessment.

The question for metaethics is whether moral judgments are context-sensitive, assessment-sensitive, or neither. The traditional view of folk morality is that ordinary moral judgments are neither context-sensitive nor assessment-sensitive (e.g., Mackie, 1977; Dancy, 1986; Brink, 1989; Smith, 1994a; Darwall, 1998; Shafer-Landau, 2003; Huemer, 2005). Call such a view 'invariantism.' The view that moral judgments are context-sensitive—like sentences (2.1) and (2.2)—is often referred to as 'indexical moral relativism' (hereafter 'IMR') (e.g., Beebe, 2010; Björnsson & Finlay, 2010; Khoo & Knobe, 2018).[2] The view that moral terms are assessment-sensitive like sentences (3.1) and (3.2) is sometimes called 'truth value relativism.'[3]

I will argue that the patterns of folk judgments reported in the empirical literature on folk moral objectivism are best explained by a broad folk commitment to the context sensitivity of moral terms. In Section II, I examine recent attempts to study folk moral objectivism, explaining important difficulties with the probes that have been used in this area of research and showing how the theoretical perspective articulated by Khoo and Knobe (2018) points toward a reinterpretation of much of this research as actually measuring folk exclusionary judgments. In Section III, I argue that the distribution of the reinterpreted data is best explained by a semantic theory on which moral terms such as 'right,' 'wrong,' 'good,' 'bad,' etc. are context-sensitive and that invariantist, subjectivist, noncognitivist, or truth-relativist perspectives fall short. I conclude by briefly considering the implications of my thesis for the wide consensus within philosophy concerning folk metaethics.

---

[2] Whether Harman (1975) should be included in this camp is not a simple or straightforward matter.
[3] MacFarlane (2005a; 2005b; 2007; 2009; 2014) has done the most work to develop the idea that truth is assessment-sensitive, although he has given less attention to moral relativism in particular.

II.

The construct validity of an empirical measure concerns the extent to which it actually measures what it purports to measure. In this section, I explain how most of the measures constructed to study folk moral objectivism fail to be good direct measures of folk objectivism but can often be viewed as decent measures of folk exclusionary judgments.

In one of the first uses of the disagreement paradigm to study folk metaethics, Cecilia Wainryb and her collaborators (Wainryb et al., 2004) had five-, seven-, and nine-year-old children consider two situations of moral disagreement, one of which was the following:

> Do you think it is okay or not okay to hit and kick other children? Sarah believes that it's okay to hit and kick other children, and Sophie believes that it's wrong to hit and kick other children. (p. 691)

A second scenario asked whether it is okay to break other children's toys. The participants were then given the following task:

> (5)   Do you think that only one belief [what Sophie believes] is right, or do you think that both beliefs [what both Sophie and Sarah believe] are right?

At several points, Wainryb et al. (2004, p. 687, et passim) explain that they interpret the 'both beliefs are right' answer to represent moral relativism and the 'only one belief is right' answer to represent the rejection of relativism. They also interpret task (5) to represent a choice between objectivism and subjectivism, taking the choice between relativism and nonrelativism to be equivalent to that between subjectivism and objectivism (Wainryb et al. 2004, pp. 689, 695). However, task (5) fails to be a good test for participants' intuitions about either moral relativism or moral objectivism because it is silent about whether the factors that might make only one belief right are culturally variable or mind-independent ones.

Consider how a committed moral relativist would respond to Wainryb et al.'s (2004) task (5). Cultural moral relativism is perhaps the most commonly discussed alternative to moral objectivism. On this view, there is not a single set of moral principles that apply to all people at all times. Rather, moral judgments are made true by the variable values, norms, and practices of different cultures at different times and places. Since Sarah and Sophie are not represented as belonging to different cultures, the default expectation is that they most likely inhabit the same culture. Yet if this is true, then according to relativism, only one of their beliefs will be right—the one that accords with the ethos of their shared culture. And this will be true for nonobjectivist reasons—viz., because the normative attitudes and practices of their culture have made it so. Thus, the 'only one belief is right' answer cannot represent the rejection of relativism when it is in fact what relativism says is the correct answer. If Wainryb et al.'s task (5) cannot represent a basic distinction between moral objectivism and its most common alternative, it cannot be a good measure of moral objectivism.

The weaknesses of Wainryb et al.'s (2004) task (5) can also be appreciated by considering how it would apply to disagreements about conventional matters. Conventional judgments are paradigmatic instances of judgments that are not objective in the way that moral judgments are alleged to be. This is why the Turiel school (e.g., Turiel, 1983) selected them to be the primary foil for moral judgments in their well-known investigation of the moral/conventional distinction. Thus, if Sophie believes that people should drive on the right-hand side of the road, and Sarah believes that people should drive on the left-hand side, what is the correct response to Wainryb et al.'s question 'Do you think that only one belief [what Sophie believes] is right, or do you think that both beliefs [what both Sophie and Sarah believe] are right?' If Sophie and Sarah both live in

the same society (e.g., the U.S.), it is obvious that only one of their beliefs is right.[4] Anyone in the

U.S. who thinks that drivers should drive on the left-hand side is mistaken. If the discussion were

to occur in the U.K., it would still be true that there was only one correct answer to the question

and that at least one of the disagreeing parties must be mistaken. Yet if the 'only one belief is right'

answer choice represents the correct judgment to make about conventional disagreements, it

cannot be a measure of objective truth because conventional judgments are paradigmatic instances

of judgments that can be true without being objectively true. Wainryb et al.'s (2004) task (5) thus

fails to be a good measure of objectivism.

Even though Wainryb et al.'s (2004) task falls short as a measure of objectivism, Khoo and

Knobe (2018) argue that tasks such as this can be interpreted as measures of whether people think

the moral judgments of disagreeing parties are exclusionary and that this can shed light on folk

metaethics. Thus, while (5) does not directly ask about the objectivity of Sophie's or Sarah's

judgment, it does ask about their capacity of their judgments to both be true. From this perspective,

we can interpret the findings of Wainryb et al. (2004) as telling us that five- to nine-year-old

children predominantly thought that the judgments of parties who disagreed about hitting or

kicking others or breaking their toys were exclusionary.

Most other uses of the disagreement paradigm to measure folk moral objectivism suffer

from the same problems as Wainryb et al. (2004)—the answer choices that are supposed to indicate

a commitment to objectivism are answers that committed relativists would choose and that would

uncontroversially apply to disagreements about merely conventional matters. However, most of

these other measures can also be reinterpreted as respectable measures of folk exclusionary

---

[4] The same point has been made by Ayars and Nichols (forthcoming, n. 4), who write, "Even for some cases of conventions, like which side of the road to drive on, it would be natural to say that if two people disagree, one of them must be wrong."

judgments. Shaun Nichols (2004a, pp. 7-8), for example, presented university undergraduates with the following task:

> (6.1)   John and Fred are members of different cultures, and they are in an argument. John says, "It's okay to hit people just because you feel like it," and Fred says, "No, it is not okay to hit people just because you feel like it." John then says, "Look you are wrong. Everyone I know agrees that it's okay to do that." Fred responds, "Oh no, you are the one who is mistaken. Everyone I know agrees that it's not okay to do that."
>
> Which of the following do you think best characterizes their views? (Check one and give a brief justification for your answer.)
>
> (a) It is okay to hit people just because you feel like it, so John is right and Fred is wrong.
>
> (b) It is not okay to hit people just because you feel like it, so Fred is right and John is wrong.
>
> (c) There is no fact of the matter about unqualified claims like "It's okay to hit people just because you feel like it." Different cultures believe different things, and it is not absolutely true or false that it's okay to hit people just because you feel like it.[5]

Nichols (2004a, p. 8) interpreted answer choices (a) and (b) to be objectivist and (c) to be nonobjectivist. Answer choices (a) and (b) say nothing about whether the first-order moral judgments contained within them are correct for objective reasons or for more subjective or culturally variable reasons. Presumably, the fact that answer (c) attempts to articulate some kind

---

[5] Nichols (2004a, pp. 15-16) also used the same format to ask participants about general moral rules, such as "It's wrong for anyone to torture puppies for the fun of it."

of non-factualist view is supposed to make (a) and (b) seem more objective answer choices. Yet it could be that it is not okay to hit people just because you feel like it because the values, norms, and practices of our society happen to have made it so—and not because it is an objective, mind-independent moral fact that this sort of action is wrong.

Despite the fact that Nichols' task (6.1) fails to offer a clear choice between objectivism and its contraries, it does seem to present a relatively straightforward contrast between an exclusionary perspective on John and Fred's disagreement in (a) and (b) and a non-exclusionary one in (c). We can thus interpret Nichols' (2004a) findings as telling us that adult participants made exclusionary judgments most of the time when the disagreements concerned hitting people just because you feel like it or torturing puppies for the fun of it but that they made non-exclusionary judgments a good deal of the time even in these cases.

In another study reported in the same article, Nichols (2004a, p. 19) presented undergraduate participants with the following, slightly different task:

(6.2)   John and Ted are members of different cultures, and they are in an argument about some of the scenarios described above. John says, "It was okay for Frank to hit Ben and for Lisa to shove Nancy," and Ted says, "No, those things were not okay." John then says, "Look you are wrong. Everyone I know agrees that those kinds of things are okay." Ted responds, "Oh no, you are the one who is mistaken. Everyone I know agrees that those kinds of actions are not okay."

Which of the following do you think best characterizes their views? (Check one and give a brief justification for your answer.)

(a)   It is an objective fact, independent of what different people think, that it was not wrong for Frank to hit Ben and for Lisa to shove Nancy. So John is right and Ted is wrong.

(b)   It is an objective fact, independent of what different people think, that it was wrong for Frank to hit Ben or for Lisa to shove Nancy. So Ted is right and John is wrong.

(c)   There is no objective fact, independent of what different people think, about whether it was wrong for Frank to hit Bill or Lisa to shove Nancy. These actions were "wrong for Ted" and maybe "wrong for me," but they aren't *objectively wrong* independent of what people think about them.

Task (6.2) offers a fairly clear choice between the view that moral judgments are mind-independently true or false and the view that their truth values are determined by shared cultural norms and practices. John and Ted are represented as being from different cultures, and each appeals to the consensus of their cultural peers to back up their moral claims. The issue is clouded somewhat by the fact that the task also simultaneously asks participants to make exclusionary or non-exclusionary judgments about the moral judgments of the disagreeing parties. In any case, Nichols (2004a) only applied task (6.2) to one moral issue and failed to report how many of his twenty-two participants selected one of the first two answer choices versus how many selected the third.[6]

---

[6] In another important early use of the disagreement paradigm, Nichols and Folds-Bennett (2003, p. B25) argued that research by the Turiel school on folk acceptance of the widely discussed moral/conventional distinction fails to indicate folk commitment to moral objectivism because the folk might distinguish between moral and conventional norms and yet view moral properties as 'response dependent'—i.e., as "constituted by the responses it elicits in a population" (cf. Johnston, 1989; Pettit, 1991; Wedgwood, 1998). Because a response dependent view of morality denies that the truth values of moral judgments are determined in a mind-independent fashion, such a view would not be fully objectivist. Reasoning that folk rejection of response dependence would be evidence for folk moral objectivism, Nichols and Folds-Bennett (as reported in Nichols 2004b, p. 174) gave four- to six-year-old children the following tasks:

In a paper that inspired a good deal of subsequent work on folk metaethics using the disagreement paradigm, Geoffrey Goodwin and John Darley (2008, pp. 1343-4) asked participants whether each of a number of moral judgments was a 'true statement,' a 'false statement,' or 'an opinion or attitude.' They interpreted the first two answer choices as objectivist and the third as nonobjectivist. In a second study, Goodwin and Darley (2008) asked participants 'According to you, can there be a correct answer as to whether this [a moral judgment] is true? Yes/No.' They treated 'Yes' as objectivist and 'No' as nonobjectivist. However, moral judgments can be true for objectivist reasons, relativistic reasons, or conventional ones. Questions that merely ask whether certain judgment are true or false are incapable of getting at the issue of whether those judgments are objectively or mind-independently true or false.[7] Importantly, and unlike the other tasks described above, the probes used by Goodwin and Darley (2008) also fail to ask participants whether the judgments of disagreeing parties are exclusionary.

(6.3)   Now, think about a long time ago, before there were any people. There were monkeys back then too. Way back then, before there were people, when one monkey helped another monkey that got hurt, was that good?

(6.4)   You know, I think it was good for the monkey to help the other monkey. Some people don't like it when monkeys help each other when they're hurt. They don't think it's good when monkeys do that. Would you say that when one monkey helps a hurt monkey that is good *for some people* or that it's good *for real*?

The idea behind the disagreement paradigm task (6.4) is that if an action is 'good *for real*,' it is objectively good and not merely so in the eyes of some people. However, defenders of the response dependence of morality agree that some actions are good for real. It is just that they believe the real goodness of these actions is grounded in certain types of human responses to those actions. Moreover, the kinds of responses to which response dependence theorists standardly appeal are those that are universal rather than crossculturally variable. Thus, participants who reject the 'good *for some people*' option cannot easily be interpreted as rejecting response dependence. Task (6.4) thus does not seem to do a good job of probing participants' intuitions about the response independence of moral judgments. Task (6.4) seems to be headed in the right direction when it comes to characterizing the important metaethical distinction between moral judgments being objectively or mind-independently true and their being true for more subjective or cultural reasons. Applying task (6.4) to four predominantly unspecified moral issues, Nichols and Folds-Bennett (2003) found that children gave predominantly objectivist responses to them. Nichols and Folds-Bennett (2003) do not provide any details about the particular moral issues they included in their studies. A few details about the case of monkeys helping appear in Nichols (2004b), but no details about the other three cases appear there. For our purposes, one of the most important features of tasks (6.3) and (6.4)—and one that distinguishes them from other tasks described in this section—is that they do not shed any light on participants' exclusionary judgments. This is in part a function of the fact that the moral disagreement featured in (6.4) is not an essential part of the question being asked.

[7] Such questions may do a reasonable job of probing participants' judgments about the truth-aptness of moral judgments, but this is not what Goodwin and Darley (2008) were attempting to study.

Goodwin and Darley (2008, p. 1344) also asked participants to complete the following task:

(7) You circled ___ on the scale (1–6) which means that you strongly agreed/disagreed with this statement. A person who we tested, strongly agrees/disagrees with this statement, which means that he or she sharply disagrees with you. What would you conclude about this disagreement? We are interested in what you would privately think about this—the question is **not** about what you would be willing to say to this other person. Please circle the [appropriate answer.]

(a) The other person is surely mistaken.

(b) It is possible that neither you nor the other person is mistaken.

(c) It could be that you are mistaken, and the other person is correct.

(d) Other.

Goodwin and Darley (2008) take (a) and (c) to represent objectivism and (b) to represent subjectivism.[8] However, for reasons that are by now familiar, because the would-be objectivist answers in these tasks say nothing about whether one moral judgment is correct and another is incorrect because of objective factors or because of more subjective or culturally variable ones, these tasks fail to be good direct measures of folk moral objectivism. Nevertheless, these tasks can be viewed as reasonably adequate measures of exclusionary judgments, since they attempt to elicit intuitions about whether the truth of one moral judgment excludes the possibility that the judgment of a disagreeing party is also true. Unfortunately, Goodwin and Darley (2008) do not report how

---

[8] In a subsequent study reported in the same paper, Goodwin and Darley (2008, p. 1356) asked participants "what they would conclude about someone who hypothetically disagreed with them on a scale that ranged from 'neither of us need be mistaken' to 'other person is clearly mistaken.'"

participants responded to these tasks in their study but instead combine participants' responses to them with their responses to other tasks that measure neither objectivism nor judgment exclusion.

More recent investigations of folk moral objectivism using the disagreement paradigm typically feature simpler probes that fail to be good direct measures of moral objectivism for the kinds given above but that offer clearer measures of folk exclusionary judgments. Hagop Sarkissian, John Park, David Tien, Jennifer Cole Wright, and Knobe (2011), for example, gave undergraduate participants the following task after reading about two people (one of whom is said to be a classmate) who disagreed about the morality of a particular action:

> (8.1) Given that these individuals have different judgments about this case, we would like to know whether you think at least one of them must be wrong, or whether you think both of them could actually be correct. In other words, to what extent would you agree or disagree with the following statement concerning such a case: Since your classmate and [the other disagreeing party] have different judgments about this case, at least one of them must be wrong.

Participants were asked to indicate their level of agreement with the final statement on seven point scale ranging from agree to disagree. Sarkissian et al. (2011, pp. 487-8) varied the identity of one of the disagreeing parties, making them either "a fairly ordinary student at their own college," a member of "a traditional warrior culture" in the Amazon rainforest, or an alien who does not experience love, friendship, pleasure, or pain. Sarkissian et al. (2011, p. 482) took task (8.1) to be probing "a belief in objective moral truths and a rejection of moral relativism." Sarkissian et al. seem correct in thinking that the 'at least one disagreeing party must be mistaken' answer choice by itself fails to represent either moral objectivism or moral absolutism very well. Their effort to supplement this choice with additional, variable information about the cultural backgrounds of the

disagreeing parties allows them to examine whether ordinary individuals think cultural variability should influence metaethical judgments. The only question is what kind of metaethical judgment is expressed by agreeing or disagreeing that at least one of the disagreeing parties must be wrong. For reasons articulated above, it is difficult to view it as a direct judgment regarding objectivism or nonobjectivism, since a moral judgment can be true and its negation false for mind-dependent reasons. Task (8.1) can, however, effectively test for folk commitment to the idea that the moral judgments of disagreeing parties are exclusionary.

We can thus interpret Sarkissian et al.'s (2011) findings as revealing that participants were more likely to make exclusionary judgments when the disagreeing parties are from the same culture than when they are from different cultures. It is important to emphasize that participants displayed this strong tendency to let cultural considerations affect their exclusionary judgments in response to the following rather violent cases:

(8.2)   Horace finds his youngest child extremely unattractive and therefore kills him.

(8.3)   Dylan buys an expensive new knife and tests its sharpness by randomly stabbing a passerby on the street.

Sarkissian et al. (2011) intentionally chose cases where they thought cultural relativity would be least likely to be observed, and yet they found it anyway.

Khoo and Knobe (2018) replicated the results of Sarkissian et al. (2011) while also asking participants if two seemingly disagreeing parties were genuinely disagreeing. They found that participants were just as likely to think two parties were disagreeing when they were from the same culture as when they were from different cultures, but participants were less inclined to make exclusionary judgments about them when they were from different cultures. Using the same probe as Sarkissian et al. (2011), Matthew Fisher, Knobe, Brent Strickland, and Frank Keil (2017) found

that alternative argumentative mindsets can influence individuals' construal of the metaethical features of moral issues. They observed that participants who adopted cooperative (argue-to-learn) mindsets were less inclined to make exclusionary judgments than were participants who adopted competitive (argue-to-win) mindsets.[9]

In a number of studies, Beebe and his collaborators (Beebe, 2014; Beebe, 2015; Beebe, Qiaoan, Wysocki, & Endara, 2015; Beebe & Sackris, 2016) employed the following task that bears important similarities to that of Sarkissian et al. (2011):

(9)   If someone disagrees with you about whether [the moral judgment is true], is it possible for both of you to be correct or must one of you be mistaken?

(a) It is possible for both of you to be correct.

(b) At least one of you must be mistaken.

Like others working in the disagreement paradigm, Beebe and his collaborators mistakenly interpreted answer choice (a) to represent moral relativism and (b) represent moral objectivism. However, task (9) does seem to be a straightforward measure of judgment exclusion. Using this probe, Beebe and his collaborators found that folk exclusionary judgments varied significantly across ten moral issues that included donating to charity, assisting with a friend's suicide, abortion, lying, cheating, theft, and racial discrimination. They found that participants were more likely to make exclusionary judgments about moral disagreements when (i) the participants held stronger opinions about the issue in question, (ii) participants perceived there to be less societal disagreement about the issue, (iii) moral judgments concerned wrongdoing rather than good or

---

[9] In the Competitive condition participants were told, "You will be justifying your position… to another MTurk worker who has a strong stance on these issues. This is a highly competitive exchange and your task will be to outperform the other person." In the Cooperative condition, participants were told "You will be sharing your position… to another MTurk worker who has a strong stance on these issues. This is a highly cooperative exchange and your task will be to learn as much as you can from the other person" (pp. 1122-3).

morally exemplary actions, (iv) participants were older, (v) participants were asked to reflect on the extent of societal disagreement after (rather than before) making metaethical judgments, and (vi) participants were asked to imagine concretely (as opposed to more abstractly) described individuals disagreeing with them.[10] The first four findings were observed among teenaged and adult participants from the U.S., China, Poland, and Ecuador. The remaining findings were obtained with adult participants from the U.S.

Using the same probe as Beebe and his collaborators, Larisa Heiphetz and Liane Young (2017) found that 4- to 6-year-old children and adults largely agreed that the moral judgments of two disagreeing parties were exclusionary when there was widespread agreement that the judgment of one of the parties is correct and the judgment of the other is incorrect. In cases that did not feature uncontroversially correct or incorrect first-order moral judgments, participants were less likely to make exclusionary judgments. Children were more likely than adults to make exclusionary judgments in both kinds of cases. Like others working in this area, Heiphetz and Young (2017) took their probe to be measuring folk commitment to moral objectivism.

In a number of studies, Wright and her collaborators (Wright, Grandjean, & McWhite, 2013; Wright, McWhite, & Grandjean, 2014) used slight variations of probes constructed by Goodwin and Darley (2008). However, in response to published criticisms of these measures (e.g., Beebe, 2015; Pölzler, 2017, 2018), Wright (2018) has more recently developed a new measure of folk moral objectivism, in which participants were asked to indicate which of the following options best represented a number of situations of moral disagreement between them and another person:

---

[10] Goodwin and Darley (2008) observed some similar patterns in the composite scores they calculated using the tasks described above: participants were more likely to give what Goodwin and Darley consider to be objectivist responses when (i) they were responding to moral judgments than when responding to statements about social conventions or tastes, (ii) participants held stronger opinions about the moral issue in question, (iii) participants perceived there to be less societal disagreement about the issue, and (iv) the moral judgments concerned wrongdoing rather than good or morally exemplary actions.

(10.1) (a) *It would be morally (un)acceptable for that person to x.* The person would be correct because the rightness/wrongness of *x* is determined individually, by each person's beliefs, attitudes, or feelings about the act of *x*-ing or type of action that *x*-ing is

(b) *It may or may not be morally (un)acceptable for that person to x.* Whether the person was correct would depend on the community in which that person lives. The rightness/wrongness of *x* is determined by a community's collective beliefs, attitudes, or feelings about the act of *x*-ing or type of action that *x*-ing is.

(c) *It would not be morally (un)acceptable for that person to x.* The person would be mistaken (as would anyone else who made this claim). The rightness/wrongness of *x* is determined by the action or type of action it is, independently of the person's or his/her community's beliefs, attitudes, or feelings about it. That is, there is something about *x*-ing or the type of action *x*-ing is that makes it right/wrong regardless of what that person or other people think or feel about it. (pp. 125-6)

Wright classified options (a) and (b) as subjectivist and (c) as objectivist. Wright (2018) constructed a second task to measure participants' commitments to relativism and non-relativism, which began with a paragraph that introduced participants to the idea of relative vs. non-relative terms. Participants were asked which of the following judgments best characterized a number of different moral disagreements:

(10.2) (a) ONLY ONE of these statements would be correct—either it is true that it was wrong for that person to *x* or it is true that it was <u>not</u> wrong for that person to *x*,

regardless of who is making the statement or the contexts in which it is being made. Both statements cannot be correct.

(b) BOTH of these statements could be correct—whether it is true that it was wrong for that person to *x* or <u>not</u> wrong for that person to *x* depends on who is making the statement and/or the context in which it is made.

Wright took option (a) in (10.2) to represent non-relativism and (b) to represent relativism. Wright's task (10.1) appears to be the clearest and best measure of folk moral objectivism in print. Task (10.2) is also seems to be a good measure of folk intuitions about judgment exclusion.

In response to a number of moral issues that included selling children on the internet, cannibalism, theft, adultery, prostitution, racial discrimination, and helping those in need, Wright (2018) observed more objectivist responses to task (10.1) when the cases involved moral wrongdoing rather than moral goodness. With task (10.2), Wright found that participants were more likely to make exclusionary judgments when the disagreements concerned cases of wrongdoing rather than moral goodness. Wright (2018) also found that the vast majority of participants (78%) gave objectivist responses on task (10.1) to some moral judgments but not to others, while considerably fewer gave consistently objectivist or subjectivist responses across the board. A strong majority (85%) also gave a mixture of exclusionary and non-exclusionary judgments on task (10.2).

Thus, a number of researchers have in recent years attempted to investigate folk moral objectivism using measures that typically fail to be good direct measures of moral objectivism but that appear to be good measures of whether individuals think moral judgments are exclusionary. Findings obtained using these reinterpreted measures reveal that people are more inclined to make exclusionary judgments when one judgment in a moral disagreement is widely accepted as being

correct within the participant's society or concerns the wrongness of inflicting unwanted harm (Wainryb et al., 2004; Nichols, 2004a; Beebe et al., 2015; Beebe & Sackris, 2016; Heiphetz & Young, 2017; Wright, 2018). They also reveal that individuals are significantly less inclined to make exclusionary judgments when (i) there is more disagreement within a society about a moral issue (Beebe et al., 2015; Beebe & Sackris, 2016; Heiphetz & Young, 2017), (ii) the lack of agreement within a society is made more salient to participants (Beebe, 2014), (iii) there is significant cultural distance between the disagreeing parties (Sarkissian, et al., 2011; Fisher, et al., 2017), or (iv) participants simply hold weaker opinions about the moral judgments in question (Beebe et al., 2015; Beebe & Sackris, 2016). There is also significant within-person variation about judgment exclusion from issue to issue (Beebe et al., 2015; Beebe & Sackris, 2016; Wright, 2018). The question to which I turn in the following section is how best to explain these findings.

<center>III.</center>

The received view within contemporary moral philosophy is that folk morality is committed to the idea that moral judgments are mind-independently, objectively true and are neither context-sensitive nor assessment-sensitive (cf. Mackie, 1977; Dancy, 1986; Brink, 1989; Smith, 1994a; Darwall, 1998; Shafer-Landau, 2003; Huemer, 2005). J. L. Mackie (1977, p. 33) famously put the point as follows:

> The ordinary user of moral language means to say something about whatever it is that he characterizes morally, for example a possible action, as it is in itself… and not about, or even simply expressive of, his, or anyone else's, attitude or relation to it… one that is absolute, not contingent upon any desire or preference or policy or choice.

The findings reported above do not comport well with an invariantist, objectivist model of folk metaethics. If invariantism were true, every moral disagreement would be a situation where at least one disagreeing party had to be wrong. It would not matter if the parties were from different cultures, if there were significant disagreement within a society about the moral issue, or if people simply failed to hold firm opinions about the issue. If moral judgments are fully factual and invariantism is true, it should be obvious that it is impossible for more than one disagreeing party to have a true belief about the matter. It is important to note that the consensus within philosophy is not merely that folk morality is invariantist and objectivist but that it is obviously so, such that every philosophically untrained individual can recognize that this is the case with little (if any) reflection.

Subjectivist models of moral language also fail to fit the observed data. If 'x is wrong' always means 'I disapprove of x' (or something else similarly truth apt that reports the attitudes of particular individuals), then the moral judgments of disagreeing parties should always be judged to be non-exclusionary. Jules' utterance of 'Stealing is wrong' will report Jules' attitudes, and Vincent's utterance of 'Stealing is not wrong' will report Vincent's, and there will be no more logical incompatibility between these two attitude reports than there is between Jules' utterance of 'I like strawberries' and Vincent's utterance of 'I don't like strawberries.' In the data described above, however, we do not see experimental participants making non-exclusionary judgments in the way that subjectivism would predict.

Simple versions of noncognitivism (e.g., the emotivist of Ayer, 1936) also fail to provide explanatorily adequate models for the data. According to these theories, moral judgments are not capable of being true or false and express only noncognitive mental states like emotions, motivations, desires, or approval. Because judgments are exclusionary when at least one of them

must be false, judgments that are not capable of being false cannot be exclusionary. Any approach that deems the moral judgments in every case of moral disagreement to be non-exclusionary cannot provide an adequate explanatory model for data showing that individuals take moral judgments to be exclusionary some of the time but not all of the time.

A number of noncognitivists (perhaps beginning with Stevenson, 1937, p. 26ff.) maintain that they can make sense of moral disagreement by appealing to the incompatibility of the attitudes the disagreeing parties take toward the action(s) in question. It might be thought that incompatibility of moral attitudes could underwrite an explanation of judgment exclusion. If we grant that when Jules says 'Stealing is wrong' he is expressing disapproval of stealing (or perhaps approval of norms that prohibit stealing), and when Vincent says 'Stealing is not wrong' he is expressing something like tolerance of stealing (or perhaps disapproval of norms that prohibit stealing or approval of norms that do not prohibit stealing), how can noncognitivists explain the incompatibility between the Jules' and Vincent's attitudes without appealing to any incompatibility between the contents of those attitudes? One possibility is for noncognitivists to maintain that a sincere utterance of 'Stealing is wrong' requires one to have a certain attitude (or attitudes) toward stealing, an utterance of 'Stealing is not wrong' requires one to have a distinct attitude (or attitudes), and it is psychologically impossible for someone to hold both (sets of) attitudes simultaneously toward stealing. These judgments might then be said to be exclusionary in a psychological rather than a logical sense.

Unfortunately, the claim that it is psychologically impossible for someone to simultaneously possess the attitudes required for the sincere assertion of both 'Stealing is wrong' and 'Stealing is not wrong' is clearly much too strong because it is obvious that someone can have a "fractured sensibility" (Blackburn, 1984, p. 195) or a set of cognitively dissonant attitudes that

do not comport well together. Cognitivists who view the propositional contents of these judgments as being logically inconsistent will all agree that logically inconsistent beliefs are nevertheless possible—however irrational they may be.

Furthermore, attempts to replace psychological impossibility with a weaker but equally psychological notion look doubtful because even the strong notion fails to be robust enough to explain how Jules' and Vincent's moral judgments cannot both be correct in some sense. To explain this, let us first grant noncognitivists the right that many of them claim to talk about the truth and falsity of moral judgments in the sense articulated by minimalist or deflationary conceptions of truth, according to which truth is not a substantial relation or property (Blackburn, 1993; Smith, 1994b). On one version of the view (e.g., Ramsey, 1927), there is nothing more to saying 'p is true' than there is to saying 'p.' In the moral domain, this means that an utterance of ''Stealing is wrong' is true' does nothing more and nothing less than express the same noncognitive mental states as an utterance of 'Stealing is wrong.' Now we can ask whether the notion of psychological impossibility described above can help noncognitivists explain how Jules' and Vincent's moral judgments cannot both be (minimally) true and how at least one of them must be (minimally) false. It seems that it cannot. It is obvious that Jules could possess and express the kind of attitude(s) needed for sincere utterance of 'Stealing is wrong' while Vincent at the same time holds the attitude(s) required for sincere utterance of 'Stealing is not wrong.' And if Jules says 'Stealing is wrong' at one time but then undergoes a transformational experience and says 'Stealing is not wrong' at some later time, it is equally clear that he could possess the kind of attitude(s) needed for sincere utterance of the first sentence at the earlier time and yet possess the very different kind of attitude(s) required by sincere utterance of the second sentence at the later time. Thus, even if it were psychologically impossible for Jules to possess both sets of attitudes at

the same time, this would do nothing to show that it is impossible for different people at the same time or for the same person at different times. As a result, it does not appear that psychological impossibility can provide the basis for a noncognitivist explanation of how the minimal truth of Jules' moral judgment somehow excludes the minimal truth of Vincent's (and vice versa).

It should be noted that even if judgment exclusion could be successfully explained in terms of conflicting attitudes, every pair of moral judgments featured in the disagreement paradigm would be deemed exclusionary because the disagreeing parties are always described as taking disagreeing attitudes toward the same action. Yet experimental participants were observed to make exclusionary judgments about these individuals' judgments only some of the time. Thus, the current approach would not be able to explain the variations observed in the data reported above.

Noncognitivists might adopt a different approach to explaining judgment exclusion by following Simon Blackburn's (1984) proposal that embedded, non-assertive uses of moral terms can be understood as expressing higher-order noncognitive attitudes. Blackburn (1984, p. 195), for example, glosses 'If lying is wrong, telling your little brother to lie is wrong' as expressing approval of making disapproval of getting one's brother to lie "follow upon" disapproval of lying. Noncognitivists could take a similar tack and explain exclusionary judgments as expressing disapproval of certain combinations of attitudes. Thus, an exclusionary judgment about Jules' and Vincent's judgments about stealing might express disapproval of combining the disapproving attitude Jules expresses toward stealing with the permissive attitude Vincent expresses toward stealing. A non-exclusionary judgment might express a permissive attitude toward such a combination. Can such an approach provide the basis for an explanation of the data about folk exclusionary judgments reported in Section II?

An important thing to note about the noncognitivist perspective under consideration is the contingency it attributes to the attitudes expressed by exclusionary and non-exclusionary judgments. It does not mandate that the judgments of disagreeing parties should always be deemed exclusionary, nor does it mandate that they should always be deemed non-exclusionary. This approach thus manifests a kind of flexibility regarding judgment exclusion that is not present in the other theories considered above. Nevertheless, although this perspective is compatible with individuals taking the moral judgments of disagreeing parties to be sometimes but not always exclusionary, it does not explain or predict when or why they make the (non-)exclusionary judgments they do. Moreover, contemporary noncognitivist theories are almost always developed in ways that are incompatible with allowing non-exclusionary judgments. Mark van Roojen (2018, sec. 2.3) explains:

> [Contemporary noncognitivists] have been as concerned with vindicating the legitimacy of moral practice and argument as with anything else. As a result, they have put more time and energy into explaining, and in a certain sense justifying, the realist-seeming features of moral discourse in the absence of a commitment to realism (Hare 1952; Blackburn 1984, 1998; Gibbard 1990).

Contemporary noncognitivists' desire to preserve the allegedly realist-seeming features of moral discourse leads them to propound theories that make precisely the same predictions as invariantism about what we should observe in experiments using the disagreement paradigm, and this in turn means that the problems raised for invariantism above will apply to these theories as well.[11][12]

---

[11] Khoo and Knobe (2018, p. 129) make a similar point.

[12] There are other important versions of noncognitivism—e.g., the prescriptivism of Hare (1952), the expressivism of Gibbard (1990; 2003), and contemporary hybrid expressivism (e.g., Jackson 1999; Barker 2000; Ridge 2006, 2007)—that I do not have space to address here because of their significant complexity. However, the points I make here about noncognitivism generally apply to these other versions as well.

In contrast to the theoretical perspectives considered above, a number of scholars have offered indexical relativist accounts of moral language that are well suited to explain the data from Section II (e.g., Beebe, 2010; Björnsson & Finlay, 2010[13]; Khoo & Knobe, 2018). Beebe (2010) characterizes his version of IMR with the following theses:

(MR1) The central normative terms of moral judgments—e.g., 'right,' 'wrong,' 'good,' 'bad,' 'permissible,' 'impermissible'—are indexicals.

(MR2) The contents that get expressed by moral terms are determined by the moral standards that are in place at the contexts in which those terms are used.

(MR3) There exists only one (or at least one privileged) conversational record or 'semantic scoreboard' (in Lewis' 1979 sense) that reflects the set of things taken for granted by participants at any point in a conversation and that imposes requirements on the truth conditions of their utterances.

According to MR1 through MR3, if Jules' utterance of 'Stealing is wrong' and Vincent's utterance of 'Stealing is not wrong' are made in the same context, and there is agreement within that context regarding the permissibility of stealing, then these judgments will express contradictory propositions, and at least one of them must be false. If Jules and Vincent make their moral judgments in different cultural contexts, but these cultures share a common moral outlook (i.e., have equivalent semantic scoreboards) in regard to stealing, then these judgments may again express contradictory propositions that cannot both be true. However, if Jules' and Vincent's judgments are made in different contexts, and the moral standards in place in these contexts differ in regard to stealing, their judgments will express propositions that are not contradictory, and it may be possible for both of them to be true. If Jules and Vincent are in the same context but there

---

[13] The theory of Björnsson and Finlay (2010) includes additional theoretical complications that I do not have space to address here. So, I will focus only on the accounts of Beebe (2010) and Khoo and Knobe (2018) for present purposes.

is not enough agreement within that context regarding what is taken for granted about stealing, their judgments may fail to express determinate propositions. On IMR, then, the moral judgments of disagreeing parties are sometimes but not always exclusionary. Note that IMR provides this flexible (and dynamic[14]) perspective on the semantics of moral judgments by utilizing the well understood phenomenon of indexicality, in contrast to noncognitivist theories that face formidable challenges in articulating the meaning of moral terms (cf. Unwin, 1999, 2001; Dreier, 2006; Schroeder, 2008, 2010).

Khoo and Knobe (2018, p. 132) offer a similar theory, on which moral judgments are only true or false relative to a norm parameter of the contexts in which they uttered and where their assertion is seen as "a proposal to update both the norms of the context and what is common ground in that context." Khoo and Knobe (2018, p. 131) explain moral disagreement within a conversation in terms of "the speakers making incompatible proposals to update the norm parameter." If neither speaker accepts the update proposed by the other, it may be indeterminate which norms are in play in that context and thus their moral judgments may be neither true nor false. In such a situation, their judgments will fail to be exclusionary because they will lack determinate content of any kind. Although Khoo and Knobe (2018) focus on cases where the people making disagreeing judgments in the same conversational context, their account can be extended to cover situations where the speakers are in different conversational contexts.

The versions of IMR developed by Beebe (2010) and Khoo and Knobe (2018) both straightforwardly predict and explain the findings from Section II regarding folk exclusionary judgments. When there is significant disagreement about a moral issue or when disagreeing parties belong to cultures with significantly different values, customs, or norms, there will not be a shared

---

[14] I do not have space to discuss the dynamic elements of the view here, but cf. Lewis (1979; 1996), DeRose (2004), and Beebe (2010) for details.

common ground in light of which pairs of contrasting moral judgments can have determinate exclusionary contents. Given individuals' recognition of the different degrees of agreement or controversy surrounding different moral issues, these theories predict that individuals will think that the moral judgments of disagreeing parties will have exclusionary content some of the time but not all of the time. These theories can also explain why individuals will be less inclined to think there is exclusionary content when asked to reflect on the extent of societal disagreement before rather than after making judgments about exclusionary content. By making the potential lack of common ground more salient, such reflection can make it more doubtful that there are determinate contents that can be exclusionary.

Beebe (2010) takes Lewis' (1979, 1996) work on presupposition accommodation and epistemic contextualism as points of departure for developing an account of various factors that can lead individuals to be more or less inclined to modify the moral norms or presuppositions that constitute the common ground of conversational contexts that feature moral disagreement. The kinds of factors responsible for individuals' differential willingness to accommodate the assertions of disagreeing others can explain much of the remaining data summarized above. For example, individuals who hold stronger opinions about a moral issue will be more resistant and more opposed to proposals to update the norms that are (at least thought to be) in place in a context if those proposals would modify the 'initialized' or operative norms in a direction to which the individuals are strongly opposed. If a revisionary proposal has not been allowed to call into question the perceived status quo of the semantic scoreboard, a strongly opinionated person with an update-resistant mindset may think it is a determinate matter about which norms are in place and thus that contrasting pairs of moral judgments made in light of those norms will have exclusionary content. Someone with a weaker opinion who is initially inclined to think there is

agreement about what norms are operative in a context may be more easily moved into a state of doubt about this when confronted with a disagreeing party. Regarding Fisher et al.'s (2017, pp. 1122-3) findings about cooperative and competitive mindsets, if your aim is "to learn as much as you can from the other person," you will be more open to what that person has to say and thus more open to allowing any disagreeing judgments that person makes to call into question what you take to be common ground in the conversation. When someone with a weak opinion is led to doubt what the common ground is or someone with a cooperative mindset opens themselves to the possibility of calling into question or modifying the common ground, there will not be a fact about the relevant norm parameters in the conversation that can be taken for granted.

Truth value relativism about moral judgments (hereafter 'TVR') hypothesizes that moral judgments are assessment-sensitive, and like IMR, it can accommodate variability in individuals' intuitions about whether the moral judgments of disagreeing parties are exclusionary. However, TVR cannot explain all of the data described above. According to TVR, a moral judgment such as 'Stealing is wrong' will express the same proposition in all contexts of use—regardless of whether the speaker or the speaker's culture approves or disapproves of stealing. The truth value of that judgment will depend upon the context of the person assessing the truth or falsity of the judgment. (Some of the time the person making a judgment and the person assessing that judgment's truth value will be the same, but this need not always be the case.)

Suppose that Jules and Vincent are from different cultures that have conflicting moral outlooks on stealing and that their own outlooks coincide with those of their respective cultures.[15] When someone from Jules' culture assesses Jules' utterance of 'Stealing is wrong,' it will be

---

[15] For present purposes, I am setting aside the issue of how to treat cases where individuals disagree with most other members of their culture, but cf. DeRose (2004) and Beebe (2010) for details on how such cases can be handled within a relativist framework.

correctly assessed as true; but when that moral judgment is assessed by someone from Vincent's culture, it will be correctly assessed as false. Furthermore, Vincent's judgment will be assessed as true by his culture and as false by Jules'. According to TVR, if 'Stealing is wrong' is considered in abstraction from any context of assessment, there will be no fact of the matter about whether it is true or false. If a third party considers a moral disagreement between two other individuals, it is the moral standards of the third party rather than those of either disagreeing party that determine the truth values of the moral judgments in question. Thus, if Winston wants to assess the moral judgments of Jules and Vincent, and Winston's culture has norms that clearly prohibit stealing, and Winston is fully in agreement with his culture on this matter, then 'Stealing is wrong' will be true at Winston's context, and 'Stealing is not wrong' will be false at Winston's context. Importantly, if Winston inhabits a culture in which there is no agreement about the morality of stealing, and Winston's own attitudes reflect this indeterminacy, both 'Stealing is wrong' and 'Stealing is not wrong' will be neither true nor false at Winston's context—even if those judgments are true or false relative to Jules' or Vincent's context of assessment.

It is noteworthy that on TVR it will be impossible for 'Stealing is wrong' and 'Stealing is not wrong' to both be true at a given context of assessment. Thus, in contrast to IMR, which allows a number of possible truth value combinations for the moral judgments of disagreeing parties (e.g., both true, both false, both neither true nor false, one true and one false, one true and one neither, one false and one neither), TVR allows only two possible outcomes (one true and one false or both neither true nor false).

When we consider the cases of moral disagreement that appear in the disagreement paradigm literature, according to TVR it is the moral standards in place at each participant's

context that determine whether the judgments of the disagreeing parties are exclusionary.[16] Such a perspective can explain some but not all of the data described in Section II. For example, when there is significant disagreement within a participant's society about a moral issue, TVR predicts that there may not be a sufficient basis for assigning truth values to the disagreeing judgments, in which case they would fail to be exclusionary. TVR also predicts that when participants are asked to reflect on the extent of disagreement within their own society about a moral issue before making an exclusionary or non-exclusionary judgment about a moral disagreement between two other parties, they may be less inclined to think there is a basis for assigning truth values to the judgments of the parties and thus as viewing them as exclusionary. When participants have weaker attitudes or commitments regarding a moral issue, they may also be less inclined to view disagreeing judgments about that issue as exclusionary.

However, according to TVR it does not matter if the disagreeing parties belong to cultures whose values, customs, and norms differ significantly from each other or if those cultures differ significantly from that of the participant. If the participant's culture takes a clear stance on the issue, the judgments of the disagreeing parties should always be deemed exclusionary. Yet this is not what we observe in the cases of Horace (8.2) and Dylan (8.3) that were tested by Sarkissian et al. (2011) and Khoo and Knobe (2018). It should not matter that killing a child for being unattractive or testing the sharpness of a knife by randomly stabbing a passerby is considered permissible by some faraway warrior culture. Since these actions are clearly and strongly prohibited by the culture of every experimental participant, if TVR provides a good model of folk metaethical commitments, participants should view any disagreeing judgments about these cases as exclusionary. Yet Sarkissian et al. (2011) and Khoo and Knobe (2018) found that participants

---

[16] Thanks to Joshua Knobe for helpful insights about this point.

made significantly more non-exclusionary judgments in these cases when the cultures of the disagreeing parties were dissimilar. These data are better explained by IMR than by TVR.

Wright and her collaborators (Wright, Grandjean, & McWhite, 2013; Wright, McWhite, & Grandjean, 2014; Wright, 2018) have suggested that the data from studies of folk moral objectivism reveal that the folk are metaethical pluralists—i.e., that they "view some moral issues as objectively grounded, while viewing others as grounded non-objectively" (Wright, McWhite, & Grandjean, 2014, p. 28). However, there are a number of reasons why this perspective cannot explain the data as well as IMR can. The most important is that metaethical pluralism is based on the assumption that the probes used in the folk moral objectivism literature directly measure objectivism. As we have seen, this assumption is false. A second reason is that the thesis of metaethical pluralism is basically a description of the data—akin to a summary statistic—instead of an explanation of the data. Beginning from the assumption that 'at least one of the judgments of the disagreeing parties must be false' expresses a commitment to objectivism, metaethical pluralism is for the most part simply the observation that this "objectivist" answer is selected some of the time but not all of the time by experimental participants. Thirdly, the theory fails to provide an account of the meaning of moral judgments. When the folk deny that some moral judgments are objectively true or false, metaethical pluralism fails to say whether these denials are best modeled by a subjectivist, noncognitivist, indexical relativist, truth relativist, or some other perspective. And it is radically unclear how 'is wrong' could sometimes ascribe a moral property to actions in judgments that are mind-independently true or false yet at other times ascribe a property in judgments that fail to be mind-independently true or false or in judgments that fail to ascribe any moral property at all. For these reasons, Wright's metaethical pluralism fails to offer as good an explanatory model of the relevant data as IMR.

An important explanatory virtue of IMR is that it can explain the available data while preserving what Michael Gill (2009) calls 'the uniformity assumption' regarding moral language—the idea that all standard moral language has the same kind of semantics (e.g., all expressivist, all cognitivist, or all relativist)—and this in turn can block the charge that folk metaethics is confused, inconsistent, and ultimately incoherent (Loeb, 2008). Charges of internal incoherence within folk metaethics stem from the assumption that the semantics for moral terms must be invariantist. When variability is observed from this perspective, it is presumed that it is the folk rather than the semantics that is at fault. IMR's explanation of the meaning of moral terms allows for significant semantic variability across moral discourse while explaining that variability within a unified semantic framework that does not undermine the coherence of that discourse.[17]

IV.

During the last decade and a half, a number of scholars have attempted to investigate folk moral objectivism (Wainryb et al., 2004; Nichols & Folds-Bennett, 2003; Nichols, 2004a; Goodwin & Darley, 2008, 2010, 2012; Sarkissian, et al., 2011; Beebe, 2014; Beebe et al., 2015; Beebe & Sackris, 2016; Fisher, et al., 2017; Heiphetz & Young, 2017; Wright, Grandjean, & McWhite, 2013; Wright, McWhite, & Grandjean, 2014; Wright, 2018). However, for reasons explained above, most of the probes used in this area of research fail to be good direct measures objectivism. Following Khoo and Knobe (2018), I have explained how many of these tasks can be seen as

---

[17] The meaning of the simplest probes described in Section II should not have been difficult for non-specialists to decipher, even if the broader philosophical issues to which they are connected are fraught. There is thus no obvious reason to think that participants in these studies consistently misunderstood the nature of the tasks they were given. Khoo and Knobe (2018, pp. 128-9) emphasize that "we have good reason to think that our experimental participants know English and that they would correctly make judgments that follow trivially from facts about our language." Thus, while there is perhaps some theoretical possibility that participants' mistaken metaethical or metaphysical beliefs might be repeatedly overriding their actual semantic competence with regard to some very simple expressions of ordinary English, Chinese, Polish, or Spanish, it is a virtue of IMR that it can explain the relevant data without attributing widespread linguistic error to study participants.

probing whether participants think that the moral judgments of disagreeing parties are exclusionary; and on this basis I have argued that the best explanation of the reinterpreted data from these studies is broad folk endorsement of IMR.

If my conclusion is correct, most contemporary moral philosophers are mistaken about the metaethical contours of folk morality in one very important respect. It is widely assumed that folk morality is objectivist and invariantist—i.e., that moral judgments are truth apt, that their truth values are determined objectively, and that their content and truth values are independent of who assertively utters or evaluates them. However, as has been pointed out by every researcher working on folk moral objectivism, these assumptions are not based upon any solid empirical evidence. It is also widely recognized that relativism always seems to have had greater appeal to students in undergraduate philosophy courses than their instructors expect and that these instructors struggle to understand both the nature and the appeal of such views. As Beebe (2010) has argued at length, IMR provides an elegant explanation of many of the behaviors of "the average, philosophically untrained moral relativist" that philosophy instructors find puzzling. Thus, from a theoretical perspective, the fact that IMR can explain a wide variety of data makes it less revisionary than competing perspectives that need to posit widespread error in folk metaethical judgments. I hope that this contribution enriches the discussion and debate about descriptive metaethics and metaethics more generally.

References

Ayars, Alisabeth, and Shaun Nichols. (forthcoming). "Rational Learners and Metaethics: Universalism, Relativism, and Evidence from Consensus." *Mind & Language*.

Ayer, A. J. (1936). *Language, Truth and Logic*. London: Gollancz.

Barker, Stephen J. (2000). "Is Value Content a Component of Conventional Implicature?" *Analysis*, 60: 268–79.

Beebe, James R. (2010). "Moral Relativism in Context." *Noûs*, 44, 691-724.

Beebe, James R. (2014). "How Different Kinds of Disagreement Impact Folk Metaethical Judgments." In Hagop Sarkissian & Jennifer Cole Wright (eds.), *Advances in Experimental Moral Psychology*. London: Bloomsbury, pp. 167-178.

Beebe, James R. (2015). "The Empirical Study of Folk Metaethics." *Etyka*, 15, 11-28.

Beebe, James R., Runya Qiaoan, Tomasz Wysocki, and Miguel A. Endara. (2015). "Moral Objectivism in Cross-cultural Perspective." *Journal of Cognition and Culture*, 15, 386-401.

Beebe, James R., and David Sackris. (2016). "Moral Objectivism Across the Lifespan." *Philosophical Psychology*, 29, 912-929.

Björnsson, Gunnar, and Stephen Finlay. (2010). "Metaethical Contextualism Defended." *Ethics*, 121, 7-36.

Blackburn, Simon. (1984). *Spreading the Word*. Oxford: Oxford University Press.

Blackburn, Simon. (1993). *Essays in Quasi-Realism*. Oxford University Press.

Brink, David O. (1989). *Moral Realism and the Foundations of Ethics*. Cambridge: Cambridge University Press.

Dancy, Jonathan. (1986). "Two Conceptions of Moral Realism." *Proceedings of the Aristotelian Society, Supplemental Volume*, 60, 167-187.

Darwall, Stephen. (1998). *Philosophical Ethics*. Boulder, CO: Westview.

DeRose, Keith. (2004). "Single Scoreboard Semantics." *Philosophical Studies* 119: 1–21.

Dreier, James. (2006). "Negation for Expressivists: A Collection of Problems with a Suggestion for their Solution." In Russ Shafer-Landau (ed.), *Oxford Studies in Metaethics, Vol. 1*. Oxford: Oxford University Press, pp. 217–34.

Fisher, Matthew, Joshua Knobe, Brent Strickland, and Frank Keil. (2017). "The Influence of Social Interaction on Intuitions of Objectivity and Subjectivity." *Cognitive Science*, 41, 1119-1134.

Gibbard, Allan. (1990). *Wise Choices, Apt Feelings*. Cambridge, MA: Harvard University Press.

Gibbard, Allan. (2003). *Thinking How to Live*. Cambridge, MA: Harvard University Press.

Gill, Michael B. (2009). "Indeterminacy and Variability in Meta-Ethics." *Philosophical Studies*, 145, 215-234.

Goodwin, Geoffrey P., and John M. Darley. (2008). "The Psychology of Meta-Ethics: Exploring Objectivism." *Cognition,* 106, 1339-1366.

Goodwin, Geoffrey P., and John M. Darley. (2010). "The Perceived Objectivity of Ethical Beliefs: Psychological Findings and Implications for Public Policy." *Review of Philosophy and Psychology*, 1, 161-188.

Goodwin, Geoffrey P., and John M. Darley. (2012). "Why are Some Moral Beliefs Perceived to be More Objective than Others?" *Journal of Experimental Social Psychology*, 48, 250-256.

Heiphetz, Larisa, and Liane Young. (2017). "Can Only One Person Be Right? The Development of Objectivism and Social Preferences Regarding Widely Shared and Controversial Moral Beliefs." *Cognition,* 167, 78–90.

Hare, R. M. (1952). *The Language of Morals*. Oxford: Clarendon.

Harman, Gilbert. (1975). "Moral Relativism Defended." *Philosophical Review*, 84: 3–22.

Huemer, Michael. (2005). *Ethical Intuitionism*. Basingstoke: Palgrave MacMillan.

Jackson, Frank. (1999). "Non-Cognitivism, Validity and Conditionals." In Dale Jamieson (ed.), *Singer and His Critics*. 18–37.

Johnston, Mark. (1989). "Dispositional Theories of Value." *Proceedings of the Aristotelian Society, Supplementary Volume*, 63, 139–174.

Kaplan, David. (1989). "Demonstratives: An Essay on the Semantics, Logic, Metaphysics, and Epistemology of Demonstratives and Other Indexicals." In Joseph Almog, John Perry, and Howard Wettstein (eds.), *Themes from Kaplan*. Oxford: Oxford University Press, pp. 481–566.

Khoo, Justin, and Joshua Knobe. (2018). "Moral Disagreement and Moral Semantics." *Noûs,* 52, 109–143.

Lewis, David K. (1979). "Scorekeeping in a Language Game." *Journal of Philosophical Logic*, 8, 339–359.

Lewis, David K. (1996). "Elusive Knowledge." *Australasian Journal of Philosophy*, 74, 549–567.

Loeb, Don. (2008). "Moral Incoherentism: How to Pull a Metaphysical Rabbit out of a Semantic Hat." In Walter Sinnott-Armstrong (ed.), *Moral Psychology, Vol. 2. The Cognitive Science of Morality: Intuition and Diversity*. Cambridge, MA: MIT Press, pp. 355–85.

MacFarlane, John. (2005a). "The Assessment Sensitivity of Knowledge Attributions." In Tamar Szabo-Gendler & John Hawthorne (eds.), *Oxford Studies in Epistemology, vol. 1*. Oxford: Oxford University Press, pp. 197–233.

MacFarlane, John. (2005b). "Making Sense of Relative Truth." *Proceedings of the Aristotelian Society*, 105: 321–339.

MacFarlane, John. (2007). "Relativism and Disagreement." *Philosophical Studies*, 132: 17–31.

MacFarlane, John. (2009). "Nonindexical Contextualism." *Synthese,* 166: 231–250.

MacFarlane, John. (2014). *Assessment Sensitivity: Relative Truth and Its Applications*. Oxford: Oxford University Press.

Mackie, J. L. (1977). *Ethics: Inventing Right and Wrong*. New York: Penguin.

Nichols, Shaun. (2004a). "After Objectivity: An Empirical Study of Moral Judgment." *Philosophical Psychology, 17,* 5-28.

Nichols, Shaun. (2004b). *Sentimental Rules: On the Natural Foundations of Moral Judgment*. Oxford: Oxford University Press.

Nichols, Shaun, and Trisha Folds-Bennett. (2003). "Are Children Moral Objectivists? Children's Judgments About Moral and Response-Dependent Properties." *Cognition, 90,* B23-B32.

Pettit, Philip. (1991). "Realism and Response-Dependence." *Mind,* 100, 587-626.

Pölzler, Thomas. (2017). "Revisiting Folk Moral Realism." *Review of Philosophy and Psychology*, 8, 455-476.

Pölzler, Thomas. (2018). "How to Measure Moral Realism." *Review of Philosophy and Psychology*, 9, 647-670.

Ramsey, F. P. (1927). "Facts and Propositions." *Proceedings of the Aristotelian Society Supplementary Vol. 7*: 153–170.

Ridge, Michael. (2006). "Ecumenical Expressivism: Finessing Frege." *Ethics*, 116: 302–36.

Ridge, Michael. (2007). "Ecumenical Expressivism: The Best of Both Worlds?" in Russ Shafer-Landau (ed.), *Oxford Studies in Metaethics, Volume two*, Oxford University Press, pp. 51-76.

Sarkissian, Hagop, John Park, David Tien, Jennifer Cole Wright, and Joshua Knobe. (2011). "Folk Moral Relativism." *Mind & Language,* 26, 482-505.

Schroeder, Mark. (2008). "How Expressivists Can and Should Solve Their Problem With Negation." *Noûs* 42: 573–99.

Schroeder, Mark. (2010). *Noncognitivism in Ethics*. London: Routledge.

Shafer-Landau, Russ. (2003). *Moral Realism: A Defence*. New York: Oxford University Press.

Smith, Michael. (1994a). *The Moral Problem*. Oxford: Blackwell.

Smith, Michael. (1994b). "Why Expressivists About Value Should Love Minimalism About Truth." *Analysis* 54: 1-12.

Stevenson, C. L. (1937). "The Emotive Meaning of Ethical Terms." *Mind* 46: 14–31.

Turiel, Elliot. (1983). *The Development of Social Knowledge: Morality and Convention*. Cambridge: Cambridge University Press.

van Roojen, Mark. (2018). "Moral Cognitivism vs. Non-Cognitivism." In Edward N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy* (Fall 2018 Edition). Available at: https://plato.stanford.edu/archives/fall2018/entries/moral-cognitivism/.

Wainryb, Cecilia, Leigh A. Shaw, Marcie Langley, Kim Cottam, and Renee Lewis. (2004). "Children's Thinking About Diversity of Belief in the Early School Years: Judgments of Relativism, Tolerance, and Disagreeing Persons." *Child Development*, 75, 687–703.

Wedgwood, Ralph. (1998). "The Essence of Response-Dependence." *European Review of Philosophy*, 3, 331–354.

Wright, Jennifer Cole. (2018). "The Fact and Function of Meta-Ethical Pluralism: Examining the Evidence." In Tania Lombrozo, Shaun Nichols, & Joshua Knobe (eds.), *Oxford Studies in Experimental Philosophy, Vol 2*. New York: Oxford Press, pp. 119-150.

Wright, Jennifer Cole, Piper T. Grandjean, and Cullen B. McWhite. (2013). "The Meta-Ethical Grounding of our Moral Beliefs: Evidence for Meta-Ethical Pluralism." *Philosophical Psychology, 26*, 336–361.

Wright, Jennifer Cole, Cullen B. McWhite, and Piper T. Grandjean. (2014). "The Cognitive Mechanisms of Intolerance: Do our Meta-Ethical Commitments Matter? In Joshua Knobe, Tania Lombrozo, and Shaun Nichols (eds.), *Oxford Studies in Experimental Philosophy, Vol. 1*. Oxford: Oxford University Press, pp. 28–61.