

Motivational internalism and folk intuitions

Gunnar Björnsson, Umeå University, University of Gothenburg

John Eriksson, University of Gothenburg

Caj Strandberg, University of Gothenburg

Ragnar Francén Olinder, University of Gothenburg

Fredrik Björklund, Lund University

1. Introduction

Suppose that you and a friend are debating whether the two of you, and others in your situation, ought to donate money to charity. After a while, she says that she agrees with you that there is such an obligation. However, it turns out that she is not the least motivated to donate even the smallest amount of money. This may lead you to think that she wasn't really convinced by your argument and that her agreement with you was insincere. Her lack of motivation suggests a lack of moral conviction. At least under normal circumstances, we expect moral judgment and moral motivation to be intimately connected.

The exact nature of this connection is one of the most debated issues in metaethics. Internalism takes the connection between moral judgments and motivation to be metaphysically or conceptually necessary: a person *cannot* be making a moral judgment unless she is at least somewhat motivated to act accordingly. The following is a simple version of internalism about judgments of moral wrongness:

Internalism: It is necessary that if a person judges that it is morally wrong for her to φ , then she is, at least to some extent, motivated to refrain from φ -ing.

This formulation might in turn be modified and complicated in various ways. For instance, according to a *conditional* version of internalism, the connection between moral judgment and motivation holds only if the person fulfills a certain further condition, such as being fully rational or psychologically normal.¹ Externalism can be understood as the claim that there is no necessary connection of this kind.

Internalism is assumed to have significant implications for the nature of moral thought, the meaning of moral terms, and the possibility of objective truth and knowledge in morality. Most importantly, it has been maintained that internalism in combination with the so-called Humean theory of motivation entails that cognitivism is false—that moral judgments cannot consist in ordinary representational beliefs—and provides support for non-cognitivist theories according to which moral judgments consist in desire-like states. Conversely, it has been maintained that if externalism is correct, then moral judgments cannot consist in desire-like states, as such states are necessarily motivational (e.g. Shafer-Landau 2003: ch. 6).

In arguing for and against internalism, metaethicists traditionally appeal to intuitions about whether people really do make moral judgments when the relevant kind of motivation is absent, i.e. appeal to intuitions about whether “amoralists” are possible: internalisms of various kinds rule out the corresponding kind of amoralist. However, metaethicists’ fundamental intuitions about such cases often come in conflict, and there are at least two ways in which empirical investigations of folk intuitions might be helpful in resolving that conflict: by revealing theoretical bias among metaethicists, and by indicating whether people

¹ For an overview of various internalist positions, see Björklund et al. (2012). For different versions of conditional internalism, see e.g. Dreier (1990); Smith (1994); Korsgaard (1996), Blackburn (1998: Ch. 3), Björnsson (2002), and Gibbard (2003: ch. 7). For defenses of externalism, see e.g. Brink (1989: ch. 3), Strandberg (2011), Svavarsdóttir (1999), and Zangwill (2008).

in general operate with the sort of internalist conception of moral judgment that would be required for an understanding of moral disagreement as disagreement in attitude, i.e. the sort of understanding that noncognitivists stipulate.

Consider first the possibility of *revealing theoretical bias*. Conflicting intuitions raise the possibility that one party of the disagreement is making a mistake in thinking about putative amoralists. Perhaps people who think that amoralists are possible are (a) unconsciously assuming some *background connection* to motivation (see Bedke 2009: 191–95; Blackburn 1998: 59–68; Dreier 1990: 9–14; Tresan 2006: 149–52, 2009: 185–93), or (b) conflating the *possibility* of amoralists with the *pragmatic appropriateness* of extending talk of moral judgments to cases that do not involve moral judgments proper but are otherwise similar (Björnsson 2002: §6; cf. Hare 1952: 124–25, 164–65). Or perhaps people who think that amoralists are impossible are (c) not really thinking about *moral judgments* proper, but rather about certain *emotionally laden reactions* that characteristically accompany moral judgments in paradigmatic cases of the latter (Kauppinen forthcoming), or (d) conflating the *unlikeliness* of sincere amoralists with their *impossibility* (Svavarsdóttir 1999), or (e) conflating the pragmatic inappropriateness of making moral claims when lacking the corresponding motivation with the impossibility of moral judgments unaccompanied by motivation (Finlay 2004, 2005, Strandberg 2012a).

Such mistakes would not be entirely surprising given that metaethicists often have theoretical reasons for favoring either internalist or externalist positions. Noncognitivists and Kantian constructivists build their theories around strong assumptions about the practical consequences of moral judgments, seemingly leaving little room for amoralists. Realists instead build their accounts around beliefs representing moral facts that exist independently of our concerns and interests, leaving it unclear how motivation would enter the picture in a privileged way required by internalism. Given the general human propensity for confirmation

bias, we might expect a tendency for both parties to latch on to intuitions about the relevant cases that seem to conform to their theoretical preconceptions, being somewhat insensitive to the assumptions on which these intuitions are based, or the objects they are directed at.

To minimize the role of confirmation bias, the standard argumentative strategy is to make cases of putative amoralists even more explicit, and to see whether the various error theories proposed for internalist or externalist intuitions can be plausibly spelled out (e.g. Blackburn 1998: Chs. 1 & 3, Svavarsdóttir 1999). But another, complementary, strategy, and one that we will pursue here, is to consult the intuitions of people who might have less strong metaethical views and thus be less subject to theoretical bias. If a substantial majority of non-philosophers are willing to attribute moral judgments to judges without the corresponding motivation, this might indicate that internalist intuitions among metaethicists are theoretically biased; likewise for externalist intuitions if a majority is *unwilling* to do this.

Much of the debate over internalism seems to presuppose both the existence of a well-defined object—judgments of moral wrongness, goodness, or obligation, say—and the appropriateness of *a priori* methods for identifying and determining the nature of that object. This sort of empirical investigations just mentioned can be seen as falling within such a paradigm: though itself *a posteriori*, it might reveal a defect in *a priori* methods. But one might doubt that the parties of the debate employ co-referential concepts of moral judgments, rather than concepts with largely but not completely overlapping extensions (Francén Olinder 2010). One might also suspect that at least some rejections of amoralists are based not on the perception that amoralists are conceptually or *a priori* impossible, but on the sense that they are synthetically or *a posteriori* impossible because the core psychological function of moral judgments is, *as a matter of empirical fact rather than as a conceptual truth*, essentially motivational (Björnsson 2002, Björnsson & Francén Olinder 2012). In line with this, one might think that what matters for metaethics is not primarily what is implicated by our

(perhaps diverging) concepts of moral judgments. Instead, what matters is the nature of the actual processes of moral thinking and moral discourse, and the role of acts of judgment and motivationally relevant states in these processes (cf. Brink 1997: 22, n 21).

If this is correct, internalist theories will have to be backed up by substantial *a posteriori* theorizing about moral thinking. The full vindication of internalism or externalism would then require a thorough understanding of the cognitive, motivational, emotional and communicative processes at work in normal moral thinking and discourse, and their relations to various cognitive, motivational, and communicative deficiencies.² A comprehensive model of moral thinking informed by such an understanding might vindicate an internalist or an externalist model of moral judgments, depending on how essential motivational factors are in normal moral thinking and how theoretically fruitful it is to categorize states and processes not involving motivation or emotion with the ordinary cases.³

Even within such a deeply *a posteriori* approach, however, empirical studies of folk attributions and conceptions of moral judgments can potentially be highly relevant. For example, to understand the actual workings of moral discourse and our sense of moral agreement and disagreement, we need to know how participants in moral discourse and those engaged in moral thinking conceptualize moral judgments and moral claims. If people's attributions of moral judgments tend to be insensitive to the non-existence of motivation, this would make it less likely that they take one another to be in moral disagreement based on the motivational or practical import of moral judgments and moral claims. If people do treat

² See e.g. discussions of psychopaths and patients with damaged ventromedial prefrontal cortex in Björnsson (2002), Kennett & Fine (2008a, b), and (Roskies (2003, 2006, 2008).

³ Among the best known of numerous recent attempts to provide such general accounts are those of Brink (1989), Gibbard (1990) and Prinz (2007).

motivational connections as necessary, on the other hand, this might instead suggest that these connections are essential to people's conception of moral disagreement. This, then, is another way in which empirical investigations might help us resolve questions about internalism.

In this paper, we present a series of studies of folk conceptions of moral judgment, designed to make modest progress on both the issue of folk conceptions of moral judgments, and the issue of theoretical bias. We begin by looking at the results of a pioneering study by Shaun Nichols (2002, 2004: Ch. 3). Nichols took these results to undermine a certain kind of conditional internalism, as nearly all subjects were willing to attribute moral understanding to an amoralist psychopathic criminal. However, our replication of Nichols' study produced quite different results, and we identify several important problems in its design. To avoid these problems, we presented subjects in a series of studies with a variety of more elaborated scenarios in which an agent classifies a certain action as "wrong" but clearly lacks all motivation to refrain from performing it, and asked them whether they would attribute states like moral belief and moral understanding to that agent. Taken together, the results from these studies (a) suggest that a majority of subjects operate with some internalist conceptions of moral belief, and (b) are compatible with the hypothesis that an *overwhelming* majority of subjects do this. The results also seem to suggest that if metaethicists' intuitions are theoretically biased, this bias is more prominent among externalists.

2. Nichols' study

The main target of Nichols' study is Michael Smith's (1994) version of conditional internalism. In Smith's view, there is a conceptually necessary connection between a person's moral judgment and her motivation to act on condition that she is practically rational. In order to test whether moral judgment ascriptions require motivation under conditions of rationality,

Nichols presented the following scenario to “philosophically unsophisticated undergraduates:”

John is a psychopathic criminal. He is an adult of normal intelligence, but he has no emotional reaction to hurting other people. John has hurt and indeed killed other people when he has wanted to steal their money. He says that he knows that hurting others is wrong, but that he just doesn’t care if he does things that are wrong (Nichols 2004, p. 74).

The case of psychopaths, it would seem, is particularly apt to cause problems for a rationalist version of internalism like Smith’s, as literature and film contain a plethora of psychopathic characters that seem rational and aware of their wrongdoing, but unmoved by moral considerations. The results that Nichols obtained seem to confirm this picture. After being presented with the scenario, subjects ($N=26$) were asked whether John really understands that hurting others is morally wrong: 85% of the subjects responded “Yes” to this question, whereas 15% responded “No.” Nichols took these results to undermine Smith’s internalism, as people seem strongly inclined to attribute moral understanding to John, in spite of his apparent rationality and lack of motivation (Nichols 2002, 2004).

There are a number of problems with Nichols’ study, however. First, the intended target of Nichols’ argument is Smith’s conditional internalism according to which there is a necessary connection between moral judgments and motivation in individuals that are *practically rational* in a certain qualified sense. However, as Richard Joyce (2008) points out, Nichols’ scenario does not make it explicit that John is practically rational in that sense. Even if it may seem likely that subjects assume that John conforms to the movie stereotype of the rational psychopath, it is thus unclear whether the study shows what it is intended to show. Nevertheless, if Nichols’ study is reliable this might indicate that whatever connection between motivation and moral judgment people require has to be rather sophisticated. If the

connection is indirect enough, this might be enough to undermine the noncognitivist view that that moral judgments are inherently motivational states, as well as views that take moral claims to conventionally convey that the speaker is motivated to act accordingly (cf. Copp 2008). Proponents of such forms of internalism might still be able to explain away these intuitions, but such explanations would have to be independently motivated.

Second, the description of John might leave too much implicit to test even for more straightforward, unconditional, internalist constraints on people's attributions of moral judgment. The description does not clearly exclude the possibility that John is *somewhat* motivated to refrain from doing what he understands is wrong. The formulation that he has "no emotional reaction to hurting other people" does not entail that he completely lacks motivation to refrain from performing such actions. This is important since most forms of internalism merely entail that a person's moral judgment necessarily is accompanied by *some* motivation to act accordingly, not that this is her strongest motivation. If it were made very clear that the target agent completely lacks motivation, this could well cause fewer people to attribute moral judgments to John.⁴

Third, given how the scenario is constructed, refusals to attribute moral judgment to the psychopath might not be indicative of their internalist conceptions of moral judgment. The standard externalist explanation of why we are reluctant to attribute moral judgment in absence of moral motivation is that we have a general defeasible expectation that people are at least somewhat motivated to act in accordance with their moral judgments. To make sure that subjects are not just taking the absence of motivation to contingently indicate the absence

⁴ It can be argued that a study by Strandberg and Björklund (2013) suffers from the same difficulty, as the scenarios used in their studies leave open that the agents involved have *some* motivation to act according to their moral judgments.

of the cognitive state that externalists standardly think constitute moral judgments, the relevant sort of scenario should stress that the characteristic cognitive states and processes of moral judgment are present.⁵ Given that only 15% of subjects withheld attributions of understanding, this is not a pressing worry for the interpretation of Nichols' results. However, since a study designed to avoid other problems listed here might yield lower attributions of moral judgment, it should better be able to distinguish between internalist and externalist accounts of why people withhold attributions of moral judgment in the absence of motivation.

Fourth, one might worry that the seriousness of John's misdeeds might lead to motivated reasoning on part of subjects. Reading that he kills people to steal their money, subjects want to hold him accountable, and might thus be motivated to assume that he has whatever is required for accountability, including an understanding of the moral import of his actions.

Fifth, the question Nichols asked his subjects concerns whether John "really understands" that hurting others is wrong. A worry with this question is that it does not adequately test internalism, as usually understood. What metaethicists have been concerned with is whether *the judgment* that is expressed (conventionally indicated) by a verbal endorsement of a moral sentence (used assertorically) is necessarily accompanied by

⁵ The study by Strandberg and Björklund (2013) might suffer from a related difficulty. It tries to get at the modal component of internalism by asking subjects whether it *could* be the case that the agent without motivation thinks that she is morally required to perform a certain action. The difficulty here is that the modal locution "could" has an epistemic reading on which "it could be the case that P" means, roughly, "P is not ruled out by what we know". Subjects taking the absence of motivation to make moral judgment sufficiently unlikely might answer in the negative without therefore taking it to conceptually or metaphysically ruled out. To avoid problems like this, the questions in our studies asked for attributions of actual rather than possible moral judgments.

motivation. The concern has thus not been with *understanding*, but something more like a moral *belief*. One might think that for someone to understand that an act is wrong implies believing that it is wrong, but we hypothesized that people would be more inclined to attribute understanding to John, than to attribute belief. Moral belief ascriptions, it seemed to us, would be more closely connected to motivation than ascriptions of moral understanding.

Finally, even if Nichols' results were statistically significant, the study only had 26 subjects. To test the robustness of the findings, we replicated Nichols' study word by word, with 93 participants recruited from Amazon's Mechanical Turk.⁶ Whereas 85% of Nichols' subjects had answered that John really understands that hurting people is wrong, only 48% of our subjects did, the remaining 52% denying that John really understands this.⁷ In other words, Nichols' results were not reproduced.⁸ Even if there were no other problems with

⁶ Unless indicated otherwise, the subjects participating in our studies were recruited from Amazon's Mechanical Turk. Mechanical Turk is an online labor system connecting potential workers with tasks, including participation in research projects for a small pay. Mechanical Turk participants are slightly more demographically diverse than standard Internet samples, provide self-report data of standard reliability, and respond similarly as traditional college students samples on tasks related to judgment and decision making (Buhrmester, Kwang, & Gosling, 2011; Goodman, Cryder & Cheema, 2013). There was no effect of age and only a single weak effect of sex in our data, i.e., only as much as would be expected by chance.

⁷ We also replicated Nichols' study on 40 US college students, where 50% responded that John really understands that hurting people is wrong.

⁸ One might think that this would be enough to refute Smith's claim that the rationalist connection between moral judgment and moral belief is a *platitude*, i.e. captured by the inferential and judgmental dispositions of someone who has mastered the concept or term in question (Smith 1994: 38–9). (We thank an anonymous reviewer for pressing this point.) Things are not that simple, however. Subjects attributing moral understanding to John might be implicitly committed to think of John's judgment as lacking in some characteristics essential to

Nichols' study, externalists and internalists alike would need to explain why at least a substantial minority answers in line with the contrary view.

3. Psychopathy revisited

In an attempt to obtain more reliable results and avoid problems identified in Nichols' study, we ran a study with a considerably larger set of subjects and with vignettes and questions designed to rule out any motivation to act in accordance with the judgment expressed, to minimize the risk of motivated reasoning, and to distinguish questions of understanding from questions of belief. All subjects were presented with the following scenario: (The numbers within brackets refer to different parts of the vignette. For ease of exposition, when presenting variations of the vignette, we will only present the parts that are different from the vignette below. Subjects did not read the title of the scenario: it is only included here to allow easy reference.)

SCENARIO 1: PSYCHOPATH

[1] Most of us distinguish between actions that we think are morally right and actions that we think are morally wrong (i.e. immoral). Moreover, our classifications into right and wrong actions often influence how we act. For example, those who think that eating meat is wrong more often choose vegetarian alternatives, and those who think that we have a moral duty to donate money to famine victims are more prone to do so. When we do what we classify as wrong, we often feel a sting of guilt.

moral wrongness-judgments *proper* (this is what Smith's argument from fundamental evaluative change is concerned to establish). Results from one of our studies, presented in section 4, seem to support this contention.

[2] Now, consider Anna. Just as the rest of us, Anna classifies actions using expressions like “morally right” and “morally wrong”. The only difference is that her classification of actions into these two categories does not in any way influence her choices. One example is the following situation. Anna has to buy a new cell phone and she has found the model she wants. At the counter, the cashier shows her two phones of this model. They are identical, and sell for the same price. But the cashier explains that if Anna chooses the left phone, \$20 go to starving children in Sudan.

[3] This makes Anna think to herself that not choosing the left phone would be a typical example of an action she classifies as “morally wrong”. But as always, this sort of thought has no effect whatsoever on her choice. She does not have even the slightest impulse or desire to buy the left phone. She says to the cashier that she does not care which phone she gets. This does not make her feel the least regret or guilt. The explanation of why Anna’s classification into “morally right” and “morally wrong” does not affect her choices at all is that she lacks compassion and empathy. She is strikingly egoistic, often displays antisocial behavior, and does not hesitate to expose herself and others to risk.

This scenario improves on Nichols’. It rules out that Anna is even a little motivated to refrain from doing what she classifies as “morally wrong,” both by explicitly excluding “even the slightest impulse or desire” and by describing a situation where she has no personal gain from the morally worse action, thus ensuring that she does not perform it because self-interest overrides some weak moral motivation. Moreover, the scenario stresses similarities between Anna and paradigmatic moral judges, to highlight that lack of motivation is the only difference. Because of this, it is at least *prima facie* plausible that subjects’ unwillingness to attribute moral judgment to Anna stems only from her lack of motivation, and not from the sense that this lack reveals a lack of other aspects that are necessary for moral judgment.

Finally, the scenario involves a much less dramatic wrongdoing than Nichols' scenario to minimize the role of motivated reasoning.

We also wanted to test the hypothesis that questions about whether the agent *believes* or *herself thinks* that the action is morally wrong, would prompt different answers than questions about whether she *understands* this. We therefore randomly assigned subjects to one of the three versions of the following question:

UNDERSTANDS / BELIEVES / HERSELF THINKS: 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?

The results for PSYCHOPATH were:

UNDERSTANDS (N=83):	Yes 76%	No 24%
BELIEVES (N=79):	Yes 46%	No 54%
HERSELF THINKS (N=86):	Yes 49%	No 51% ⁹

⁹ We replicated the study on US college students (N=164, mean age=25.15, 49% males) using a between subjects design and the UNDERSTANDS and BELIEVES questions, and received the following results: UNDERSTANDS, 50% Yes 50% No; BELIEVES, 34% Yes 66% No. We also replicated the study on a convenience sample in Sweden (commuters, students etc., N=77, mean age=23.5 years, 33% males) using a within subjects design. There were no order effects and the responses were clearly asymmetric on UNDERSTANDS (“förstår”) (62% Yes and 38% No) and HERSELF THINKS (“själv tycker”) (32% Yes and 68% No).

A reviewer worried that the mention, in part [1] of the vignette, of how “classifications into right and wrong actions often influence how we act” might prime the participants to respond in way that supports

Consider first the UNDERSTANDS question. The result we obtained is similar to Nichols', but different from the result we obtained replicating his study. In any event, the answers to this question were predominantly along the lines of an externalist prediction. The answers to the BELIEVES and HERSELF THINKS questions, by contrast, tended to fall in line with the internalist prediction. (With 95% confidence, the difference in positive answers between UNDERSTANDS and BELIEVES falls within a 15–44% interval, while that between UNDERSTANDS and HERSELF THINKS falls within a 12–40% interval). This indicates that the choice of question is not innocuous. Moreover, since internalism is standardly understood as a claim about what is required for moral belief or for what agents themselves think is morally right or wrong, rather than for moral understanding, we should primarily focus on the BELIEVES and HERSELF THINKS questions.

The answers to these questions give us no ground for saying that intuitions of defenders of internalism are more theoretically biased than those of externalists, as a majority of subjects were unwilling to attribute moral belief to Anna. Taking the answers of this study at

internalism. To say that people are often influenced is of course theoretically innocent, as it merely spells out what is common ground between parties in internalism debates: the question is whether there is a stronger, necessary connection. Moreover, one might even suspect that the remarks would prime externalist ways of thinking, as talk about what is *often* the case seems to pragmatically imply that it is not always the case. However, to bring clarity to the issue, we ran the three versions of PSYCHOPATH (UNDERSTANDS, BELIEVES, HERSELF THINKS) without any mention of ordinary motivational or emotional effects of moral judgments ($N=241$). Comparing the results to the original versions, we saw no evidence of priming in any of the versions (all $ps > .32$).

face value, the folk are as divided in their intuitions as are philosophers.¹⁰ Moreover, since internalist constraints seem to be widely operative in folk attributions of moral belief, it seems at least *possible* that folk intuitions of moral disagreement are driven by an understanding of the practical role of moral beliefs and moral claims, as noncognitivist internalists would suggest. The study thus seems to leave us where we started: internalists will have to explain why many are willing to attribute belief in the absence of motivation, just as externalists will have to explain why many are unwilling to do so.

4. The inverted commas hypothesis

Internalists often acknowledge that there are states of mind naturally expressed in moral language that are parasitic on moral judgments proper. One suggestion is that people who lack moral motivation can mimic the judgmental dispositions of others and make judgments in an “inverted commas” mode, in effect judging what others would label, say, “morally wrong” without themselves really thinking that it is morally wrong (see e.g. Hare 1952, pp. 124-126). If most people’s conceptions of moral judgments proper were indeed internalist, but if some people are willing to say that agents who make judgments in the inverted commas sense believe that certain actions are morally wrong, this could explain why almost half the subjects were willing to attribute moral belief to Anna. Moreover, if people are more prone to attribute understanding than belief to agents who make inverted commas judgments, this could explain why more subjects answered “Yes” to the UNDERSTANDS question.

¹⁰ In a PhilPapers.org survey of the position on motivational internalism among metaethicists ($N=102$), 45 accepted or leaned towards internalism, whereas 37 accepted or leaned towards externalism; others rejected the question, were undecided, or accepted some third position (<http://philpapers.org/surveys/results.pl>).

To explore the inverted commas hypothesis about seemingly externalist intuitions, we presented subjects with a scenario in which it is clear that Anna uses ethical terms to express her understanding of other people's moral views. The scenario was similar to the first one, but instead of describing Anna as a psychopath, it contained the following information (where "[2i]" replaces [2] in the scenario above):

SCENARIO 2: INVERTED COMMAS

[2i] Now, consider Anna. Anna knows that there are actions that most other people in her society would classify as "morally right", and others that they would classify as "morally wrong". When she classifies acts into "morally right" and "morally wrong" she does so only on the basis of how most other people in her society use these expressions. For example, she classifies murder and theft as "morally wrong" only because she thinks that most other people in her society do so.

After the scenario, subjects were asked one of three questions, similar to those in previous questionnaires:

We have seen that when Anna classifies acts as "morally wrong" this is solely based on how she thinks that most other people use such expressions, and that this never makes her even the least inclined to avoid performing 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?

Given the inverted commas hypothesis, attributions of moral judgment should remain relatively high when the inverted commas reading is made explicit. Moreover, if the differences in answers between the UNDERSTANDING question and the BELIEVES and HERSELF THINKS questions to SCENARIO 1 were due to the inverted commas readings of the former, one

might expect that difference to increase when the scenario makes inverted commas reading more salient.

The results for INVERTED COMMAS were:

UNDERSTANDS (N=77):	Yes 51%	No 49%
BELIEVES (N=88):	Yes 45%	No 55%
HERSELF THINKS (N=75):	Yes 20%	No 80%

Notice four aspects of these results, as compared to those of PSYCHOPATH: (1) Positive answers to UNDERSTANDS questions were significantly lower (Confidence Interval for difference, at 95%: 10 to 39%). (2) Positive answers to HERSELF THINKS were significantly lower (CI for diff.: 14 to 42%). (3) BELIEVES answers remained the same (CI for diff.: -15 to 15%). (4) The difference between UNDERSTANDS and BELIEVES was reduced to non-significance (CI for diff.: -10 to 20%).

One plausible explanation of (1) is that the INVERTED COMMAS scenario made it clear that there is a sense in which Anna does not think for herself and thus lacks one kind of understanding of moral wrongness. Similarly, (2) seems plausibly explained by the fact that the INVERTED COMMAS scenario highlighted how Anna's judgments rely on what *others* think, and that the HERSELF THINKS question stresses the contrast between what Anna thinks and what others think by use of the reflexive *herself*. (Notice that BELIEVES does not similarly stress the contrast.)

Whether these explanations are accurate or not, the decrease in positive answers to both these questions contrast with (3), i.e. with the unchanged proportion of positive answers to the BELIEVES question. Apparently attributions of belief that something is wrong can be largely insensitive to whether attributees *themselves think* that it is wrong. This suggests that subjects see Anna's beliefs in the INVERTED COMMAS scenario as in some way parasitic. There are also

antecedent reasons not to take Anna's beliefs as moral beliefs proper for the purpose of metaethical theorizing: any metaethical theory will accept the possibility of beliefs of *this* kind. Because of this, the continued high level of positive BELIEVES answers supports the internalist hypothesis that attributions of moral belief to putative amoralists are of limited evidential value. It is an open question whether attributions of moral belief in the PSYCHOPATH scenario are attributions of moral belief proper.

Finally, we had hypothesized that the difference between UNDERSTANDING and BELIEVES answers in PSYCHOPATH was due to subjects' higher willingness to attribute understanding than belief to agents who make inverted commas judgments. Based on this, we had predicted that the difference should remain or be even more pronounced in an explicit inverted commas scenario; instead we saw the difference reduced to insignificance (4). This might not falsify the hypothesis, however. Suppose that subjects confronted with the PSYCHOPATH scenario attributed understanding because they sensed that Anna displays *some* kind of understanding—in effect the kind involved in inverted commas judgments. Then maybe subjects confronted with the INVERTED COMMAS scenario were less willing to attribute understanding because that scenario made salient how Anna's understanding *differed* from another, perhaps more important, kind of understanding that others have but Anna obviously lack.¹¹ Even if the hypothesis remains, however, it doesn't tell us *why* subjects would be more willing to attribute understanding than belief in such scenarios. Before further exploring the seemingly internalist-friendly results above, we try out one seemingly plausible explanation of the belief-understanding difference.

¹¹ We thank an anonymous reviewer for suggesting this possibility.

5. The blaming hypothesis

An interesting difference between explicit attributions of moral understanding and explicit attributions of moral belief is that the former are more central to our practices of holding people morally responsible, i.e. of blaming and punishing. To mitigate blame, we frequently say that the agent didn't understand what they were doing, or didn't understand that it was wrong. Perhaps, then, people are more inclined to attribute understanding to agents if it is important to hold them responsible. In order to test this hypothesis, we first designed a pair of scenarios, varying the moral severity of the action performed by Anna: in one scenario Anna knowingly shoots at and kills a person, in the other she knowingly shoots at and damages someone's motorcycle. If the hypothesis is correct, we might expect more attributions of understanding in the first, more severe case. Both scenarios were variations of the original PSYCHOPATH scenario. The first substituted the following for [2] and [3] in that scenario:

SCENARIO 3: PSYCHOPATH KILLS HIKER

[2ii] Now, consider Anna. Just as the rest of us, Anna classifies actions using expressions like "morally right" and "morally wrong". The only difference is that her classification of actions into these two categories does not in any way influence her choices. One example is the following situation. Anna has bought a new hunting rifle and she needs to try out the telescope sight before going hunting. Through the telescope she sees two, for her purpose, equally suitable targets about 200 yards away: a cactus and a man, probably a hiker, sitting close to it.

[3ii] Anna thinks to herself that shooting the hiker would be a typical example of an action she classifies as "morally wrong". But as always, this sort of thought has no effect whatsoever on her choice. She does not have even the slightest impulse or desire not to shoot the hiker. Since the rifle happens to be pointing towards the hiker, she decides that

he will be her target and fires, killing him with her first shot. This does not make her feel the least regret or guilt. The explanation of why Anna’s classification into “morally right” and “morally wrong” does not affect her choices at all is that she lacks compassion and empathy. She is strikingly egoistic, often displays antisocial behavior, and does not hesitate to expose herself and others to risk.

The second variation (SCENARIO 4: PSYCHOPATH SHOOTS AT BIKE) was identical to this, but instead of a person next to the cactus there is a motorcycle parked next to it that Anna fires at, the bullet penetrating its fuel tank. When subjects had been presented with one of these scenarios, they were asked a question similar to the ones in previous studies, in terms of either *understanding* or *belief*:

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 target. In light of this, would you say that she [*understands* / *believes*] that it is morally wrong not to choose the cactus?

We received the following answers for these two scenarios:

Table 1. Responses to the hiker and bike scenarios, across question type

Scenario		<i>N</i>	Yes	No
PSYCHOPATH KILLS HIKER	UNDERSTANDS	81	58%	42%
	BELIEVES	79	37%	63%
PSYCHOPATH SHOOTS AT BIKE	UNDERSTANDS	82	62%	38%
	BELIEVES	80	44%	56%

Contrary to what one might have expected if attributions of understanding were triggered by a willingness to blame, subjects were not more inclined to attribute understanding in the hiker scenario.¹² Moreover, they were *less* inclined to attribute understanding in either of these two scenarios than in the telephone scenario, where Anna's wrongdoing is likely to have seemed much less blameworthy.¹³

These results seem to undermine the blaming hypothesis. But one might worry that the severity of wrongdoing in the HIKER case might have simultaneously undermined attributions of understanding: it might just seem much less likely that anyone could grasp more severe wrongdoing without being moved by their understanding. If so, this might have counteracted and masked any tendency to attribute more understanding to justify willingness to blame. The fact that attributions of understanding had been higher for the original psychopath scenario might not be as plausibly explained away, but to avoid this problem entirely we wanted to test more directly for correlations between attributions of responsibility on the one hand, and belief and understanding on the other. We thus presented subjects with the original PSYCHOPATH scenario and asked them, as before, either whether they would say that Anna *understands* that it is wrong of her not to choose the left phone, or whether she *believes* that it is wrong. However, we also asked subjects to what extent they agreed with the statement that Anna *deserves blame* for not choosing the left phone, with answers given on a 7-point scale ranging from "not at all" to "completely." If attributions of understanding are increased by attributions of blameworthiness, we should expect a significant positive correlation between

¹² Neither UNDERSTANDS nor BELIEVES answers differed significantly between the two scenarios ($p=.59$ and $p=.37$, respectively, two-tailed).

¹³ 95% confidence interval for difference between positive answers to UNDERSTANDS questions here and in original PSYCHOPATH case: HIKER: -2.6 to -32% ; BIKE: 1.2 to -28% .

the two. Since no such correlation was found, we have further reason to reject the hypothesis that willingness to attribute understanding is affected by willingness to blame.¹⁴

Our best remaining hypothesis about the difference between attributions of understanding and attributions of beliefs is that belief, unlike understanding, is taken to involve sufficient *endorsement* of some kind. Strength of endorsement of a thought or claim has various aspects, including degrees of subjective certainty in it, dispositions to reasons from it, defend it, and act from it in characteristic ways. For normative or evaluative claims, the latter sort of disposition is likely to be particularly salient. When this type of endorsement is missing, this would undermine subjects' willingness to attribute moral belief more than their willingness to attribute understanding, as the latter is not as closely associated with endorsement. Thus far, however, we have not found any straightforward way of testing this hypothesis.¹⁵

6. Are subjects answering at random?

One striking fact about the results presented in the last two sections is that the distribution of both “Yes” and “No” answers have stayed fairly close to 50%. Given that the scenarios were described at some length, one might therefore worry that some subjects had skipped ahead, providing answers at random. Judging from free form comments that subjects had given after

¹⁴ Answers for UNDERSTANDS ($N=77$, 64% Yes, 36% No) and BELIEVES ($N=85$, 42% Yes, 58% No) were not significantly different from those in our original study ($p=.091$ and $p=.81$, respectively, two-tailed). To measure the correlation we coded “Yes” answers to UNDERSTANDS / BELIEVES questions as “2” and “No” answers as “1”. The correlation between answers for UNDERSTANDS and attributions of blame was .06, whereas the correlation between answers for BELIEVES and attributions of blame was .17 ($p<.12$).

¹⁵ Cf. Murray et al. (2013), who argue on empirical grounds that belief implies conviction. Conviction can be seen as a kind of endorsement, making the present suggestion a generalization of theirs.

answering the questions, many subjects had given at least some thought to their answers. Moreover, when subjects were asked how confident they were about their answers on a 7-point scale ranging from “very low” to “very high”, they generally indicated a high confidence (mean well over 5 for all questions). This suggests that subjects took the questions and vignettes to be clear enough. To further confirm this impression, however, we wanted to provide subjects with a scenario of roughly the same complexity as before, but where both externalism and most forms of internalism would take attributions of moral judgments to be relatively unproblematic. For this purpose, we constructed a scenario with accompanying UNDERSTANDS and BELIEVES questions in which Anna acted against her putative moral judgment, but where she experienced a struggle in doing this. Only forms of internalism requiring that moral judgments come with *overriding* motivation to act would rule out attributing moral judgments in such a scenario:

SCENARIO 5: INNER STRUGGLE

[2iii] Now, consider Anna. Just as the rest of us, Anna classifies actions using expressions like “morally right” and “morally wrong”. Normally these classifications influence what she chooses to do. However, sometimes this is not enough to make her act accordingly. One example is the following situation. Anna has to buy a new cell phone and she has found the model she wants. At the counter, the cashier shows her two phones of this model. They sell for the same price and are identical except that one is black and the other is white. Anna prefers the black phone, but the cashier explains that if Anna chooses the white phone, \$20 go to starving children in Sudan.

[3iii] Anna thinks to herself that not choosing the white phone would be a typical example of an action she classifies as “morally wrong”. However, since she prefers to have a black phone she experiences an inner struggle about the choice of phone. After

some hesitation she says to the cashier that she wants the black phone. Later she sometimes feels a sting of guilt about not choosing the white phone.

As before, we randomly assigned subjects to either an UNDERSTANDS or a BELIEVES question:

We have seen that Anna classifies some actions as “morally wrong”. This typically influences what she chooses to do, but not always enough to make her act accordingly.

We saw this when she bought her cell phone. In light of this, would you say that she [*understands / believes*] that it is morally wrong not to choose the white phone?

If subjects were not paying attention or if they were confused by the complexity of the scenarios, we should expect in the neighborhood of 50% “Yes” and 50% “No” answers. If they were paying attention and understanding the scenario as intended, however, a large majority should be willing to assign both understanding and belief in this case. This is indeed what we saw:

UNDERSTANDS (N=33):	Yes 85%	No 15%
BELIEVES (N=44):	Yes 80%	No 20%

Judging from this, the reason that subjects are hesitant to attribute moral belief and moral understanding to Anna in the prior scenarios is that her motivation is completely missing, not that the scenarios are too complex.

7. Defaults and explanations

We also wanted to test one final hypothesis, namely that subjects operate with a default assumption that moral judgments are accompanied by motivation, and are willing to treat putative moral judgments unaccompanied by motivation as moral judgments *to the extent that they take there to be a special explanation for the deviation from the norm*. Suppose that

subjects understand moral judgments as states disposing judges to act and be motivated to act in certain ways, or states the function of which is to produce motivation and action, as many internalists have thought (e.g. Gibbard 1990, Björnsson 2002, Bedke 2009, Björnsson & McPherson *in press*, Eriksson *in press*). Then we should expect them to take the absence of moral motivation to indicate the absence of moral judgment unless there is reason to think that something blocks the disposition, or prevents the state from performing its function.

To test this hypothesis, we devised two versions of our original PSYCHOPATH scenario. In that scenario, we had explained Anna's lack of moral motivation in terms of her lack of empathy. Now we added two scenarios, one identical to PSYCHOPATH except that no reason was given for Anna's lack of motivation, and one in which her lack of motivation is explained by temporary listlessness induced by a personal crisis. Subjects were assigned randomly to one of these new scenarios:

SCENARIO 6: NO REASON

[2iv] Now, consider Anna. Just as the rest of us, Anna classifies actions using expressions like "morally right" and "morally wrong". The only difference is that her classification of actions into these two categories does not in any way influence her choices. One example is the following situation. Anna has to buy a new cell phone and she has found the model she wants. At the counter, the cashier shows her two phones of this model. They are identical, and sell for the same price. But the cashier explains that if Anna chooses the left phone, \$20 go to starving children in Sudan.

[3iv] This makes Anna think to herself that not choosing the left phone would be a typical example of an action she classifies as "morally wrong". But as always, this sort of thought has no effect whatsoever on her choice. She does not have even the slightest

impulse or desire to buy the left phone. She says to the cashier that she does not care which phone she gets. This does not make her feel the least regret or guilt.

Subjects who read NO REASON were assigned randomly to either an UNDERSTANDS or a BELIEVES question:

We have seen that Anna classifies some actions as “morally wrong”. But this never makes her even the least inclined to avoid these actions. We saw this indifference when she bought her cell phone. In light of this, would you say that she [*understands / believes*] that it is morally wrong not to choose the left phone?

SCENARIO 7: LISTLESSNESS

[2v] Now, consider Anna. Just as the rest of us, Anna classifies actions using expressions like “morally right” and “morally wrong”. Normally these classifications influence what she chooses to do. However, as a result of a personal crisis Anna has lost her zest for life. She is listless and apathetic, no longer caring about things that used to be important to her, such as going to work and meeting her friends.

Anna continues to classify actions as “morally right” and “morally wrong”, but at present this does not in any way influence her choices. One example is the following situation. Anna has to buy a new cell phone and she has found the model she wants. At the counter, the cashier shows her two phones of this model. They are identical, and sell for the same price. But the cashier explains that if Anna chooses the left phone, \$20 go to starving children in Sudan.

[3v] This makes Anna think to herself that not choosing the left phone would be a typical example of an action she classifies as “morally wrong”. But because of her listless condition this thought has no effect whatsoever on her choice. She does not have even the

slightest impulse or desire to buy the left phone. She says to the cashier that she does not care which phone she gets. This does not make her feel the least regret or guilt.

Subjects who read LISTLESSNESS were assigned randomly to either an UNDERSTANDS or a BELIEVES question:

We have seen that Anna classifies some actions as “morally wrong”. But because of her listlessness and apathy this no longer 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*] that it is morally wrong not to choose the left phone?

The prediction was (a) that subjects would be more willing to assign belief and understanding in LISTLESSNESS than in PSYCHOPATH, as it is clearer in the former case that the judgment made is of a type that normally comes with corresponding motivation, and (b) that subjects would, by the same token, be least willing to assign belief and understanding in NO REASON. This is indeed what we saw:

Table 2. Responses to the background-related scenarios, across question type

Scenario		<i>N</i>	Yes	No
LISTLESSNESS	UNDERSTANDS	43	81%	19%
	BELIEVES	40	70%	30%
PSYCHOPATH	UNDERSTANDS	83	76%	24%
	BELIEVES	79	46%	54%
NO REASON	UNDERSTANDS	42	55%	45%
	BELIEVES	36	36%	64%

Though far from conclusive,¹⁶ this supports the suggestion that attributions of moral judgments rely on a default assumption that moral judgments are accompanied by motivation.¹⁷ Obviously, more needs to be said about the nature of the hypothesized default assumption, and about whether it is best explained given some form of externalism, some form of conditional internalism, or by some other kind of account entirely. But the results seem to be most obviously in line with what could be expected given certain forms of conditional internalism.

8. Concluding remarks

We started with the ambition to make progress on two sorts of questions. First, we wanted to see whether folk attributions of moral judgment could give us reason to think that the intuitions of either internalist or externalist metaethicists were theoretically biased. Second, we wondered whether people in general operate with the sort of internalist conception of moral judgment that would be required for an understanding of moral disagreement as disagreement in attitude, i.e. the sort of understanding that noncognitivists stipulate.

In the studies presented here, a large portion of subjects withhold attributions of moral belief in the absence of moral motivation. A straightforward explanation for this would be that they require moral judgments to be motivating, or to be states of mind of a kind that motivates under normal circumstances. When motivation was present but trumped by stronger

¹⁶ At the 95% confidence level, differences between PSYCHOPATH and LISTLESSNESS on UNDERSTANDS and between PSYCHOPATH and NO REASON on BELIEVES were not statistically significant.

¹⁷ We might also note that in response to INNER STRUGGLE, where it is even clearer than in LISTLESSNESS that the effects of moral judgment on action is blocked by conflicting desires, 85% of subjects attributed understanding and 80% belief. Though this is in line with the prediction, the differences between these results and those of LISTLESSNESS were not statistically significant.

motivation (in INNER STRUGGLE), 80% of subjects attributed moral belief to Anna, and when motivation was absent without any explanation (in NO REASON), only 36% attributed moral belief: between those extremes were cases where motivation was missing for more or less familiar reasons (LISTLESSNESS 70%, PSYCHOPATH 46%). Together, these results suggest that at least a substantial majority of people operate with an internalist understanding of moral belief. We have of course also seen that a significant minority of subjects were willing to attribute moral belief in the absence of moral motivation. But it is not clear what weight to give to this fact, as subjects were equally willing to attribute moral belief in the INVERTED COMMAS scenario, where Anna arguably does not make moral judgments proper.

All in all, then, our studies tentatively suggest a positive answer to our second question: they provide some positive reason to think that a majority of people operate with an internalist understanding of moral judgments proper, and no strong reason against expecting the same about most of the remaining people. Nichols' conclusion that internalism is undermined by folk psychology does not seem to hold up.

Similarly, in response to our first question, our studies suggest that the worry about theoretical bias is a more pressing one for externalists, as most subjects seemed to have broadly internalist intuitions, and as there is reason to mistrust or reinterpret the seemingly externalist tendencies that yielded the same attributions of moral belief in the PSYCHOPATH and INVERTED COMMAS scenarios.

None of this shows that externalism is mistaken, of course. Externalists might be able to explain away seemingly internalist answers, perhaps in pragmatic terms, and for all we have said, a substantial minority might be operating with an externalist conception of moral judgments proper. Moreover, if there are independent viable arguments for some form of absolutist moral cognitivism, that might provide independent reasons for thinking that such an externalist conception is the philosophically more interesting: what unifies the class of

putative moral judgments would be their content rather than their motivational role. But since it is far from obvious how these internalist answers can be explained away, and an open question what independent reasons there are for accepting absolutist cognitivism, the study does put some pressure on defenders of externalism.

The results presented here also leave open whether the constraint leading subjects to withhold attributions of moral belief in the absence of motivation is strong enough to support noncognitivism. One might think that what noncognitivism requires is something like the simple form of internalism with which we began our discussion, but the results in Section 7 might suggest that the operative constraint is largely conditional. It is not entirely obvious how this can be enough for noncognitivism.¹⁸

References

- Bedke, M. S. (2009). Moral judgments purposivism: Saving internalism from amorality. *Philosophical Studies*, 144, 189–209.
- Björklund, F., Björnsson, G., Eriksson, J., Francén Olinder, R., & Strandberg, C. (2012). Recent work on motivational internalism. *Analysis*, 72, 124–137.
- Björnsson, G. (2002). How emotivism survives immoralists, irrationality, and depression. *Southern Journal of Philosophy*, 40, 327–344.
- Björnsson, G. & Francén Olinder, R. (2012). Internalists beware – we might all be amorality!, *Australasian Journal Of Philosophy*, 91, 1–14.
- Björnsson, G., & McPherson, T. (in press). Moral attitudes for non-cognitivists: Solving the specification problem. *Mind*.
- Blackburn, S. (1998). *Ruling passions*. Oxford: Oxford University Press.
- Brink, D. O. (1989). *Moral realism and the foundation of ethics*. Cambridge: Cambridge University

¹⁸ Brink (1997) argues that certain forms of conditional internalism are incompatible with noncognitivism, and Strandberg (2012b) presses similar problems. For contrary arguments, see Björnsson 2002 and Eriksson *in press*. For a noncognitivist account of the default assumption that fits neither the externalist nor internalist bracket, see Björnsson and McPherson *in press*.

Press.

Brink, D. O. (1997). Moral Motivation, *Ethics*, 108, 4-32.

Buhrmester, M., Kwang, T., & Gosling, S. D. (2011). Amazon's Mechanical Turk: A new source of inexpensive yet high-quality, data? *Perspectives on Psychological Science*, 6, 3-5.

Copp, D. (2008). Realist-Expressivism and Conventional Implicature. In R. Shafer-Landau (Ed.) *Oxford Studies in Metaethics* Vol. 4 (pp. 167–202). Oxford: Oxford University Press.

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

Eriksson, J. (in press) Elaborating expressivism: moral judgments, desires and motivation, *Ethical Theory and Moral Practice*.

Finlay, S. (2004). The conversational practicality of value judgement, *The Journal of Ethics*, 8, 205–23.

Finlay, S. (2005). Value and implicature, *Philosophers' Imprint*, 5, 1–20.

Francén Olinder, R. (2010). Moral Motivation Pluralism, *The Journal of Ethics*, 14, 117–48.

Gibbard, A. (1990). *Wise Choices, Apt Feelings*. Oxford: Clarendon Press.

Gibbard, A. (2003). *Thinking how to live*. Cambridge, Mass.: Harvard University Press.

Goodman, J. K., Cryder, C. E., & Cheema, A. (2013). Data collection in a flat world: The strengths and weaknesses of Mechanical Turk Samples. *Journal of Behavioral Decision Making*, 26, 213–224.

Hare, R. M. (1952). *The language of morals*. Oxford: Oxford University Press.

Joyce, R. (2008). What neuroscience can (and cannot) contribute to metaethics. In W. Sinnott-Armstrong (Ed.), *Moral Psychology*, Vol 3 (pp. 371–394). Cambridge, Mass.: The MIT Press.

Kauppinen, A. (forthcoming). Intuition and belief in moral motivation. In G. Björnsson, R. Francén Olinder, C. Strandberg, J. Eriksson, F. Björklund (Eds.), *Motivational Internalism*, New York: Oxford University Press.

Kennett, J., & Fine, C. (2008a). Internalism and the evidence from psychopaths and “acquired sociopaths.” In W. Sinnott-Armstrong (Ed.), *Moral psychology*, Vol. 3 (pp. 173–190). Cambridge, Mass.: The MIT Press.

Kennett, J. & C. Fine. (2008b). Could there be an empirical test for internalism? In *Moral Psychology*, Vol. 3, ed. W. Sinnott-Armstrong, 217–26. Cambridge, MA: The MIT Press.

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

Nichols, S. (2002). How psychopaths threaten moral rationalism: Is it irrational to be amoral? *The Monist*, 85, 285–304.

Nichols, S. (2004). *Sentimental rules: On the natural foundations of moral judgment*. Oxford: Oxford University Press..

Prinz, J. (2007) *The Emotional Construction of Morals*. Oxford: Oxford University Press.

Roskies, A. (2003). Are ethical judgments intrinsically motivational? Lessons from “acquired

- sociopathy.” *Philosophical Psychology*, 16, 51–66.
- Roskies, A. (2006) Patients with ventromedial frontal damage have moral beliefs, *Philosophical Psychology*, 19, 617–27.
- Roskies, A. (2008). Internalism and the evidence from pathology. In W. Sinnott-Armstrong (Ed.), *Moral psychology*, Vol. 3 (pp. 191–206). Cambridge, Mass.: The MIT Press.
- Shafer-Landau, R. (2003). *Moral realism. A defence*, Oxford: Oxford University Press.
- Smith, M. (1994). *The moral problem*. Oxford: Blackwell.
- Strandberg, C. (2011). The pragmatics of moral motivation. *Journal of Ethics*, 15, 341–369.
- Strandberg, C (2012a). A dual aspect account of moral language, *Philosophy and Phenomenological Research*, 84, 87–122.
- Strandberg, C (2012b). Expressivism and dispositional desires, *American Philosophical Quarterly*, 49, 81-91.
- Strandberg, C., & Björklund, F. (2013). Is moral internalism supported by folk intuitions? *Philosophical Psychology*, 26, 319–335.
- Svavarsdóttir, S. (1999). Moral cognitivism and motivation. *The Philosophical Review*, 108, 161–219.
- Tresan, J. (2006). De dicto internalist cognitivism. *Noûs*, 40, 143–65.
- Tresan, J. (2009). The challenge of communal internalism. *The Journal of Value Inquiry*, 43, 179–99.
- Zangwill, N. (2008). The indifference argument, *Philosophical Studies*, 138, 91–124.