Moral internalism, amoralist skepticism and the factivity effect

Kenneth Shields

Routledge
Taylor & Francis Group

Moral internalism, amoralist skepticism and the factivity effect

Kenneth Shields (iD)

Department of Philosophy, University of Missouri, Columbia, MO, USA

ABSTRACT

Philosophers are divided over moral internalism, the claim that moral judgement entails some motivation to comply with that judgement. Against moral internalism, externalists defend the conceptual coherence of scenarios in which an individual makes genuine moral judgements but is entirely unmoved by them. This is amoralist skepticism and these scenarios can be called amoralist scenarios. While the coherence of amoralist scenarios is disputed, philosophers seem to agree that the coherence of amoralist scenarios is not affected by whether the amoralist is described as having moral knowledge or mere belief. But recent experimental research challenges this assumption. When evaluating amoralist scenarios, people's intuitions lean towards externalism when the amoralist is described as knowing that X is morally wrong, whereas people's intuitions lean towards internalism when the amoralist is described as believing that X is morally wrong. Call this the *factivity effect*. In this paper, I argue that the factivity effect is unlikely to be explained as an experimental artifact and that as a consequence, the traditional dispute over moral internalism and amoralist skepticism may need a major overhaul. The results of three studies testing the factivity effect provide support for this thesis. Implications of these results for the traditional debate over moral internalism are discussed.

1. Introduction

Philosophers are divided over moral internalism, the claim that moral judgement entails some motivation to comply with that judgement. For example, if someone claims that they view eating meat as immoral but they lack any inclination to stop eating meat, we may doubt their sincerity or their grasp of the term "immoral". How could they view meat-eating as immoral if they lack even the slightest motivation to alter their behavior? Call this an *amoralist scenario*. Some philosophers deny the coherence of amoralist scenarios, while others defend their coherence. How one explains (or explains away) these amoralist scenarios has important implications for one of the most fundamental metanormative questions – Why be moral?

The notion that someone could make genuine moral judgements without being moved (even slightly) is called *amoralist skepticism* (Brink, 1989, p. 46). Philosophers who view amoralist skepticism as either incoherent or in need of qualification are *internalists*. These philosophers take motivation to be internal to the judgement itself. This approach offers a relatively simple explanation for the practical aspect of moral reasoning – we make moral judgements to determine what we should (and shouldn't)

---

do. But internalists are hard-pressed to explain why amoralist skepticism isn't even *conceptually* possible, given that many philosophers take it to be obvious that amoralist skepticism is possible (even if it's not very common). Those who take amoralist skepticism to be possible are *externalists*. These philosophers take motivation to be separate from moral judgement. This approach retains the sense in which amoralist skepticism is conceptually possible. But externalism must then bear the full brunt of the *why-be-moral* question. How *could* this question be satisfactorily answered without appealing to something like the internalist's connection between judgement and motivation?

Despite these different approaches to amoralist skepticism, both sides seem to share a crucial assumption regarding amoralist scenarios: describing the amoralist as having moral *belief* is equivalent to describing the amoralist as having moral *knowledge*, in the sense that these doxastic states can be interchanged without loss (or gain) in the scenario's overall conceptual coherence. Let's call this assumption *descriptive* e*quivalence*.

> **Descriptive equivalence (DE)**. Describing the amoralist as having moral *belief* is equivalent to describing the amoralist as having moral *knowledge*, in the sense that these doxastic states can be interchanged without loss (or gain) in the scenario's overall conceptual coherence.

It's likely the standard view that the coherence of amoralist scenarios doesn't depend on whether the amoralist is said to know, or just merely believe, some moral requirement. Still, knowing something differs from simply believing something in at least one important sense. Knowing is *factive*, whereas believing is *non-factive*. In this context, a factive is a term that semantically implies that the judgement is true. Thus, in an amoralist scenario, to say that the amoralist knows the morally right thing to do is to imply that the amoralist's judgement accurately captures the moral facts. Non-factives lack this implication. Non-factive judgements may address a domain without accurately capturing a fact within that domain. For example, a person's belief that abortion is immoral would still address the moral domain even if abortion were not, in fact, immoral. So DE can now be restated in the terminology of factives and non-factives.

> **Descriptive equivalence (DE)**. Factive descriptions of the amoralist are equivalent to non-factive descriptions of the amoralist, in the sense that they can be interchanged without loss (or gain) in the scenario's overall conceptual coherence.

This broader formulation of DE still seems to capture a basic assumption held by most metanormative theorists.

But recent experimental work in this area suggests that DE is false. People's intuitions lean towards externalism when the amoralist's moral judgement is described factively as understanding or knowing that she shouldn't do X, whereas people's intuitions lean towards internalism when the amoralist's judgement is described non-factively, as believing or thinking that she shouldn't do X. Call this the *factivity effect*.

> **Factivity effect (FE)**. People's intuitions lean toward externalism when evaluating factive amoralist scenarios, whereas people's intuitions lean toward internalism when evaluating non-factive amoralist scenarios.

FE seems to raise a challenge to DE, thus challenging a foundational assumption within the traditional debate over moral internalism and amoralist skepticism.

In this paper, I argue that FE is unlikely to be explained away as an experimental artifact; and that as a consequence, the traditional dispute over moral internalism and amoralist skepticism may need an overhaul. The results of studies testing three different explanations of FE provide support for this thesis. I conclude by discussing the implications of these results for the traditional dispute.

In section 2, I briefly survey characterizations of the amoralist by both externalists and internalists for evidence of a commitment to DE and a foreshadowing of FE. In section 3, I present the *factivity effect* (FE): people's intuitions lean towards externalism when the amoralist's moral judgement is described factively, as understanding or knowing that she shouldn't do X, whereas when the amoralist's judgement is described non-factively, as believing or thinking that she shouldn't do X, people's intuitions lean towards internalism. In section 4, I present three studies testing different explanations of

FE. The first two studies test the hypothesis that FE is merely an experimental artifact. The final study tests the hypothesis that FE reflects a feature of the structure of folk thought. In section 5, I consider three possible philosophical ramifications of the claim that FE reflects a division within folk thought. I end by highlighting some limitations of the present work that could be addressed in future research.

2. Describing the amoralist

Philosophers in the traditional debate over moral internalism appear to tacitly assume DE. A brief consideration of four characterizations of amoralism, two from prominent externalists and two from prominent internalists, will help reveal this tacit commitment to DE. It will also help to introduce the difference between factive and non-factive descriptions of the amoralist.

Consider the externalist David Brink's introduction of amoralist skepticism:

> But another traditional kind of skepticism accepts the existence of moral **facts** and concedes that we have moral **knowledge**, and asks why we should care about these **facts**. Call this *amoralist skepticism*. Amoralists are the traditional way of representing this second kind of skepticism; the amoralist is someone who **recognizes** the existence of moral considerations and remains unmoved (1989, p. 46).[1]

Notice that Brink relies exclusively on factives to describe the amoralist's moral judgement. Now if DE is true, then the coherence of Brink's characterization shouldn't be affected by replacing his factives with non-factives. So Brink's "someone who *recognizes* the existence of moral considerations and remains unmoved" can become "someone who *believes* in the existence of moral considerations and remains unmoved".

Another example comes from the externalist Sigrún Svavarsdóttir's description of her amoralist Patrick: "He *knew* what was right to do in the circumstances, but could not have cared less" (1999, p. 178). If DE is true, then the coherence of her descriptions shouldn't be affected by the following non-factive equivalent: "He *believed* that this was right to do in the circumstances, but could not have cared less". If DE is correct, these translations are equivalent to their original factive versions in terms of overall conceptual coherence.[2]

Now notice the internalist Jamie Dreier's use of non-factives when referencing an alleged amoralist politician:

> And suppose he now says, "What my friends believe is wrong: not individualism but a life in the service of others is really good". But the politician has no inclination to serve the less fortunate; instead he advances the cause of self-reliance whenever he can … It seems to me that what we want to say about the new politician is that he is using the word "good" either insincerely or incorrectly. We will not take his assertion at face value and attribute to him the **belief** "Life in the service of others is good" (1990, p. 13).

Contra Brink and Svavarsdóttir, Dreier characterizes amoralism (or its impossibility) using non-factives. But just as DE allows converting factive descriptions into non-factive descriptions, DE allows converting non-factive descriptions into factive descriptions. Dreier's assessment of the apathetic politician now becomes, "we will not take his assertion at face value and attribute to him the *knowledge* 'Life in the service of others is good'".

Finally, consider the internalist Michael Smith's practicality requirement. "If an agent *judges* that it is right for her to Φ in circumstances C, then either she is motivated to Φ in C or she is practically irrational" (1994, p. 65). If DE is true, then Smith's requirement has the following factive equivalent: "If an agent *knows* that it is right for her to Φ in circumstances C, then either she is motivated to Φ in C or she is practically irrational". So, if DE is correct, these translations are equivalent to their original non-factive versions in terms of overall conceptual coherence.[3]

Aside from this tacit commitment, there's a standard view concerning the relationship between knowledge and belief that seems to provide strong support for DE – the view that knowledge entails belief. This view has come to be called the *entailment thesis*.[4] It is likely (and plausibly) assumed that moral knowledge and moral belief both fall within the scope of this thesis. If this is right, and if the entailment thesis is true, then part of DE seems to follow. If an amoralist is described as knowing that

X is morally wrong, then that same amoralist must also believe that X is morally wrong. But recent experimental results seem to cast doubt on even this part of DE.

3. The factivity effect

Experimental investigation into moral internalism and amoralist skepticism began around the time of the advent of experimental philosophy, with a study carried out by Nichols (2002). Nichols presented participants with an amoralist scenario describing John, a murderous psychopath, who says that he understands that killing and hurting people is wrong, but he just doesn't care. The majority of participants (nearly 85%) reported that John "really understood that hurting others is morally wrong, despite the absence of motivation" (p. 289). While Nichols viewed these results as evidence for the coherence of amoralist skepticism, a particular feature of his study has recently gained attention. Like the externalists' characterizations discussed in the previous section, Nichols' description of the amoralist relies exclusively on factive terminology. Within his vignette, the amoralist is described as saying that "he knows that hurting others is wrong" (p. 289). His test question also retains this factive use, "Does John really understand that hurting others is morally wrong?" (p. 289).

If DE is correct, then it seems reasonable to predict that Nichols' results would remain unchanged if the factive terms were replaced with non-factive counterparts, "He believes that hurting others is wrong" and "Does John really believe that hurting others is wrong?" Roughly a decade after Nichols' preliminary study, Björnsson, Eriksson, Strandberg, Olinder, and Björklund (2015) ran follow-up studies testing this prediction. The prediction failed. In fact, participants were significantly more inclined to report relatively internalist intuitions when confronted with amoralist scenarios described using non-factives.

While Björnsson and colleagues ran numerous studies, it will suffice to summarize one in detail, the case of amoralist Anna. This study was designed as a more sophisticated version of Nichols' study. They replaced Nichols' moral transgression with a much less violent wrongdoing in an effort to block participants from granting understanding to amoralists merely out of a desire to hold them accountable. They also added a brief introduction describing how people normally classify some actions as morally right and others as morally wrong. And, crucially, they varied their test question from factives to non-factives. Their choice to include non-factives was out of a concern that Nichols' original, factive question "does not test internalism, as usually understood" (2015, p. 720). In their view, the term "belief" seemed to better capture internalism's focus on moral *judgement*, whereas "understand" seemed to downplay this aspect.

Participants in Björnsson and colleagues' study received a somewhat lengthy vignette. The first part described the ordinary practice of classifying actions as morally right or morally wrong. The second part introduced the reader to the amoralist, Anna, a woman who "classifies actions using expressions like 'morally right' and 'morally wrong'", but stipulating that her classifications don't "influence her choices" (2015, p. 721). Finally, the vignette presents the reader with a scenario in which Anna must choose one of two cell phones. While identical in type and price, only one includes a free donation of $20 to help address starvation in Sudan. While Anna does classify purchasing the non-donation phone as "morally wrong", and the donation-phone as "morally right", Anna's classification does not influence her choice of the phones in the slightest. Each participant was then randomly assigned to one of three test question variations:

> We have seen that Anna classifies some actions as "morally wrong." But because she lacks compassion and is strikingly egoistic, this never makes her even the least inclined to avoid these actions. We saw this indifference when she chose her cell phone. In light of this, would you say that she [understands/believes/herself thinks] that it is morally wrong not to choose the left phone? (Björnsson et al., 2015, p. 722)

Participants given the *understands* variation responded similarly to Nichols' participants—76% granted that Anna understands, but in both the *believes* and *herself thinks* variations, the responses leaned towards internalism relative to the understand variation—only 46% granted that Anna believes, and only 49% granted that she herself thinks that the choice was morally wrong.

Replacing factives like "understand" with non-factives like "believe" while leaving everything else in the study unaltered significantly changed how subjects responded. Participants leaned toward externalism when the amoralist was described as *understanding* that X is morally wrong but then leaned toward internalism when she was described as *believing* or *thinking* that X is morally wrong. This is the factivity effect.

This switch between factive and non-factive terminology appears to affect the degree to which the amoralist scenario seems conceptually coherent. Thus, FE seems to be in tension with DE. To what extent DE remains plausible depends on what best explains FE. If FE is just an experimental artifact, then DE survives unscathed. However, if FE reflects a feature of the structure of folk thought about moral judgement, then DE may seem less plausible. And this is no small matter. If DE is incorrect, the metanormative discussion over moral internalism and amoralist skepticism may need reframing (see section 5).

I present three studies in the following section, each testing different explanations of FE. The first two studies address what we can call *deflationary* explanations of FE. These explanations account for FE by appeal to deflationary causes (for example, that FE is merely an experimental artifact). Proponents of DE will presumably bet on some deflationary explanation being confirmed. The final study addresses what we can call a *substantive* explanation of FE. This explanation accounts for FE by appeal to substantive causes (for example, that FE is part of the structure of folk thought). But since it's unclear whether DE survives a substantive explanation of FE, some important implications for metanormative inquiry seem to follow.

4. Three explanations of the factivity effect

4.1. Deflationary explanation 1: The inverted commas response hypothesis

4.1.1. Background

Some philosophers give an error-theoretic account of externalist intuitions called the *inverted commas response* (ICR). In its traditional formulation, ICR involves claiming that alleged amoralists use moral language only in an "inverted commas" sense.[5] Rather than making an evaluative moral judgement about some behavior, an amoralist is merely making a sociological judgement concerning how that behavior is generally perceived. On this account, for example, if the amoralist utters the sentence "torturing babies for fun is morally wrong", the amoralist is merely referring to the *general consensus* that torturing babies for fun is morally wrong, *not* the moral fact itself. ICR can be understood as the combination of two claims:

(1)   Amoralists can't really make evaluative moral judgements.
(2)   At most, amoralists make sociological judgements about the general moral consensus (thus, externalist intuitions must be tracking this, if anything).

While an argument can be made for weakening 2 (see discussion), let's focus for now on this traditional formulation of ICR.

As it happens, ICR suggests a nice explanation of FE. Assuming that folk internalism is true, confusing sociological judgements with moral judgments might be what leads participants to (mistakenly) grant understanding to the amoralist. Specifically, FE can be explained by the following set of claims:

(1)   The folk conception of moral judgement is internalist.
(2)   Given ICR, factives like "understand" in experimental studies *mis*lead participants by confusing sociological knowledge with moral knowledge proper.
(3)   Non-factives like "believe" *correctly* lead a substantial number[6] of participants to deny genuine moral judgement to amoralists.
(4)   (Thus) FE is an experimental artifact; factives elicit apparent externalist responses only because of the ambiguity of "understand" in this context.

This explanation of FE can be called the *ICR hypothesis*. Is the ICR hypothesis correct? One attempt at testing this hypothesis was carried out by Björnsson and colleagues (2015). Their approach was to

explain ICR to the participants and then stipulate that the amoralist's moral expressions are always used in an "inverted-commas" sense. Participants were expected to grant that the amoralist understands that everyone around her judges certain behaviors to be morally wrong. Their prediction was that FE should thus be *more* pronounced if the ICR hypothesis is correct. But FE was not more pronounced. In fact, FE disappeared (p. 724)! So this study fails to confirm the ICR hypothesis.

Maybe the participants took the stipulation to apply to *all* moral judgements made by Anna, not just the factive ones like Anna's understanding that the act is wrong. If participants think they are being told to assume that Anna always refers to the social norm when she claims that she understands, believes or even herself thinks that something is morally right or wrong, then this would explain why FE disappeared.[7] But if this is the case, then responses to all three variations of the test question should have been significantly higher than in the original study. Yet, the opposite result occurred. "Yes" answers to both *understands* and *herself thinks* variations of the test question were significantly lower in terms of frequency (responses to the belief variation remained the same) (Björnsson et al., 2015, p. 724).

Perhaps this direct, explicit approach is partly to blame. If factives are ambiguous in this context, as ICR seems to predict, then creating further senses of "understand" seems unwise. Explicitly stipulating that "understand" will now refer to the "inverted-commas" sense seems too demanding (recall that participants must first learn about ICR from the study itself). This is *not* to say that the folk lack the relevant competence for determining whether amoralists make genuine moral judgements, any more than the denial that the folk have much of an explicit grasp of the nature of grammar must imply that they lack the relevant competence for determining whether certain sentences in their native language are grammatical.[8]

An implicit approach would eliminate, in the participant's mind, the possibility that the amoralist has any knowledge about the general moral consensus. If successful, this could focus participant attention on moral judgement. If achieved, and if the ICR hypothesis is correct, then participants should deny *both* understanding and belief to the amoralist. FE should disappear. Such a result would provide evidence for the ICR hypothesis, and thus go some way towards deflating FE and its impact on metanormative inquiry. The following study takes this indirect approach to test the ICR hypothesis.

4.1.2. Two preliminaries

First, to eliminate in the participant's mind the possibility that the amoralist has knowledge of the general moral consensus, it was important to select actions that currently lack any moral consensus. Two actions were chosen to fit this bill – ending a pregnancy simply to avoid parenthood and smoking marijuana recreationally. Showing that the amoralist recognizes the lack of a moral consensus on these cases was thought to be sufficient to block participants from attributing the "inverted-commas" sense of understanding.

Second, because factive terminology suggests (semantically or pragmatically) that something is actually the case, it would have been infelicitous to ask participants whether the amoralist understands that X is morally wrong if the participants *themselves* deny that X is morally wrong. Thus, it was necessary to devise a sorting procedure to remove participants from the data who did not personally view the behavior in question as morally wrong.

The following study tests the ICR hypothesis. If this hypothesis is right, then FE should disappear when participants evaluate an amoralist confronted with an action that even the amoralist realizes lacks any moral consensus.

4.1.3. Method

4.1.3.1. Participants. Four hundred participants completed a study on Amazon Mechanical Turk, spanning a wide age range, and 77 of these were used in the study (see below).[9] The mean age was 35, with 31 female participants and 46 male participants. Half of the participants received $0.15 for their participation, the other half received $0.30 (please see endnote for explanation).[10]

4.1.3.2. *Procedures and materials.*   Subjects were split into four groups.[11] Participants in each group were assigned one of two vignettes, *abortion* or *marijuana*, and one of two attitudes, *understands* or *believes*. For example, participants assigned to the abortion groups received the following vignette[12]:

> Jane is a 24-year-old graduate student. She and her friend Sandy are outside a clinic that performs abortions. Jane intends to end her pregnancy because she doesn't want to be a parent, but Sandy argues that an abortion for that reason is morally wrong. Jane replies that it's controversial whether abortion to avoid parenthood is immoral, and Sandy concedes that there's currently no consensus on the matter, but Sandy continues to make her case. After listening to Sandy's argument, Jane says, "I agree, Sandy – ending my pregnancy to avoid parenthood is morally wrong; however, I don't care at all if I do what's morally wrong." After saying this, Jane enters the clinic and a doctor ends her pregnancy as she requests.

Those assigned to the marijuana groups received a similar vignette, except the abortion transgression was replaced with a case of recreational marijuana use and a discussion between Chris and Bill on its moral status. After reading their vignette, participants were presented with a series of statements concerning whether the amoralist agent in the story understands or believes that the transgression is morally wrong. Specifically, participants were asked to indicate to what degree on a scale from 1 (strongly disagree) to 7 (strongly agree) they agreed or disagreed with the following statements[13]:

(1) Jane understands/believes that ending her pregnancy to avoid parenthood is morally wrong.
(2) Jane understands/believes that ending her pregnancy to avoid parenthood is unethical.
(3) Jane understands/believes that she shouldn't end her pregnancy to avoid parenthood.

Finally, participants indicated their agreement or disagreement with a statement concerning their personal view of the transgression in the story:

(1) In my personal view, it's morally wrong for anyone to end their pregnancy to avoid parenthood.

Responses were cut from the data if the participant responded with anything lower than a five on this personal view statement – the lowest degree of agreement allowed by the scale.

4.1.4. Results

Out of 400 participants, only 77 gave a response of 5 or above on the personal view statement (marijuana, $N = 25$; abortion, $N = 52$). If the ICR hypothesis were correct, then these participants should have denied both understanding and belief to the amoralists. If they had done this, FE would have disappeared, thus providing some reason to think FE is simply an experimental artifact.

But this is not what happened. Even with the dramatic loss in power resulting from removing participants based on their responses to the personal view statement, FE was still detectable (Items Averaged: $F(1, 77) = 16.32$, $p < .001$, $\eta^2 = 183$, $\alpha = .72$). The Items Averaged results are illustrated in Figure 1.[14]

4.1.5. Discussion

FE remained despite the manipulations. Participants continued to attribute moral understanding, but deny moral belief, to the amoralist. This result conflicts with the ICR hypothesis, thus providing no support for a deflationary explanation of FE.

One might reformulate ICR by removing any positive claims about the nature or content of the judgement that the amoralist succeeds in making.[15] The ICR hypothesis could then be reformulated to reflect this more minimal explanation of FE: participants mistakenly grant understanding to the amoralist for *some* reason, not necessarily because they confuse it with sociological knowledge of a moral consensus.

But what this minimal version gains in plausibility, it lacks in testability. Since nothing definite is said about where the mistake may lie, it's difficult to translate this explanation into a testable hypothesis.[16]

At any rate, this first follow-up study failed to confirm one plausible, though deflationary, account of FE. The next follow-up study tests a different deflationary explanation that relies on assuming a kind of folk *externalism*.
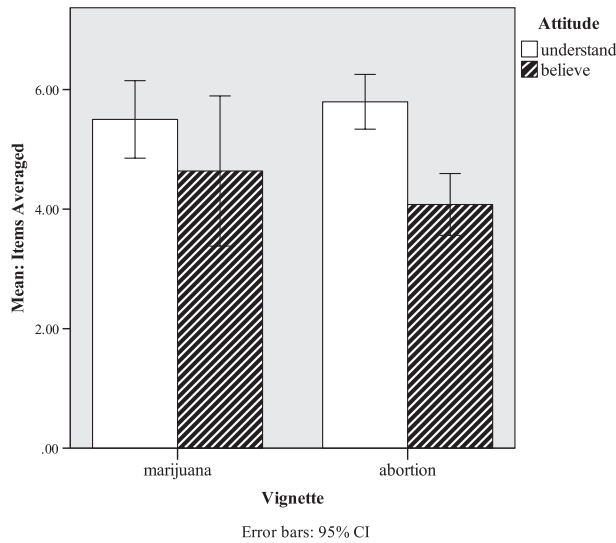
Error bars: 95% CI

Figure 1. Items averaged.

4.2. Deflationary explanation 2: the dispositional belief hypothesis

4.2.1. Background

A deflationary explanation for FE can be drawn from recent experimental work on what is called the *entailment thesis*: knowledge entails belief. Myers-Schulz and Schwitzgebel (2013) offer counterexamples to the entailment thesis by way of four studies. One study, called *unconfident examinee*, resulted in a significant majority of participants ascribing to Kate (the examinee) the knowledge that Queen Elizabeth died in 1603 *while denying* that Kate believes this proposition. The asymmetry in folk ascriptions in this study looks eerily similar to FE. The participants appear to be granting a factive mental state while denying the traditionally entailed non-factive mental state.

In response to Myers-Schulz and Schwitzgebel (2013), Rose and Schaffer (2013) argue that none of these four studies present genuine counterexamples to the entailment thesis when conceived in a *dispositional* sense, that is, that knowledge entails *dispositional* belief. For instance, concerning *unconfident examinee*, Rose and Schaffer show that participants were willing to ascribe *both* knowledge and belief to Kate when they were first asked to evaluate a sleeping individual's doxastic states (p. S37).

The apparent similarity between FE and the above asymmetry suggests a common cause. If Rose and Schaffer (2013) have discovered the cause of the asymmetry in Myers-Schulz and Schwitzgebel (2013), then it is worth testing whether this cause also explains FE. This account presumes a kind of folk externalism which explains ascriptions of understanding to the amoralist. The reluctance to ascribe moral belief is then explained as a myopic focus on the amoralist's *occurrent* mental states. This can be called the *dispositional belief hypothesis*. If correct, FE should then disappear when participants are primed with examples of dispositional belief.

4.2.2. Method

4.2.2.1. Participants.
Three hundred and seven participants completed a study on Amazon Mechanical Turk, spanning a wide age range.[17] The mean age was 31; with 106 female participants, 198 male participants, two participants of other genders, and one not specified.

4.2.2.2. Procedures and materials.
Subjects were split into four groups. Participants in each group were assigned one of two valences (*rightness* or *wrongness*) and one of two attitudes (*understands* or *believes*).[18] For example, participants assigned to the wrongness groups received the following vignette:

Jane is a 24-year-old graduate student. She and her friend, Sandy, are in line to buy some expensive headphones. Unbeknownst to the clerk, the headphones mistakenly ring up as $15 instead of $150. Sandy argues that buying the headphones at the mistaken price is morally wrong. After listening to Sandy's argument, Jane says, "I agree, Sandy – buying the headphones at the mistaken price is morally wrong; however, I don't care at all if I do what's morally wrong." After saying this, Jane pays $15 and leaves with the headphones. When they get back to their apartment, Sandy takes a nap while Jane listens to music with her new headphones.

Those assigned to the rightness groups received a similar vignette, except that the wrongness judgement was converted into a rightness judgement ("telling the clerk about the mistake is the morally right thing to do"). After reading the vignette, participants were presented with a series of statements concerning whether both the sleeping agent and the awake amoralist in the story understand or believe that the transgression is morally wrong or right. Specifically, participants were asked to indicate to what degree on a scale from 1 (strongly disagree) to 7 (strongly agree) they agreed or disagreed with the following statements[19]:

(1) Sandy (despite being asleep) understands/believes that buying the headphones at the mistaken price was morally wrong.
(2) Sandy (despite being asleep) understands/believes that buying the headphones at the mistaken price was unethical.
(3) Sandy (despite being asleep) understands/believes that Jane shouldn't have bought the headphones at the mistaken price.
(4) Jane understands/believes that buying the headphones at the mistaken price was morally wrong.
(5) Jane understands/believes that buying the headphones at the mistaken price was unethical.
(6) Jane understands/believes that she shouldn't have bought the headphones at the mistaken price.

4.2.3. Results

Unlike in Rose and Schaffer (2013), the introduction of a sleeping agent did not alter the results in any significant way. FE remained despite this manipulation. Participants granted that sleeping Sandy both understands and believes that X is morally right or morally wrong (Sleeping Sandy Averaged: $F(1,307) = .048$, $p = .83$, $\eta^2 < .001$, $\alpha = .7$), but they continued to deny that Jane believes (Amoralist Jane Averaged: $F(1,307) = 94.16$, $p < .001$, $\eta^2 = .237$, $\alpha = .7$). Figures 2 and 3 illustrate the difference in the Items Averaged responses to Sleeping Sandy and Amoralist Jane.

4.2.4. Discussion

This is a rather surprising result for two reasons. First, it's surprising that the Rose and Schaffer-inspired sleeping manipulation didn't cause FE to disappear, given how similar FE looks to the effect in Myers-Schulz and Schwitzgebel (2013). Secondly, this result seems to suggest that even the dispositional entailment thesis fails when it comes to moral knowledge and moral belief. A denial of the entailment thesis means that the amoralist might know, but still not really believe, that X is morally wrong. Oddness notwithstanding, the results of the final study arguably suggest taking seriously the possibility that the entailment thesis fails within moral contexts. This possibility will be discussed in section 5.

Let's take stock. Thus far, two deflationary explanations of FE have failed to be confirmed.[20] The final study we will consider tests a substantive explanation of FE. As will be seen, this is the only study of the three to receive confirmatory results.

4.3. A substantive explanation: the moral emotions hypothesis

4.3.1. Background

Deflationary concerns notwithstanding, the factivity effect suggests that folk thought regarding amoralist skepticism is divided. Amoralism seems more coherent if the amoralist is described as
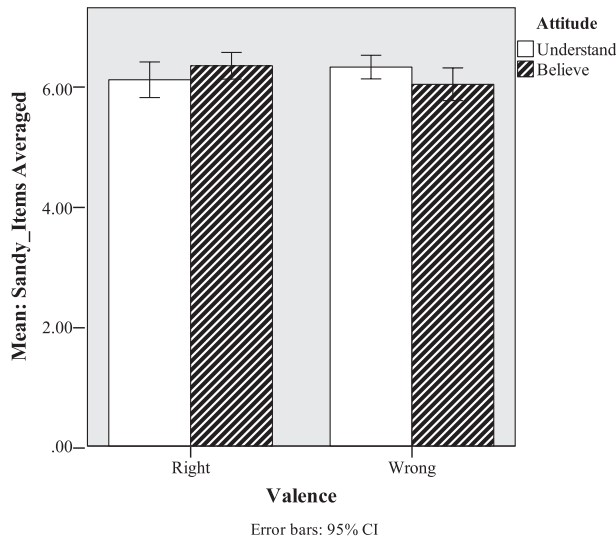
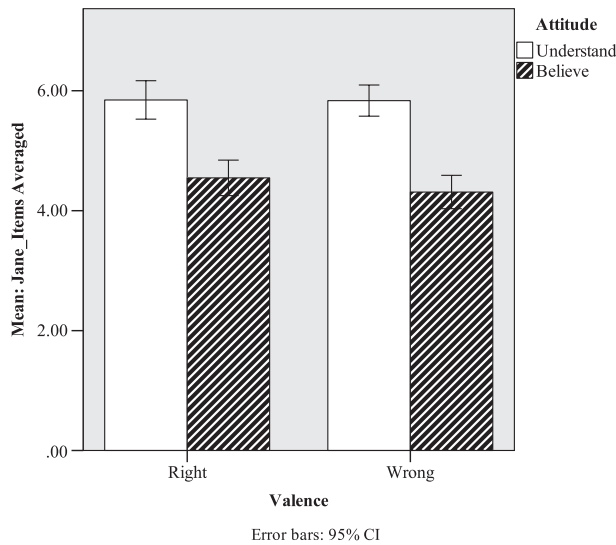**Figure 2.** Sleeping Sandy – Items averaged.

Figure 3. Amoralist Jane – Items averaged.

understanding or knowing that the behavior is morally wrong, but less coherent when described as believing or thinking that the behavior is morally wrong. This division could be explained if it reflected a feature of the structure of folk thought *itself*. But what might this feature be?

One suggestion is that this feature involves a folk association between moral emotion and moral belief that is absent for moral knowledge.[21] When we realize we've done something wrong, we normally feel some twinge of guilt or regret. We feel bad about what we have done. But an amoralist doesn't feel these feelings of regret and guilt precisely because of her amoralism. If people associate these traditional moral feelings with moral belief, then perhaps this would explain their reluctance to ascribe moral belief to the amoralist. And if people don't associate these feelings with factive mental states like knowledge, then we would seem to have a good explanation of FE. On this hypothesis, the folk should be more willing to attribute classic moral feelings to amoralists when they are stipulated

as believing, but not when they are stipulated as understanding, that X is morally wrong. This can be called the *moral emotions hypothesis*.

4.3.2. Method

4.3.2.1. Participants.
One hundred and thirty-nine participants completed a study on Amazon Mechanical Turk, spanning a wide age range.[22] The mean age was 31, with 46 female participants, 92 male participants, and one participant of another gender.

4.3.2.2. Procedures and materials.
Subjects were split into four groups. Participants in each group were assigned one of two vignettes (*$20 bill* or *headphones*) and one of two attitudes (*understands* or *believes*). To illustrate, participants assigned to *headphones* received the following vignette:

> Unbeknownst to the clerk, the headphones Jane wants to buy mistakenly ring up as $15 instead of $150. Jane [understands/believes] that buying the headphones at the mistaken price is morally wrong. Nevertheless, Jane proceeds to pay $15 and leaves with the headphones.

Those assigned to *$20 bill* received a similar vignette, except the relevant transgression was replaced by the theft of $20. Notice that the agent's cognitive attitude is explicitly *stipulated* in the vignette. This was an effort to gather ascriptions of the agent's emotional state rather than her cognitive state (unlike the previous studies).[23] After reading the vignettes, participants were presented with a series of statements concerning how the amoralist agent in the story felt. Specifically, participants were asked to indicate to what degree on a scale from 1 (strongly disagree) to 7 (strongly agree) they agreed or disagreed with the following statements[24]:

(1) Jane felt bad about buying the headphones.
(2) Jane felt regret about buying the headphones.
(3) Jane felt guilty about buying the headphones.

4.3.3. Results

The hypothesis was confirmed, though the effect size was relatively small; Items Averaged: ($F(1,139) = 6.54$, $p = .01$, $\eta^2 = .046$, $\alpha = .91$). Figure 4 illustrates the results of the Items Averaged.

While an effect appears to be trending in *$20 bill*, *headphones* shows a clearer effect in the direction predicted by the hypothesis.
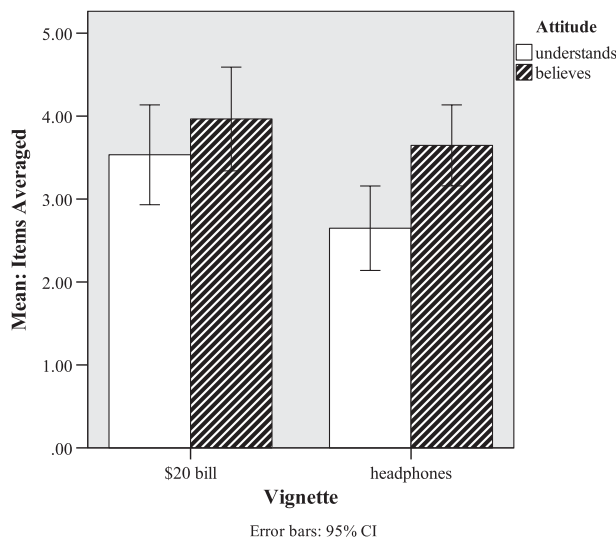

Error bars: 95% CI

Figure 4. Items averaged.

4.3.4. Discussion

Unlike the previous studies, this study's hypothesis was confirmed. Although the effect size was relatively small, the fact that this substantive explanation succeeds where the two deflationary explanations failed is intriguing. It seems to suggest that perhaps FE cannot be brushed aside as merely an experimental artifact. This seems to have serious implications for metanormative inquiry regarding moral internalism and amoralist skepticism.[25]

5. General discussion

Three conclusions seem supported by the experiments addressed in this paper:

(1)   There is a factivity effect. People's intuitions lean toward externalism when evaluating factive, amoralist scenarios, whereas people's intuitions lean towards internalism when evaluating non-factive, amoralist scenarios.
(2)   Two deflationary explanations of FE were not confirmed.
(3)   There is some evidence that FE occurs because people associate moral emotions with non-factive states but *not* with factive states.

These experimental results seem in tension with DE. Nevertheless, one might respond to these results in one of three ways:

(1)   People *mistakenly* attribute factive states to the amoralist.
(2)   People *mistakenly* deny non-factive states to the amoralist.
(3)   People *correctly* attribute factive states while denying non-factive states to the amoralist.

In this section, I will discuss each way of responding to the experimental results. While the current data admittedly underdetermines all three of these positions, they still seem to deserve attention. At the end of this section, I will discuss some limitations of the present research and offer suggestions for future research.

5.1. People mistakenly attribute factive states to the amoralist

Internalists deny or qualify amoralism. So regarding FE, they would likely view people's ascriptions of factive states to the amoralist as mistaken. But how might the data support this view? One way is to show that such results seem more easily explained by internalism than externalism. To see this, it's helpful to note that the participants in Björnsson and colleagues (2015) study mostly withheld attributions of moral belief to agents who lacked appropriate motivation *unless there was some condition that explained the lack of motivation*. A straightforward explanation of these results, they claim, would be a default commitment to a conditional form of internalism (they suggest something like Smith's (1994) practicality requirement).[26] Moreover, the results of the studies in the present paper could arguably support internalism. The moral emotion study in particular could reasonably be construed as identifying an internalist component within folk cognition. Perhaps some inclination towards regret is necessary when one fails to live up to one's own moral beliefs, provided those beliefs are genuine. This seems much more in-line with a broadly internalist picture of the relationship between doxastic states and morality, at least on the plausible assumption that moral emotions are not motivationally inert.

5.2. People mistakenly deny non-factive states to the amoralist

To show that people mistakenly deny non-factive states to the amoralist, one might appeal to the entailment thesis that knowledge entails belief. Insofar as participants are willing to ascribe certain factive states to the amoralist, they should be willing – on pain of *incoherence* – to ascribe non-factive states to the amoralist.

Now, one might question whether the entailment thesis is true, or whether its scope includes any and all factive and non-factive states. But on the strongest reading of the entailment thesis, any factive state that an agent is in entails that the agent is in a corresponding non-factive state. And it's traditionally held to be more difficult to obtain a factive state than a non-factive, given that presumably *more* must be added to the non-factive state to make it factive.

However, there appears to be some evidence that knowledge is more easily attained, and more easily ascribed, than belief. This is a highly unexpected result if the entailment thesis is true. One would predict that ascribing belief would be quite simple in cases in which the agent clearly has knowledge; the entailment thesis allows for belief to be easily read off such scenarios. But current experimental work on this prediction suggests that the matter is more complex (Phillips, Knobe, & Cushman, 2015). At any rate, it's not clear that an appeal to the entailment thesis alone will be enough to show that people are mistaken to deny non-factive states to the amoralist.

5.3. People correctly attribute factive states while denying non-factives states to the amoralist

One might think that people aren't making any mistakes: they're right to attribute factive states to the amoralist and they're also right to deny non-factive states to the same amoralist. As the results from the moral emotion study suggest, people seem to take some moral emotions to be requirements for non-factive states. However, they don't seem to take these same emotions to be required for factive states. So when presented with an individual that seems to have moral knowledge while lacking this emotional component, people describe this person as having moral knowledge without having the associated moral belief.

But to make sense of this account of moral judgement, we will need to make sense of how it can be that non-factive states aren't entailed by factive states. A start might be to show that moral knowledge and belief do not, or need not, fall within the scope of the entailment thesis – the view that knowledge entails belief. But since knowledge and belief seem to fit quite naturally with truth-apt, declarative claims, it seems this task will be more challenging for some accounts of morality (e.g. cognitivism) than others.

But perhaps the results from the moral emotion study can be instructive here. Attention to the idea that non-factive states require moral emotions, whereas factive states do not, may reveal how factive ascriptions to an amoralist can be correct even when non-factive ascriptions to that same amoralist are false. Consider again the scenario in which the amoralist, Chris, keeps for himself a $20 bill that he just saw someone in front of him accidently drop. In this case, people are inclined to say that Chris knows that this act of theft is morally wrong, yet they want to deny that Chris really believes, or thinks, that this action is wrong. And as the results of the moral emotion study suggest, people want to say that Chris needn't feel any regret or guilt if he knew that the action was wrong, but he would need to if he believed that the action was wrong.

Maybe what's happening is that some normative evaluations of Chris's actions don't require that Chris feel a certain way about his behavior, whereas others do require that Chris have certain relevant moral emotions. Suppose we wanted to blame Chris for what he did. In that case, Chris's knowledge may be enough for him to be morally blameworthy; it's just not relevant whether Chris feels guilty or not about the behavior (at least, on the folk conception). And, more importantly, it's irrelevant if Chris doesn't *believe* it's wrong.[27]

However, if it's true that Chris thinks that this theft is morally wrong, then perhaps the normative evaluation shifts from *moral responsibility* to *rationality*: if Chris thinks this, and does it anyway, then he will feel guilty for what he's done (on pain of rationality). On this account, participants find that the only way to really make sense of Chris, if he really believes that this theft is wrong, is to assume he must feel some of the standard moral emotions that come with such moralizing thoughts, insofar as they are genuine and sincere.

But if it's possible to hold someone morally accountable without their having any of the traditional moral emotions insofar as they have knowledge, and if it's not possible to attribute genuine moral beliefs to someone without their experiencing such states, it does seem that the entailment thesis fails in this context. To be sure, metanormative theorists may wish to offer a revisionist account of some of these concepts—either moral responsibility, moral belief, or both—that would respect the entailment thesis. But if the results of the moral emotions study suggest something like the sketch above, and if metanormative theorists are trying to capture commonsense normative concepts and relationships, then perhaps the plausibility of the entailment thesis deserves more scrutiny within moral contexts.

5.4. Limitations and future research

One limitation of this research (and much experimental philosophical research) is the relatively small number of vignettes used in each study. For instance, given the relatively small effect size for the results in the moral emotions hypothesis study, questions of whether the effect would survive more numerous stimuli seem appropriate. Increasing the statistical power of these experiments should be a focus of future research.[28]

Speaking of these vignettes, a further limitation of this research concerns the variety of moral transgressions covered. For the most part, the studies in this paper address only *fairness*-based transgressions such as stealing money or not paying full-price for headphones (the abortion vignette notwithstanding). Showing that FE isn't merely tied to some peculiar feature of fairness-based moral transgressions would speak to its robustness. Thus, another focus of future research might be to employ *harm*-based and *rights*-based moral transgressions in future vignettes.

Finally, there is still some conceptual work to do regarding *how* exactly FE conflicts with DE. DE and FE are *not* mutually exclusive, so their conflict is not likely one of logical inconsistency. However, DE seems to warrant the prediction that the intuitive reactions of people competent with the relevant concepts will remain *consistent* across the factive/non-factive divide. For instance, assuming DE is true, if participants have externalist reactions in factive contexts, one may reasonably expect, *ceteris paribus*, that these externalist reactions will remain in non-factive contexts. But as Björnsson and colleagues (2015) seem to show, this prediction fails to be confirmed. Instead, what these researchers found was FE. So while FE does not contradict DE, if DE warrants the above prediction, the presence of FE should count as *inductive* evidence against DE. In this way, FE seems surprising on the assumption that DE is true. So if DE is true, and DE does in fact warrant the above prediction, then FE must have some deflationary explanation. Yet, as I've shown in this paper, two initially promising deflationary accounts of FE failed to be confirmed. Instead, a *non*-deflationary account of FE, the moral emotions hypothesis, was the only account that received confirmation. To be sure, the moral emotions hypothesis doesn't necessarily conflict with DE any more than FE does. But non-deflationary explanations of FE are arguably less sympathetic to DE than deflationary accounts, provided that the participants are competent with the relevant concepts.[29,30]

6. Conclusion

Both internalists and externalists traditionally assume descriptive equivalence (DE), that is, that factive descriptions of the amoralist are equivalent to non-factive descriptions of the amoralist, in the sense that they can be interchanged without loss (or gain) in the scenario's overall conceptual coherence. However, externalists seem to favor factive descriptions whereas internalists seem to favor non-factive descriptions.[31] Interestingly, non-philosophers lean toward externalism when the amoralist scenario is described using factives, yet lean toward internalism when the same scenario employs non-factives instead. This is the factivity effect (FE). Because FE seems to be in tension with DE, defenders of DE may wish to explain away FE as an experimental artifact. But two deflationary explanations of FE failed to receive confirmation. A substantive explanation of FE – that a feature within folk cognition requires moral emotion for genuine moral belief – was the only hypothesis of the three that received

confirmation. Implications of these results for the traditional debate were considered. While some implications seem to leave the traditional debate intact (for example, that externalism may be more theory-driven than internalism), others seem to call for a radically different approach (for example, that the entailment thesis fails for moral knowledge). That the factivity effect appears to have some crucial implications, and that it doesn't appear (so far) to have a deflationary explanation, should be of interest to those engaged in metanormative inquiry and moral psychology.

Notes

1. Bold italics were added; non-bold italics are in the original text.
2. For a third externalist example, see Boyd (1988, p. 216).
3. For an example in which non-factives are used when describing internalism, but factives used when describing externalism, see Korsgaard (1986, pp. 8–9).
4. See section 4.2.
5. Hare, 1969, p. 164, 1981, p. 183.
6. But what about the sizable minority of participants that *do* ascribe belief to the amoralist? Presumably all experimental studies have some unexpected results. When we're reasonably optimistic about the study, these unexpected results are traditionally called *noise* – unanticipated variations in the data that do not necessarily tell against the effect borne out by the data. Thus, (as a first pass) the divergent responses in the above follow-up studies are likely noise.
7. Thanks to an anonymous reviewer for this explanation.
8. Note that it needn't follow from this that there is an innate moral grammar analogous to the Chomskian notion of innate grammar. See Alfano and Loeb (2014) for further discussion of this analogy.
9. Because these studies were between-subject, multiple responses from the same IP address were cut. The cutting procedure was effectively random; the IP addresses for all responses in the study were ordered from lowest to highest, and the first response of a set of identical IP addresses was kept while the others were cut. For this study, 18 responses were cut due to redundant IP addresses.
10. Given the large number of responses cut because participants didn't personally view either of the moral transgressions as morally wrong, the study was run a second time to make up for this dramatic decrease in statistical power. And because participants took longer, on average, to complete the survey than anticipated, the compensation was increased on the second run.
11. Note that these studies all employ *between-subject* designs. FE also occurred in a within-subject variation of a preliminary study, with test questions counter-balanced to avoid possible order effects.
12. For the complete materials for this study, please see the supplemental materials.
13. Participants received two reading comprehension questions: "Is it stated in the paragraph above that Jane says the following: 'I agree, Sandy – ending my pregnancy to avoid parenthood is morally wrong'? Y/N" and "Is it stated in the paragraph above that Jane understands/believes that ending her pregnancy to avoid parenthood is morally wrong? Y/N" The original intent of adding these questions to each study was to cut participants from the study that replied "no" to the first question or "yes" to the second question. While almost every participant replied "yes" to the first question, very few participants replied "no" to the second question. For instance, only nine participants answered the second question correctly, whereas only two failed the first. Counterbalancing the order of the presentation of the questions did not affect this result. After this occurred for all three studies, it was decided that there was a problem with the question itself (perhaps an ambiguity in the word "stated"). For this reason, responses to these questions were ignored for all three studies. The use of the comprehension question in the moral emotions study was primarily to focus participant attention on which word – factive or non-factive – was employed in the vignette.
14. While there may have been little to no significant difference between the two attitude levels for the marijuana factor, this is likely because the sample size was only 25, which is admittedly quite small.
15. Smith (1994) says, "The very best we can say about amoralists is that they try to make moral judgments but fail" (p. 64).
16. An anonymous reviewer offers another possibility: the reference of "morally wrong" in the mouth of Jane needn't be the general consensus on what's morally wrong, but rather what the general population *ought* to believe given society's commitments on other moral matters. For example, suppose Jane knew that the general consensus on infanticide was that it was morally wrong, and then suppose that Jane believed that, for people to have *consistent* moral views, they should view abortion as morally wrong. If participants attributed something like this reasoning to Jane, they may still have interpreted Jane's use of "morally wrong" in an inverted commas sense *even if* my attempt at blocking participants from attributing knowledge of a general consensus about abortion to Jane was successful. One worry about this explanation is that it requires attributing a relatively complex reasoning process to participants. To conclude that Jane may still be using "wrong" in an inverted commas sense while

granting Jane's knowledge that there's no general moral consensus on abortion, participants presumably must engage in some fairly sophisticated reasoning about Jane's thoughts on the matter. But it seems unlikely that participants would have done this.

17. 93 responses were cut from an original 400 due to redundant IPs (see endnote 9).

18. The focus on valence in this study is a residual from previous preliminary experiments in which it was hypothesized that a *rightness effect*, analogous to the factivity effect, may also be present in people's responses to amoralist scenarios. While this hypothesis did not receive confirmatory results, the presence of interaction effects made analysis difficult.

19. Participants received three reading comprehension questions: "Is it stated in the paragraph above that Jane says the following: 'I agree, Sandy – buying the headphones at the mistaken price is morally wrong'? Y/N", "Is it stated in the paragraph above that Jane believes that buying the headphones at the mistaken price is morally wrong? Y/N", and "Is it stated in the paragraph above that Sandy argues that buying the headphones at the mistaken price is morally wrong? Y/N". All participants failed the third question, whereas only 38 failed the second and 37 failed the first. But see endnote 13.

20. One further deflationary hypothesis would be that the factivity effect is merely localized to the words "understand" and "believe". Maybe FE would disappear if different factive and non-factive words were used. This hypothesis was tested by presenting a vignette similar to the headphones scenario, but the test question replaces "understands" and "believes" with "knows" and "thinks", respectively. Ninety-three participants completed this study on Amazon Mechanical Turk, spanning a wide age range. After reading the vignette, participants were asked to indicate to what degree, on a scale from 1 (strongly disagree) to 7 (strongly agree), they agreed or disagreed with the following statement: "Jane knows/thinks that she shouldn't buy the headphones at the mistaken price". The hypothesis was not confirmed. FE persisted despite the replacement factive and non-factive words (One-way ANOVA: $F(1,93) = 73.46$, $p < .001$, Cohen's $d = 1.8$; Knows, $N = 44$, Mean = 6.00, S.D. = 1.49; Thinks, $N = 49$, Mean = 2.71, S.D. = 2.11).

21. Laura King suggested this hypothesis.

22. Sixty-one responses were cut from an original 200 due to redundant IPs (see endnote 9).

23. While the moral emotions served as dependent variables in this study and the cognitive attitude as an independent variable, there wasn't an experimental or philosophical reason for this assignment. The hypothesis and prediction for this study would remain the same if the roles were switched.

24. Participants also received one reading comprehension question: "Is it stated in the paragraph above that Jane believes that buying the headphones at the mistaken price is morally wrong? Y/N". Only two participants failed to give the correct answer to this question.

25. To be sure, it is worth testing whether participants will, *upon reflection*, give up their initial responses for responses that more clearly respect the entailment thesis. Kenny Boyce suggested this deflationary explanation.

26. See section 2 for Smith's practicality requirement. I presented one study in detail earlier in the paper, but Björnsson and colleagues (2015) did run a number of other studies, results of which seem to suggest (though by no means conclusively) this folk default reliance on something like conditional internalism (that is, a conceptual connection between judgement and motivation, absent certain conditions such as listlessness and depression).

27. Precedence for this idea can be found within Christianity. In Romans 1:18-21, Paul seems to argue that atheists have no excuse because God has made Himself known to them. "Since what may be known about God is plain to them, because God has made it plain to them" (1:19; New International Version). Another interpretation of Paul here is that he views God as having provided sufficient evidence regarding His existence, such that the atheist's denial of God is either not genuine or the result of epistemic negligence.

28. Thanks to an anonymous reviewer for pressing this limitation.

29. One might deny that the participants are competent with the relevant concepts. See Kauppinen (2007) for an example, but see Nadelhoffer and Nahmias (2007) for a reply to this general concern. While the worry likely gets traction with some experimental philosophy projects, I think the present project concerning moral judgement is an exception. While space limitations prevent a developed reply, here's a quick sketch. Participants are likely competent in their folk psychological ability to attribute moral responsibility, and this capacity arguably requires the capacity to determine if the agent *knew* that what they were doing was morally wrong. This latter capacity arguably requires competence with the concept of moral judgement, particularly in terms of attributing (or withholding attributions of) moral judgments to agents.

30. Thanks to an anonymous reviewer for pressing me on the relationship between DE and FE. The reviewer may have had a stronger worry in mind – that *no* amount of inductive evidence could affect the plausibility of DE. While I'm sympathetic to this kind of concern generally, I think there's reason to view some metaethical contexts as exceptions. For more on this issue, see the exchange between Loeb (2008a, 2008b), Gill (2008), and Sayre-McCord (2008) over metaethical questions and empirical inquiry.

31. See section 2, and endnotes 2 and 3. Admittedly, without an exhaustive sampling of the literature, this appearance is only suggestive.

Acknowledgements

Disclosure statement

ORCID

Kenneth Shields 🆔 http://orcid.org/0000-0003-2596-8168

References

Alfano, M., & Loeb, D. (2014). Experimental moral philosophy. In E. N. Zalta (Ed.), *The stanford encyclopedia of philosophy* (Summer ed.). Retrieved from http://plato.stanford.edu/archives/sum2014/entries/experimental-moral/

Björnsson, G., Eriksson, J., Strandberg, C., Olinder, R. F., & Björklund, F. (2015). Motivational internalism and folk intuitions. *Philosophical Psychology, 28*, 715–734.

Boyd, R. (1988). How to be a moral realist. In G. Sayre-McCord (Ed.), *Essays on moral realism* (pp. 187–228). Ithaca, NY: Cornell University Press.

Brink, D. (1989). *Moral realism and the foundations of ethics*. New York, NY: Cambridge University Press.

Dreier, J. (1990). Internalism and speaker relativism. *Ethics, 101*, 6–26.

Gill, M. B. (2008). Metaethical invariability, incoherence, and error. In W. Sinnott-Armstrong (Ed.), *Moral psychology, vol. 2 – The cognitive science of morality: Intuition and diversity* (pp. 387–402). Cambridge, MA: MIT Press.

Hare, R. M. (1969). *The language of morals*. Oxford: Oxford University Press.

Hare, R. M. (1981). *Moral thinking: Its levels, method, and point*. Oxford: Clarendon.

Kauppinen, A. (2007). The rise and fall of experimental philosophy. *Philosophical Explorations, 10*, 95–118.

Korsgaard, C. M. (1986). Skepticism about practical reason. *The Journal of Philosophy, 83*, 5–25.

Loeb, D. (2008a). Moral incoherentism: How to pull a metaphysical rabbit out of a semantic hat. In W. Sinnott-Armstrong (Ed.), *Moral psychology, vol. 2 – The cognitive science of morality: Intuition and diversity* (pp. 355–386). Cambridge, MA: MIT Press.

Loeb, D. (2008b). Reply to Gill and Sayer-McCord. In W. Sinnott-Armstrong (Ed.), *Moral psychology, vol. 2 – The cognitive science of morality: Intuition and diversity* (pp. 413–422). Cambridge, MA: MIT Press.

Myers-Schulz, B., & Schwitzgebel, E. (2013). Knowing that p without believing that p. *Nous, 47*, 371–384.

Nadelhoffer, T., & Nahmias, E. (2007). The past and future of experimental philosophy. *Philosophical Explorations, 10*, 123–149.

Nichols, S. (2002). How psychopaths threaten moral rationalism: Is it irrational to be amoral? *Monist, 85*, 285–303.

Phillips, J., Knobe, J., & Cushman, F. (2015). *Knowledge before belief: Response times indicate evaluations of knowledge prior to belief*. Paper presented at the annual meeting of the Society for Philosophy and Psychology, Durham, NC. June 2015.

Rose, D., & Schaffer, J. (2013). Knowledge entails dispositional belief. *Philosophical Studies, 166*, 19–50.

Sayre-McCord, G. (2008). Moral semantics and empirical enquiry. In W. Sinnott-Armstrong (Ed.), *Moral psychology, vol. 2 – The cognitive science of morality: Intuition and diversity* (pp. 403–412). Cambridge, MA: MIT Press.

Smith, M. (1994). *The moral problem*. Cambridge, MA: Blackwell.

Svavarsdóttir, S. (1999). Moral cognitivism and motivation. *The Philosophical Review, 108*, 161–219.