

Gripped by authority

Terry Horgan and Mark Timmons

Department of Philosophy, University of Arizona, Tucson, AZ, USA

ABSTRACT

Moral judgments are typically experienced as being *categorically authoritative* – i.e. as having a prescriptive force that (i) is motivationally gripping independently of both conventional norms and one's pre-existing desires, and (ii) justificationaly trumps both conventional norms and one's pre-existing desires. We argue that this key feature is best accommodated by the meta-ethical position we call 'cognitivist expressivism', which construes moral judgments as sui generis psychological states whose distinctive phenomenological character includes categorical authoritativeness. Traditional versions of expressivism cannot easily accommodate the justificationaly trumping aspect of categorical authoritativeness, because they construe moral judgments as fundamentally desire-like. Moral realism cannot easily accommodate the aspect of inherent motivational grip, because realism construes moral judgments as a species of factual belief.

ARTICLE HISTORY Received 31 December 2017; Accepted 10 January 2018

KEYWORDS Categorical authoritativeness; expressivism; moral judgment; motivation; phenomenology

Metaethical expressivism has its challenges, including those that concern embedding, negation, and making sense of contexts in which one makes a moral judgment but thinks, 'I might be wrong.' In some of our previous work we have had something to say in response to these challenges and in defense of our own non-reductive brand of expressivism, 'cognitivist expressivism'.¹ However, another serious challenge to irrealist versions of expressivism like ours concerns the *authority* of moral reasons and the moral judgments they ground.² We have begun to address this challenge, too, in some of our most recent work,³ but there is more to do. In particular, there is the phenomenon that we like to refer to as the *experienced authority* of moral reasons and associated moral judgments. Some philosophers argue that the experienced authority of moral reasons provides pro tanto reason to favor either some robust form of moral realism or moral error theory (the latter type of view perhaps accompanied by a fictionalist story about moral judgment). The argument we have in mind is overtly phenomenological, and in this chapter we argue that careful attention

CONTACT Terry Horgan  thorgan@u.arizona.edu; Mark Timmons  mtimmons@u.arizona.edu

© 2018 Canadian Journal of Philosophy

to the phenomenology of the experienced authority of moral reasons makes trouble for competing metaethical views, but is fully compatible with our brand of non-reductive irrealist expressivism.⁴

Here is our game plan. Because we are concerned with a phenomenologically grounded challenge to expressivism, we begin in Section 1 with a specific scenario involving an agent's morally tinged decision and its psychological aftermath that will be put to use later in the paper when we respond to the challenge. Then, in Section 2, we turn to some of the work by Jean Hampton who provides a detailed phenomenological description of what it is like to experience the authority of reasons, a description nicely complemented by remarks from J. L. Mackie.⁵ What we point out is that the experienced authority of moral reasons is Janus-faced: it has a motivational dimension and a normative justificatory dimension, which we use the work of Hampton and Mackie to highlight. Properly accommodating these dimensions of concrete moral experience, we claim, sets a challenge for metaethical theories, including our own brand of expressivism. With our concrete scenario, a phenomenological description of reasons experiences, and the challenge on the table, we turn in Section 3 to select metaethical theories (including versions of 'reductive' expressivism), arguing that each of them has trouble in the face of the phenomenological challenge. Either they do not acknowledge the phenomenology, or they do but must pay a high price for doing so. In Section 4, we articulate a *non-reductive* form of expressivism that we call 'cognitivist expressivism,' and then in Section 5, we return to our concrete scenario and explain how cognitivist expressivism can meet the challenge we set for metaethical theories, and do so without tears. Section 6 is our conclusion.

1. A concrete scenario: Clive's cowardice

The (fictional) scenario is from Ian McEwan's 1998 novel *Amsterdam*, in which the character Clive Linley, a Londoner and a composer of some notoriety, is struggling to complete a symphony, celebrating the new millennium, for an upcoming concert in Amsterdam. His aim is to compose a masterpiece that will become the crowning achievement of his already illustrious career. Frustrated with a succession of failed attempts to compose the finale, Linley decides to seek inspiration by spending time away from the city, hiking in the Lake District located in a mountainous region of northwest England. While hiking, Linley is suddenly struck with an idea for the finale, and stops to scribble notes, attempting to work out the melody. His concentration is interrupted when he hears the nearby voices of a man and a woman quarreling. Clive climbs to the top of a large rock where he can see the quarreling couple standing face to face in a small clearing about thirty yards away. As the confrontation continues, loud talking soon gives way to shouting. The man grabs the woman's elbow, violently pulling her in his direction. Witnessing all this, Clive, with pencil and notebook in hand, sighs and ponders what to do:

Was he really going to intervene? He imagined running down there. The point at which he reached them was when the possibilities would branch: the man might run off; the woman would be grateful Even this least probable of outcomes would destroy his fragile inspiration. The man was more likely to redirect his aggression at Clive while the woman looked on, helpless. Or gratified, for that was possible too; they might be closely bound, they might both turn on him for presuming to interfere What was clear now was the pressure of choice: he should either go down and protect the woman, if she needed protection, or he should creep away ... He could not remain here doing nothing. (93)

Clive closes his eyes and tries to concentrate on the elusive melody he is after. But at the sound of angry voices he takes another look. The woman breaks loose of the man's hold with a sharp downward jerk of her arm, and turns to run, but the man tackles her from behind. They fall to the ground, the woman trying to crawl away, the man holding onto her ankles. The man, having gotten up, is now dragging the woman, both hands on her left ankle; she screams. Clive now understands the seriousness of the unfolding event; he thinks for a moment that he absolutely *must* intervene, come to the woman's aid. But the significance, the importance of his work! Clive hurries away from the scene, trying hard to recall those few notes of the melody he was so desperately trying to work out. 'He was trying to call it back, but his concentration was being broken by another voice, the insistent, interior voice of self-justification: ... if he had approached the couple, a pivotal moment in his career would have been destroyed' (95). Back from his hike, Clive decides to leave immediately for London, certain that he can work out the entire finale on the train. As he excitedly waits for his taxi ride to the train station, he reflects: 'He wanted the anonymity of the city again, and the confinement of his studio, and – he had been thinking about this scrupulously – surely it was excitement that made him feel this way, not shame' (96).

For our purposes here, let us construe this scenario – by stipulation – as conforming to the following interpretation, which is suggested implicitly by McEwan's own characterization. Clive experiences a specific consideration – viz., that the woman is being assaulted and needs help – as an authoritative moral reason to intervene, despite deciding not to do so. The envisioned possibility of the man redirecting his aggression toward Clive, perhaps with the woman joining in, enters his mind primarily as a potential excuse for not intervening whose flimsiness he already appreciates (he would be in no serious danger of significant bodily harm) – and which in any case, he realizes, is clearly outweighed by a compelling moral reason once he sees the woman try to escape and the man then dragging her by the ankle. And the subsequent feeling that keeps him scrupulously thinking about what happened, rather than being able to concentrate on working out the finale of his symphony, really is shame rather than excitement.

We offer this story as a fictional, but reasonably concrete, case involving what we go on to characterize as someone who is gripped by the experienced authority of moral reasons.⁶ We shall return to this story in Section 5 when we explain

how our irrealist version of metaethical expressivism is able to accommodate the experience of the normative authority of moral reasons as featured in our story. In the meantime, our next order of business is to dwell on the phenomenology of moral-reasons experiences, guided primarily by the pertinent phenomenological observations we find in some of Jean Hampton's work.⁷

2. The authority of moral reasons

We turn now to a characterization, in general terms, of the phenomenology of moral-reasons experience, leaning heavily on a chapter from Jean Hampton's posthumously published book, *The Authority of Reason*, from which we quote liberally. Her chapter is particularly nice for our purposes, because in addition to offering a fairly detailed characterization of the phenomenology in question, she also explains why she thinks only a robust form of non-naturalist moral realism is consistent with this phenomenology. We also quote from Mackie, whose phenomenological observations complement Hampton's. Our aim, then, in consulting these authors is to assemble a collection of phenomenological observations about moral-reasons experience that we think any plausible metaethical theory ought to 'accommodate,' in a sense to be explained in the following section.

In the third chapter of her book, entitled 'Reason's Authority,' Hampton's stated goal is to argue that there is something about ethics 'that appears to make it scientifically problematic,' viz., 'a certain thesis held by moral objectivists about moral norms and the reasons they generate that fails to pass scientific muster' (83). Her argument is phenomenological, focusing on introspectively salient aspects of moral experience (Note well: that such aspects are available to introspection will be crucial when we eventually address the phenomenological challenge, appealing to our version of expressivism). Hampton's argument proceeds in two stages. In the first stage she offers what she says is a 'minimalist and metaphysically neutral' initial characterization of how the normativity of moral reasons is experienced (83). In the second stage, she describes two metaphysically different ways of 'embellishing' the initial characterization (83), and she argues that only the second way – a version of non-naturalist moral realism – comports with the phenomenology of moral-reasons experience. In this section we briefly summarize her argument, in a way that closely follows her own text.

1.1. The first stage

Hampton begins by saying that the notion of authority in question is not decisiveness, because the reasons in support of a particular act can carry authority even if they are outweighed by reasons in support of some other act; and that the pertinent notion is not strength of motivational commitment either, because the reasons in support of a particular act can be experienced as stronger, qua reasons, than considerations upon which one chooses to act. Rather, she says,

a theory of the authority of norms tries to explain *what it means* for a norm to be 'applicable' to us. The word 'applicable' is a poor one, because norms don't merely 'apply' to us, they *direct* us. Indeed, we use all sorts of words to elaborate on this applicability: besides 'authority,' we speak of a norm's 'prescriptivity' or its 'obligatory force' over us, its 'compelling nature' or its 'pull,' its status as an 'order' or a 'command' (and not a mere 'suggestion').... That is, normative authority presupposes... that it is correct to say that it specifies a reason for x-ing for an agent. (88)

Here, she does not say explicitly what she takes a 'norm' to be, and for present purposes this does not much matter.⁸ The crucial thing presently, we take it, is this: to *experience* a norm 'as specifying a reason for x-ing for an agent,' where consideration C is the specified reason, is to experientially *regard* consideration C as a reason for x-ing for an agent (Presumably, the agent in question might or might not be oneself. Often enough, one experiences a moral reason as universalizable: a reason for x-ing for *anyone* who might be in the pertinent circumstantial situation).

She points out that 'this compelling quality of reasons is not ... the same as the feeling of liking or approving the directive of a reason. It is easy to give examples of people who know they have a reason to x, and who not only do not like the action that the reason directs, but even despise it' (89).

She also points out that not all reasons have the same kind of directive force. 'In particular, some reasons command us, and thereby give us mandates, and others direct us in ways that indicate permissions, rather than commands' (90). Nonetheless, 'what permissive and mandatory reasons have in common, such that they are all *reasons*, is their authority – the sense in which they have for us a "compelling rightness"' (91).

She takes all this to be phenomenological description, a matter of isolating 'the distinctive "feel" of reasons' (93). She also takes it to be so far metaphysically neutral; she denies that she has meant 'to suggest that this compelling quality must be understood as somehow "in the world," and thus a part of our reality' (93). But next comes stage 2 of her argument, in which she articulates and compares two competing theses about normative authority – specifically, about moral normative authority – each of which she considers metaphysical.

1.2. The second stage

The first thesis she considers might naturally be regarded as compatible with a broadly naturalistic metaphysical worldview, and as not being 'scientifically problematic.' As a prelude to introducing it, she begins with the following observations:

Consider a norm in the sport of dressage, requiring owners of horses performing dressage tests in a competition to present the horse in the show ring with its mane braided. It purports to give those who recognize it a reason to act in a certain way ... [T]o explain the authority of this or any similar norm, such as rule in baseball, or a norm of etiquette, or a norm about taste in foods, we would tell a story locating

the social forces that generated the norm, and the psychological responses to those forces by certain people that give these norms their power. In this view, the norm is the (mere) invention of particular human beings by virtue of their interests, and the sense such people have that a norm authoritatively applies to them ... is entirely a matter of social and psychological contingencies This sort of norm, which I will call culture-dependent, is ubiquitous, and includes norms of etiquette, rules of various sports, and ideals of physical beauty. (94)

She now introduces the first thesis as follows, by way of contrast with norms that are obviously and unproblematically culture-dependent:

[S]ome theorists believe that *all* norms, including all moral norms, are culture-dependent. Those who believe this explain the authority of all norms as (what I will call) a 'psycho-social' phenomenon. (94)

(She is here using the expression 'culture-dependent' broadly enough to encompass psychological factors that might be innate or otherwise deeply ingrained in human nature, as well as factors resulting from socialization.)

After explaining why she takes Gibbard's expressivism, Mackie's error theory, and Boyd-style naturalist realism to be versions of the psycho-social thesis, she summarizes as follows what she takes to be the common thread that runs through the various respective versions of the psycho-social metaphysical approach:

All these versions of the psycho-social thesis accept the same basic strategy for explaining the authority of reasons – that authority is understood to be merely in the head (explicated as a feeling, or a cognitive state, or a theoretical belief), and its origins are explicable by virtue of human psychology, human biology, and/or human sociology. (96)

She turns next to a second metaphysical 'elaboration' of the features of moral-authority experience that she described in the first stage of her argument – an alternative elaboration that she takes to be incompatible with the psycho-social thesis. She calls it the 'objectivist' thesis, which she characterizes as follows (we insert numbered brackets into this quoted passage, and ones following it, in order to flag the various key phenomenological observations).

Those who are normative objectivists maintain that some norms (but not all norms – for example, not norms of dressage) are examples of what I will call culture-independent or objective norms. The authority of these norms is supposed to be [1] independent of social and psychological contingencies [O]bjectivists argue that such (independent) authority is the reason society has (or ought to have) such norms as part of its culture, and they insist that no matter the facts of our society or our psychology, we ought to recognize, accept, and obey them. (96)

The notion of objectivist authority, she says, figures in the objectivist view in a number of ways:

First, and most importantly, it appears in the theory's explanation of how it is that moral norms 'apply' to us [T]hese reasons, no matter what we may think or how we have been raised, have authority over us. [2] Such authority is 'outside' the agent, and that to which she is responding when she says that she understands that she ought to act from them [T]he authority is not the invention of the

agent, nor of human communities, but something to which agents and human communities respond.

Second, the moral objectivist assumes that the notion of authority is one that human beings can 'see' or (in some way) discover [W]e usually 'feel' or 'comprehend' its (objective) authority, which means [3] experiencing a sense of its pull, such that we take it to be something that we are in certain circumstances bound to act upon.

Third, the moral objectivist claims that having felt this authority, it is – at least sometimes – an authority for the sake of which we can act, so that it is [4] motivationally efficacious. (98–99)

She further elaborates the aspect of 'outsiderness' as involving a distinctive kind of necessity, about which she says:

Let us say that moral reasons generated by objectively authoritative moral norms are necessary in the sense that [5] their *governance* over us is inescapable. And by 'inescapable' here I mean that these reasons 'apply' to us 'no matter what.' According to this way of thinking about objective authority, no matter what we may do or think, we are directed by these reasons – either in the form of permissions or in the form of mandates. And the governance is inescapable or necessary because there is no way that we can throw it off, or change it by our actions, beliefs, or social systems So understood, normative necessity is still a metaphysical concept because it is supposed to hold regardless of whether or not we know about it or are aware of it. It is just not the metaphysical concept that is usually referred to by the term 'necessity.' (105–106)

And she urges that the 'outsiderness' of moral normative authority, with its aspect of necessity or inescapability, is central to the actual phenomenology of moral-reasons experience:

This way of thinking about authority is, I think, closest to the way that the authority of reasons *feels* to us – that is, it approximates what the experience of 'having a reason' is like for those who understand and act from reasons [R]easons feel like orders – strong in the case of mandates, weak in the case of permissions, but directives nonetheless, with an [6] inescapable rightness about them. (106)

Although she does not say so explicitly, her overall discussion of moral authority clearly implicates – especially in light of the just-quoted passage – that the objectivist thesis fits people's actual moral-authority phenomenology better than does the psycho-social thesis.

This feeling that reasons have an inescapable rightness about them and, moreover, one that is irreducible, is captured nicely by Mackie's well-known characterization. Mackie famously observed that were it true that there are instantiated objective values they would be experienced as [6] having 'to-be-pursuedness somehow built into them.' He adds that 'if there were objective principles of right and wrong any wrong (possible) course of action would have not-to-be-doneness somehow built into it' (1977, 40). This sense of to-be-pursuedness and not-to-be-doneness as reflected in one's moral experiences captures both the categorical authoritativeness and independence of those reasons that are

experienced as mandating the actions they favor, and it also reflects [7] the seeming irreducibility of the pertinent categorical authoritativeness of such reasons.⁹

Let us now itemize, for purposes of subsequent citation, various interconnected features that Hampton (with some help from Mackie) has identified as elements of the phenomenology of being gripped by the authority of moral reasons. For simplicity, we restrict attention to reasons that one experiences as morally *requiring* a certain action, and as requiring such an action *by oneself*. (Her discussion can be generalized to cover reasons that one experiences as favoring a certain action without requiring it, and to cover reasons that one experiences as pertaining to other persons – or to anyone who might find oneself in certain circumstances.) And we focus, as does Hampton, upon experiences of *non-normative* considerations as being reasons for a certain kind of action – non-normative considerations in virtue of which such an action is experienced as being inescapably, authoritatively required. Such experiences, we suggested at the outset of this paper, have a Janus-faced character, involving a blend of the normative and the motivational. The key phenomenological elements of such fittingness-experience are these¹⁰:

- *Pull*: Such reasons are experienced as pulling one toward a certain specific moral judgment and corresponding action (perhaps a refraining). [3]
- *Independence*: The pull of such reasons is experienced as being independent of pre-existing desires or human conventions. [1]
- *External source*: The source of the independence is thus experienced as ‘outside’ oneself. [2]
- *Inescapable necessity*: Such reasons are experienced as inescapable in the sense that they necessarily apply to one; their conditions of application are not contingent. [4]
- *Grip*: Such reasons, when not experienced as being outweighed by other reasons of the same kind that pull toward some incompatible moral judgment and corresponding action, normally exert an involuntary phenomenological grip upon oneself that is experienced as binding – where becoming thus experientially bound toward performing the particular action constitutes a moral judgment that one ought to perform that action, an action that is experienced as something to-be-done (or in the case of wrong actions, not-to-be-done). [7]
- *Compelling rightness*: In cases where one is gripped by a reason for performing a certain action, thereby judging that one ought to perform it, one experiences the envisioned action as compellingly demanded, something one *must* judge and something one feels *ordered* to do. [6]
- *Motivation*: Experiencing the authority of moral reasons is *inherently motivational*; one is typically motivated to judge and act accordingly (although

their motivational strength can be outweighed by other psychological states such as pre-existing desires). [5]

This collection of phenomenological observations about experiencing moral reasons, then, is supposed to capture the experienced authority of moral reasons (If the moral authority of a reason is experienced as a pull without (yet) being experienced as binding, then it is a pull toward both making a moral judgment and performing a corresponding action. But if and when the authority gets experienced as binding – i.e. as exerting a grip upon oneself – this experience constitutes *making* the moral judgment (for that reason).) Thus, as noted at the outset, the experienced authority of a moral reason for some doing or refraining is Janus-faced: it has both normative elements and motivational elements that combine to give the experience the rich phenomenological character it has. So, although *Motivation* has mainly to do with the motivational dimension of the experienced authority of moral reasons, the remaining features either have to do with the normative dimension (*Independence, External source, Inescapable necessity*), or they feature a fusion of the motivational and the normative (*Pull, Grip, Compelling Rightness*). With these phenomenological observations in view, we turn next to a particular challenge facing any metaethical theory, based on these observations.

3. The challenge: a looming trilemma

The challenge associated with the rich phenomenal character of the experienced authority of moral reasons, as we have described such experience, can be explained as figuring in a seeming trilemma for metaethical views that involves three metaethical desiderata. *First*, it counts in favor of a metaethical theory that it be able to acknowledge the phenomenology of moral experience, including experiencing the authority of moral reasons as they bear on an individual's choice and action. *Second*, it counts in favor of a metaethical view is that it avoids problematic metaphysical and associated epistemological commitments. *Third*, in light of the fact that moral experience, thought, and discourse seem humanly unavoidable, it counts in favor of a metaethical view that its commitments are not in tension with preserving morality. So, ideally, one wants a metaethical theory that meets these desiderata:

Acknowledgement: acknowledges the Janus-faced phenomenology of moral-reasons experience,

Metaphysics: avoids troublesome metaphysical views, and

Preservation: is not in tension with the preservation of the practices of morality.¹¹

The clearest way to satisfy the first and third desiderata simultaneously – and arguably the only adequate way – is to *accommodate* the phenomenology, in this sense: (i) acknowledge its various aspects as summarized at the close of

Section 2 above, and (ii) treat those features as not ubiquitously misrepresenting the world – rather than treating them as ubiquitously non-veridical.

The looming trilemma is that these three desiderata seem hard to satisfy simultaneously. And insofar as a metaethical view fails to satisfy one or another desideratum it loses what David Enoch (2011, 14) aptly refers to as ‘plausibility points.’ We now want to press this looming trilemma by briefly considering a handful of familiar metaethical views, namely: two types of reductive view, reductive ethical naturalism and reductive expressivism, and two types of non-reductive view, ethical non-naturalism and moral error theory. We refer to the trilemma as ‘looming’ because we acknowledge that creative advocates of these views might think that they can find ways to avoid it. Nevertheless, the challenge to do so remains. We begin with reductive versions of ethical naturalism and ethical expressivism.

According to reductive ethical naturalism – a version of moral realism – there are metaphysically robust moral properties (including relations) that are a species of some type of metaphysically robust, objective (i.e. ‘stance-independent’) natural properties and facts.¹² This sort of view is partly motivated by wanting to avoid troublesome metaphysical commitments, and seems compatible with preserving morality. However, from what we can tell, this view succumbs to the first horn of the looming trilemma: it ultimately fails to adequately accommodate both aspects of the Janus-faced moral phenomenology we described in the previous section. In her book, Jean Hampton remarks in a number of places that all such versions of ethical naturalism ‘leave out the guts of morality’ (1998, 47, 120 n.9), by which she means the kind of inescapable necessity that she thinks is an aspect of much moral experience, including experiencing the authority of moral reasons. The problem is that such views end up not acknowledging that moral reasons (and the obligations they ground) have the kind of inescapable necessity that people experience them as having. That is, Hampton claims (and we agree) that versions of ethical naturalism end up treating the authority of moral reasons as a contingent matter, and so fail to acknowledge the phenomenological character of moral-reasons experiences. (See the quotation in Section 2 about Boyd-style ethical naturalism.) Moreover, such views typically embrace some form of motivational externalism, thus denying the phenomenological aspect of reasons-experiences that seems inherently motivational. The result is that reductive ethical naturalism has trouble with *Acknowledgement*, at least with respect to the phenomenology of the authority of moral reasons.

Another type of reductive view includes certain versions of expressivism (including old-time non-cognitivism). Both advocates and opponents of expressivism often characterize moral experiences and moral judgments by reference to states of mind which quite clearly do not purport to represent metaphysically robust, instantiated, properties or relations or metaphysically robust facts. Sometimes the reference to such states of mind is put forth merely as an instructive analogy; sometimes the suggestion is that although prototypical such states

of mind are not moral experiences, nevertheless the pertinent mental category includes moral experiences or moral judgments as a sub-species; and sometimes it is not made fully clear whether the first or the second construal is being suggested. At one time or another, expressivists have compared moral experiences and moral judgments to prototypical non-moral mental states of the following kinds, among others: desires, commands, universalized commands, attitudes of approval or disapproval, states of norm-acceptance, states of planning what to do. (Such comparisons are *strongly* reductive insofar as they treat moral experiences and moral judgments as a species of one or another of these kinds of mental state; and they are least *weakly* reductive insofar as they treat moral experiences as *not pertinently different* from prototypical instances of some such mental state.¹³) Like reductive ethical naturalism, expressivist views advertise that they satisfy *Metaphysics* and *Preservation*.

However, when it comes to meeting *Acknowledgement*, no such comparison – and no combination of such comparisons – seems theoretically satisfying. The problem is this: for each such comparison-category, prototypical instances of that category are mental states that simply do not seem to have the phenomenological aspect of inescapable authoritativeness. Ordinary desires don't seem to have it, because categorical authoritativeness is experienced as being independent of one's pre-existing desires. The mental states expressible as ordinary commands don't seem to have it, because their phenomenology is not as-of a state of mind expressible by uttering a command, but rather (a) as-of being 'commanded' oneself, and (b) as of this command's having 'compelling rightness.' Ordinary states of norm-acceptance or action-planning don't seem to have it, because typically they are experienced either as straightforwardly voluntary (as in voluntarily playing a game and thereby subjecting oneself to its rules), or at any rate as states one is in by virtue of one's contingent social circumstances (as in the case of the fictional character Ivan Denisovich [Solzhenitsyn 1962], who accepts the norms of bricklaying upon having been sentenced, in Stalin's post-war Soviet Union, to 25 years of bricklaying in Siberia).¹⁴

So, reductive versions of both ethical naturalism and ethical expressivism arguably fail to acknowledge one or both aspects of the Janus-faced experience of being gripped by the authority of moral reasons.

We turn now (and again, very briefly) to non-reductive metaethical views we have identified: ethical non-naturalism and error theory. Both views take seriously the irreducible character of concrete moral experience, and so they eschew any attempt to reduce putative moral properties and facts to properties and facts of some other kind, whether natural or supernatural. The difference, of course, is that the non-naturalists (of the sort we are concerned with)¹⁵ attempt to fully accommodate the phenomenology of the authority of reasons, whereas error theorists do not. Let us take a closer look at both types of theory.

Consider, first, non-naturalism. In light of the phenomenological character of moral-reasons experience, a non-naturalist like Hampton will hold that there

are metaphysically robust properties and facts that are inherently normative – they provide categorically authoritative normative reasons for choice and action that do not depend on one's desires, valuings, or intentions, and do not depend on human conventions – which nevertheless are part of the fabric of the world, to which we have access. Of course, the non-naturalist who wants to fully accommodate moral-authority phenomenology must also claim that the non-normative properties and facts that possess categorically authoritative normativity also are intrinsically motivating; otherwise she won't fully accommodate the motivational dimension of such experiences. This makes the view look doubly queer, which is certainly a theoretical cost for the theory (and explains why Mackie's queerness objection combines the normative with the motivational dimensions of moral experience). In short, although non-naturalist realism thoroughly accommodates the phenomenology of moral-reasons experience, thereby simultaneously satisfying both *Acknowledge* and *Preservation*, it fares quite badly with respect to *Metaphysics*.

According to moral error theory, the phenomenology of ordinary moral experience purports to represent as instantiated the sorts of metaphysically robust properties and facts posited by non-naturalism; likewise for ordinary moral thought and moral discourse. In view of this presumed ontological commitment, moral error theorists are ontological skeptics, charging ordinary moral experience, moral thought, and moral discourse with systematic error. Focusing just on moral-authority phenomenology, the error theory characterizes itself as not accommodating (in the sense stipulated above) this phenomenology. This poses a serious *prima facie* problem with respect to satisfying *Preservation*. Of course, an error theorist might nevertheless aim to preserve morality – as did Mackie, and as do some contemporary error theorists like Jonas Olson (2014). But one theoretical cost for such a view is its apparent 'bad faith' problem – viz., (on one hand) recommending that people continue to use moral concepts, to make moral judgments, and to take moral experience seriously, all the while (on the other hand) believing that it is all deeply error ridden. In other words, the tension referred to in *Preservation* remains, because error theory only *acknowledges* the phenomenology of moral-reasons experience but explicitly disavows *accommodating* this phenomenology.

The upshot of the preceding discussion is that the looming trilemma arising in light of the phenomenology of moral authority causes serious problems for each of the standard metaethical positions. Reductive moral realism apparently fails to satisfy the desideratum *Acknowledgement*, as do standard, reductive, versions of expressivism. Nonreductive moral realism apparently fails to satisfy the desideratum *Metaphysics*. And moral error theory apparently fails to satisfy the desideratum *Preservation*.

4. The neutrality thesis

In our view, each of the positions in the familiar menu of theoretical options in metaethics really does fail to satisfy at least one of those three desiderata. The way forward, we maintain, is to embrace a non-standard version of expressivism – a *non-reductive* version that not only acknowledges the phenomenology of moral authority but also (unlike error theory) claims to accommodate it too.

The possibility of such a position in the space of theoretical options in metaethics has thus far been largely obscured from view because of a widespread presupposition among those who have acknowledged the phenomenology of moral authority, viz., non-reductive moral realists like Hampton and error theorists like Mackie and Olson. Both non-naturalism and error theory take seriously the irreducible character of concrete moral experience, and so they eschew any attempt to reduce putative moral properties and facts to properties and facts of some other kind, whether natural or supernatural. (The difference, of course, is that the non-naturalists [of the sort we are concerned with] attempt to fully accommodate the phenomenology of the authority of reasons, whereas error theorists do not.) The presupposition can be formulated this way, as an entailment claim:

- (1) *Experiencing the authority of moral reasons*: Ordinary moral experiences of moral reasons have an inherent aspect of authoritative compelling rightness.

entails:

- (2) *Ontological purport*: This aspect of authoritative compelling rightness of moral reasons purports to represent a metaphysically robust, objective, relation of *being a reason for* as instantiated in the world.

Two features of the allegedly entailed proposition (2) are crucially important. First (as emphasized earlier), the modifier ‘metaphysically robust’ is intended to signal that the notion of ‘a relation instantiated in the world’ is to be understood *non-minimalistically*, as carrying ontological commitment to such a putative relation.¹⁶ Second, the label ‘objective’ is intended to entail that this putative relation is not a psychological one – and in particular, is not the in-the-world psychological relation *being regarded as a reason*, which obtains between a non-normative consideration and an experiencing agent.

Both non-naturalists and error theorists accept this (purported) entailment by simply not noticing that (1) and (2) make distinct claims, and by thus supposing that one can determine on the basis of introspection alone that (2) is true. We ourselves maintain, however, that (2) is not something one can reliably determine to be true on the sole basis of introspection; nor is (2) a claim that provides the only viable potential explanation of those elements of the phenomenology of being gripped by moral reasons that themselves are reliably discernable by

direct introspection. So in our view, the inference from (1) to (2) is fallacious; we call it *the non-naturalist fallacy* (see also Horgan and Timmons *forthcoming*).

By calling it a fallacy we are not claiming that the pertinent phenomenology does not have ontological purport. This is because we are not claiming that one can determine whether or not (2) is true on the sole basis of introspection. That one cannot tell either way is what we are calling our *neutrality* thesis regarding introspection directed toward one's experiencing the authority of moral reasons.¹⁷

Nor will we be claiming that our expressivist treatment qualifies, abductively via 'inference to the best explanation,' as the clearly best hypothesis about the nature of the inescapable-authoritativeness phenomenology in experiencing moral reasons. We grant, therefore, that non-naturalist moral realism and error theory provide alternative *prima facie* viable, potential explanations of the pertinent phenomenology. Insofar as one restricts oneself to phenomenological considerations – either direct deliverances of introspection, or abductive considerations concerning the potential explanation of introspective phenomenological data – we contend only that the expressivist approach we will describe below constitutes *one* *prima facie* viable way to theoretically embrace and explain the categorical-authoritativeness phenomenology that figures in moral deliberation, alongside of the non-naturalist way (and the error-theoretic way). This will suffice to undermine the move from (1) to (2) on the basis of an appeal to phenomenology, leaving our version of expressivism in the running as a viable option for coming to terms with the phenomenology in question. Wider, largely non-phenomenological, considerations – including how the competing theories fare with respect to the three desiderata *Acknowledgement*, *Metaphysics*, and *Preservation* – thereafter can be brought to bear in doing comparative cost-benefit evaluation of non-naturalist moral realism, error theory, and our non-reductive version of expressivism.¹⁸

We emphasize that we are not presupposing the neutrality thesis at the outset of the discussion below. Rather – and granting as we do, that non-naturalism (and moral error theory) can also acknowledge the pertinent phenomenology – the case for neutrality will emerge as the discussion unfolds.

5. Nonreductive expressivism

Expressivists focus their metaethical theorizing, first and foremost, on the pertinent *states of mind* associated with matters ethical: moral experiences and moral judgments. Moral assertions are then treated as *expressing* moral judgments but not as *describing them*. A key tenet of any version of expressivism is that moral experiences and moral judgments do not purport to attribute metaphysically robust, instantiated, moral properties or relations, or to describe metaphysically robust moral facts. (This does not preclude the use of property-talk or

relation-talk or fact-talk or truth-talk in moral discourse, but it does construe such talk as operating minimalistically rather than carrying ontological purport.)

Typically when one forms a moral judgment, one experiences certain considerations as *reasons* for that judgment – and indeed, as categorically authoritative reasons. Such reasons-experiences should be construed by expressivists in the same kind of way that moral judgments themselves are construed – viz., as states of mind that do not purport to attribute metaphysically robust relations (say, *being fitting in light of*, *being a reason for*, or *being required by*, or the like). This point was emphasized by Charles Stevenson (1961), and we ourselves heartily concur.

So the task we face has three intertwined aspects. First is the need to articulate, at least in broad outline, a version of expressivism that differs from standard versions in being clearly non-reductive. Second is the need to provide, within this wider expressivist theoretical framework, an intelligible and plausible expressivist construal of moral reasons-experiences themselves, over and above the moral judgments to which they give rise. Third is the need to do so in a way that also accommodates the phenomenological aspect of categorically authoritative moral normativity that non-naturalists like Hampton so rightly emphasize is central to moral experience. We address the first item in the remainder of the present section, and the second and third items in Section 6.

In a number of our past writings we have articulated and defended a non-reductive and metaphysically irrealist metaethical position that we call ‘cognitivist expressivism.’ It is so labeled because on our view, moral judgments are a species of genuine *belief* – although not a species of belief that purports to represent metaphysically robust moral properties or moral facts. For present purposes, however, what matters is not our claim that moral judgments are beliefs – a claim that one might question even if one accepts our position otherwise – but rather our contention that moral judgments are, in important respects, states of mind that cannot be reductively analyzed (Although we do claim that they are beliefs, we also claim that they are a distinctive and irreducible *species* of belief, significantly different from the states of mind ordinarily classified as beliefs.) We will now briefly characterize the position, in a way that emphasizes the irreducible character of moral judgments but leaves aside the question of whether or not they are a species of belief.

An ordinary belief is a certain kind of psychological commitment state with respect to a potential way the world might be – viz., a commitment to the world’s actually *being* that way. We therefore call ordinary beliefs *is-commitments*. According to our non-reductive metaethical expressivism, a moral judgment too is a psychological commitment-state with respect to a potential way the world might be – an *ought-commitment*, as we call it. On this picture, to believe that Bertie will apologize to Madeline is to be *is-committed* with respect to the (non-moral, descriptive) way-the-world-might-be *that Bertie apologizes to Madeline*. Similarly, to judge morally that Bertie ought to apologize to Madeline

is to be ought-committed with respect to that same way-the-world-might-be, viz., *that Bertie apologizes to Madeline*. (As metaethical irrealists, we maintain that there is no such way-the-world-might-be as the putative belief-content *that Bertie ought to apologize to Madeline*. Likewise, there is no such metaphysically robust relational fact as the putative fact *Bertie's having failed to keep his promise being an objective moral reason for Bertie to apologize to Madeline*; and there is no such metaphysically robust relation as *moral fittingness* between Bertie's apologizing and his having failed to keep his promise.)

Further delineating ought-commitments as distinctive, irreducible, psychological states is not a matter of trying to analogize them to, or to subsume them under, states such as desires, commands, universalized commands, plans, norm-acceptances, or the like.¹⁹ This reductive approach, as already emphasized, looks incapable of accommodating the authoritativeness aspect of moral phenomenology. How then should one go about giving an illuminating expressivist characterization of these mental states? Well, largely by 'triangulating' them *vis-à-vis* other kinds of mental states, both with respect to phenomenology and with respect to functional roles in thought and action-guidance: for instance, underscoring their phenomenological and functional-role similarities to ordinary non-descriptive beliefs, while also emphasizing important phenomenological and functional-role differences too; underscoring their phenomenological and functional-role similarities to ordinary desires, while again also emphasizing important differences too; etc. This we have done at some length in prior writings.

We will not rehearse here our prior discussions of such matters. Instead we will proceed directly to the task at hand: extending our non-reductive expressivism to incorporate moral-reasons experiences, and arguing that the resulting position smoothly accommodates the phenomenology of moral authority.

6. Addressing the challenge

We are now prepared to defend our neutrality thesis, making our case for claiming that careful attention to moral-authority phenomenology does not reveal whether or not such experiences carry, as an aspect of their content, commitment to metaphysically robust, non-natural, properties, relations, and facts. Focusing largely on the Clive example from Section 1, we aim to make clear why moral-reasons experiences can possess all the elements of moral-authority phenomenology that Hampton (together with Mackie) correctly attributes to them (and more besides) without purporting to represent metaphysically robust, categorically authoritative, fittingness-relations or fittingness-facts. The methodology will be *the phenomenological method of similarity and contrast* (as we will call it) – viz., comparing moral-reasons experiences to various other kinds of mental states, noting both important phenomenological similarities and important phenomenological differences.

Consider the onset of Clive's non-normative belief that the woman is in serious danger of being harmed. He experiences what he sees happening – including, in particular, the woman's screaming while the man drags her by the ankles – as pulling him toward believing this, i.e., pulling toward an is-commitment *vis-à-vis* that content (This is the aspect of *Pull*.) He experiences this as an 'external' pull, emanating from what he sees happening outside of himself. (This is the aspect of *External Source*.) He finds himself involuntarily gripped, in virtue of the strength of this pull, by an is-commitment *vis-à-vis* the likelihood of her being harmed. (This is the aspect of *Grip*.) This is-commitment arises independently of any pre-existing desires he has, and independently of human conventions – indeed, in this case it occurs despite his pre-existing desire *not* to believe that she is in serious danger. (This is the aspect of *Independence*.) Since his becoming thus is-committed is both involuntary and independent of his pre-existing desires, it is experienced as inescapable, given his current evidential situation. (This is the aspect of *Inescapability*.)

Consider now the onset of Clive's judgment that he is morally obligated to intervene on the woman's behalf – with this judgment being construed as an ought-commitment *vis-à-vis* the non-normative content *that I intervene on the woman's behalf*. This experience is phenomenologically similar, in each of the ways lately noted, to the onset of an is-commitment. He experiences her being in serious danger of being harmed (something about which he now has an is-commitment) as pulling him toward an ought-commitment with respect to his intervening. (This is the aspect of *Pull*.) He experiences this as an 'external' pull, since her being in danger of harm is something outside of himself. (This is the aspect of *External Source*.) He finds himself involuntarily gripped, in virtue of the strength of this pull, by an ought-commitment *vis-à-vis* intervening on her behalf. (This is the aspect of *Grip*.)²⁰ The onset of this ought-commitment occurs independently of any pre-existing desires he has, and independently of human conventions – indeed, in this case it occurs despite his pre-existing desire *not* to intervene. (This is the aspect of *Independence*.) Since his becoming thus ought-committed is both involuntary and independent of his pre-existing desires, it is experienced as *inescapable*, given his current evidential situation. (This is the aspect of *Inescapability*.) Finally, he experiences the contemplated action of intervening as compellingly right – something he must do. (This is the aspect of *Compelling rightness*.)

The onset of Clive's is-commitment to the woman's being in danger of harm and the onset of his ought-commitment to intervening on her behalf are thus similar to one another in all the ways lately noted. (These similarities, we contend, are strong enough and broad enough to render the ought-commitment a species of the genus *belief* alongside the is-commitment – although our argumentation in this paper does not require agreement on this point.) In each case, the non-normative consideration experienced as *reason* for the commitment exhibits the same set of elements that Hampton identifies as figuring in

the phenomenology of categorical authoritativeness – viz., *Pull*, *Source*, *Grip*, *Independence*, and *Inescapability*. And this is so even if – as we ourselves maintain – Clive’s experiencing the woman’s danger of harm as reason for his intervening does not purport to represent a metaphysically robust relation of ‘fittingness.’

On the other hand, the onset of Clive’s ought-commitment also differs in an important respect from the onset of his is-commitment regarding the woman’s being in danger of harm – viz., the ought-commitment has a motivationally ‘hot’ role within Clive’s psychology, all by itself.²¹ The is-commitment, by contrast – like other is-commitments *vis-à-vis* non-normative contents – plays no motivational role by itself apart from other pertinent psychological states with which it might combine, such as pre-existing desires. Thus, one’s reason-experience in the case of the ought-commitment also has the final phenomenological element on the list at the end of Section 3, viz., *Motivation*. (This element can be present and operative without being motivationally dominant – as in Clive’s case, since he does not act in accordance with his experienced moral obligation.)

The phenomenological element *Motivation* is also present, of course, in prototypical experiences of desire. In that respect at least, moral experiences are similar to ordinary desires – a point often emphasized by metaethical expressivists. Yet the differences from desire are palpable too. For one thing, desires are not commitment-states, whereas moral judgments are. And although a desire often is experienced as a reason for a specific action (given the belief that the action will lead to desire-satisfying consequences) – and thus exhibits *Pull* toward the action – the source of this pull is the desire itself, and so is not experientially external; thus, *Source* is not present. Also, acting on the basis of a desire, or forming an intention to act on the basis of a desire, typically is a voluntary matter, and thus does not exhibit *Inescapability*. And since desires themselves are the sources of the actions or intentions to which they sometimes give rise as reasons, such desire-based reasons do not exhibit *Independence*.

Yet further elements of the categorically-authoritative-fittingness aspect of moral phenomenology, beyond those explicitly mentioned by Hampton, can be brought into view by considering self-directed reactive attitudes that typically arise when one fails to act in a way that one experiences as morally obligatory – and by comparing these with self-directed reactive attitudes that typically arise when one violates an ongoing intention of a non-moral kind, such as an intention to stick to one’s diet until one has lost ten pounds. In both kinds of case, one is apt to experience a sentiment of guilt or of shame, which in turn is apt to motivate one to take compensatory remedial action as best one can. And in both cases, such a sentiment is apt to arise because the pertinent state is experienced as exerting a governing authority over oneself – an authority that one has contravened. But there is a crucial difference. Although a voluntarily formed intention, such as the intention to follow a specific diet because of one’s desire to lose ten pounds, is apt to exert an experienced authority over oneself that will induce guilt and/or shame in circumstances where one has violated

that intention, this authority and resultant sentiment(s) are experienced as *contingent and desire-dependent*: the authority one has contravened is experienced as operative upon oneself only in virtue of one's voluntarily-formed intention, and thereby only in virtue of the pre-existing desire that motivated that intention in the first place. By contrast, the phenomenological authority of moral reasons, and the resultant self-directed reactive sentiment(s), are experienced as *inescapable and desire-independent*: the authority one has contravened is experienced as operative upon oneself independently of any medium-term or long-term intentions one might have voluntarily formed, and independently of one's pre-existing desires.

We submit that everything we have said in the present sub-section about the inherent aspect of the-authority-of-moral-reasons phenomenology is compatible with the contention that moral experience does not purport to represent a metaphysically robust, instantiated, moral-reasons relation.²² We are claiming no more than that, because as we emphasized earlier, the pertinent phenomenology also is compatible with the contention that moral experience *does* purport to represent metaphysically robust, instantiated, moral-reasons relations. (We claim that direct introspection cannot reliably settle this issue either way, and that the phenomenology can be acknowledged by each of three competing metaethical positions: Hampton-style non-naturalist realism, Olson-style error theory, and our own non-reductive version of expressivism.) But the fact that our own expressivism can acknowledge and accommodate the phenomenology of moral authority is all we need, for our dialectical purposes.

Note finally, that if we are correct in arguing that cognitivist expressivism satisfies *Accommodation* (at least as far as the phenomenological character of moral-authority experience is concerned), then since it also satisfies *Metaphysics* and *Preservation*, it piles up serious plausibility points.

7. Conclusion

Careful examination of the rich phenomenology of experiencing moral reasons reveals that such experiences include an inherent aspect of categorical authority that has both motivational and normative elements. Accommodating this phenomenology sets a challenge for both standard reductive and non-reductive metaethical theories, which makes trouble for such theories. Reductive naturalism and reductive expressivism have trouble accommodating the pertinent phenomenology. Non-naturalism and error theory – both of which are non-reductive in what they attribute to ordinary moral experience, thought and discourse – nevertheless commit what we call the 'non-naturalistic' fallacy insofar as they take for granted that one can reliably determine on the basis of introspection alone that the phenomenology in question has non-naturalist, moral realist, ontological purport. That one cannot reliably determine this matter about ontological purport on the sole basis of introspection is our neutrality

thesis. Our preferred non-reductive version of expressivism – cognitivist expressivism – can smoothly and without distortion accommodate the aspects of the phenomenology of moral-authority experience, as illustrated by our story of Clive.

We also advocate a wider claim: our non-reductive version of expressivism can accommodate moral authority as a characteristic of a broad range of moral experiences (including, but not restricted to, experiencing the authority of moral reasons) that can be reliably detected on the sole basis of introspection. But a full defense of this claim requires careful examination of other species of moral experience, which we have begun to do in some of our other work (Horgan and Timmons 2008, 2015, 2017, 2018, *forthcoming*).²⁴

Notes

1. We address embedding in Horgan and Timmons (2006); negation in Horgan and Timmons 2009; and the possibility of deep moral error in Horgan and Timmons 2015.
2. We are using ‘moral reasons’ in a broad sense to refer not only to reasons whose characterization involves moral terms (e.g. that such and so action is wrong), but also to non-normative reasons (e.g. that such and so action would cause much harm) of the sort that purport to explain why actions, attitudes, and other objects of moral evaluation have the moral status they have, and which one appeals to in supporting a moral judgment.
3. Horgan and Timmons 2018 (*forthcoming*).
4. We understand phenomenology to be a largely descriptive field of study whose methodology is introspection and whose subject matter is the concrete ‘what-it-is-likeness’ of experience. So, we do *not* include (as some do) within the scope of phenomenological inquiry all of the deeply embedded aspects of ordinary moral thought and discourse.
5. Hampton’s discussion is actually broader in scope than morality; her description also is meant to capture the authoritative grip of epistemic reasons.
6. As we use the germ ‘grip’ here and throughout, being ‘gripped’ by the experienced authority of a certain consideration *C*, as a reason for (or against) performing act Φ , constitutes being in a state of *judging*, on the basis of *C*, that one ought to (or ought not to) perform Φ . When one is thus gripped, one experiences the reason as moral-normatively decisive; but it need not be also motivationally decisive (and in Clive’s case, is not). The judgment need not be consciously explicit; instead, its content might be implicit in the specific phenomenological character of one’s current experience – in much the same way that appreciation of pertinent background information often is implicit in the specific phenomenological character of the experience of understanding a culturally topical joke, even though that information is not being explicitly consciously rehearsed. ‘Chromatic illumination’ is our expression for such implicit conscious appreciation of content that is not being explicitly represented in consciousness. In Horgan and Timmons (*forthcoming*) we discuss chromatic illumination at length, with specific attention to its operation in moral experience. The notion of chromatic illumination was originally introduced, in connection with the contention that the justification-

status of a belief often depends in part upon implicit conscious appreciation of pertinent background evidence, in Horgan and Potrč (2010).

7. Moral-reasons experiences need not include an explicit *judgment* that such-and-such considerations constitute moral reasons for thus-and-such action. Often enough, it seems, one experiences certain considerations as moral reasons for a certain action Φ – one experiences those considerations ‘moral-reasonably *vis-à-vis* Φ ’, so to speak – without forming such an explicit conscious judgment whose content is that considerations C constitute moral reasons for action Φ . Indeed, much might be operative in consciousness only implicitly, by way of chromatic illumination (cf. note 6): perhaps certain considerations that are figuring as authoritative moral reasons, perhaps one’s appreciation of those considerations *as* authoritative moral reasons, and perhaps even one’s being *gripped* by that appreciated authority. The phenomenology to be described below comports with these observations.
8. Hampton does provide a conception of norms on pp. 49–53 of her book.
9. The irreducibility of such experience is nicely put by David Enoch when he observes that with regard to normative facts generally, ‘Normative facts are just too different from natural ones to be a subset thereof’ (2011, 4), moral reasons being a special case.
10. One can easily add to this list. In our Forthcoming, we discuss the phenomenology of moral deliberation which, as Nagel (1986, 49) rightly observes, often includes the thought that whatever decision one comes to on the basis of such deliberation, *one might still be mistaken*.
11. We acknowledge that an advocate of metaethical error theory might challenge the contention that *Preservation* is a legitimate theoretical desideratum – perhaps by claiming that the goal of preserving the practice of morality is the ‘wrong kind of reason’ for accepting a metaethical theory. But at the very least, this goal constitutes a legitimate and important reason to *seek* a credible metaethical theory that preserves morality while also satisfying the desiderata *Acknowledgement* and *Metaphysics*. And of course if such a theory can be found, then its availability will undermine the principal motivation for error theory – *viz.*, the contention that moral-authority experience purports to represent putative ‘do-be-done-ness’ facts and properties that are metaphysically queer.
12. Here and throughout, we use the modifier ‘metaphysically robust’ to signal that the contextually operative use of the terms ‘property’ and ‘fact’ is not a minimalistic use. Although we recognize that these terms are sometimes used minimalistically – in which case, for instance, asserting ‘Abortion is morally wrong’ is essentially equivalent to asserting ‘That abortion is morally wrong is a fact’, or to asserting ‘Abortion has the property of moral wrongness’ – we deny that such minimalistic uses are the *only* ones that are ever contextually operative. We also deny that they have any privileged ‘default’ status.
13. We say ‘not pertinently different from’, rather than ‘pertinently similar to’, because our own non-reductive version of expressivism, as described in Sections 5 and 6 below, does emphasize certain pertinent similarities between moral-authority experiences and various kinds of mental state not involving morality. Even so, our brand of expressivism is neither strongly reductive nor weakly reductive, because it also emphasizes pertinent *differences* between moral-authority experiences and each of those other kinds of mental state.
14. Experiencing a state of mind like norm-acceptance or action-planning as voluntary, or as contingent upon one’s social circumstances, need not be a matter of an explicit higher-order judgment that attributes voluntariness or contingency

to one's first-order state of norm-acceptance or action-planning. Instead, it can be a matter of how the first-order mental state is chromatically illuminated; cf. notes 6 and 7 above.

15. Recently Cuneo and Shafer-Landau (2014); have proposed what they call 'conceptual non-naturalism' one aim of which is to avoid troublesome metaphysical and epistemological difficulties that arise for robustly metaphysical versions of the sort we are considering. For a trenchant critique of this view, see Copp 2018a. See also Copp (2018b), for a critique of the metaethical views of Parfit and Scanlon, whose versions of non-naturalism are also supposed to avoid these same troublesome commitments.
16. Words like 'relation', 'property', 'fact', and 'true' all have minimalistic, disquotational uses in ordinary discourse that need not be ontologically committal – a point rightly emphasized, for instance, by metaethical expressivists like Simon Blackburn who embrace 'quasi-realism' about ordinary moral thought and discourse. But it is a serious mistake, we maintain, to infer from this that the *only* actual or legitimate uses of such words are minimalistic. That mistake is a dangerous first step down a looming garden path to metaethical quietism (cf. Horgan and Timmons 2015) (As we say in that paper, Beware of becoming beHorwiched!).
17. Note that ironically, if we are right, then when it comes to the introspectible aspects of experiencing the grip of moral reasons, the moral error theory remains unmotivated. It must therefore rely on appealing to other features of moral thought and discourse in order to claim that such thought and discourse is error-ridden.
18. We acknowledge that a hybrid theory that combines aspects of naturalist moral realism with aspects of expressivism might be dialectically in the running too, insofar as it embraces a construal of moral-authority phenomenology like the one we propose below. But an advocate of such a position would bear the dialectical burden of providing an adequate theoretical motivation for the two claims (1) that moral experience and moral judgment purport to represent metaphysically robust moral properties and moral facts, and (2) that these are identical to certain *natural* properties and facts whose essence is non-normative (Discussion with David Copp during the 'Representation and Evaluation' conference prompted the present note.)
19. Nor, of course, is it a matter of construing an ought-commitment as a higher-order belief that attributes a first-order state to oneself, e.g. the psychological state *regarding non-normative consideration C as a moral reason for action A*.
20. The following objection arises, which was pressed upon us at the 'Representation and Evaluation' conference by our commentator Bruno Guindon and also by several audience members including Paul Bloomfield and Bill FitzPatrick:

Although cognitivist expressivism recognizes and accommodates the externality of the non-normative considerations that one experiences as authoritative reasons, it does not recognize and accommodate the externality of the experienced status of those considerations *as authoritative reasons*; and in this respect, cognitivist expressivism is really a version of metaethical error theory.

Our response is to urge that introspection alone cannot reliably ascertain whether or not the experiential externality of moral reasons, *qua* authoritative moral reasons, involves anything more than the further phenomenological features that we describe just below – in particular, involuntariness and independence of pre-existing desires. Our own view is that the phenomenology

as we describe it already *constitutes* experiencing the pertinent non-moral considerations as morally authoritative.

21. This motivational role might be direct, or might be a matter of generating a new desire with the same content as the ought-commitment. We ourselves find the former possibility more plausible, phenomenologically and psychologically.
22. Concerning the phenomenological aspect *External Source*, the following remarks bear emphasis. Hampton's own formulation, as expressed in the pertinent bullet point at the end of Section 2, is ambiguous. On one hand, it can be construed as including both (i) that the non-normative factual consideration which one experiences as a reason is external to oneself, and (ii) that this fact is experienced as authoritative over oneself in a categorical way, independently of one's pre-existing desires, one's contingent social roles, etc. On the other hand, it can be construed as including not only features (i) and (ii) but also this feature: representing a putative, external, independent, metaphysically robust, *moral-reasons relation*. We, of course, are construing *External Source* the former way in claiming that our non-reductive expressivism is compatible with what is *reliably introspectible* about this aspect of moral phenomenology. To contend that introspection reliably reveals that moral phenomenology satisfies the stronger construal of *External Source*, we contend, is to commit the non-naturalist fallacy.
24. For helpful comments and discussion, we thank the participants at the 2017 conference 'Representation and Evaluation' at the University of British Columbia, especially Bruno Guidion (our commentator), Paul Bloomfield, David Copp, Bill FitzPatrick, and the two conference organizers Matt Bedke and Stefan Sciarraffa.

Notes on contributors

Terry Horgan is professor of philosophy at the University of Arizona. He publishes in various areas of philosophy, including philosophy of mind, epistemology, metaphysics, and metaethics. He is author of *Essays on Paradoxes* (Oxford University Press, 2017), and is co-author with John Tienson of *Connectionism and the Philosophy of Psychology* (MIT Press, 1996), co-author with Matjaž Potrč of *Austere Realism: Contextual Semantics Meets Minimal Ontology* (MIT Press, 2008), and co-author with David Henderson of *The Epistemological Spectrum: At the Interface of Cognitive Science and Conceptual Analysis* (Oxford University Press, 2011).

Mark Timmons is professor of philosophy at the University of Arizona, who publishes in the areas of metaethics and Kant's ethics. He is the author of *Morality without Foundations* (Oxford University Press, 1999) and *Moral Theory: An Introduction* (2012), editor of *Oxford Studies in Normative Ethics*, published annually, and co-editor of a newly inaugurated series *Oxford Guides*. Some of his work on Kant's ethics is collected together in *Significance and System: Essays on Kant's Ethics* (Oxford University Press, 2017). He and Terry Horgan are working on a book manuscript *Illuminating Reasons: An Essay in Moral Phenomenology*, under contract with Oxford University Press.

References

- Copp, D. 2018a. "Are There Substantive Moral Conceptual Truths?" In *Moral Skepticism: New Essays*, edited by D. E. Muchaca, 91–114. London and New York: Routledge.

- Copp, D. 2018b. "A Semantic Challenge to Non-Realist Cognitivism." *Canadian Journal of Philosophy* 48: 554–575.
- Cuneo, T., and R. Shafer-Landau. 2014. "The Moral Fixed Points." *Philosophical Studies* 171: 399–443.
- Enoch, D. 2011. *Taking Morality Seriously: A Defense of Robust Realism*. Oxford and New York: Oxford University Press.
- Hampton, J. 1998. *The Authority of Reason*. Cambridge: Cambridge University Press.
- Horgan, T., and M. Potrč. 2010. "The Epistemic Relevance of Morphological Content." *Acta Analytica* 25: 155–173.
- Horgan, T., and M. Timmons. 2006. "Cognitivist Expressivism." In *Metaethics after Moore*, edited by T. Horgan and M. Timmons, 255–298. Oxford and New York: Oxford University Press.
- Horgan, T., and M. Timmons. 2008. "What Can Moral Phenomenology Tell Us about Moral Objectivity?" *Social Philosophy & Policy* 25: 267–300.
- Horgan, T., and M. Timmons. 2009. "Expressivism and Contrary-Forming Negation." *Philosophical Issues* 19: 92–112.
- Horgan, T., and M. Timmons. 2015. "Modest Quasi-Realism and the Problem of Deep Moral Error." In *Passions and Projections: Themes from the Philosophy of Simon Blackburn*, edited by R. N. Johnson and M. Smith, 190–209. Oxford and New York: Oxford University Press.
- Horgan, T., and M. Timmons. 2017. "Sentimentalist Moral-Perceptual Experience and Realist Pretensions: A Phenomenological Inquiry." In *Ethical Sentimentalism*, edited by K. R. Stueber and R. Debes, 86–106. Cambridge: Cambridge University Press.
- Horgan, T., and M. Timmons. 2018. "The Phenomenology of Moral Authority." In *Moral Skepticism: New Essays*, edited by D. E. Muchaca, 115–140. London and New York: Routledge.
- Horgan, T., and M. Timmons. *Forthcoming*. "The Phenomenology of Deliberation and the Non-Naturalistic Fallacy." In *The Naturalistic Fallacy*, edited by N. Sinclair. Cambridge: Cambridge University Press.
- Horgan, T., and M. Timmons. *Forthcoming*. *Illuminating Reasons: An Essay in Moral Phenomenology*. Oxford: Oxford University Press.
- McEwan, I. 1998. *Amsterdam*. New York: Anchor Books.
- Mackie, J. L. 1977. *Ethics: Inventing Right and Wrong*. Harmondsworth: Penguin Books.
- Nagel, T. 1986. *The View from Nowhere*. Oxford and New York: Oxford University Press.
- Olson, J. 2014. *Moral Error Theory: History, Critique, Defense*. Oxford and New York: Oxford University Press.
- Solzhenitsyn, A. 1962. *One Day in the Life of Ivan Denisovich*. Rutherford, NJ: Signet Classics.
- Stevenson, C. L. 1961. "Relativism and Non-Relativism in the Theory of Value." *Proceedings and Addresses of the American Philosophical Association* 35: 25–44 (reprinted as Chapter 5 of his *Fact and Value*. New Haven: Yale University Press, 1963).