

# Moral Judgment and the Content–Attitude Distinction

Uriah Kriegel

Forthcoming in *Philosophical Studies*

*Abstract* :: Let cognitivism be the view that moral judgments are cognitive mental states and noncognitivism the view that they are noncognitive mental states. Here I argue for *moral judgment pluralism*: some moral judgments are cognitive states and some are noncognitive states. More specifically, according to my pluralism some judgments are moral because they carry a moral content (e.g., that genocide is wrong) and some are moral because they employ a moral attitude (e.g., indignation, or guilt); the former are the cognitive moral judgments and the latter the noncognitive ones. After explaining and motivating the view, I argue that this kind of pluralism handles quite elegantly several of the core issues that have structured the debate on cognitivism vs. noncognitivism.

## 1 Introduction

Let cognitivism be the view that moral judgments are cognitive mental states (roughly: belief-like, truth-apt states) and noncognitivism the view that they are noncognitive states. Because each view has some clear attractions, but also faces serious difficulties, much recent work in metaethics has focused on developing hybrid accounts according to which moral judgments are peculiar states that combine elements of the cognitive and of the noncognitive. In this paper, I recommend a different tack: moral judgment *pluralism*, according to which some moral judgments are cognitive and some are noncognitive.

The paper is in two parts. The first part introduces the view I want to defend, the second part presents a preliminary case for it. In §2, I introduce the relevant notion of moral judgment, the distinction between cognitive and noncognitive states, and the resulting debate between cognitivism and

noncognitivism about moral judgments. In §3, I present the view I call moral judgment pluralism and motivate it with two general considerations concerning our mental life. In §4, I show how pluralism handles two core challenges to noncognitivism: the Frege–Geach problem and the objective phenomenology of moral judgments. In §5, I bring out pluralism’s resources in affecting the dialectic around a core challenge for cognitivism, namely, the threat of error theory. I close, in §6, with a discussion of the alleged internal link between moral judgment and motivation, arguing that pluralism is better placed than cognitivism and noncognitivism to accommodate some of the most prominent “data” in this area.

In a way, it is only natural to suspect a pluralism here. Indeed, *wherever* there are strong reasons to deny that all As are Bs, but also strong reasons to deny that all As are non-Bs, it is natural to suspect that *some* As are Bs and some are non-Bs. The problem is that this will often seem facile and ad hoc. The “philosophical hypothesis” I want to float here is that there is an *independent reason* to expect pluralism in the domain of moral judgment. As I will use the term, a moral judgment is any mental state that embodies moral evaluation. Whenever we morally evaluate something in our mind, we are making a moral judgment. Now, mental states have both content and attitude, and, I will argue, moral evaluation can occur through either. Accordingly, we should *expect* there to be two kinds of moral judgments: ones that morally evaluate in virtue of carrying a morally evaluative content and ones that morally evaluate in virtue of deploying a morally evaluative attitude. This is the “philosophical hypothesis” I want to defend.

## *2 Cognitivism and Noncognitivism*

Cognitivism is the view that moral judgments are cognitive states, noncognitivism the view that they are noncognitive states. But how are we to understand the distinction between cognitive and noncognitive mental states?

One approach appeals to ostension of paradigms. Often it is assumed that belief is the paradigmatic cognitive state and desire the paradigmatic noncognitive state. In this approach, cognitivism holds that moral judgments are beliefs, or at least belief-like, while noncognitivism holds that they are desires,

or desire-like. Another approach appeals to *direction of fit*: cognitive states have mind-to-world direction of fit, noncognitive ones a world-to-mind direction of fit. From this perspective, cognitivism holds that moral judgments have a mind-to-world direction of fit whereas noncognitivism holds that they have a world-to-mind direction of fit. A third approach relies on functional role: cognitive states' functional role features centrally *inference* (its core is *inferential* role) while noncognitive states' functional role features centrally *motivation* (its core is *motivational* role). Thus cognitivism holds that moral judgments are "inferentially integrated" states, while noncognitivism holds that they are "motivationally hot" states. A fourth approach focuses on correctness or fittingness conditions: cognitive states are ones it is fitting to enter when their contents are *true*, whereas the fittingness of entering a noncognitive state is not a matter of its content being true. Here cognitivism converges on the idea that moral judgments are truth-apt states, while noncognitivism is understood as the view that they are not.

It is epistemically possible that all these approaches will turn out to be extensionally equivalent, with mental states forking neatly into truth-apt, inferentially integrated, belief-like states with mind-to-world direction of fit, on the one hand, and truth-inapt, motivationally hot, desire-like states with world-to-mind direction of fit, on the other.

It is also possible to integrate all four approaches (and potentially others) into a *cluster account*, whereby a mental state would, say, (a) count as cognitive just if it exhibits at least three of the four proposed marks of the cognitive, (b) count as noncognitive just if it exhibits three of the proposed marks of the noncognitive, and (c) count as a "hybrid state" if it exhibits, say, two marks of the cognitive and two of the noncognitive.

In what follows I remain neutral on which of these ways of understanding the debate on cognitivism vs. noncognitivism is "best" (most fruitful, most accurate, or whatever). The view I want to develop here does not become unavailable, or invisible, on any of the above conceptions of the cognitive/noncognitive divide. So there is really no need for us to take a stand on the matter, so long as we are thinking clearly about what is at issue.

I turn now to making some preliminary remarks on the notion of moral judgment as it features in the debate. Note, first, that the term “judgment” as used in this debate is supposed to be neutral on the status of a mental state as cognitive or not – obviously, since both cognitivism and noncognitivism are proposed as accounts of moral *judgments*. Personally, this strikes me as an aberrant use, insofar as “judgment” clearly names a paradigmatically cognitive state. However, this use is so entrenched in the relevant literature, and the matter is after all only terminological, so I will indulge here the entrenched terminology.

One element of instability in the entrenched usage is that “moral judgment” is sometimes used to denote a type of mental state but sometimes to denote a type of *linguistic utterance* or *speech act*. Very roughly, a “cognitive” judgment is an utterance in the *indicative* mood, a “noncognitive” one is an utterance in the imperative mood. Here I focus on moral judgments as mental states, though it is entirely possible that much of what I have to say could be “translated” into claims about utterances.

It is a nontrivial question what earns a judgment the qualification *moral*. Clearly, a moral judgment embodies a certain sort of evaluation. What sort? Well, *moral* evaluation, clearly. But what marks off moral from other kinds of evaluation is a difficult question. I want to stay neutral on this and rely on our intuitive, pretheoretic grasp of what it is to evaluate something *morally*, because philosophers with different commitments here will be equally in need of a notion of moral judgment, and should be able to speak to each other *about* moral judgments regardless of their substantive takes on the mark of the moral.

In my informal presentation of cognitivism and noncognitivism, I have formulated them as *generics*, with “moral judgments” used as a so-called bare plural. As is their wont, however, philosophers typically think of cognitivism and noncognitivism as universally quantified theses: *All* moral judgments are cognitive/noncognitive. There is, of course, logical room for a thesis rejecting both universals, namely, the thesis that *some* moral judgments are cognitive states and *some* are noncognitive states. I call this conjunction of existentials *pluralism* about moral judgment. A version of this thesis is what I want to defend here.<sup>1</sup>

### 3 *Moral Contents and Moral Attitudes*

My basic motivation for pluralism about moral judgment may be put as follows. A moral judgment is a mental state that embodies moral evaluation, that is, an event or state consisting in morally evaluating something (or someone) in one's mind. Now, mental states have both content and attitude, and – this is the key claim – *moral evaluation can occur through either*. So, there are (a) mental states that morally evaluate as part of their content and (b) mental states that morally evaluate through their attitude. Thus we have reason to *expect* there to be two different kinds of moral judgment. Moreover, there are reasons to suspect that the states which morally evaluate through their content are *cognitive*, while the states which morally evaluate through their attitude are *noncognitive*.

Consider first an analogy. It is obviously two different things to be afraid of flying and to believe that flying is dangerous. It is possible to have either without the other. Both states, however, embody a similar kind of evaluation: they both evaluate flying as dangerous. What, then, is the difference between them? My proposal is that the belief is a mental state that evaluates flying as dangerous through its content, whereas the fear is a mental state that evaluates flying as dangerous through its attitude. The content of the belief is that flying is dangerous; the concept of dangerousness is thus a constituent of this content. In contrast, dangerousness need not show up in the content of the fear – *what one fears* is typically not the dangerousness of the flying, but the *flying itself* – and accordingly a creature may experience fear that does not possess the concept of danger. If this is right, then there is nothing about the *content* of fear that evaluates the feared object in any way. Rather, it is the very *attitude* of fearing that constitutes the evaluation of the object as dangerous. In taking the fear attitude toward flying (rather than some other attitude), we *thereby* evaluate the flying as dangerous.

We might say that the belief represents flying-as-dangerous, whereas the fear represents-as-dangerous flying. In the belief, dangerousness is an element in *what is represented*, part of the state's *subject matter*. In the fear, what is represented is only the flying; the element of dangerousness characterizes rather the *manner* in which the flying is represented. This is why it goes to the

very nature of fear, as the type of mental state it is, that it presents its object “under the guise of the dangerous,” so to speak; whereas it is in no way essential to belief that it concerns the dangerous. Thankfully, most our beliefs are not about danger.

Belief has its own distinctive manner of representing its subject matter, of course. Its proprietary manner of representation is *as true*: the belief that *p* is a mental state that represents-as-true the content  $\langle p \rangle$ . This is why it is in the nature of belief, as the kind of mental state it is, to present its object *under the guise of the true* – in the same sense it is in the nature of fear to present its object under the guise of the dangerous. The fundamental contrast between the belief that flying is dangerous and the fear of flying may thus be put as follows:

The belief	::	represents-as-true $\langle$ flying is dangerous $\rangle$
The fear	::	represents-as-dangerous $\langle$ flying $\rangle$

Here the content is in chevrons and the attitude is the hyphenated expression preceding the chevrons. Dangerousness shows up in both states’ intentional structure, but shows up in different “parts” of that structure in each.

To be clear, I am not denying that it is possible to fear dangerousness – nor, for that matter, that it is possible to believe that *p* is true. For example, one may believe *that Fermat’s last theorem is true*, and one may fear *dangerous spiders*. But this is not the basic form in either case: typically, we just believe that *p*, not *that p is true*; and likewise, we just fear *x*, not *x’s dangerousness*. Nonetheless, belief by its nature frames *p* as true, while fear by its nature frames *x* as dangerous.

What kinds of properties are these representing-as-F properties? I want to stay neutral on this as well. They may turn out to be functional-role properties (perhaps characterizable through boxes-and-arrows diagrams), or normative properties (perhaps characterizable in terms of fittingness conditions, or appropriateness conditions), or phenomenological properties (perhaps ineffable and only *nameable*, not *characterizable*), or something else. What matters is that they are *attitudinal* properties, as opposed to *content* properties. That is, they are properties mental states have not in virtue of the representational content they carry, but in virtue of the attitude they take toward that content – not in virtue of *what* they represent, but in virtue of the

way they relate representationally to whatever they represent (the “manner” or “guise” of representation, understood to have nothing to do with Fregean modes of presentation, since the latter are constituents of content).

Danger evaluation is of course not *moral* evaluation (except when moral danger is concerned). Fear is in that sense not a moral emotion. But some emotions *are* moral emotions, in that their distinctive manner of presenting their object is moral. Take guilt. It is again two different things to feel guilty about lying to Smith, say, and to believe that it was wrong to lie to Smith. Both states embody a similar evaluation: they both evaluate lying to Smith as wrong. The difference, I contend, is that the belief evaluates the lying as wrong through its content, whereas the guilt does so through its attitude. Wrongness shows up in the content of the belief, but need not show up in the content of the guilt. What one feels guilty *about* may be simply the lying; it is in taking the guilt attitude *toward* the lying that we evaluate the lying as wrong. Again we may present the contrast as follows:

The belief	::	represents-as-true	<lying to Smith was wrong>
The guilt	::	represents-as-wrong	<lying to Smith>

In the belief, wrongness is an element in *what is represented*, a constituent of the belief’s representational content. In the guilt, what is represented is typically only the lying; the element of wrongness characterizes rather the *manner* in which the lying is represented: as-wrong is an aspect of *how* the subject is aware of the lying in guilt. This is why it goes to the very nature of guilt, as the type of mental state it is, that it presents its object *under the guise of the wrong*.

The point here is that the belief that it was wrong to lie evaluates the lying as wrong through its content, whereas the guilt about lying evaluates the lying as wrong through its attitude. Now, the wrongness in question does seem to be *moral* wrongness – at least in paradigmatic cases of guilt, and perhaps in *all* cases of guilt (there is some debate on this in the guilt literature, but that need not concern us here – see, e.g., Greenspan 1992 for discussion). Thus we have here two mental states that *morally* evaluate, but in importantly different styles: one through its content, the other through its attitude.

It is not difficult to find other pairs of mental states, one an emotion and the other a belief, such that both embody moral evaluation but one through its

attitude and the other through its content. Compare feeling indignant about the Hutu's mass slaughter of the Tutsi and believing that the Hutu's mass slaughter of the Tutsi was extremely unjust (or whatever). These are clearly two different mental states. After enough time has passed, and after having felt indignant about it for a long time, you might find yourself emotionally exhausted, or just distracted, and no longer experience occurrent indignation, while still believing it was extremely unjust. Conversely, we can imagine a scenario where new information about a case may make you reconsider or even suspend your belief that some event was extremely unjust, without yet ridding yourself of your indignation about it – such is the recalcitrance of emotions. Still, both states have something important in common: they embody a moral evaluation of something (e.g., the Hutu's slaughter of the Tutsi) as extremely unjust. My proposal is that the difference is due to the “location” of this moral evaluation in these states' intentional structure: the indignation represents-as-extremely-unjust <the Hutu's slaughter of the Tutsi>, while the corresponding belief represents-as-true <the Hutu's slaughter of the Tutsi was extremely unjust>. (To be clear, I am using the expression “extremely unjust” as a dummy here, standing in for whatever the fuller, more textured description would be that would express more accurately the kind of moral evaluation characteristic of indignation.)

Note, now, that a belief with the content <the Hutu's slaughter of the Tutsi was extremely unjust> has all the hallmarks of a cognitive state, while an indignation with the content <the Hutu's slaughter of the Tutsi> has many of the hallmarks of a noncognitive state. Perhaps most clearly, the belief that the Hutu's slaughter of the Tutsi was extremely unjust, being a belief, is certainly *truth-apt*; whereas the corresponding indignation is not. If there is no such property as being unjust, then error theory would kick in and the belief would come out false – which is only possible, of course, for a truth-apt state. In contrast, it is a category mistake to describe an indignation as true (*or* false) in the relevant sense. An indignation can of course be fitting or unfitting, appropriate or inappropriate. But what determines which fittingness or appropriateness conditions attach to a mental state is its attitudinal profile. The reason belief's fittingness conditions are *truth* condition is that belief represents-as-*true*. For indignation's fittingness conditions to be truth conditions, indignation would have to likewise represent-as-true; whereas what



indignation does is to represent-as-extremely-unjust (or something like that). So while indignation can be evaluated for fittingness, it would not be evaluation for truth.

In addition, the belief about the Hutu slaughter of the Tutsi is also *inferentially integrated*: it can be inferred, for instance from the beliefs that (a) senseless slaughters fueled by ethnic hatred are extremely unjust and (b) the slaughter of Tutsi by the Hutu was a senseless slaughter fueled by ethnic hatred. In contrast, the functional role of indignation is clearly more motivational than inferential. While feeling indignant about a mass slaughter I feel pulled, at a minimum, to refrain from joining in.

Relatedly, indignation also has a world-to-mind *direction of fit*: the world must change – unjust slaughter must end – for a fit to be established between the world and my mind. Belief, in contrast, is the paradigmatic state with mind-to-world direction of fit: if my mind is in a state of disbelief or suspension of judgment with respect to the proposition <the Hutu's slaughter of the Tutsi was extremely unjust>, then my mind must change into a state of belief in order for fit with the world to be established.

This leaves only the issue of being belief-like or desire-like. The slaughter belief is trivially belief-like, of course. Whether indignation is desire-like is a more complicated question, but it is certainly belief-*unlike*, given its truth-inaptness, motivational role, and direction of fit.

I conclude that indignation is both (i) a noncognitive state (being as it is a truth-inapt, motivationally hot, belief-unlike state with world-to-mind direction of fit) and (ii) a moral judgment (being as it is a way of morally evaluating something in our mind). It follows that there are moral judgments that are noncognitive states. Likewise, a belief with the content <the Hutu's slaughter of the Tutsi was extremely unjust> is both (i) a cognitive state and (ii) a moral judgment (for the same kinds of reasons). It follows that there are moral judgments that are cognitive states. Hence moral-judgment pluralism.

The case of indignation about a mass slaughter and a corresponding belief is not isolated. The same points just made will apply to the belief that it was wrong to lie to Smith and the guilt about having lied to Smith: the former has all

the hallmarks of a cognitive state, the latter the hallmarks of a noncognitive state.<sup>2</sup> And there will presumably be other belief/emotion pairs exhibiting a similar duality of moral evaluation – more on this shortly. So the ensuing pluralism is thorough and systematic, not a case where a conjunction of existentials is true of a class of items because all but one of them are F while the last one is  $\sim F$ .

In light of this, the “philosophical hypothesis” I would like to air is this: there are mental states – moral beliefs – that have morally evaluative contents, and there are also mental states – I suspect: moral emotions – that employ morally evaluative attitudes. Both are forms of morally evaluating something in one’s mind, and so are moral judgments in our sense. But they are different *kinds* of moral judgment, insofar as they embody evaluation in different ways. One represents-as-morally-(dis)valuable some content  $\langle x \rangle$ , the other represents-as-true a content of the form  $\langle x \text{ is } M \rangle$ , where M is a moral (dis)value predicate. The second moral judgment is a cognitive state, the first a noncognitive state.

Incidentally, there is a principled reason, I suspect, why we can expect a number of such moral emotion/moral belief pairs in our psychological repertoire. Suppose A and B have been married happily, and happily monogamously, for 15 years, but that to “spice up” their life they have now decided together to “open” their relationship for a year. A month later, A sleeps with C, such that  $C \neq B$ , and the next morning A feels tremendous guilt. There may be versions of this scenario where A takes the guilt to reveal a deep truth about RAB (I am trying to be as unromantic as possible here!) and goes back to B pleading for reverting to a “closed” relationship. But there are also versions where A recognizes the feeling of guilt but simply does not endorse it. Perhaps A sees the guilt as a vestige of an emotional setup that no longer serves to further either A’s or B’s flourishing inside RAB, and is thus still committed to the open-relationship project. There is clearly a difference between the mind of the person who feels guilty but does not endorse the feeling and the mind of the person who feels guilty and in addition endorses the feeling. What does this difference consist in? My inclination is to say: the person who endorses her guilt has (thereby?) formed a corresponding moral belief, whereas the person who does not has not. More generally, endorsement of moral emotions yields corresponding moral

beliefs. If so, wherever we have a moral emotion, and the capacity to endorse it, there will be a moral belief we can pair with that emotion. And so there will be a moral emotion/moral belief pair for any “endorsable” moral emotion we have.<sup>3</sup> (This allows, of course, for the epistemic possibility of moral emotions which cannot be endorsed, and of moral beliefs that do not correspond to any moral emotions, e.g. because their contents are too abstract, or too fine-grained, or whatever. There is thus a many-many relation between moral emotions and moral beliefs, despite the existence of systematic pairs.)

It is an open question, of course, just which emotions are the “moral” emotions. Various claims have been made on behalf of different emotions for this status, and various notions of moral emotion have been presupposed in these discussions. In this paper, the relevant notion of a moral emotion is the notion of an emotion which embodies moral evaluation in virtue of the attitude it deploys. It is up for debate just which emotions are like this, but there are some paradigms we can all agree on, such as guilt and indignation. There may also be positive counterparts of these, perhaps certain types of pride and appreciation, though the matter is complicated (see Kriegel forthcoming). Other natural candidates include certain types of respect (see notably Darwall 1977 on “recognition respect”) and its negative counterpart in contempt (see Mason 2003), as well as certain types of love (Velleman 1999). It is not necessary to take a substantive position on these matters to appreciate that there is likely going to be a plurality of moral emotions, many of which admitting of potential endorsement resulting in a moral belief.

#### *4 Pluralism, the Frege-Geach Problem, and Moral Phenomenology*

The basic motivation for pluralism about moral judgment is the fact that moral evaluation can occur either through content or through attitude, and a moral judgment just is a mental moral evaluation. Elsewhere I have argued that there are also *architectural* reasons to expect the mind to employ two different kinds of moral judgment (Kriegel 2012). In the remainder of this paper, I want to show that pluralism brings with it novel resources for addressing the central challenges facing cognitivism and noncognitivism. I start, in this section, with two central challenges to noncognitivism.

Perhaps the greatest dialectical embarrassment for noncognitivism is the Frege–Geach problem: the problem, very roughly, of understanding how moral judgments can embed in inference and reasoning if they are, as noncognitivists hold, noncognitive states that are not truth–apt (see Geach 1960). Consider this piece of perfectly good reasoning: if nobody visited grandma last week, then I ought to do so this week; nobody did visit grandma last week; so, I ought to visit grandma this week. Something like this reasoning could perfectly well take place in someone’s mind, and it would be a very good piece of reasoning indeed. What makes it good? Part of the answer is that it is a *valid* inference, in the familiar sense that it is *truth–preserving*: if the premises are true, then the conclusion *must* be true. This presupposes, however, that the premises and the conclusion be truth–apt, which by most accounts means they are cognitive states.

Various moves are available to noncognitivists (e.g., fashioning a notion of validity that is not understood in terms of truth–preservation). Such moves are available to pluralists as well, of course. But pluralism is in reality entirely unthreatened by the Frege–Geach problem. The apparent lesson of the Frege–Geach problem is that it is possible for us to infer moral judgments (and to infer *from* moral judgments). For this to be possible for us, we must, it would seem, have some moral judgments that are truth–apt. And this is something pluralism not only allows but *entails*. Crucially, there is nothing in the Frege–Geach “datum” – the fact that we can infer to and from moral judgments – to suggest that *all* moral judgments must be truth–apt. In other words, the fact that we have moral judgments that enter into inferential relations does not show that we lack moral judgments that do not.

In conclusion, once we frame clearly what the Frege–Geach datum is, we see that it presents no problem for pluralism and is if anything a feather in its cap, insofar as pluralism *predicts* that some moral judgments could enter into inferential relations.

I want to say something similar about another traditional source of resistance to noncognitivism. This is the idea, close to many cognitivists’ heart, that moral judgments have “objective pretensions.” The idea is at bottom phenomenological: that judging that genocide is wrong *feels* like mentally asserting an objective fact. When one wonders whether abortion is morally permissible, it is liable to feel to one as though there is a correct answer, an

answer one is trying to *find out*. Reaching a judgment may then feel like *coming down* on the issue, thus committing to the truth of one position at the expense of the other.

Here too, I want to say that the phenomenological datum is, strictly speaking, that *some* moral judgments feel objective. Certainly whether genocide is wrong or not is something most of us have *beliefs* about. We do not just have emotions like indignation when it comes to genocide, we also *endorse* these emotions. We should only expect, then, that such beliefs will exhibit the objective phenomenology cognivist point to. Thus the pluralist *predicts* that some moral judgments have objective pretensions.

Unlike the cognivist, however, the pluralist has room for moral judgments that lack this objectivistic phenomenology. The clearest examples would be the moral emotions we experience but do not endorse. Going back to A's sorry experiment in bold unconventional lifestyle, we can envisage A feeling instinctively guilty about sleeping with C but taking distance from that feeling, actively "dis-endorsing" it so to speak. In that scenario, A's disendorsed guilt will likely *not* feel like a representation of an objective fact, the objective fact that it was wrong to sleep with C; on the contrary, it will feel like a subjective hang-up to get over. Both kinds of moral evaluation occur in our mental life, I contend: those that feel like they home in on an objective moral fact (these will include prominently paradigmatic moral beliefs) and those that feel much more subjective (these will include prominently unendorsed – or perhaps *disendorsed* – moral emotions). All of which is precisely as the pluralist predicts.

Might the cognivist insist that, even if we have emotions that embody moral evaluation, only *beliefs* that embody moral evaluation deserve to be called "moral judgments" – say because the emotions are somehow "less than a judgment"?

She could, of course, but if this is intended as an objection to pluralism, it faces a dilemma. For the claim is either (a) that moral emotions should not be called moral judgments given the *everyday* sense of "judgment" or (b) that they should not be called that given the *technical* sense of "judgment." Option (b) is not really viable, however, given that the term "moral judgment" is used in the relevant technical literature as neutral between the cognitive and the

noncognitive (since cognitivism and noncognitivism are both theories of a phenomenon referred to in that literature as “moral judgment”). The idea must be, then, that the term “judgment” *as commonly used* denotes a doxastic rather than emotional state. But if this is the idea, I can only concur wholeheartedly, since as noted I think the use of “judgment” as neutral between the cognitive and the noncognitive is aberrant. In this form, however, the objection becomes unsubstantive. What the pluralist is giving a theory of is the phenomenon of morally evaluating something in one’s mind, and this objector appears to contest no aspect of the pluralist’s theory of this phenomenon. In particular, she does not contest the idea that we can morally evaluate something in our mind by having emotions that morally evaluate that thing. She just chooses to reserve the label “moral judgment” for moral *beliefs*, which is of course her prerogative. As long as the pluralist’s underlying picture of how the mind morally evaluates is left intact, though, the question of what to grant the label “moral judgment” to becomes a bookkeeping question.

## *5 Pluralism and the Specter of Error Theory*

One of the main worries about cognitivism is the looming specter of error theory. The thought goes something like this. Suppose moral judgments are beliefs. A belief is true just when there is something in the world that *makes* it true – a *truthmaker*. Thus, if evaluating the Armenian genocide as morally wrong in one’s mind is just having the belief that the Armenian genocide was morally wrong, then there must be something in the world that makes true this belief – a state of affairs, presumably, consisting in the Armenian genocide instantiating the property of being morally wrong. For a variety of reasons, however, many philosophers have suspected that there could be no such properties and therefore no such states of affairs. That is, there can be no moral truthmakers. If so, beliefs to the effect that those states of affairs obtain are erroneous one and all. Our entire moral-belief system is, in a sense, a grand illusion.

This problem may appear to affect moral judgment pluralism just as much as cognitivism, since pluralism is committed more or less to all the moral beliefs cognitivism is committed to; it just adds to them non-belief states that also embody moral evaluation. In this section, however, I want to argue that pluralism

provides two new kinds of input into the dialectic that make error theory significantly less pernicious.

Obviously, the simplest solution for the cognitivist is to insist that there *are* moral facts that make moral beliefs true. This move is of course available to the pluralist as well. What I want to convince you in this section is that the pluralist has certain moves available to her that are *not* available to the cognitivist – in other words, that the pluralist is in a better dialectical position than the cognitivist.

Start with the following question: What is so bad about error theory? Suppose there are in fact no moral truthmakers, so our moral beliefs are untrue. That would be unfortunate for us qua moral believers, but what difference would it make to us qua moral *theorists*? What is the *theoretical* embarrassment in error theory? It might be true, after all, that there are no moral truthmakers, but that we have moral beliefs.

There are two *theoretical* problems, I want to suggest, with error theory – and pluralism helps with both.

The first problem has to do with the fittingness, or appropriateness, or correctness conditions of belief. It is fitting to have a belief just when the belief is true. If  $p$  does not hold, then belief is the wrong attitude to take toward  $p$ . And if you *know* that  $p$  does not hold, then you should definitely not adopt the attitude of belief toward  $p$ . Accordingly, if evaluating the Armenian genocide as morally wrong in your mind is a matter of having the belief that the Armenian genocide is morally wrong, and the belief is untrue, then given that it is never fitting to have an untrue belief, it follows that it is unfitting to evaluate the Armenian genocide as morally wrong in your mind. And if you happen to be a moral anti-realist, rejecting moral truthmakers, then you should definitely not evaluate the Armenian genocide as wrong in your mind. The first problem with error theory is that such consequences are manifestly implausible. In reality, few things are more fitting than evaluating the Armenian genocide as wrong in one's mind. And it is very odd to think that, having been convinced by one argument or another for moral antirealism, one should stop evaluating the Armenian genocide as wrong in one's mind.

Here pluralism offers a clear dialectical advantage as compared to cognitivism. For suppose that it is never fitting to adopt the *belief* attitude toward the proposition that the Armenian genocide was morally wrong. Still, pluralism allows that there are *other ways* to morally evaluate the Armenian genocide as wrong in one's mind. We can also adopt the right *emotional* attitudes toward the Armenian genocide – say, feel *indignant* about it. Accordingly, if I am a moral antirealist who rejects moral truthmakers, I can still rationally and fittingly evaluate the Armenian genocide as morally wrong in my mind, for instance by feeling indignant about it. In this way, pluralism can embrace error theory about cognitive moral judgments and still account for the manifest fact that many of the mental evaluations we perform are fitting or appropriate.

It is also significant for the fittingness status of moral beliefs themselves that certain corresponding moral evaluations are fitting. It is true that for a belief that torturing kittens for fun and profit is wrong to be fitting, the belief must be true; and that if there are no moral truthmakers, then it is *untrue*. Nonetheless, there is an important difference between the belief that torturing kittens for fun and profit is wrong and the belief that writing poetry on a Wednesday is wrong. The error theorist needs to be able to account for this difference. The pluralist has a ready explanation. Although the belief that torturing kittens for fun and profit is wrong is not itself fitting, it *corresponds to* a moral judgment that *is* fitting, namely, indignation about torturing kittens for fun and profit. (What is it for a belief B to “correspond” to a fitting moral judgment J? Perhaps for B and J evaluate in the same way the same thing.) In contrast, not only is the belief that writing poetry on a Wednesday is wrong unfitting, it corresponds to no moral judgment that is fitting. This marks a crucial difference between the two beliefs.

To signal the different status of these two beliefs, we might with some legitimacy introduce a notion of “secondary fittingness.” A moral belief B to the effect that some *x* is wrong (or is M for any moral predicate M) is secondarily fitting just if there is a moral emotion E, such that E represents-as-wrong *x* (or more generically, such that E represents-as-M *x*). The justification for calling B “secondarily fitting” here is simply that B and E are in an important sense the same *kind* of moral evaluation, and E is fitting. The kitten-torture belief is “secondarily fitting,” then, because (i) it evaluates kitten torture as morally wrong and (ii) there is a fitting – *primarily* fitting – mental state that evaluates



kitten torture as morally wrong (namely, indignation about kitten torture). In contrast, Wednesday-poetry is both primarily and secondarily unfitting.

This is important, because a second theoretical problem with error theory is that it fails to make crucial discriminations. It is all well and good to say that all moral beliefs are false, but surely we must find some merit in the belief that genocide is morally wrong and a corresponding demerit in the belief that genocide is morally unproblematic. Compare: the grand illusion theory of color perception holds that all color perception is illusory, since there are no color properties in the world; but such a theory is often supplemented with an account of the difference between (a) a person who undergoes bluish experiences when looking at the sky on a sunny day and (b) a person who when looking at the sky has all manner of random color experiences, from the purplish to the orangish, regardless of the state of the sky. (For a relatively well-known attempt to capture this difference, see Chalmers 2006, esp. §7.) The error theorist likewise must find a way to draw crucial distinctions between moral beliefs. The moral judgment pluralist can do so by distinguishing those moral beliefs that are secondarily fitting from those that are secondarily unfitting. This is because she allows (primarily) fitting non-belief moral judgments that embody the same kind of moral evaluation as some moral beliefs. The cognitivist, in contrast, has no similar resource in her theory, since for her *all* moral judgments are beliefs.

To repeat, I am well aware that one way the cognitivist has out of error theory is to embrace the sort of robust moral realism that comes with an ontology of objective moral facts, which can then serve as truthmakers for moral beliefs. No part of our discussion is intended to question (nor endorse!) this kind of move. Our discussion is once removed from the substantive question of whether to embrace moral realism in the metaphysics of morals. Our point here is only that this sort of move is available equally to the cognitivist and to the pluralist, but that in addition, there are important moves available to the pluralist which are *unavailable* to the cognitivist. Bringing out this asymmetry, and the resulting increased stability of the pluralist's dialectical position, is the only purpose of the present discussion. Its upshot may be stated thus: *If* it turns out that the world contains no moral facts that could serve as truthmakers for moral beliefs, then the cognitivist (i) must apparently hold that morally evaluating genocide negatively in one's mind is always and everywhere inappropriate and

(ii) cannot produce any distinction between believing in the permissibility of genocide and believing in the impermissibility of genocide; whereas the pluralist can (i) allow for *some* perfectly fitting negative moral evaluations of genocide and (ii) distinguish in a principled way between beliefs in the permissibility versus impermissibility of genocide (namely, in terms of what I have called secondary fittingness). In other words, if it turns out that the world contains no moral truthmakers, cognitivism leads directly to some fairly intolerable consequences which are avoided by pluralism.

There is an objection to this line of thought that may be put as follows. For indignation about the Armenian genocide to be fitting, it would have to be the case that the Armenian genocide is wrong. If it is not wrong, what makes indignation about it fitting? But now, once we admit that that the Armenian genocide is wrong, it would seem that a belief to the effect that it is wrong would be true after all, and hence fitting. So there cannot be any real dialectical advantage here for pluralism over cognitivism, insofar as it could never be the case that indignation about *x* is fitting but belief that *x* is morally wrong is unfitting. For what makes indignation about *x* fitting is that *x* is morally wrong, and if *x* is morally wrong, then a belief that *x* is morally wrong is also fitting.

My response is that although it is certainly one legitimate view that something's being morally wrong is what makes indignation about it fitting, that is not the only legitimate view in the area. One notable alternative is that what makes indignation about the Armenian genocide fitting is a collection of "descriptive" or "empirical" facts about the Armenian genocide: that it involved the death of many innocent people, that these innocents were slaughtered due to their ethnicity, that suffering in Armenian families persisted many decades thereafter, and so on (compare Scanlon 1998 Ch.2). If this is right, then indignation about the Armenian genocide can be fitting *even if there are no moral facts*, as the error theorist claims.<sup>4</sup> Here lies the dialectical advantage of the pluralist over the cognitivist, then: if (a) there are no moral facts, but (b) it is in virtue of certain empirical facts that certain moral emotions are fitting, then the cognitivist is stuck with the unseemly consequence that mentally evaluating the Armenian genocide as wrong is always and everywhere unfitting, whereas the pluralist avoids this consequence. Now, it is true that if one rejects (a) or (b), then the pluralist has no advantage over the cognitivist: if one rejects (a),

then the cognitivist has truthmakers for moral beliefs, which can therefore be fitting; if one rejects (b), then the pluralist cannot account for the fittingness of moral emotions (without, that is, making moral beliefs fitting as well). Still, the fact that on some perfectly defensible constellation of positions in this area the pluralist has an advantage over the cognitivist means that the pluralist brings something new to the dialectic.

To be clear, none of the above is intended as *endorsing* (a) and (b). My discussion is once removed from the vicissitudes of the relevant first-order debates and pertains rather to the overall dialectical landscape. What I have attempted to bring out in this section is the extra resources the pluralist has for dealing with one of the central issues in the extant literature on the fate of moral cognitivism. In the next section, we do the same exercise on another issue central to the dispute between cognitivists and noncognitivists: the connection between moral judgment and motivation.

## *6 Pluralism and Moral Motivation*

One of the chief arguments against cognitivism, and thus *for* noncognitivism, goes roughly like this: 1) moral judgments are inherently motivating; 2) cognitive states are not; therefore, 3) moral judgments are not cognitive states. The noncognitivist maintains that mental states such as desire and emotion move one to action by their nature, whereas mental states such as belief and perception merely record how things are, without in themselves propelling us to *do anything about it*. But moral judgments, by their very nature, do propel us to action (*or* inaction), and are therefore better thought of as noncognitive states cut of the same cloth as desire or emotion.

Although it is certainly possible to contest Premise 2 of the above argument, most cognitivists have denied Premise 1. David Brink (1989 Ch.3), for instance, argues that 1 is incompatible with the conceivability of an amoralist: a person who judges that she ought to do something but is simply unmoved to do it. You may find yourself lying on your couch, thinking to yourself you should really go to that dinner party you are expected at, and somehow just feel no pull to actually do it. Now, the noncognitivist may certainly insist that in such circumstances you do feel an ever so slight tug to go to the party, or else you

cannot be seriously described as judging that you ought to go. Bracketing these further debates, I want to point out that the pluralist can easily take at face value the apparent conceivability of an amoralist: It would have to be someone who has certain moral beliefs without having the corresponding moral emotions (someone, e.g., who believes it would be wrong not to go to the party but does not feel guilty about not going). This might not be a *typical* “someone,” but it is certainly a *conceivable* “someone.”

One influential argument for an internal, non-contingent connection between moral judgment and motivation comes from Michael Smith (1994 Ch. 3) and goes something like this. There is a reliable correlation in good and strong-willed persons between making moral judgments and being motivated to act accordingly. This correlation calls for explanation. The noncognitivist’s explanation is that moral judgment is inherently motivating: it is part of what makes a mental state a moral judgment that it motivates. There is a kind of cognitivist explanation, in contrast, that Smith finds unacceptable: that the moral judgment in itself is not motivating, but in good and strong-willed persons it is paired with another mental state that is, namely, the desire *de dicto* to do the right thing (i.e., a desire to do the right thing whatever that turns out to be). This explanation casts moral motivation as *fetishizing* morality, as though people who avoid genocidal acts care not a jot about the suffering of genocide victims, concerning themselves only with what they take to be right and wrong, regardless of the underlying affairs of human and other sentient beings.

The pluralist has a third and distinct explanation, however. It is that in good and strong-willed persons, moral beliefs and moral emotions tend to converge. Certainly classical portrayals of the virtuous person emphasize the harmony of “reason and emotion,” and intuitively, when a good and strong-willed person believes that it is wrong not to go to a dinner party at which she is expected, she feels guilty about not going. Lying there on her couch, totally not in the mood for this party but keenly aware of being expected at it, her guilt increases and increases until she gets up and goes. Because of the reliable correlation between her moral beliefs and moral emotions, and the inherent motivational character of the latter, there is also a reliable correlation between her moral beliefs and her motivation. On this explanation, then, the reason good and strong-willed persons who judge that it would be morally right to  $\phi$  tend to

$\phi$  is neither that all moral judgments are inherently motivating, nor that such persons desire *de dicto* to do the right thing; rather, it is that such persons tend, on the one hand, to have emotions that represent-as-morally-right  $\phi$ -ing, which emotions are inherently motivating, and, on the other hand, tend to have beliefs that represent-as-true the moral rightness of  $\phi$ -ing when they have emotions that represent-as-morally-right  $\phi$ -ing.

In this way, pluralism can capture some of the alleged “data” that motivate both the cognitivist’s and the noncognitivist’s view on the connection between moral judgment and motivation. The cognitivist’s “datum” (or one of them) is the conceivability of the amoralist; (one of) the noncognitivist’s is the reliable correlation between moral judgment and motivation in good and strong-willed persons. Pluralism is not just consistent with these “data” – it *predicts* them. Cognitivism and noncognitivism, in contrast, each predict only one of them. Note well: I do not mean to endorse either datum here! The claim I am making is again once removed from the ground-level debate about the actual connection between judgment and motivation, and pertains only to the resources different views of the nature of moral judgment have for accommodating certain considerations that have proved alluring to a significant number of moral philosophers.

## *Conclusion*

One of the liveliest debates in the philosophy of mind circa 1990 was between the “theory theory” and the “simulation theory” of folk psychology. The organizing question was: to understand each other, do we apply an implicit theory of what makes people tick, or do we “put ourselves in other people’s shoes” and simulate their inner lives? It took only a little over a decade for many of this debate’s protagonists to start converging on a simple answer: *Both!* There is no theoretical reason why we should be assumed to use only one of these two types of mechanism – theory application and mental simulation – in trying to understand each other. On the contrary, there is every reason to expect both mechanisms to be operative in us. And so the real issue had become what specific *roles* one should give to each of these mechanisms in one’s overall approach to folk psychology.

My recommendation in this paper has been to take a similar stance toward the debate between cognitivism and noncognitivism about moral judgment. Do we mentally morally evaluate aspects of the world through cognitive states like belief, or through noncognitive states like emotion? I recommend we answer “Both!” here as well. I see little more theoretical pressure to assume a single manner of moral orientation on the world in our psychological repertoire.<sup>5</sup> I have argued that once we adopt the pluralist approach, the considerations that have structured much of the debate on cognitivism vs. noncognitivism lose much of their force. Crucially, however, the motivation for pluralism does not come simply from the theoretical benefits attendant to its adoption. The basic – and entirely independent – motivation for pluralism is that moral evaluation can occur either through content or through attitude, and so *it stands to reason* that there should be these two fundamentally different modes of morally evaluating something in one’s mind. In other words, the content–attitude distinction, combined with the observation that both content and attitude can embody moral evaluation, should make us *expect* there to be both cognitive and noncognitive moral judgments.<sup>6</sup>

## References

- Brink, D. 1989. *Moral Realism and the Foundations of Ethics*. Cambridge: Cambridge University Press.
- Chalmers, D. 2006. ‘Perception and the Fall from Eden.’ In T. Gendler and J. Hawthorne (eds.), *Perceptual Experience*. Oxford: Oxford University Press.
- Darwall, S.L. 1977. ‘Two Kinds of Respect.’ *Ethics* 88: 36–49.
- Geach, P.T. 1960. ‘Ascriptivism.’ *Philosophical Review* 69: 221–225.
- Greenspan, P.S. 1992. ‘Subjective Guilt and Responsibility.’ *Mind* 101: 287–303.
- Johnston, M. 1989. ‘Dispositional Theories of Value.’ *Proceedings of Aristotelian Society* 63: 139–174.
- Kriegel, U. 2012. ‘Moral Motivation, Moral Phenomenology, and the Alief/Belief Distinction.’ *Australasian Journal of Philosophy* 469–486.
- Kriegel, U. Forthcoming. ‘Indignation, Appreciation, and the Unity of Moral Experience.’ *Ethical Theory and Moral Practice*.
- Leslie, S-J. and A. Lerner 2016. ‘Generic Generalizations.’ *Stanford Encyclopedia of Philosophy*.

- McDowell, J. 1985. 'Values and Secondary Qualities.' In T. Honderich (ed.), *Morality and Objectivity*. Boston: Routledge & Kegan Paul.
- Mackie, J.L. 1946. 'A Refutation of Morals.' *Australasian Journal of Psychology and Philosophy* 24: 77–90.
- Mackie, J.L. 1977. *Ethics: Inventing Right and Wrong*. New York: Penguin.
- Mason, M. 2003. 'Contempt as a Moral Attitude.' *Ethics* 113: 234–272.
- Nickel, B. 2008. 'Generics and the Way of Normality.' *Linguistics and Philosophy* 31: 629–648.
- Olson, J. 2014. *Moral Error Theory: History, Critique, Defence*. Oxford : Oxford University Press.
- Scanlon, T. 1998. *What We Owe to Each Other*. Cambridge MA: MIT Press.
- Smith, M. 1994. *The Moral Problem*. Oxford: Blackwell.
- Velleman, J.D. 1999. 'Love as a Moral Emotion.' *Ethics* 109: 338–374.

---

<sup>1</sup> If one formulates cognitivism and noncognitivism as generics rather than universals, it becomes trickier to *formulate* pluralism. But this is mostly because the exact degree of freedom involved in withdrawing from a universal to a generic is unclear at present, with many different theories proposed but none commanding consensus (see Leslie and Lerner 2016 for a partial review). On the simplest theory, generics are disguised universals about *typical* or *normal* instances (see, e.g., Nickel 2008). Thus, to say that moral judgments are cognitive states is to say that all typical moral judgments are cognitive states. Pluralism would then become the thesis that some *typical* moral judgments are cognitive and some are noncognitive. On other views of what a generic exactly says, pluralism would have to be formulated differently. Most generically, so to speak, pluralism could be formulated as the conjunction of two negations of generics: “moral judgments are cognitive” is false & “moral judgments are noncognitive” is false.

<sup>2</sup> I am assuming here that emotions cannot be exhaustively reduced to belief or some other cognitive state. There are of course theories of emotion that do reduce them entirely to beliefs, but against the background of those theories the debate between cognitivism and noncognitivism becomes considerably less interesting anyway.

<sup>3</sup> This does commit us to a specific picture of what is involved in the relevant form of endorsement. It would have to be a process that effects a transition from a mental state that represents-as-F <x> to one that represents-as-true <x is F> (e.g., from a state that represents-as-morally-wrong <sleeping with C> to a state that represents-as-true <sleeping with C was morally wrong>). What endorsement does, on this view, is to take information which is in some sense implicit in the attitude and make it explicit in the content.

---

<sup>4</sup> It might be objected that prominent error theorists such as Mackie (1977) and Olson (2014) deny that empirical facts can make moral judgments fitting. But this is because Mackie and Olson hold that *all* moral judgments are beliefs, and moreover that the moral concepts invoked in the contents of those beliefs are concepts that purport to pick out non-natural, irreducibly normative properties. We take no stand here on how moral concepts present the properties they purport to pick out, but in any case we deny that all moral judgments are beliefs, and therefore that making a moral judgment implies applying moral concepts. How our moral concepts present the putative fittingness-makers is irrelevant to what might make fitting a moral judgment that does not involve any moral concept.

<sup>5</sup> The only pressure toward a unitarian account here comes from the demand that moral judgment be intimately linked to motivation, but as we saw the nature of the link is quite ambiguous and claims to accommodate it best are made by the cognitivist, the noncognitivist, and indeed pluralist alike.

<sup>6</sup> For comments on a previous draft, I am grateful to Abe Roth and an anonymous referee for *Philosophical Studies*. I have benefited from presenting material relevant to this paper at Brown University and Rice University, as well as at the Chapel Hill Normativity Workshop. I am grateful to the audiences there, in particular Mustafa Aghahosseini, Christian Blac t, Gwen Bradford, Alex Campbell, Jamie Dreier, Anna Giustina, Eric Guindon, Max Khan Hayward, Richard Heck, Thomas Hofweber, Christa Johnson, Hitkarsh Kumar, Andrew Lee, Chad Marxsen, Daniel Pinto, Gideon Rosen, Reuben Sass, Josh Schechter, Sarah Stroud, Patricia Thornton, Anna Tsvetkov, Alex Worsnip, and Orfeas Zormpalas.