



# Moral motivation and the affective appeal

Jennifer Corns<sup>1</sup> · Robert Cowan<sup>1</sup>

Published online: 29 January 2020  
© The Author(s) 2020

**Abstract** Proponents of “the affective appeal” (e.g. Dancy in *Ethics* 124(4):787–812, 2014; Zagzebski in *Philos Phenomenol Res* 66(1):104–124, 2003) argue that we can make progress in the longstanding debate about the nature of moral motivation by appealing to the affective dimension of affective episodes such as emotions, which allegedly play either a causal or constitutive role in moral judgements. Specifically, they claim that appealing to affect vindicates a version of Motivational Internalism—roughly, the view that there is a necessary connection between moral judgment and motivation—that is both more empirically respectable and less theoretically controversial than non-affective versions. We here argue that the affective appeal fails: versions of Internalism which appeal to affect are neither more empirically supported, nor clearly less controversial, than versions of Internalism which make no such appeal. Although affect doubtless has an important role to play in explaining moral motivation, we are sceptical that establishing any such role advances the debate.

**Keywords** Moral motivation · Affect · Moral psychology · Emotions · Ethics

Our starting point is a widely recognised feature of moral discourse and thought:

*Practicality* moral judgments (understood as mental episodes)<sup>1</sup> are importantly connected with motivation to act.

---

<sup>1</sup> In contrast to speech acts.

---

✉ Robert Cowan  
robert.cowan@glasgow.ac.uk

Jennifer Corns  
jennifer.corns@glasgow.ac.uk

<sup>1</sup> Philosophy, School of Humanities, University of Glasgow, Glasgow G12 8QQ, UK

Consider an illustrative case. After considering the environmental impact of air travel, Anna judged that *air travel is morally wrong*. Previously a keen air traveller, she henceforth avoided flying. For example, when invited to international conferences she would either give her talk by Skype or travel overland. Occasionally, however, Anna's desire for far flung parts of the world would prevail and she would fly. But in these instances, she was conflicted. Indeed, guilt would often take the shine off her enjoyment.

Note three features of this case relevant to Practicality. First, adopting a new moral judgment is typically associated with alteration in the agent's motivation. Second, the relatively direct and explanatory way in which moral judgments seem connected with motivation contrasts with other judgments, e.g. about mathematics. Third, a moral judgment can impact an agent's motivational dispositions without resulting in action.<sup>2</sup>

Practicality is common ground among ethicists. Debate concerns how precisely to capture the connection between moral judgment and motivation. Consider:

*Unconditional Motivational Internalism* (U-Internalism) Necessarily, if S judges that she morally ought to  $\phi$ , then S is motivated (at least somewhat) to act in accordance with her judgment.<sup>3</sup>

U-Internalism's necessity claim is typically understood as either asserting *metaphysical* or *conceptual* necessity, i.e. as about moral judgments themselves, or as concerning our concept MORAL JUDGMENT. Our arguments target both, but we'll address worries that our discussion fails to connect with both claims as they arise. For either modal interpretation, U-Internalism is compatible with moral judgments being either extrinsically or intrinsically motivating, i.e. with a necessary connection between moral judgments and motivation, or with moral judgments themselves being motivational. Although concerning *moral ought* judgments U-Internalism can be—and often is—adapted to encompass other moral predicates, e.g., *cruelty*, *rightness*. Importantly, the intended sense in which S is motivated is defeasible but not cancellable, i.e., it's compatible with S not  $\phi$ -ing, but is incompatible with S lacking any motivation to  $\phi$ . Finally, S's being motivated to act *in accordance with her judgment* could be made true by a motivation to  $\phi$ , taking steps towards  $\phi$ -ing, or encouraging others similarly situated to  $\phi$ . The precise nature of the motivation, e.g. when, how, and in what ways one is motivated, awaits precisification depending on the judgment at issue, e.g. whether it concerns obligation.

U-Internalism appears to have rich explanatory resources vis-à-vis Practicality. It explains, e.g., why Anna's motivation changes when her moral judgment does. Had Anna not been motivated to act in accordance with her judgment, then this would entail that the judgment wasn't genuine.

However, U-Internalism faces significant objections. For example, morality seems cognitive, such that moral judgments appear to be belief-like states.

<sup>2</sup> For overviews see "Introduction" in Björnsson et al. (2015) and Rosati (2016).

<sup>3</sup> E.g. Eggers (2015) and Prinz (2015), both in Björnsson et al. (2015) and Lenman (1999).

Conjoining U-Internalism with Cognitivism entails that there are necessary connections between belief-like states and motivation. But this apparently conflicts with the popular Humean Theory of Motivation, which denies such connections.<sup>4</sup> There are also alleged counterexamples: e.g. agents who (apparently) sincerely judge that *it is wrong to kill humans*, but are nevertheless unmotivated to act in accordance with this, seem conceivable. If conceivability entails possibility then such beings—“amoralists”—are apparently possible. Indeed, psychopaths or depressives may constitute actual cases.

In light of these considerations, some endorse Motivational Externalism, whose core claim is the denial of U-Internalism.<sup>5</sup> On this view, the reliable connection between moral judgment and motivation is contingent. As a simple example, perhaps Practicality is explained by the fact that most people are raised to want to do as they morally ought.<sup>6</sup> It's just a contingent feature of typical human psychology that moral judgment and motivation are connected. For instance, if Anna had an unusual upbringing she could make a genuine moral judgment about the wrongness of air travel that failed to motivate her.

In response, some attracted to the spirit of U-Internalism modify it. One well-known version conditionalises Internalism to apply only to agents who are *practically rational*.<sup>7</sup> Arguably, this avoids problems allegedly besetting U-Internalism, e.g. perhaps amoralists are practically irrational. But against this, some<sup>8</sup> claim to conceive of agents who sincerely make moral judgments, fulfil the relevant condition, yet lack motivation, i.e. endorse Externalisms denying various conditionalised Internalisms.

While distinct thinkers are attracted by distinct strategies for advancing the debate,<sup>9</sup> it remains generally unresolved, and has the appearance of an impasse. Indeed, some suggest that Internalists and Externalists—along with non-specialists—have different MORAL JUDGMENT concepts.<sup>10</sup>

Recently, however, some have sought to advance the debate by appealing to affective episodes, i.e. mental episodes that are paradigmatically pleasant or unpleasant, such as emotions, and which are widely thought to be motivational by Cognitivists and Non-Cognitivists alike.<sup>11</sup> They are then claimed to play either a causal or constitutive role in moral judgments. Though sometimes only implicit, a common thought is that the appeal to affective episodes vindicates a version of

<sup>4</sup> This may be avoided by requiring that a belief be accompanied by motivation to be *classified* as a moral belief. See Tresan (2006, esp. pp. 147–148).

<sup>5</sup> E.g. Brink (1989).

<sup>6</sup> Or perhaps moral utterances conversationally implicate motivation because we're aware of morality's action-guiding function. E.g. Strandberg (2011, esp. section 5).

<sup>7</sup> E.g. Smith (1994, esp. p. 61).

<sup>8</sup> For discussion see van Roojen (2010, esp. p. 499).

<sup>9</sup> E.g. “Deferred Internalisms” which further attenuates the necessary connection between judgment and motivation. See, e.g. Dreier (1990, esp. pp. 9–14).

<sup>10</sup> E.g. Francen (2010, esp. p. 130).

<sup>11</sup> For examples of the former see footnotes 13–15. For the latter see works by C20th Emotivists like Charles Stevenson, and, more recently, Blackburn (1998, esp. pp. 17–20).

Internalism that is both more empirically respectable and less theoretically controversial than non-affective versions. Call this *the affective appeal*.

Consider some examples. Jesse Prinz<sup>12</sup> defends Internalism on the basis of what he calls *Sentimentalism*, i.e., the view that all genuine moral judgments are constituted by emotions. Prinz's view is partly based on empirical evidence which purportedly supports the claim that emotions are both necessary and sufficient for moral judgments. Crucially, he thinks that emotions are paradigms of motivational phenomena. Elsewhere, Antti Kauppinen<sup>13</sup> and Linda Zagzebski<sup>14</sup> argue that there is a plurality of kinds of moral thoughts, so-called *Moral Thought Pluralism*, and that among these are a class of moral judgments which are intimately connected to emotions. For instance, Zagzebski thinks that the most developmentally and explanatorily basic moral judgments deploy thick concepts that are emotional. Pluralists think their view provides a nuanced moral psychology which vindicates a version of Internalism (with respect to *affective* moral judgments) whilst doing justice to Externalism (some version may be true of at least some *non-affective* moral judgments). Again, the idea is that if we accept that some moral judgments are affective, then Internalist-friendly claims about motivation come for free. Finally, Moral Intuitionists like Jonathan Dancy<sup>15</sup> argue that some moral emotions involve the *affective presentation* of reasons for action, such that some version of Internalism holds.<sup>16</sup>

We here critically assess the affective appeal, i.e. the attempt to vindicate a more empirically respectable and less theoretically controversial Internalism by appealing to affective episodes. We specifically focus on appeal to the *affective aspect* of affective mental episodes as opposed e.g. to any sensory-discriminative or non-affective cognitive aspects that they may also have. Accounts of the motivational force of paradigmatically affective episodes, such as emotions, which *don't* appeal to affect are therefore strictly irrelevant. Narrowing our focus to affect is appropriate for three reasons. First, it might seem *prima facie* plausible that emotions (and other affective episodes) motivate in virtue of their affective feel. Second, the predominant view in the philosophy of affect is that there's a necessary connection between affect and motivation, suggesting that the *affect* of affective episodes is a

<sup>12</sup> See his (2015, section 4.2.1).

<sup>13</sup> Kauppinen (2013) and (2015, esp. p. 366, section 2.3), "Intuition and Belief in Moral Motivation" in Björnsson, G. et al, esp. sections 12.1–12.1.1.

<sup>14</sup> Zagzebski (2003, esp. 118–123).

<sup>15</sup> Dancy (2014, esp. p. 802).

<sup>16</sup> Doesn't restricting Internalism in the ways proposed amount to victory for Externalism? They seem no more of a capitulation to Externalism than more familiar non-affective Conditional Internalisms. We here charitably grant that extant versions of the affective appeal could support Internalism.

plausible locus for their connection to motivation.,<sup>17,18</sup> Third, some ethicists—most notably Zagzebski<sup>19</sup>—have explicitly claimed that it is affect’s motivational nature that yields states which constitute counterexamples to the Humean claim that there are no states which are both cognitive and motivational. Thus is the affective appeal not only supposed to vindicate a version of Internalism, but one that is compatible with Cognitivism.

We argue that the affective appeal fails: affective Internalisms are neither more empirically supported, nor clearly less controversial, than non-affective versions. Although affect doubtless plays an important role in moral motivation, we’re sceptical that establishing any such role advances the Practicality debate. Not in any straightforward way. At the very least, we take ourselves to establish that the affective appeal is far more controversial than hitherto assumed.

We proceed as follows. In Sect. 1, we consider and criticise the affective appeal for U-Internalism. In Sect. 2, we do the same for three Conditional versions of Internalism. We close in Sect. 3 by briefly identifying a plausible connection between affect and motivation which survives our arguments. But this connection doesn’t support Internalism. Thus, the affective appeal fails.

## 1 Unconditional Internalism

Recall:

*Unconditional Motivational Internalism* (U-Internalism): Necessarily, if S judges that she morally ought to  $\phi$ , then S is motivated (at least somewhat) to act in accordance with her judgment.

To consider the affective appeal we henceforth stipulate that ‘moral judgment’ in the versions of U-Internalism we’re considering refers to a judgment that is either caused or constituted by affect. So, for Sentimentalists like Prinz, U-Internalism applies to *all* moral judgments, while for Pluralists like Zagzebski, U-Internalism applies only to affective moral judgments.

In order to ground an affective U-Internalism proponents need to defend something like the following view:

*Hedomotive*<sup>20</sup>: Necessarily, if S is the subject of an occurrent affective mental episode that X, then she is motivated (at least somewhat) concerning X.

<sup>17</sup> See Aydede (2014, esp. p. 123).

<sup>18</sup> Beyond ethics and the philosophy of affect, our arguments have broader significance. E.g. something similar to the affective appeal is made in accounts of the nature of suffering (see, e.g. Brady 2019, esp. p. 44), and is implicit in pleasure-based theories of desire, e.g. Morillo (1990, esp. pp. 176–177).

<sup>19</sup> (2003, p. 116). See also Dancy (2014, p. 803).

<sup>20</sup> We take ‘Hedomotive’ from Bain (2011, p. 164). The term is introduced to capture affect’s presumed motivational nature. Hedomotive assumes that affective mental episodes have intentional content. Intentional accounts of affective mental episodes, particularly the *affect* of these episodes, are currently mainstream, e.g. see works by Aydede, Bain, Brady, Cohen, Cutter, Heathwood, Fulkerson, Jacobson,

Occurrent<sup>21</sup> affective mental episodes include any which have an affective aspect, e.g. emotions, pleasures, (painful) pains, and (un)pleasant thoughts. As in U-Internalism, Hedomotive can be understood metaphysically or conceptually. Again, the relevant motivation is defeasible, though not cancellable, and awaits precisification depending upon the relevant type of affective mental episode, e.g. for painful pain the motivated behaviour is plausibly that which is “apt to avoid or limit bodily damage,”<sup>22</sup> while for fear it is fight-or-flight behavior.<sup>23</sup>

The clause “concerning X” is to be understood in such a way as to potentially support U-Internalism, i.e. it includes motivation to act in ways semantically encoded in X, taking steps to act in such ways, etc. Whether Hedomotive should be construed more broadly, e.g. to include cases of *referred* motivation where I am angry at one person but “take it out” on someone else, and whether this would support Internalism, is a point we set aside. Indeed, there are general issues concerning the scope of the connection between judgment and motivation that independently impact versions of Internalism and Hedomotive, and are too large to address here. We henceforth assume the relatively restrictive views introduced, and endorsed by at least some proponents.<sup>24</sup> While more permissive versions of both views may survive some of our arguments, such views risk triviality and theoretical insignificance.

Keeping these complexities in mind, why believe Hedomotive? We identify three reasons. First, empirical evidence apparently identifies a substantial connection between affect and learning. As Prinz<sup>25</sup> writes:

it has been known since the early days of psychology that emotions can be used to condition behaviour. Humans and other animals will avoid behaviours that have been associated with negative emotions in the past.

However, this evidence is compatible with a merely typical, or reliable, connection between affect and motivation—and thus fails to establish the required necessary connection. A similar problem besets appealing to the association of basic

---

Footnote 20 continued

Klein, Martinez, and Tye. Without intentional content, we’re doubtful that appealing to affect could ground U-Internalism.

<sup>21</sup> As is standard, we take affect to be episodic. This is consistent with the fact that some paradigmatically affective phenomena are sometimes realized not only episodically, but dispositionally, e.g. love for one’s children.

<sup>22</sup> Bain (2013, p. 71).

<sup>23</sup> Zagzebski (2003) may be interpreted as endorsing Hedomotive, e.g. her claim that affect is intrinsically “pushy” (p. 116), and her defence of an Internalism “that is not relative to the judgments of persons who are virtuous or even especially rational” (p. 123).

<sup>24</sup> See Eggers (2015, p. 92) for U-Internalism. See Bain (2013) S71-2 for Hedomotive. In more recent work Bain (2019), he includes experience-directed motivation, e.g. as when I take a painkiller, as falling under the scope of Hedomotive. But experienced-directed motivation is the *wrong* kind of motivation for Internalists (see sections 1.2 and 2.1).

<sup>25</sup> (2015, p. 74).

emotions with characteristic action tendencies;<sup>26</sup> association doesn't entail necessary connection.

Second, it may seem to be an obvious part of folk psychology that affect is necessarily motivational. This assumption is often implicit in philosophical discussions,<sup>27</sup> though sometimes explicit. For example, Zagzebski references the "pushy-ness"<sup>28</sup> of affect, and Bain thinks that acceptance of something like Hedomotive constrains any adequate affective theorising.<sup>29</sup> In what follows, we provide reasons for doubt.

Before doing so, note finally that Bain<sup>30</sup> offers a putative additional reason to accept Hedomotive that isn't open to those making the affective appeal. Here, Hedomotive is taken to be plausible because affect has evaluative intentional content, e.g. represents things *as bad*, combined with the presumption that some kind of evaluative Internalism is true.<sup>31</sup> This move is circular if made in *defence* of Internalism.

We now provide reasons for doubting Hedomotive, but most importantly, for doubting that affective versions of U-Internalism have an advantage over non-affective versions.

### 1.1 Sub-personal-level evidence against Hedomotive

Empirical evidence concerning affective and motivational processing tells against Hedomotive. We here focus on empirical evidence concerning sub-personal states/processes as contrasted with the personal-level states/processes. One rough characterisation stemming from Daniel Dennett's introduction of the distinction is that personal-level states/processes are appropriately offered to explain the behaviour of the person by appealing to states/processes of the person as a whole, whereas sub-personal explanations appeal to states/processes of parts of persons.<sup>32</sup> If, e.g., I appeal to Anna's wanting the water when explaining her lifting the glass, I am appealing to a personal-level state of wanting; Anna, the person, wants the water. If I instead appeal to activity in Anna's motor cortex when explaining her lifting the glass, I am appealing to sub-personal processing; Anna's motor cortex, a part of Anna, is active. How to more precisely draw the distinction between personal-level mentality and underlying sub-personal processes to which they are related is controversial—as is the relationship between them.<sup>33</sup> The empirical research which we here focus on is paradigmatically sub-personal. There is strong evidence that sub-personal affective and motivational processing doubly dissociates.

<sup>26</sup> Frijda (1986). See also Prinz (2015 p. 74).

<sup>27</sup> Kauppinen (2013, p. 371).

<sup>28</sup> (2003, p. 116).

<sup>29</sup> (2013, S71).

<sup>30</sup> See, e.g., Bain (2013, S8.1).

<sup>31</sup> For discussion see Cohen and Fulkerson (2014, esp. §2.4).

<sup>32</sup> See Dennett (1969, p. 93).

<sup>33</sup> For a useful overview of the distinction see Drayson (2014).

As one of us has argued directly for this claim at some length elsewhere<sup>34</sup> we here only briefly summarise, and update, the evidence. As further explained below, the research reported here at least shifts the burden to proponents of Hedomotive in both its metaphysical and conceptual forms.

Dissociation between affective and motivational processing began to ascend to acceptance beginning with Robinson and Berridge's (1993)<sup>35</sup> introduction of *incentive sensitization theory*. Further details aside, according to this theory, there are two distinct sub-systems involved in reward: the sub-personal processing system involved in positive hedonics, dubbed 'liking,' and the sub-personal processing system involved in motivation, dubbed 'wanting.' 'Wanting' and 'liking' were intentionally constrained in quotes to emphasise that the relevant processing systems are sub-personal, such that their activities are distinct from any conscious, personal-level experience of positive affect or desire. The focus in the experimental paradigms is on the sub-personal processes, e.g. neurochemical release and transmission, posited as underlying the personal-level episodes. As one of us has argued elsewhere, 'liking,' accordingly, may be best understood as immediate hedonic impact and 'wanting' as what we might call motivational *oomph*, where neither affective or motivational signaling should be conflated with the conscious experience of affect or motivation.<sup>36</sup> The two sub-personal processing systems for hedonic impact and motivational *oomph* often work together, such that one 'likes' what one 'wants' and 'wants' what one 'likes.' Moreover, activity in these systems will often both be modulated by other internal states and processes, in particular homeostatic processes, such that e.g. when hungry, one 'likes' *and* 'wants' food more than one either 'likes' or 'wants' food when sated. The central claim of interest for present purposes is that these two systems are nonetheless distinct—involving distinct neurochemicals, pathways and, structures—and can be independently modulated.

Almost twenty-five years on, it's widely accepted that the sub-personal processing for positive hedonic impact and motivational *oomph* are indeed distinct and dissociable. It should be admitted that terminology and conceptualisations of these two systems, or processes, remains torturous. Particular difficulties have plagued interpretations of 'wanting' and 'liking', how to operationalise these, and how to limn the processing streams.<sup>37</sup> Nonetheless, and despite these difficulties, the distinction and dissociation between these processes is now so widely accepted that it's difficult to know whom best to cite. Robinson et al.,<sup>38</sup> summarise some of the evidence, including evidence for 'liking' without 'wanting' and 'wanting' without 'liking' (whatever terms we prefer). Pool et al.<sup>39</sup> review a wide range of impressive evidence supporting the existence of these two distinct and dissociable

---

<sup>34</sup> Corns (2014).

<sup>35</sup> Robinson and Berridge (1993).

<sup>36</sup> For elaboration, see again Corns (2014).

<sup>37</sup> For discussion see Barbano et al. (2009) and Tibboel et al. (2015).

<sup>38</sup> Robinson et al. (2015).

<sup>39</sup> Pool et al. (2016).



processes in both non-humans and, suitably conceptualised, in humans. We here merely report and cannot defend these claims.

While research to date has focused on dissociation between motivation and *positive* hedonics, there's reason to accept a parallel dissociation between motivation and *negative* hedonics. If there is a single affective system for both positive and negative hedonics, then existing evidence is already sufficient. Whether there is a single affective system, or if there are instead two, is currently debated.<sup>40</sup> Even if there are two distinct affective processing systems, however, the dissociation between positive hedonics and motivational processing would seem to provide at least *prima facie* support for the dissociation between negative hedonics and motivational processing, and there is some direct evidence. For example, as dopamine is the key neurochemical in motivational *oomph*, its release and seemingly similar role in processing negative stimuli provides some support.<sup>41</sup> Similarly, there is evidence that there are similarities in both systems, and their interaction, during both positive and negative affective experiences.<sup>42</sup> Thus, while there is less evidence concerning dissociating systems for *negative* hedonic impact and motivational *oomph*, the evidence that we do have supports dissociation.

We think this empirical evidence is persuasive and that philosophers should conclude, along with the affective science community, that there are double-dissociations between hedonic impact and motivational *oomph* at the level of sub-personal processing. For positive affect, the evidence for this claim is well established and widely accepted. For negative affect, the existing evidence is at least *prima facie* extendable, with additional evidence accumulating.

One might nonetheless worry that the dissociation between affective and motivational processing canvassed above is irrelevant to Hedomotive. We identify three reasons for this.

First, one might be sceptical that the sub-personal processes identified in affective science as affective and motivational processes, e.g. 'liking' and 'wanting', are relevant to the claims about affect and motivation in which philosophers are interested. Such scepticism is difficult to generally alleviate. Although affective scientists stand behind their conceptual frameworks and experimental paradigms, they are continually updated in the light of ongoing work, and open to challenge. Some may doubt that the 'wanting'/approach behaviour of the mouse in some one experiment illuminated motivational processing, or whether the rat's 'liking' facial expression illuminates affective processing in any. We take this to be the general case when philosophers engage with scientific paradigms. In this case, the scientific community has converged on these claims concerning affective and motivational processing through a number of experiments over the last

---

<sup>40</sup> On the former view, a single affective system is responsible for the full range of hedonic processing, e.g. an affective keyboard or workspace consisting mostly of hedonically neutral neurons, voxels, and areas. See e.g. Reynolds et al. (2008). The latter view posits two distinct and independent affective processing systems, each involving mostly hedonically specialised neurons, areas, and neurochemicals. See e.g. Norman et al. (2011).

<sup>41</sup> E.g. Schultz (2002) and Waddell (2013).

<sup>42</sup> See Leknes and Tracey (2008).

thirty years. For present purposes, we can only further say that we're willing to accept the dominant experimental paradigm in affective science unless and until there is some undermining challenge against it.<sup>43</sup>

Second, one might hold that, even if compelling, the empirical findings are irrelevant to metaphysical Hedomotive, e.g. cases of 'liking' without 'wanting' entail nothing about whether there is ever a personal-level mental episode of affect without motivation. While we concede that there are no straightforward entailments between the sub-personal evidence and the motivational profile of affect at the personal-level, one's antecedent commitments concerning their relationship may fill the gap. One may, e.g., accept a reductively mechanistic approach according to which personal-level phenomena are decomposed into sub-personal mechanisms as revealed by scientific inquiry.<sup>44</sup> In any case, sub-personal dissociation is sufficient to shift the argumentative burden onto those positing necessary connections at the personal-level. Specifically, if the mechanisms whose activity underwrites affect and motivation doubly dissociate, then one would at least like strong positive reasons for thinking that they're nonetheless necessarily linked at the personal-level. Further, in the following section, we offer examples of person-level dissociation.

Third and similarly, one might hold that, even if compelling, the empirical findings are irrelevant to *conceptual* Hedomotive, e.g. cases of 'liking' without 'wanting' don't fall under the category AFFECT. We are, however, sceptical of the view that conceptual claims are immune from empirical evidence. If the evidence suggests that what would otherwise be classified as AFFECT does not motivate, then this provides reason for conceptual revision. Admittedly, this is a complicated issue, not least because one's conceptual commitments will influence how one interprets empirical evidence.<sup>45</sup> As noted above, the conceptual frameworks deployed by the scientists are always open to challenge. We grant that, for some readers, reasons against conceptual Hedomotive require considering "intuitive" responses to personal-level examples of affect. We turn to these in the following section.

Despite these important concerns, we think—at the very least—this sub-personal evidence undermines the claim that the affective appeal yields a more empirically respectable or less theoretically controversial Internalism. The empirical research supports dissociation between affective and motivational processing, not necessary connection. At the very least, the burden is on the defender of Hedomotive. One wants reason to think that the relevant experimental paradigms or conceptual frameworks are flawed, that what is true of the sub-personal processing is false of the personal-level episodes, or that the relevant conceptual claims are totally immune from empirical revision.

We now directly consider personal-level evidence against Hedomotive.

<sup>43</sup> One of us has discussed these experiments as relevant to present philosophical claims in more detail elsewhere (Corns 2014).

<sup>44</sup> For some among many, see Bechtel (2008), Piccinini, esp. ch.1, G., and Craver (2011).

<sup>45</sup> See Kauppinen (2008, esp. p. 4).

## 1.2 Personal-level evidence against Hedomotive

It might appear that a promising way of establishing a personal-level version of Hedomotive (metaphysical or conceptual) is to defend a personal-level view of affect according to which it's at least partly constituted by some intrinsically motivational state, such as a desire. One might adopt a first-order view<sup>46</sup> according to which what makes, e.g., a sensation (un)pleasant, i.e. what (at least partially) constitutes its affect, is the existence of a desire for the intentional object of the sensation. Or one might endorse a second-order view<sup>47</sup> according to which what makes a sensation or thought pleasant is the existence of an intrinsic desire that the sensation or thought be currently occurring. If either view is correct, then assuming that desires are intrinsically motivating, affect will be also.

Before considering problem cases for both, note that appealing to second-order views is problematic for the affective appeal. If correct, then an affective moral judgment would be necessarily connected to motivation, but the relevant motivation would be *internally*-directed towards the judgment, and not *externally*-directed towards its object. To illustrate: suppose someone witnesses a cat being set on fire and judges that it's wrong. On the views under consideration, this is an unpleasant affective judgment. According to a second-order view, the affect is constituted by a desire that it (the judgment) not be occurring. Any motivation to do anything about the cat or its torturers is derivative of this desire. But that is an implausible and unattractive picture of moral motivation. It's implausible because it falsely implies that agents seek to eliminate negative affective moral assessments when they act in accordance with them. It's unattractive because it implies that moral agents care first-and-foremost about features of themselves, rather than the weal or woe of others. Thus, adopting the second-order desire view doesn't support the affective appeal.

We now present examples evidencing personal-level dissociation between affect and motivation. As with non-affective amoralist cases, we don't regard the examples as clear-cut. Instead, they demonstrate that the affective appeal leaves the Internalist embroiled in the same kind of dialectical controversy regarding intuitive responses to examples. Hence, the affective appeal doesn't advance the debate.

Consider the following two cases:<sup>48</sup>

*Unwarranted Guilt:* If someone knows that they're prone to feeling unwarranted guilt after nights out, then they may be unmotivated by guilty feelings experienced on subsequent occasions.

Here there is no personal-level evidence suggesting that the person is motivated by their guilt, e.g. they aren't motivated to apologise, gather evidence about wrong-doing, or to encourage others not to act as they have. They know that their affective episode is chemically-induced and that their behaviour was not actually wrong, and the episode

<sup>46</sup> E.g. Jacobson (2018, esp. section 3).

<sup>47</sup> E.g. Heathwood (2007, esp. p. 37).

<sup>48</sup> Space constrains our discussing non-moral dissociation cases in text, but we think they exist, e.g. cases of addiction wherein one may have an increase of approach motivation coupled with a decrease in positive affect; cases of uncontrolled pleasure wherein one may have a positive affect experience yet lack any motivation to continue. Such examples aren't clear-cut, but seem no more controversial than non-affective counterparts.

consequently lacks any motivational force. It is thus a case of cancelled affective motivation.

*Moral Despair* Due to over-exposure to 24-hour news coverage, someone morally despairs about the world and thereby lacks motivation to do anything as a result.

Here a moral emotion with negative affect fails to motivate action concerning the intentional object of the affective episode as needed for Hedomotive, e.g. they're not motivated to engage in activism (to expect this misunderstands the nature of the despair).

These cases provide reason for thinking that motivation and the affect of moral mental episodes dissociate at the personal-level. In conjunction with the sub-personal evidence, they threaten Hedomotive.

Against this, one may claim that one of two responses must hold for all such cases: either there is appropriate motivation of the predicted kind, e.g., it's merely overridden, or the relevant affect is absent. To illustrate, in *Unwarranted Guilt*, the subject *must* have some, if weak, motivation to apologise, despite knowing that they've done nothing wrong, which is simply overridden. In *Moral Despair*, the agent *must* be motivated: specifically, to withdraw from their despair's object, i.e. from the world.

We make two replies. First, we think the behavioural and linguistic evidence would support our dissociative description of these cases as against either kind of response.<sup>49</sup> Denying the relevant affect is unsupported in both. However, a proponent of conceptual Hedomotive might maintain that, since a mental episode gets to count as falling under the concept AFFECT (or perhaps MORAL AFFECT, i.e. the kind of affective state that could cause or constitute moral judgment) only if it is accompanied by motivation, our cases don't involve affect. However, the intuitive plausibility of our cases shows that conceptual Hedomotive isn't obviously less controversial than a non-affective Internalism which only counts something as a MORAL JUDGMENT if accompanied by moral motivation.<sup>50</sup> Thus, this affective appeal doesn't advance the debate.

Denying the appropriate motivation in our cases seems likewise dubious, if less obviously so. In *Unwarranted Guilt*, were you to ask the agent "aren't you at all motivated to apologise?" they would almost certainly say "of course not!"<sup>51</sup> They

<sup>49</sup> Aren't appeals to subjects' reports and behaviour-based inferences unreliable bases for interpreting cases? We don't think the unreliability is such that reports/inferences have *no* evidential weight. Also, aside from empirical evidence about affect—which favours rejecting Hedomotive—there isn't much that *is* more reliable.

<sup>50</sup> Cf. Tresan (2006, esp. pp. 147–148).

<sup>51</sup> Developing a suggestion by Eggers (2015, section 5.2), it might be thought that we can gain a reason for favouring the alternative (defeated not cancelled) interpretation by appealing to the following: (1) were we to "remove" the knowledge, we believe the agent would be motivated in the relevant ways; (2) if (1), then we have good reason to believe that in the actual case there exists a defeated, non-cancelled, motivation. However, although counterfactuals may provide us with guidance about what is the case, we're doubtful that counterfactuals about motivation always guide in the way (2) claims. Compare a structurally similar case: if the cheque *hadn't* been cancelled, it would have been cashable. But when a cheque *is* cancelled it doesn't retain some overridden cash-ability. This is anyway the same kind of dialectical controversy that predates the affective appeal.

may well be motivated to rid themselves of the unpleasant guilty feeling; but that won't vindicate U-Internalism. In Moral Despair, we doubt that the agent will be positively motivated to do nothing, e.g. if asked "are you motivated to just let go while the world goes to hell?" they would likely say "no". Moral despair (and non-moral cases of despair and depression) can ground a *lack* of motivation (cf. cases of contemplating aesthetic beauty, where agents may be motivated to do nothing but contemplate). Morally despairing agents may, however, have personal-level motivation to 'shake' the feeling, e.g. avoidance of the despair-inducing news, but this is irrelevant to U-Internalism.

For those still attracted to alternative interpretations, our second response is to highlight the kind of debate in which we're now engaged: concerning whether or not motivation, or the mental phenomenon posited as necessarily connected to it, is actually present, defeated, or cancelled in a range of cases. This is identical to a central aspect of the Practicality debate surrounding cases of putative amorality; e.g. does the amoralist make a genuine moral judgment; does the amoralist have any motivation, or is it just defeated or cancelled, etc. Given that this is one feature of the impasse between Internalists and Externalists, the affective appeal, through Hedomotive, doesn't advance the debate. At the very least, the affective appeal is a good deal less empirically supported and more theoretically controversial than proponents think.

## 2 Conditional Internalism

We here evaluate three weaker versions of Hedomotive—understood metaphysically and conceptually—and consider of each whether it can vindicate a concomitant Conditional Internalism or advance the Practicality debate. Though these conditionalised versions of Hedomotive and Internalism have yet to be explicitly discussed in the moral motivation literature they (1) mirror three main versions of non-affective Conditional Internalisms currently offered; (2) they're plausibly suggested by accounts explicitly discussed in the moral motivation literature; and (3) the first two proposals are based on conditionalised versions of Hedomotive which are explicitly considered in the affect literature. More generally, as precise detail concerning the nature of the intended connection between affect and motivation is largely absent from the moral motivation literature, conditional versions of Hedomotive developed from proposals in the affect literature may provide new resources for those seeking to vindicate a Conditional Internalism.

### 2.1 Well-functioning Hedomotive

Consider first:

*Well-Functioning Motivational Internalism* (F-Internalism): Necessarily, if S judges that she morally ought to  $\phi$ , then S is motivated (at least somewhat) to act in accordance with this judgment, if S is well-functioning.<sup>52</sup>

<sup>52</sup> E.g. Björnsson (2002).

Note again that ‘moral judgment’ here refers exclusively to judgments that are either caused or constituted by affect. As with all Conditional Internalisms we consider, it can be understood as a conceptual or metaphysical thesis. Our arguments here, and in the following sections, are intended to be relevant to both.

It might be thought that F-Internalism would be true if the following claim about the connection between affect and motivation were:

*Well-Functioning Hedomotive* (F-Hedomotive): Necessarily, if S is the subject of an occurrent affective mental episode that X, then S is motivated (at least somewhat) concerning X, if S is well-functioning.

F-Hedomotive is to be understood exactly as Hedomotive, with the sole difference of the further condition that S is well-functioning. The condition expresses the teleological idea that at least one function of affect is to motivate.<sup>53</sup>

Is F-Hedomotive plausible? Prima facie, it evades the personal-level counterexamples. Perhaps what goes awry in Unwarranted Guilt and Moral Despair is that the agent isn’t functioning properly.

However, at least two considerations are cause for concern.

First, especially when emotions are properly distinguished from their affective aspect, there is to our knowledge no clear empirical evidence supporting the teleological claim that affect has a distinctively motivational function, and hence no clear evidence favouring metaphysical F-Hedomotive. This functional claim shouldn’t be confused with weaker claims about the important roles that affect plays, e.g. in learning. Although there is good evidence that affect does play an important role in learning, so do non-affective representational states, like garden-variety assertoric beliefs. If the teleology at issue is natural, then empirical evidence that affect not only plays an important role in motivation, but that it has the function of motivating, should be forthcoming. The dissociations of the sub-personal mechanisms underlying affect and motivation discussed earlier provide reason to doubt that such evidence will be discovered.

Second, we have doubts about the *particular* functional connection needed to vindicate F-Internalism in its metaphysical and conceptual variants, i.e. a functional connection between affective episodes that X and motivation concerning X, in contrast to a general motivational affective function. Let us grant that when one judges, e.g., that they morally ought to refrain from eating meat, that if this judgment is affective and they’re well-functioning, then it will motivate them. Even granted, further reason needs to be provided for thinking that they will be motivated (e.g.) to *not eat meat*, as opposed to being motivated with respect to the unpleasant mental episode itself, e.g. to get rid of it, and only derivatively motivated to refrain from meat-eating. To vindicate F-Internalism, F-Hedomotive must require that affect has a motivational function concerning the intentional targets of affective episodes. But the motivation to act which is most plausibly identified as affect’s function appears to be *internally*-directed, as against the *externally*-directed

<sup>53</sup> See, e.g. Aydede and Fulkerson (2018, esp. section 3.3).

motivation needed to vindicate F-Internalism.<sup>54</sup> So even if affective episodes have intentional targets, and it's granted that the affect of those episodes has a motivational function, we're sceptical that the *affect* of affective episodes has the function of motivating action concerning those episode's intentional targets.

Despite these concerns, we remain open to the empirical and conceptual possibility that affect has the particular function of motivating action concerning the intentional targets of affective mental episodes. So, suppose that either metaphysical or conceptual F-Hedomotive is true.

Does F-Hedomotive constitute a success for the affective appeal? The functional integration of the mental provides strong reasons for doubt.

For any version of Hedomotive to support a concomitant version of Internalism that advances the Practicality debate, the posited connection between motivation and affect must be different from that which holds between motivation and non-affective cognition, but the functional integration of the mental is such that it's far from obvious that F-Hedomotive posits such a connection. On the proposed view, the affective moral judgments of the well-functioning agent are motivating; presumably because affect has a motivational function. But this is as plausible for non-affective cognition (this holds whether we think of the functional integration in metaphysical or conceptual terms). Many philosophers accept this. Uriah Kriegel, e.g., in defending a version of Moral Thought Pluralism, draws from the empirical literature to identify two cognitive processing systems.<sup>55</sup> While the details don't concern us, *non-affective* cognitions play a key role in the well-functioning of both systems. On this empirically-based model, non-affective cognition is as architecturally and functionally integrated with the motor system and motivation as affect. Kriegel's affective moral thoughts are explicitly based upon Tamar Gendler's aliefs.<sup>56</sup> While Gendler of course distinguishes aliefs from beliefs, their difference doesn't lie in whether they're functionally integrated with non-affective cognition in the well-functioning agent. Both are.

The relevant systems of a well-functioning agent who is motivated by their moral judgments apparently involve the functional integration of both affect and non-affective cognition. Indeed, such integration is plausibly required for well-functioning. Consequently, it's no more empirically supported or less theoretically controversial to posit affect than non-affective cognition as having a motivational function in the well-functioning agent. Appeal to affect through F-Hedomotive to vindicate an affective version of F-Internalism is thus as plausible, and for the same reasons, as appeal to the role of non-affective beliefs in the well-functioning agent to vindicate a non-affective F-Internalism. If so, then the affective appeal through F-Hedomotive doesn't advance the Practicality debate.

---

<sup>54</sup> Cf discussion of second-order theories on p. 15.

<sup>55</sup> Kriegel (2012, esp. p. 478).

<sup>56</sup> Gendler (2008).

## 2.2 Rational Hedomotive

Continuing to limit interpretation of the relevant moral judgments to the supposedly affective, consider a further Conditional Internalism:

*Rational Motivational Internalism* (R-Internalism) Necessarily, if S judges that she morally ought to  $\phi$ , then S is motivated (at least somewhat) to act in accordance with this judgment, if S is practically rational.

It might be thought that R-Internalism would be true if the following were:

*Rational Hedomotive* (R-Hedomotive) Necessarily, if S is the subject of an occurrent affective mental episode that X, then S is motivated (at least somewhat) concerning X, if S is practically rational.

R-Hedomotive's distinctive condition expresses the idea that undergoing an affective mental episode imposes a rational requirement on the agent, such that they're practically irrational if they're not (at least somewhat) motivated by the episode.

The most promising proposal for understanding R-Hedomotive is that affect *presents* subjects with reasons for action.<sup>57,58</sup> Whereas beliefs, e.g., *represent* their objects, other mental phenomena, such as sensory experiences, allegedly *present* their objects. To illustrate, compare believing that there is a red mail-box on your street on the basis of testimony, with having a visual experience as of the red mail-box. Some claim that in the latter but not the former case, there is a distinctive *presentational phenomenology* with respect to the proposition that there is a red mail-box.

There is, unfortunately, disagreement concerning clarifications of presentational phenomenology. Here is a representative list:

- (i) A feeling of passivity, i.e. of being insensitive to evidence.<sup>59</sup>
- (ii) The feeling of being made aware of a truth-maker.<sup>60</sup>
- (iii) The feeling of being "impressed" upon by reality.<sup>61</sup>

A further feature associated with presentational phenomenology is:

- (iv) Being beyond the space of reasons or not being rationally integrated.<sup>62</sup>

<sup>57</sup> An alternative view is that affect itself constitutes a reason for action. We worry that this either collapses into F-Hedomotive, or else only generates internally-directed reasons.

<sup>58</sup> Dancy (2014, esp. pp. 796, 803) defends the view that some affective episodes present reasons for action. See also Kauppinen (2013, esp. p. 370).

<sup>59</sup> Brogaard and Chudnoff (2016, esp. p. 72).

<sup>60</sup> Brogaard and Chudnoff (2016, esp. p. 69). Reiland (2015, esp. p. 525).

<sup>61</sup> Bain (2019, p. 484).

<sup>62</sup> Sellars (1956, esp. section 36).



Though further clarification would be ideal, we can use these characterisations to consider the proposal that affective mental episodes have presentational phenomenology with respect to reasons for action. If we accept (iii), for instance, then the current proposal is that undergoing an affective episode, e.g. an unpleasant pain, involves the feeling of being *impressed upon* that there are reasons for action, e.g. to do something about the pain's associated bodily location.<sup>63</sup>

Consider two initial worries for the proposal.<sup>64</sup> First, some advocates of presentational phenomenology for sensory experiences have argued directly against it for affective episodes—at least for emotions.<sup>65</sup> Brogaard and Chudnoff, for example, argue that emotions fail to meet criteria (i) and (ii) above such that, at best, emotions are *representational* with respect to reasons (or values). Second, and repeating a familiar point, even if affective episodes present reasons, they may not present the reasons needed to vindicate R-Internalism; the reasons presented by affect mightn't be the *externally*-directed reasons required. To illustrate: according to Neo-Sentimentalism, X is admirable just in case there are reasons (of the right kind) to admire X. If correct, affect may simply present reasons to be experiencing affective episodes, e.g. the affect of admiration may present reasons to feel admiration. But this is the wrong kind of reason to vindicate R-Internalism; moral judgments are posited as being importantly connected with reasons to act that go beyond reasons to feel.

Setting these concerns aside, we doubt that being presented with reasons constitutes any rational pressure on the agent to be motivated. If so, R-Hedomotive is false.<sup>66</sup>

To begin to see why, consider again Unwarranted Guilt. Here (let's suppose) my persistent guilt presents me with reasons to apologise, even though I know that there are no such reasons, because I know that I did nothing wrong. In this case, the presentation of reasons to  $\phi$  doesn't constitute any rational pressure to be motivated (even somewhat) to  $\phi$ .

It might be objected that the subject *is* practically irrational in such cases. In Unwarranted Guilt the guilt is 'recalcitrant' after-all. However, while such cases may involve irrationality, it isn't of the right kind to vindicate R-Hedomotive. The subject may be irrational in virtue of a conflict between their emotion and their judgment. Any practical failing here rests in their guilt, given their judgment. But we deny that they're irrational in virtue of an absence of affective motivation. Thus, recalcitrant cases remain counterexamples.

R-Hedomotive is presumably best supported by cases in which subjects has no antecedent knowledge that conflicts with the reasons presented by affect. For example, consider a case where I lie to my partner and consequently feel guilty. Let's suppose that I have no defeaters and that the affective component of the guilt presents me with reasons to apologise. However, even in this case, the presented

<sup>63</sup> See Bain (2019, §8) for defence of this view about painful pains.

<sup>64</sup> We set aside worries concerning presentational phenomenology's existence.

<sup>65</sup> E.g. Brogaard and Chudnoff (2016, esp. 68–71) and Dokic and Lemaire (2013).

<sup>66</sup> We merely assume that an agent is something which can act for (normative) reasons.

reasons are “non-insistent”, i.e., subjects are not rationally criticisable for not endorsing those reasons, or for being unmotivated to act in accordance with those reasons. This isn’t to deny that being presented with reasons rationally *permits* the agent to be motivated. But rational permission isn’t rational pressure, and it’s the latter which R-Hedomotive requires.

We offer two arguments that support our interpretation of the above cases.

First, an analogy with theoretical rationality. Consider a sensory experience of a red ball before one and suppose that the subject who undergoes this experience is thereby presented with the ball, its colour, and its location. We grant that, absent known defeaters, it would thereby be rationally permissible for the subject to be inclined to believe, or indeed to believe, that there is a red ball before them. However, the subject isn’t under rational pressure either to believe, or to be inclined to believe, that there is a red ball before them. Generally, reasons to believe provided by presentational phenomenology are non-insistent.

Why think this? First, many epistemologists take it as a *conceptual* truth about propositional justification (i.e. justification for belief that *p*) *in general* that it merely makes belief that *p* permissible.<sup>67</sup> Even rejecting this, the most commonly defended rivals to the permissibility account—that justification for *p* makes an agent blameless for believing *p*, makes it good to believe *p*, or some conjunction/disjunction of these—are consistent with the general claim that justification for believing falls short of rational requirement.<sup>68</sup> Second, it’s especially plausible that presentational phenomenology with respect to *p* merely makes belief that *p* *permissible*, if presentational phenomena are taken to stand beyond the space of reasons (see condition iv).

By analogy now, the plausibility of non-insistent reasons from presentational phenomenology in the theoretical case constitutes a *prima facie* reason to accept non-insistence in the practical case—for affective episodes in particular. In both cases, we’re considering the normative profile of presentational phenomenology. If presentational phenomenology is beyond the space of reasons in the theoretical case, we expect the same in the practical case. By analogy then: insofar as affect presents subjects with reasons, and thus rationalises motivation, it does so by making motivation permissible for the subject. It doesn’t make motivation obligatory, such that the subject must be motivated on pain of practical irrationality. Thus, R-Hedomotive is false.

It might be objected that, were subjects to *reflect* in theoretical cases, then in the absence of known defeaters, they would be rationally criticisable if they weren’t at least inclined to endorse their experiential presentation. The same point applies, *mutatis mutandis*, in practical cases. In reply, there will be many cases in which a subject *doesn’t* reflect on any practical reasons which affect may present. In response, our objector might appeal to the following principle: that *S would* be rationally criticisable were they to reflect on their reasons entails that *S is* rationally

<sup>67</sup> E.g. Goldman (1986).

<sup>68</sup> E.g. Silva (2017, esp. §5).

criticisable even in the absence of such reflection.<sup>69</sup> But this principle is dubious. For instance, it's doubtful that one is criticisable for not forming a particular judgment about a hard philosophical problem that one hasn't reflected upon, even if one *would* be criticisable for not forming such a judgment after reflection.

In reply, our objector may further attenuate R-Hedomotive:

*Reflective, Rational Hedomotive* (RR-Hedomotive) Necessarily, if S is the subject of an occurrent affective mental episode that X, then S is motivated (at least somewhat) concerning X, if S is *both* practically rational *and* reflects on the practical reasons with respect to X which the affective mental episode presents.

We make two responses. First, we doubt that mere reflection on the reasons presented by these experiences transforms a permission into a requirement. Second, it's worthwhile to consider what it is about reflection which supposedly generates rational requirements. Reflection often prompts active engagement with presented reasons by the reflecting agent, leading to either rejection or endorsement of those reasons. We grant that *endorsement* may generate rational requirements. However, to appeal to endorsement of the reasons which affect presents is to appeal to an active, non-affective, cognitive judgment, not to affect or its phenomenology.

This last consideration on what reflection may prompt connects to our second argument against R-Hedomotive based explicitly on Kantian considerations. Consider the difference between being presented with a reason to  $\phi$  and judging that I have reason to  $\phi$ . While we're sympathetic to the Kantian point that I can give myself reasons which exert rational pressure on me when (perhaps even merely because) I give them to myself, no such principle plausibly applies when I am merely *presented* with reasons. Accordingly, even if affect presents the agent with reasons to act, this doesn't exert any rational pressure on the agent.

This Kantian point may be approached by considering some of the criteria for presentational phenomenology. For instance, how does passively being presented with a reason to  $\phi$ , by itself, put rational pressure on me, e.g. to be motivated to  $\phi$ ? The Kantian thought is that it doesn't; active endorsement of reasons is required before the agent is rationally criticisable. How does a feeling of being *impressed upon* that there is a reason to  $\phi$  make me rationally criticisable for failing to be at all motivated to  $\phi$ ? The Kantian thought is that unless or until I *endorse* the feelings, they don't. Rational criticism only becomes appropriate *after* the subject has endorsed reasons to  $\phi$ .

Similarly, we might approach this Kantian argument against R-Hedomotive by focusing on the associated feature (iv) of presentational phenomenology: *being beyond the space of reasons* or *not being rationally integrated* with cognition. The only mental phenomena which put rational pressure on an agent are those which are rationally integrated in their psychology. This isn't to deny that there may be mental phenomena beyond the space of reasons, outside or prior to rational integration, which make things rationally *permissible*, e.g., sensory experiences may rationally

<sup>69</sup> Cf. Döring (2010, p. 295).

permit formation of particular judgments. But again, rational permissibility isn't rational requirement.

We may put consideration of feature (iv) as a dilemma for proponents of R-Hedomotive: either affect's presentation of reasons lacks rational integration or it doesn't. If feature (iv) of presentational phenomenology is *critical*, then taking this first horn is obligatory. On this horn, however, there is no reason to accept that affect's presented reasons exert any rational pressure on the agent; affect isn't integrated into the space of reasons in which the requirements of rationality apply. Affect's presented reason may at best be *inputs* to such a space, i.e. they may make mental episodes *permissible*. Thus, if we take this horn R-Hedomotive fails. To take the second horn, one must reject feature (iv) as *critical*. Crucially, however, holding that affect's presented reasons *are* rationally integrated, renders the appeal to affect unable to advance the Practicality debate. Being rationally integrated would entail that the reasons which affect presents are not independent of the other things that a subject believes, desires, etc. If so, then the way in which affect exerts rational pressure on the agent does not seem clearly distinct from the way in which any non-affective cognitions do so; rational pressure is enacted only after the affective and non-affective have been integrated. But then it's unclear why appealing to *affective* moral judgments (and the reasons they present) to vindicate R-Internalism is any less (or any more) controversial than appealing to *non-affective* moral judgments; either may generate rational requirements on action when integrated into the subject's psychology. Thus appealing to R-Hedomotive fails to advance the Practicality debate.

### 2.3 Moral perceptiveness Hedomotive

Briefly consider a final conditionalised version of Internalism:

*Morally Perceptive Motivational Internalism* (P-Internalism): Necessarily, if S judges that she morally ought to  $\phi$ , then S is motivated (at least somewhat) to act in accordance with this judgment, if S is morally perceptive.

It might be thought that P-Internalism would be true if following were:

*Morally Perceptive Hedomotive* (P-Hedomotive): Necessarily, if S is the subject of an occurrent affective moral mental episode that X, then S is motivated (at least somewhat) concerning X, if S is morally perceptive.

P-Hedomotive's distinctive condition expresses the idea that undergoing an affective *moral* mental episode is motivating for the morally perceptive agent.

Moral perceptiveness admits of at least two relevant interpretations. First, moral perceptiveness may involve responding to moral scenarios by having affective moral episodes, in contradistinction to forming non-affective moral beliefs.<sup>70</sup> So construed, however, this view collapses into a version of U-Internalism (restricted to affective judgments), and thus faces the same objections.

<sup>70</sup> E.g. Tolhurst (1995, section 5).

Second, moral perceptiveness may be a property of the virtuous agent, such that they *perceive* situations distinctively.<sup>71</sup> In particular, such that they're reliably sensitive to reasons for action that are (at least) intimately connected to their conception of how to live, which moreover involves perceiving those reasons as overriding and not in competition with others. So interpreted, we make three points. First, insofar as the virtuous agent is here taken to be *presented* with reasons, P-Hedomotive collapses into R-Hedomotive, and is above problematised. Second, if the virtuous agent isn't only presented with reasons, but if their moral perceptiveness itself involves a motivational state, then P-Hedomotive (and P-Internalism) is tautological. This renders the view non-substantive. Third, and most immediately relevant, P-Hedomotive and P-Internalism are apparently neither more nor less plausible than their non-affective counterparts, i.e. moral perceptiveness as involving non-affective moral judgments.<sup>72</sup> If so, then this affective appeal doesn't advance the debate.

### 3 Conclusion

We have argued that appealing to affect to vindicate a version of Internalism is problematic: at the very least, the resultant affective Internalisms are no less theoretically controversial and no more empirically supported than their non-affective counterparts.

The reader may wonder what connection we think *does* obtain between affect and motivation. In closing, consider:

*Normalcy Hedomotive* (N-Hedomotive): Normally, if S is the subject of an occurrent affective mental episode that X, then she is motivated (at least somewhat) concerning X.

If 'normally' is a claim about frequency, i.e., normally Y iff frequently Y, then N-Hedomotive is plausible. It certainly withstands our arguments. In support, note that we often intelligibly offer affective mental episodes as reasons for our actions, suggesting that affect is normally connected to motivation.

Although much more would need to be said in defence, N-Hedomotive is anyway too weak to support Internalism. An affective appeal vindicating Internalism requires more than a merely frequent, contingent connection between affect and motivation. Indeed, notice that non-affective moral thoughts that X are also normally connected to motivation concerning X. In support, note that citing non-affective beliefs, perhaps non-affective moral beliefs in particular, as reasons for action is just as intelligible as citing affective episodes. So, even if true, N-Hedomotive is apparently neither more nor less controversial than a parallel

<sup>71</sup> E.g. McDowell (1998).

<sup>72</sup> Although space constraints preclude discussion, this now familiar point—that an affective Internalism isn't more plausible than non-affective counterparts—can be made against an affective version of Deferred Internalism, i.e. Necessarily, if S makes a moral judgment that X, then the moral judgment that X is appropriately connected to a motivating (affective) moral judgment concerning X.

claim about the connection between motivation and non-affective beliefs. Hence, this affective appeal, though plausible, doesn't help Internalists.

Thus, the affective appeal fails. There are many rich and important connections between affect and motivation, and it's plausible that affect plays an important role in morality. These same claims about affect, however, are also true of non-affective cognition. Little wonder that the Practicality debate continues.

**Acknowledgements** Many thanks to an anonymous referee for their incisive comments. Thanks also to David Bain, Michael Brady, and audiences at Glasgow and the Hebrew University of Jerusalem for feedback on previous drafts.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Aydede, M. (2014). How to unify theories of sensory pleasure: An adverbialist proposal. *Review of Philosophy and Psychology*, 5(1), 119–133.
- Aydede, M., & Fulkerson, M. (2018). Reasons and theories of sensory affect. In D. Bain, M. Brady, & J. Corns (Eds.), *Philosophy of pain: Unpleasantness, emotion, and deviance* (pp. 27–59). London: Routledge.
- Bain, D. (2011). The imperative view of pain. *Journal of Consciousness Studies*, 18(9–10), 164–185.
- Bain, D. (2013). What makes pains unpleasant? *Philosophical Studies*, 166(1), 69–89.
- Bain, D. (2019). Why take painkillers? *Nous*, 53(2), 462–490.
- Barbano, M. F., Le Saux, M., & Cador, M. (2009). Involvement of dopamine and opioids in the motivation to eat: Influence of palatability, homeostatic state, and behavioural paradigms. *Psychopharmacology (Berl)*, 203, 475–487.
- Bechtel, W. (2008). *Mental mechanisms: Philosophical perspectives on cognitive neuroscience*. Routledge: Taylor & Francis.
- Björnsson, G. (2002). How emotivism survives immoralists, irrationality, and depression. *Southern Journal of Philosophy*, 40, 327–344.
- Björnsson, G., et al. (Eds.). (2015). *Motivational Internalism*. Oxford: Oxford University Press.
- Blackburn, S. (1998). *Ruling passions: A theory of practical reasoning, Chap. 1*. Oxford: Oxford University Press.
- Brady, M. (2019). *Virtue and suffering*. Oxford: Oxford University Press.
- Brink, D. (1989). *Moral realism and the foundations of ethics* (pp. 45–50). Cambridge: Cambridge University Press.
- Brogaard, B., & Chudnoff, E. (2016). Against emotional dogmatism. *Philosophical Issues*, 26(1), 59–77.
- Cohen, J., & Fulkerson, M. (2014). Affect, rationalization, and motivation. *Review of Philosophy and Psychology*, 5, 103–118.
- Corns, J. (2014). Unpleasantness, motivational oomph, and painfulness. *Mind and Language*, 29(2), 238–254.
- Craver, C. (2011). Integrating psychology and neuroscience: Functional analyses as mechanism sketches. *Synthese*, 183(3), 283–311.
- Dancy, J. (2014). Intuition and emotion. *Ethics*, 124(4), 787–812.

- Dennett, D. (1969). *Content and consciousness*. London: Routledge & Kegan Paul.
- Dokic, J., & Lemaire, S. (2013). Are emotions perceptions of value? *Canadian Journal of Philosophy*, 43(2), 227–247.
- Döring, S. (2010). Why be emotional? In P. Goldie (Ed.), *The oxford handbook of philosophy of emotion* (pp. 283–301). Oxford: Oxford University Press.
- Drayson, Z. (2014). The personal/subpersonal distinction. *Philosophy Compass*, 9(5), 338–346.
- Dreier, J. (1990). Internalism and speaker relativism. *Ethics*, 101, 6–26.
- Eggers, D. (2015). Unconditional motivational internalism and Hume’s lesson. In G. Björnsson, et al. (Eds.), *Motivational internalism*. Oxford: Oxford University Press.
- Francen, R. (2010). Moral motivation pluralism. *The Journal of Ethics*, 14(2), 117–148.
- Frijda, N. H. (1986). *The emotions*. Cambridge: Cambridge University Press.
- Gendler, T. (2008). Alief and belief. *Journal of Philosophy*, 105(10), 634–666.
- Goldman, A. I. (1986). *Epistemology and cognition* (p. 59). Cambridge: Harvard University Press.
- Heathwood, C. (2007). The reduction of sensory pleasure to desire. *Philosophical Studies*, 133(1), 23–44.
- Jacobson, H. (2018). Not only a messenger: Towards an attitudinal-representational theory of pain. *Philosophy and Phenomenological Research*, 99(2), 382–408.
- Kauppinen, A. (2008). Moral Internalism and the brain. *Social Theory and Practice*, 34, 1–24.
- Kauppinen, A. (2013). A Humean theory of moral intuition. *Canadian Journal of Philosophy*, 43(3), 360–381.
- Kriegel, U. (2012). Moral motivation, moral phenomenology, and the alief/belief distinction. *Australasian Journal of Philosophy*, 90(3), 469–486.
- Leknes and Tracey. (2008). A common neurobiology for pain and pleasure. *Nature Reviews Neuroscience*, 9(4), 314.
- Lenman, J. (1999). The externalist and the amoralist. *Philosophia*, 27(3–4), 441–457.
- McDowell, J. (1998). Virtue and reason. In J. McDowell & J. H. McDowell (Eds.), *Mind, value, and reality*. Cambridge: Harvard University Press.
- Morillo, C. (1990). The reward event and motivation. *Journal of Philosophy*, 87, 169–186.
- Norman, G. J., Norris, C. J., Gollan, J., Ito, T. A., Hawkey, L. C., Larsen, J. T., et al. (2011). Current emotion research in psychophysiology: The neurobiology of evaluative bivalence. *Emotion Review*, 3(3), 349–359.
- Pool, E., Sennwald, V., Delplanque, S., Brosch, T., & Sander, D. (2016). Measuring wanting and liking from animals to humans: A systematic review. *Neuroscience and Biobehavioral Reviews*, 63, 124–142.
- Prinz, J. (2015). An empirical case for motivational internalism. In G. Björnsson, et al. (Eds.), *Motivational internalism*. Oxford: Oxford University Press.
- Reiland, I. (2015). Experience, seemings, and evidence. *Pacific Philosophical Quarterly*, 96(4), 510–534.
- Reynolds, A., Reynolds, S. M., Richard, J. M., & Berridge, K. C. (2008). Mesolimbic dopamine in desire and dread: Enabling motivation to be generated by localized glutamate disruptions in nucleus accumbens. *Journal of Neuroscience*, 28(28), 7184–7192.
- Robinson, T. E., & Berridge, K. C. (1993). The neural basis of drug craving: An incentive-sensitization theory of addiction. *Brain*, 116, 247–291.
- Robinson, M. J. F., Fischer, A. M., Ahuja, A., Lesser, E. N., & Maniates, H. (2015). Roles of “wanting” and “liking” in motivating behavior: Gambling, food, and drug addictions. In E. Simpson & P. Balsam (Eds.), *Behavioral neuroscience of motivation. Current topics in behavioral neurosciences* (Vol. 27, pp. 105–136). Berlin: Springer.
- Rosati, C. (2016). Moral motivation. In E. N. Zalta (Ed.), *Stanford encyclopedia of philosophy*. <https://plato.stanford.edu/entries/moral-motivation/>.
- Schultz, W. (2002). Getting formal with dopamine and reward. *Neuron*, 36(2), 241–263.
- Sellars, W. (1956). Empiricism and the philosophy of mind. *Minnesota Studies in the Philosophy of Science*, 1(19), 253–329.
- Silva, P. (2017). The composite nature of epistemic justification. *Pacific Philosophical Quarterly*, 98(1), 25–48.
- Smith, M. (1994). *The moral problem, Chap 3*. Hoboken: Blackwell.
- Strandberg, C. (2011). The pragmatics of moral motivation. *The Journal of Ethics*, 15(4), 341–369.
- Tibboel, H., De Houwer, J., & Van Bockstaele, B. (2015). Implicit measures of “wanting” and “liking” in humans. *Neuroscience and Biobehavioral Reviews*, 57, 350–364.
- Tolhurst, W. (1995). Moral experience and the internalist argument against moral realism. *American Philosophical Quarterly*, 32(2), 187–194.

- Tresan, J. (2006). De dicto internalist cognitivism. *Nous*, 40, 143–165.
- van Roojen, M. (2010). Moral rationalism and rationalist amoralism. *Ethics*, 120, 495–525.
- Waddell, S. (2013). Reinforcement signalling in *Drosophila*; dopamine does it all after all. *Current Opinion in Neurobiology*, 23(3), 324–329.
- Zagzebski, L. (2003). Emotion and moral judgment. *Philosophy and Phenomenological Research*, 66(1), 104–124.

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.