The informational profile of valence: The metasemantic argument for imperativism

Manolo Martínez
Luca Barlassina

Abstract

Some mental states have valence—they are pleasant or unpleasant. According to imperativism, valence depends on imperative content, while evaluativism tells us that it depends on evaluative content. We argue that if one considers valence's informational profile, it becomes evident that imperativism is superior to evaluativism. More precisely, we show that if one applies the best available metasemantics to the role played by (un)pleasant mental states in our cognitive economy, then these states turn out to have imperative rather than evaluative content, since: (i) they are much more informative about behaviour than they are about the world; and (ii) they occupy a stage in the information-processing chain that is closer to behaviour production than it is to the uptake of sensory information. This is our metasemantic argument for imperativism.

1 Introduction

Tasting some yummy food, having an awful backache, imagining your favourite fragrance: these mental states have valence—they are pleasant or unpleasant. There are two main philosophical theories of valence: imperativism (Barlassina & Hayward [2019*a*]; Martínez [2011], [2015*a*], [2022]) and evaluativism (Carruthers [2018], [forthcoming]). Both theories offer an intentionalist account of valence—they claim that mental states are (un)pleasant in virtue of having a certain intentional content—but while imperativism says that (un)pleasantness depends on imperative content, evaluativism says it depends on evaluative content.[1,2]

In this article, we propose that we can make progress on the nature of valence by considering its *informational profile*. More precisely, we argue that if one applies the best available metasemantics to the role played by (un)pleasant mental states in our cognitive economy, then these states turn out to have imperative rather than evaluative content, since: (i) they are much more informative about behaviour than they are about the world;[3] and (ii) they occupy a stage in the information-processing chain that is closer to behaviour production than it is to the uptake of sensory information. This is our *metasemantic argument* for imperativism.

Before giving the argument, we need to get clearer on the distinction between imperativism and evaluativism.

2 The divide

2.1 Imperativism

Consider these two sentences:

(1) The door is closed.
(2) Close the door!

---

[1] It is important to distinguish intentionalist theories that attempt to account for the valence of all affective states—namely, the theories put forward by Barlassina & Hayward, Carruthers, and Martínez—from those that focus exclusively on pain, like Bain's ([2013]) and Klein's ([2015]). This article deals with the former, more encompassing type of theory.

[2] Different authors assign different strengths to this in-virtue-of relation, ranging from supervenience to identity. Given our aims, we needn't settle on any particular notion of metaphysical dependence. Pick the one you like best.

[3] In this paper, by 'the world' we always mean the extramental world as seen from the input side (that is to say, as it affects our sensory organs). Imperatival states also carry information about the world as seen from the output side (as it is modified by our behaviour), but here we are ignoring this aspect of their informational profile.

Sentence (1) has indicative content, that is to say, content that aims to state how things are, and hence can be true or false. The content of sentence (2), in contrast, is not truth-evaluable: it can be satisfied (if its addressee closes the door) or not satisfied (if they don't), but cannot be true or false. Sentence (2) has imperative content. The same distinction applies to mental states (Millikan [1984]; Shea [2018]). Some of them have indicative content. For example, a visual experience of a red ball has an indicative content along the lines of *There is a red ball*. Some other mental states have imperative content: a content that commands, or advises, rather than describes. The main tenet of imperativism is that (un)pleasant mental states have imperative content, and it is their imperative content that grounds their being (un)pleasant.[4]

Imperativism's main tenet has been developed by different authors in different ways. Martínez ([2011], [2015b], [2022]) defends *first-order imperativism*, according to which mental states have valence in virtue of having first-order imperative content—a type of imperative content that isn't directed at any mental state. In particular, pleasant experiences are supposed to command us to get more of a certain worldly object/state of affairs, while unpleasant experiences are supposed to command us to get less of it.[5] For example, the pleasure associated with eating an apple has an imperative content along the lines of (3), while a backache has an imperative content along the lines of (4):

(3) *Get more of this apple!*
(4) *Get less of this bodily damage in the back!*

Some other imperativists defend forms of *higher-order imperativism*: mental states have valence in virtue of having mind-directed imperative content. More precisely, *reflexive imperativism* says that a pleasant experience P has the content *Get more of P!*, and an unpleasant experience U has the content *Get less of U!* (Barlassina & Hayward [2019a]).[6]

In this article, we use 'imperativism' as an umbrella term that covers both the first- and higher-order versions. By the same token, we shall often be intentionally ambiguous as to whether the 'this' in '*Get more/less of this!*' refers to a worldly object or to a mental state. Our aim is not in fact to determine which version of imperativism is better (for a discussion, see Barlassina & Hayward [2019b], and Martínez [2022]), but rather to show that any of them is superior to evaluativism. But what is evaluativism?

_____

[4] We like to think of contents as intentional entities that are constitutively endowed with mood—hence the distinction between imperative and indicative contents (Hanks [2007], [2015]). You might not like that; perhaps you prefer to think about the imperativism-evaluativism divide as concerning valence's direction of fit: according to imperativism, valence has a world-to-mind direction of fit, while evaluativism says it has a mind-to-world direction of fit. None of our arguments hinges on this.

[5] To be precise, Martínez ([2011]) accepts that some sophisticated (un)pleasant mental states target other mental states, but concerning the kinds of state we consider in this paper, 'first-order' is an appropriate label for his view.

[6] Strictly speaking, Barlassina & Hayward characterise their view as *same-order* rather than higher-order, but we can safely ignore this subtlety.

2.2 Evaluativism

Just as imperative contents come in many shapes and forms, so do indicative contents. In particular, while some indicative contents are *descriptive*, others are *evaluative*. Consider these two sentences:

> (1) The door is closed.
> (5) The door is beautiful.

Even though both sentences express truth-apt contents, there is a significant difference between the content of (1) and the content of (5): the former is descriptive—it simply describes how things are; the latter is evaluative—it evaluates them.

Evaluativism (Carruthers [2018], [forthcoming]) is the view that mental states have valence in virtue of having indicative, evaluative content. More precisely, evaluativism's central claim is that pleasant mental states evaluate their objects as good, and unpleasant mental states evaluate their objects as bad.[7] Nothing prevents one from being a higher-order evaluativist—for example, one could maintain that an experience P is pleasant in virtue of having the evaluative content *P is good*. However, in the contemporary literature, evaluativism has been developed as a first-order theory only: according to Carruthers, pleasant/unpleasant mental states represent certain worldly objects (or states of affairs) as good/bad. A backache represents the condition of one's back as bad; the pleasant sensation associated with eating an apple represents the apple as good; and so forth.[8]

But in what sense is valence a first-order representation of goodness/badness? Here's Carruthers's answer: valence is a representation of the adaptive (dis)value of a worldly object—that is, pleasant/unpleasant mental states represent their worldly objects as fitness-enhancing/fitness-reducing.

2.3 Looking ahead

Now that imperativism and evaluativism, and the main differences between them, have been presented, it is time to establish which of these two theories is better. To do so, we need a neutral criterion of what counts as having evaluative rather than imperative content, and vice versa. That is to say, we need an explicit metasemantics. The next section proposes one.

---

[7] It is of course crucial that evaluativists take evaluative content to be indicative in nature. Suppose they instead thought: 'A mental state that evaluates object O as good has the content *More of O!*' In this case, evaluativism would collapse into imperativism.

[8] If one goes back far enough in time, one will find something that looks like a form of higher-order evaluativism. For example, Textor [manuscript] discusses Lotze's view that (un)pleasant experiences are "measures of the value of *impressions* for the individual being" (Lotze [1852], p. 242, emphasis added). Also, Brentano defended same-order evaluativism: "The simplest act, for example the act of hearing, has as its primary object the sound, and for its secondary object, itself, the mental phenomenon in which the sound is heard. Consciousness of this secondary object is threefold: it involves a presentation of it, a cognition of it and a feeling toward it. The feeling is pleasure or pain, but it is not a distinct further mental act" (Brentano, 1874/1973, p. 119). Our arguments also apply, *mutatis mutandis*, to these versions of evaluativism.

3 The metasemantic framework

3.1 Two principles

If there is a consensus view in the metasemantic study of mental representation, it is that intentional content emerges in the context of the following architectural motif—the *sender-receiver* motif, Fig. 1 (Skyrms [2010]; Godfrey-Smith [2014]; Millikan [1984]): a sender prepares information (say, incoming from the world) for transmission and transformation into a signal; the signal is then again transformed by a receiver at the other end, so that it can be put to use downstream (say, in the production of behaviour); and as a result, the signal ends up carrying information both about the source variable (typically, the world) and about the destination variable (typically, behaviour).
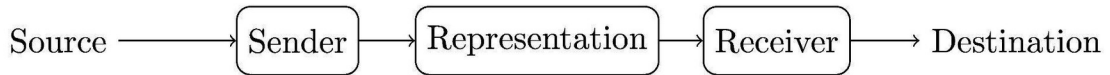

Source ⟶ Sender ⟶ Representation ⟶ Receiver ⟶ Destination

Fig. 1: The sender-receiver motif

When we say that the consensus view is that intentional content emerges in the context of the sender-receiver motif, we wish to remain neutral about whether the transmission and transformation of information suffices for intentional content (see Martínez [2019] for discussion). But information transmission is all we need here anyway, as the field has recently been converging towards the idea that it is possible to distinguish between indicative and imperative contents solely on the basis of their informational profiles. There are two interrelated ways to do this.[9] The first is to appeal to the following principle (see Martínez & Klein [2016]; Zollman [2011]):

> Informational Asymmetry Principle: If a mental state is more informative about the source variable (say, the world) than it is about the destination variable (say, behaviour), then it has indicative content but lacks imperative content. In contrast, if a mental state is more informative about the destination variable (say, behaviour) than it is about the source variable (say, the world), then it has imperative content but lacks indicative content. (If a mental state is equally informative about the source and destination variables, then it has both indicative and imperative content, and following standard usage we call it a 'pushmi-pullyu state' (Millikan [1995]).)

We can cash out the *Informational Asymmetry Principle* in terms of the standard information-theoretic idea (Cover & Thomas [2006]; MacKay [2003]), which philosophers associate most directly with Dretske's ([1981]) work, according to which informational dependencies among

---

[9] We should point out that we do not regard the two principles to follow as definitional, but merely as providing *defeasible evidence* of indicativeness and imperativalness. In other words, we take these two principles to be two *heuristics* that, given data concerning the informational profile of a mental state and its position in the processing chain, yield *reliable* but by no means infallible content ascriptions.

variables should be quantified in terms of expected reduction of uncertainty—say, mutual information. Accordingly, mental states that only have indicative (imperative) content are more informative about the world (behaviour) to the extent that their content reduces uncertainty about the source (destination) variable more than they do about the destination (source) variable.[10]

Consider a perception of a lush tree. This mental state significantly reduces one's uncertainty about what the world (in this case, the tree) is like, but it doesn't do so with respect to one's subsequent behaviour—to be told that there is a tree is not to be told much about what we will subsequently do. The *Informational Asymmetry Principle* thus ascribes indicative rather than imperative content to this mental state. In contrast, motor programs have imperative content (say, *Move your hand in this and that way!)* rather than indicative content, because they significantly reduce one's uncertainty about behaviour, while carrying little information about the world.[11]

A second way in which the distinction between indicative and imperative contents can be drawn is in terms of their positions in the information-processing chain. We can think of Figure 1 as an idealised depiction of the process that takes sensory information in on one side and produces behaviour on the other side. It is natural to suggest that a certain state is more imperatival (indicative) the further along (closer to the beginning of) this process it is. Take, for example, the motor instruction to move your right hand a certain way. This is imperatival rather than indicative because it is one of the latest stages in the process that eventuates in your hand moving. Your perception of a tree, on the other hand, is indicative rather than imperatival because it happens much earlier than that, and much closer to the reception of sensory information. Otherwise put, information processing goes from what there is to what is to be done about it—and the indicative/imperative distinction maps onto this evolving role of information processing, in its journey from sensation to behaviour. Hence the following principle:

> Processing Stage Principle: In an information-processing system, if a mental state is closer to the intake of sensory information than it is to behaviour production, then it has indicative rather than imperative content. In contrast, if a mental state is closer to behaviour production than it is to the intake of sensory information, then it has imperative rather than indicative content. (Pushmi-pullyu states occupy an intermediate position in

---

[10] The very popular *teleosemantic* approach to mental representation connects the indicative-imperative distinction not simply with informational asymmetries, but also with *explanatory* asymmetries. For example, in Shea ([2018]: chap. 7), imperative contents explain the production of the destination variable, while indicative contents are explained by the production of the source variable. We can, and will, safely gloss over the difference between explanatory and informational asymmetries in what follows, given that, in the kinds of cases that will interest us here, the explanatory asymmetry always depends on a prior informational asymmetry.

[11] The relevant uncertainty here is that of some (possibly subpersonal) cognitive mechanism that consumes the mental state in question (call this mental state 'M'), where the function of M is to guide action (for imperative M), or to construct a suitably accurate picture of the world (for indicative M). A worry is often voiced in relation to informational approaches to representation, about precisely what background knowledge or information should be assumed when calculating how, and how much, uncertainty is reduced. The straightforward answer is: whatever information is available to the consumer of M. There will of course be substantial questions about how to ascertain this empirically, but there are no particularly pressing questions of principle. We thank an anonymous reviewer for querying about these issues.

the information-processing chain.)[12]

These two principles go hand in hand. Mental states that have imperative content, but lack indicative content, are primarily involved in guiding behaviour. Accordingly, they need to transmit significant information about behaviour, after having integrated information about the source variable with other sources of information—including, for example, one's goals. This makes them both distant from the sensory periphery and not very informative about what is happening there. In contrast, mental states that have indicative content, but lack imperative content, are closer to the sensory periphery because their main function is to carry as much information about the source variable as possible.

3.2 Going back to the debate

It is easy to see how the metasemantic framework just described bears on the imperativism vs evaluativism debate. If imperativism is correct, then we should expect that (un)pleasant mental states occupy a late position in the information-processing chain, and that they are less informative about the world than they are about behaviour—in particular, imperativism predicts that pleasant/unpleasant mental states carry a lot of information about actions aimed at getting more/less of something (for brevity, we call these 'more-of/less-of actions'). In contrast, given evaluativism, valence should be closer to the sensory periphery than it is to behaviour production, and it should reduce one's uncertainty about the source variable to a greater extent than it does with respect to the destination variable—more precisely, valence should carry significant information about the value of worldly variables.

One might respond that things are in fact not so straightforward, given that the *Informational Asymmetry Principle* and the *Processing Stage Principle* were originally intended to draw a distinction between imperative and indicative states, and not between imperative and evaluative states in particular.[13] This is true, but since evaluativists insist that they take evaluative states to be a type of indicative state (see Section 1.2 above), applying the two aforementioned principles to the evaluativism vs imperativism debate should not raise any eyebrows: if a certain state is not indicative, then *a fortiori* it is not evaluative. To challenge this conditional, an evaluativist should either give up the claim that evaluative contents are indicative (although this would amount to giving up their original theory!), or put forward a novel metasemantics in which although evaluative states are indicative, they have a *sui generis* informational profile nonetheless. However, in the absence of such a metasemantics, the only way to assess whether valence is evaluative or not is to assess whether it is indicative or not. This is why, in building and defending our two metasemantic principles, we have focused on indicative states that aren't evaluative (such as the perception of a lush tree). To the best of our knowledge, no explicit metasemantics for evaluative states has ever been proposed.

---

[12] It is important not to read this principle as saying that imperative, but not indicative, states occur after sensory intake—this is true of virtually all mental states. The principle should also *not* be read as simply saying that imperatives occur far from the sensory periphery. This is true of many indicative states as well, e.g. very abstract beliefs.

[13] We thank an anonymous reviewer for pressing us on this point.

With this out of the way, we can finally give our metasemantic argument for imperativism. In a nutshell, the argument contends that valence's informational profile is as predicted by imperativism. Less briefly, it proceeds as follows. Valence plays a major role in large chunks of our cognitive economy. In particular, it is key to behaviour production, decision-making and learning. We will now consider valence's informational profile in each of these cognitive domains, and we will show that the metasemantic framework delineated above assigns imperative rather than indicative content to valence in each of them.

4 Valence and behaviour production
4.1 Valence as a behaviour controller

It is a platitude that (un)pleasant mental states bring about behaviours. Cognitive scientists sometimes make this point by saying that (un)pleasant mental states are "behaviour controllers" (Schlader et al. [2011], p. 233)—but a platitude remains. Consider, for example, these two mental states:

(A) A pleasant taste of marmite.
(B) An unpleasant smell of rotten fish.

If one tokens (A), then one will *ceteris paribus* seek another bite of marmite; given (B), one will *ceteris paribus* move away from the source of the smell.

This platitude, however, suffices to show that valence carries significant information about behaviour. The reason is simple: if the presence of an (un)pleasant mental state reliably correlates with a certain future behaviour, then this means that this state significantly reduces the uncertainty associated with what one is going to do next. Notice also that the behaviours that correlate with valence are exactly those predicted by imperativism—that is, more-of/less-of actions. In particular, first-order imperativism predicts that the likelihood that you will seek more marmite given (A), and steer clear from the source of the smell given (B), goes up because you are tokening the commands *Get more of this marmite!* and *Get less of these pathogens!*, respectively. Reflexive imperativism tells a slightly different story: you act the way you do because the imperative content of your experience commands you to get more/less of the experience itself and, of course, eating more marmite/getting away from the rotten fish is an excellent way to satisfy these commands. Either way, imperativism fits the behavioural data well.

On its own, the tight relation between valence and behaviour doesn't suffice to conclude that valence doesn't also carry information about the source variable—after all, valence could be a pushmi-pullyu representation, with approximately equal downstream- and upstream-looking informational connections, and in such a case one should conclude that valence has both imperative and indicative content. There is, however, ample evidence against the claim that valence also carries significant information about the world. Consider Mower's ([1976]) classical study on the relation between judgements of "thermal (un)pleasantness" and temperature (either of an external stimulus or of the body). As Figure 2 shows, (un)pleasantness carries almost no information about external temperature, or about bodily temperature: all pleasure ratings, high and low (in the y-axis), are compatible with high and low external temperatures (x-axis), depending on bodily temperature; and all bodily temperatures (the three lines) are compatible with high and

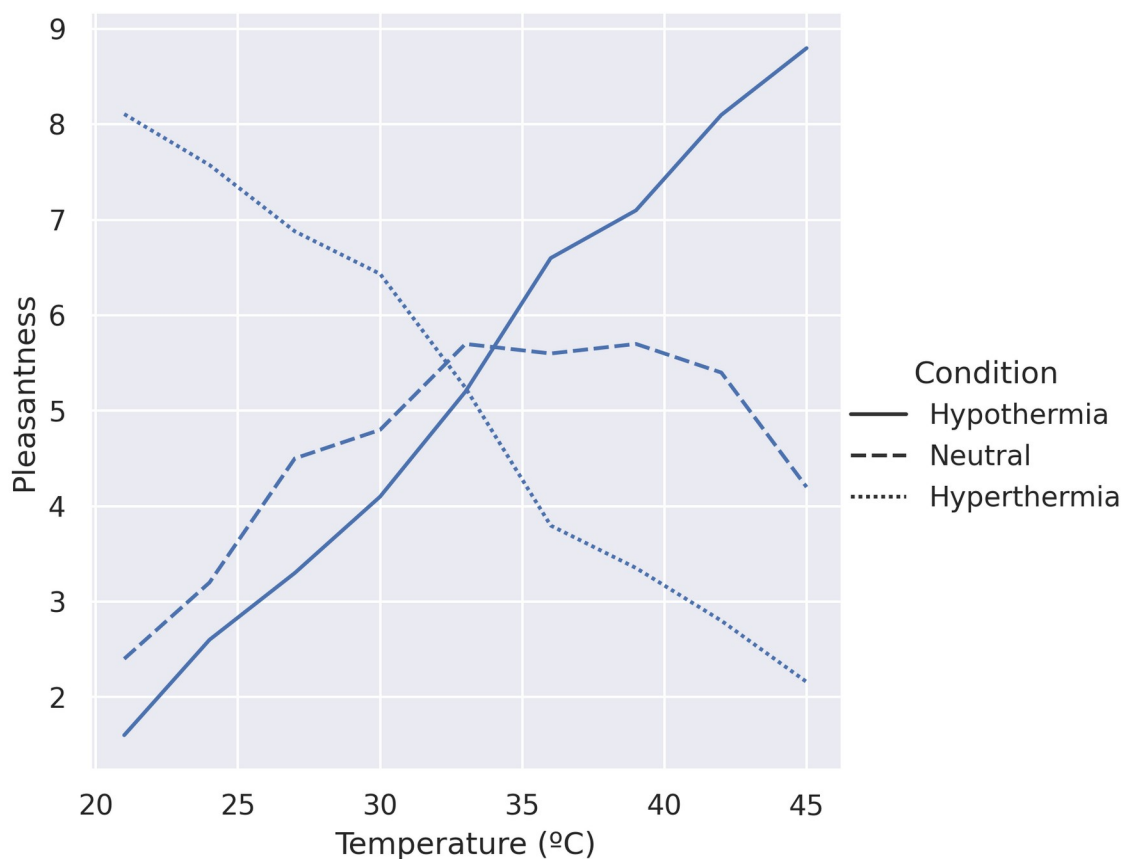low pleasure ratings, depending on the external temperature.

Figure 2: Redrawn from Mower ([1976], p. 1154)

An evaluativist is likely to respond that we are missing the point: "Fine, thermal (un)pleasantness is decoupled from worldly properties such as bodily or external temperature, and thus carries little information about either of them. But this is consistent with thermal (un)pleasantness carrying information about a more complex worldly property, namely, the fitness-enhancing/reducing profile of the combination of bodily and external temperatures. Accordingly, Mower's results don't impugn the hypothesis that thermal (un)pleasantness has the evaluative content *This bodily + external temperature is bad/good*."

We don't find this proposal convincing. In Mower's study, you can be in one of two internal temperatures (hypo- or hyperthermic), and the external temperature can take three values (low, medium, or high). Valence depends on the interaction of these two types of temperature—for example, one would experience pleasantness if the temperature of one's body is low and the external temperature is high, or if the temperature of one's body is high and the external temperature is low. The evaluativist interprets this fact as follows: pleasantness evaluates the following worldly property as good: [low body temperature and high external temperature or high bodily temperature and low external temperature]. But here's the rub: there is something glaringly disjunctive about this property—and one should avoid appealing to disjunctive properties when explaining a phenomenon. This is true in general, since evidence that one's explanatory target is a

conglomerate of disconnected disjuncts militates against its scientific significance (Putnam [1967]; Fodor [1974]; Kim [1992]), and this is also the typical stance in cognitive science: Spelke ([1990], p. 31) writes that cognition involves representing "unitary, bounded, and persisting bodies"; Feldman ([2003]) talks of cognition as concerning non-accidentally connected elements; Di Carlo and Cox ([2007], p. 333) describe cognition as bringing about the representation of unitary objects "in spite of tremendous variation in [their] appearance". The underlying point is the same: cognition is not in the business of representing disjunctive collections of properties, but rather of extracting unity out of multifarious data. A theory of valence according to which valence represents a disjunction of unrelated properties should therefore be resisted. In contrast, the content that imperativism ascribes to valence is as unified as a content could be: *More/less of this!*

But there's more to this. While Mower's experimental design stops at judgements of (un)pleasantness, there is a large body of work on behavioural thermoregulation (a recent review is Flouris & Schlader [2015]; and the foundational paper for this subfield is Weiss & Laties [1961]) showing that thermal (un)pleasantness is a powerful behavioural controller, and therefore carries a lot of information about behaviour. This is not an idiosyncratic feature of thermal (un)pleasantness. A similar picture emerges from studies on any type of (dis)pleasure, such as those associated with food consumption (Kringelbach [2004]) or sex (Georgiadis & Kringelbach [2012]): these states are very informative about behaviour and uninformative about the source variable. If we apply the *Informational Asymmetry Principle* to these data, the conclusion appears inevitable: (un)pleasant mental states have imperative content, not indicative-evaluative content.

The same conclusion can be reached if we consider the position of valence in the information-processing chain. Consider again Mower's study: low external temperature is pleasant if bodily temperature is high and unpleasant if it is low. The (dis)pleasure variable encodes this, but this means that valence is much further along than mere temperature sensing in the information-processing chain—it emerges at a stage where information about external and bodily temperature have been combined in a way that destroys information about the source variable, for the benefit of behaviour control. Again, there is nothing special about thermal (un)pleasantness in this respect. The idea that valence is much further along the processing path leading to behaviour in comparison with sensory perception is a staple of cognitive neuroscience. For example, Berridge et al. ([2009], p. 65) write that there are "many factors that alter pleasantness, such as hunger/satiety and learned preferences or aversions", while Schultz et al. ([1997], p. 1593) remark that valence "is not a static, intrinsic property of the stimulus ..., [but] a function of ... [subjects'] internal states at the time the stimulus is encountered and a function of their experience with the stimulus". Hence, as per the *Processing Stage Principle*, valence doesn't have indicative content—*a fortiori*, it doesn't have evaluative content. Rather it has imperative content.

4.2 A rejoinder

Unsurprisingly, we don't expect evaluativists to be so easily convinced. We predict that they are going to give a response along the following lines (this response is inspired by Carruthers [forthcoming]):

You argue that the reliable correlation between valence and behaviour indicates that the informational profile of valence is imperatival—more precisely: that valence has

the imperative content *Get more/less of this!* But if valence really had such a content, it should be realised in the motor cortex, which oversees the initiation and control of motor actions, or in the prefrontal cortex, since the latter is involved in goal formation and maintenance. But each of these predictions is disconfirmed by the data.

We think that this response misses the target. To begin with, imperativism does not predict any involvement of the motor cortex in valence, since *Get more/less of this!* is not a motor instruction. Rather, it is the expression of a fairly abstract request, or piece of advice, which can be implemented by different motor actions. The point runs deeper than this. It is not just that, according to imperativism, (un)pleasant mental states don't prescribe any particular motor action. It is also the case that they don't prescribe any action plan at all. The content *Get more/less of this!* tells one what to do, not how to do it.

An evaluativist might reply that imperativism then says that valence encodes what one intends to do (one's goals/intentions), and that it therefore incorrectly predicts the involvement of the prefrontal cortex. But, again, this is a mistake. There is an important difference between what-one-is-commanded-(or-advised)-to-do and what-one-intends-to-do. For imperativism, valence concerns the former type of "what", not the latter. You are running a marathon, and you hit a wall. The unpleasantness of your experience commands you to have less of this, but your intention remains the same, since you have stronger countervailing motivations. There is nothing odd about this. The commands issued by valence are just a subset of the many conative states that one can token. Still, *qua* conative states, they are closer to behaviour production than they are to sensory intake—their job is that of influencing, steering and guiding behaviour. Valence is a behaviour controller.

This, we maintain, is something that evaluativism cannot capture. If evaluativism were true, then valence should merely be in the business of carrying information about what is going on in the world. Consider the pleasure that one gets from drinking coffee. For imperativism, this positive valence is a command, hence it is intrinsically motivational. In contrast, according to evaluativism, this positive valence simply indicates that the coffee has such-and-such qualities. This means that to be motivated to drink more coffee, and to act accordingly, one would need some further pro-attitudes. Carruthers ([forthcoming], p. 7) is very clear about this: "For an imperativist [valence] is ... [an] instruction (*Do something to get more[/less] of [this]!*), [hence] its impact on action is a direct one—with facilitators, perhaps (e.g. beliefs about what might achieve more of [this]), but no intermediaries. For an evaluativist, in contrast, the motivational role of valence is indirect." That's the key point underlying imperativism. Surely, the mind contains many information-gathering states. But in the absence of imperatival states, no action is going to happen. Indicative states do the thinking; imperatival states do the moving. Valence, we say, is one such imperatival state.

5 Values without evaluations

In the previous section, we briefly discussed where valence is not realised in the brain. An unsympathetic reader might say that we were reticent about where it is realised, and that this was

intentional. Simplifying enormously, valence is realised in the orbitofrontal cortex (Grabenhorst & Rolls [2011]), and one might think that this is bad news for imperativism and excellent news for evaluativism, since many cognitive neuroscientists would agree that the orbitofrontal cortex codes for value. Carruthers again (Carruthers, forthcoming, pp. 13-14).: "Orbitofrontal cortex is thought to code for outcome values alone. ... It is conceivable that cognitive neuroscientists have gotten it badly wrong, and that orbitofrontal cortex ... [codes for *More/less of this!*] ... But it would require very strong philosophical arguments to warrant such a wholesale reworking of current scientific understanding." In this section, we argue that, *pace* Carruthers, imperativism doesn't require any "wholesale reworking" of the cognitive neuroscience of valence. The conception of valence emerging from this field is in fact already imperatival. This might sound surprising, so let us be clear about what we mean by this.

Now, Carruthers is absolutely right that value-talk is rampant in the cognitive neuroscience of valence. However, and this is our key point, one should be very careful about conflating this kind of value-talk with the notion of evaluative content proposed by the philosophers who defend evaluativism. If one takes a closer look at the cognitive neuroscience of valence, one will notice that the value-talk adopted there picks out distinctively imperatival mental states. Consider, e.g., how Paul Glimcher, one of the founders of neuroeconomics, describes the key question behind this research area: "What must the hidden internal representations which guide choice look like?" ([2014], p. 375). Analogously, Edmund Rolls, one of the most influential researchers on reward systems in the brain, writes that there is an intimate relationship between value and action, which is obtained by "bringing together information about specific rewards with information about actions, and the costs associated with actions, ... [in order to select] the correct action that will lead to a desired reward" (Grabenhorst and Rolls [2011], pp. 89-90, Box 1). A similar picture emerges from the work of Juechems and Summerfield, in that they take value to be "a summary of whether the agent is approaching or retreating from its goals" ([2019], p. 842, Box 2). Finally, the same idea can be found in the study of animal cognition. For example, in a very recent paper discussing the idea of valence as "a universal currency to compare the value of each option" (p. 9), Farnsworth and Elwood [forthcoming] describe negative value in terms of "the difference between [an animal's] current state and that sought by a goal-directed action selection system. It is therefore part of an anticipatory autonomy system. [An experience] feels bad because it is a state that is far from that desired and it motivates action in response" (p. 13). Given the *Informational Asymmetry* and the *Processing Stage* principles, it is reasonable to interpret these passages as showing that cognitive neuroscientists think of value in imperative rather than in indicative-evaluative terms.

An evaluativist might respond that discussing what cognitive neuroscientists explicitly say about value is of very limited interest. In science, the scientists do the talking, but it is the models that do the walking. Accordingly, in order to argue that the cognitive neuroscience of valence supports imperativism over evaluativism, one has to do the following: examine the neurocognitive models of valence developed in the field, and establish that these models advert to imperatival rather than to indicative-evaluative representations. This is exactly what we shall do next. In the case of valence, the cognitive capacities that get studied the most are decision-making and learning, since valence is taken to be key to both of these capacities (and we agree with that). So let's consider some well-established models of valence-based decision-making and valence-based learning in turn.

5.1 Valence-based decision-making

Valence-based decision-making is the process by which actions are chosen on the basis of imagining certain outcomes and comparing the degree of (un)pleasantness elicited by each imagined outcome. You are hungry but not thirsty, and are presented with an apple and a glass of water; you imagine choosing each item; imagining eating the apple elicits more pleasantness than imagining drinking the water; you choose the apple.

Cognitive neuroscientists tend to agree that valence-based decision-making involves the following representational-computational processes (Rangel et al. [2008]; Glimcher [2014]):

I.   A representation of the decision problem: how many states are relevant to the problem at hand, and how many actions can be taken?
II. The calculation of the value of each outcome, the results of which manifest themselves in different degrees of valence.
III. Action selection based on elicited valence.[14]

Accordingly, valence-based decision-making problems—for example, the "Apple-or-Water problem" above—can be represented in terms of a standard decision-theoretic matrix:

	Hungry, not thirsty	Thirsty, not hungry
Pick apple	+5 pleasantness	-1 pleasantness
Pick water	+1 pleasantness	+10 pleasantness

Table 1. The Apple-or-Water problem

Given all this, Carruthers is clearly right that, in these psychological models, valence is a proxy for outcome value—it indicates the value that a subject ascribes to a certain action, given a certain state of the world. However, it doesn't follow from this that if we take these models seriously, we should then conclude that valence has evaluative content. The notion of (outcome) value employed here is a technical one: in accordance with textbook decision theory, it is just a way of measuring one's preferences, and disagreement looms large as to whether preferences should be conceived of in an indicative-evaluative fashion (Broome [1991]) or not (Lewis [1988]). Therefore, Carruthers's argument from value-talk to evaluative content is again inconclusive. Whether valence is evaluative or imperatival cannot be read off from these psychological models. One should rather interpret these models in light of some principled criterion. Our metasemantic framework provides one.

Let's start with what valence is informative about in decision-making. In the psychological models under consideration, valence determines what action will be taken—for example, in the "Thirsty, not hungry" state, one will pick the glass of water rather than the apple, given that +10 pleasantness > +1 pleasantness (See Table 1). Hence valence carries perfect information about

---

[14] The computational details concerning how the option with the highest degree of pleasantness gets chosen varies from model to model. Some propose that action selection employs a winner-take-all algorithm (Glimcher [2014]), while others opt for a race-to-barrier diffusion process (Rangel et al. [2008]). These details need not detain us here.

subsequent behaviour. Of course, this is an idealisation—the actual biological process is noisy, and allows for a degree of stochasticity in the chosen action. Let's just say that these psychological models portray valence as highly informative about behaviour. In contrast, in these models, valence doesn't carry much information about the source variable, since valence is the result of combining multiple variables. In our Apple-or-Water problem, these variables are the availability of certain foods/drinks and one's own physiological states, but this is a massive simplification—any psychologically realistic model makes valence depend on a whole host of other variables, including one's risk-aversion (Glimcher & Rustichini [2004]) and one's tendency to discount future rewards (McClure et al. [2004]).[15] If we apply the *Informational Asymmetry Principle* to these psychological models of valence-based decision-making, we should then conclude that valence has imperative rather than indicative-evaluative content.

The *Processing Stage Principle* gives the same verdict, and for exactly the same reason. Recall that this principle says that mental states that are closer to the intake of sensory information than to behaviour production have indicative content but lack imperative content, while things are the other way around for mental states that are closer to behaviour production. If valence enters decision-making after information about a variety of worldly properties has been combined with information about a variety of internal states, then it is inescapable that valence emerges at a late stage of the information-processing chain, and is thus imperatival rather than indicative-evaluative.

5.2 Valence-based learning

Valence-based learning is a type of reward-based learning in which one exploits the valence generated by an action (and particularly the mismatch with predicted valence) in order to better estimate future valence: Pablo sees a carambola fruit and anticipates great pleasure from taking a bite; then he takes a bite and... meh; next time Pablo sees a carambola fruit, he will be less thrilled. Evaluativists and imperativists interpret this phenomenon differently. For the former, valence-based learning is a type of evaluative learning in which the stored evaluation of a certain worldly object (e.g. how good/bad a carambola fruit is represented to be) is updated in the light of a novel evaluation. For the latter, it is instead the strength of a certain command that gets scaled up or down depending on the strength of the command issued by a novel experience. Who is right?

Here, again, it is undeniable that cognitive neuroscientists typically resort to value-talk to describe this learning process. For example, Carruthers [forthcoming] discusses Juechems and Summerfield ([2019]), who explicitly frame the computational problem solved by valence-based learning as follows: "How is the reward value of a stimulus computed ... ?" (Juechems and Summerfield [2019], p. 838). Yet, again, it would be a mistake to conclude that this favours evaluativism over imperativism. Juechems and Summerfield are in fact adamant that valence

---

[15] Why is it the case that the more internal variables a state depends on, the less the information it carries about the input (source variable)? Consider someone choosing an apple over a glass of water: it could be that the agent is equally hungry and thirsty, but the apple looks particularly yummy; or it could be that the apple looks so-so, but the agent is a low future-reward discounter and foresees that they'll be hungry in a while; or it could be that the apple looks quite bad, but the water looks even worse; and so on. Accordingly, the decision to choose, say, an apple over a glass of water carries very little information about the quality of the apple. Given that the decision depends on a host of internal states and preferences of the agent, a lot of information about the input side is lost.

cannot be a representation of a worldly property, since valence is the result of combining information about the world with information concerning a myriad of internal factors, including physiological, cognitive, and motivational variables. This means that, for these authors, valence is uninformative about the source variable, including about the goodness/badness of the encountered worldly object, and emerges at a late stage in the information-processing chain.

But is valence informative about behaviour in this context? If we inspect the details of the computational model of valence-based learning put forward by Juechems and Summerfield ([2019]), then we can see that the answer to this question is positive. The key features of the model are that: (i) agents have a set of goals/desired states (the Θ^* in Figure 3); (ii) action (the arrows) aims to minimise one's current distance (the Θ) from these goals; (iii) the valence associated with a certain action corresponds to the distance resulting from taking that action (the various Θ') and one's goals—actions that bring about an increase/decrease in distance will result in unpleasantness/pleasantness—; and (iv) stored valence is updated on the basis of (iii). Given (i)-(iv), someone's next behaviour can be trivially computed: one will tend to perform behaviours that are predicted to minimise distance from one's goals, and will refrain from behaviours that are predicted to obtain the opposite effect. Valence is thus highly informative about behaviour.
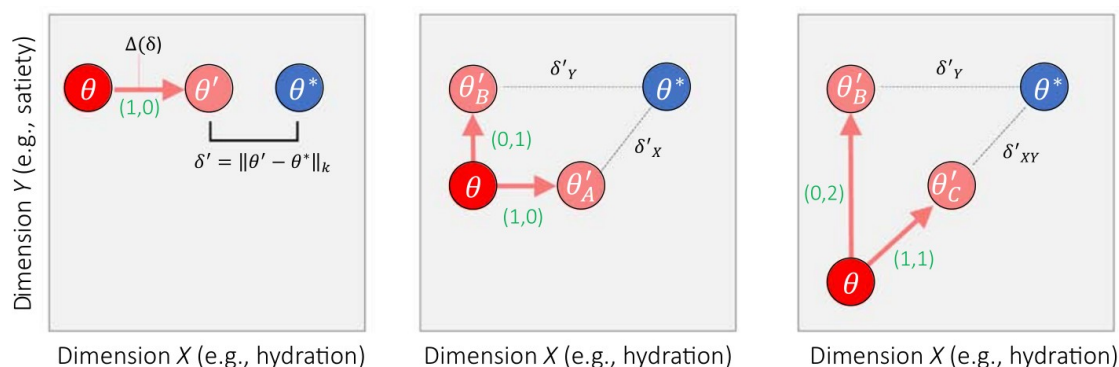

Figure 3. Juechems and Summerfield's model of valence-based learning. Reproduced from Juechems and Summerfield ([2019], Fig. 2B)

To cut a long story short, valence-based learning involves mental states that are highly informative about behaviour and uninformative about the world, and these states occur at a processing stage that is closer to behaviour production than it is to the intake of sensory information. If we apply the *Informational Asymmetry Principle* and the *Processing Stage Principle* to these data, we should conclude that these mental states have imperative rather than indicative-evaluative content.

6 Conclusion

The main philosophical divide in the study of valence is that between imperativism and evaluativism. In this article, we have argued that if one considers the informational profile of valence in the light of the best available metasemantics, then valence turns out to have imperative rather than evaluative content. In doing so, we have also shown that imperativism is not merely a

figment of the philosophical imagination, but is in fact the best way to interpret the cognitive neuroscience of valence. The importance of this point cannot be overstated, since it highlights a key difference between the cognitive neuroscience of valence and the cognitive neuroscience of perception. While the latter is the province of signal detection analysis (e.g. Gardner [2019]) and pattern recognition (e.g. Olshausen et al. [1993])—two techniques that deal with information about the source variable, and hence target mental states with indicative content—the former deals with how information about the world is combined with one's internal states and goals to guide behaviour, thereby targeting mental states with imperative content. *Pace* Carruthers, the cognitive neuroscience of valence has always been about imperatives.

Acknowledgements

Manolo Martínez
Department of Philosophy
Universitat de Barcelona
Barcelona, Spain
manolomartinez@ub.edu

Luca Barlassina
Department of Philosophy
University of Sheffield
Sheffield, United Kingdom
l.barlassina@sheffield.ac.uk

References

Ariew, A., & Lewontin, R. C. ([2004]). The confusions of fitness. *The British Journal for the Philosophy of Science*, *55*(2), 347–363. https://doi.org/10.1093/bjps/55.2.347

Bain, D. ([2013]). What makes pains unpleasant? *Philosophical Studies*, 166, S69–S89.

Barlassina, L. & Hayward, M. K. ([2019a]). More of me! Less of me! Reflexive imperativism about

affective phenomenal character. *Mind*,128(512), 1013-1044.

Barlassina, L., & Hayward, M. K. ([2019b]). Loopy regulations: The motivational profile of affective phenomenology. *Philosophical Topics*,47(2), 233-261.

Berridge, K. C., & O'Doherty, J. P. ([2014]). From experienced utility to decision utility, In P. W. Glimcher & E. Fehr (Eds.), *Neuroeconomics* (2nd ed., pp. 335–351). Academic Press. https://doi.org/10.1016/B978-0-12-416008-8.00018-8

Berridge, K. C., Robinson, T. E., & Aldridge, J. W. ([2009]). Dissecting components of reward: 'Liking', 'wanting', and learning. *Current Opinion in Pharmacology*, 9(1), 65–73. https://doi.org/10.1016/j.coph.2008.12.014

Brentano, F (1874/1973). *Psychologie vom empirischen Standpunkt*. Duncker & Humblot. English translation: *Psychology from an Empirical Standpoint*, transl. by A.C. Rancurello, D.B. Terrell, and L. McAlister. Routledge.

Broome, J. ([1991]). *Weighing Goods: Equality, Uncertainty and Time*. Oxford: Blackwell.

Carruthers, P. ([2018]). Valence and value. *Philosophy and Phenomenological Research*, 97(3), 658–680. https://doi.org/10.1111/phpr.12395

Carruthers, P. (forthcoming). On valence: Imperative or representation of value? *The British Journal for the Philosophy of Science*.

Cover, T. M., & Thomas, J. A. ([2006]). *Elements of Information Theory*. New York: Wiley.

Dretske, F. ([1981]). *Knowledge and the Flow of Information*. MIT Press.

Farnsworth, K. D., & Elwood, R. W. (forthcoming). Why it hurts: With freedom comes the biological need for pain. *Animal Cognition*. https://doi.org/10.1007/s10071-023-01773-2

Feldman, J. ([2003]). What is a visual object? *Trends in Cognitive Sciences*, 7(6), 252–256. https://doi.org/10.1016/S1364-6613(03)00111-6

Fodor, J. ([1974]). Special sciences (or: The disunity of science as a working hypothesis), *Synthese*, 28(2): 97–115. doi:10.1007/BF00485230

Flouris, A. D., & Schlader, Z. J. ([2015]). Human behavioral thermoregulation during exercise in the heat: Human behavioral thermoregulation. *Scandinavian Journal of Medicine & Science in Sports*, 25, 52–64. https://doi.org/10.1111/sms.12349

Gardner, J. L. ([2019]). Optimality and heuristics in perceptual neuroscience. *Nature Neuroscience*, 22(4), 514–523. https://doi.org/10.1038/s41593-019-0340-4

Georgiadis, J. R., & Kringelbach, M. L. ([2012]). The human sexual response cycle: Brain imaging evidence linking sex to other pleasures. *Progress in Neurobiology*, 98(1), 49–81. https://doi.org/10.1016/j.pneurobio.2012.05.004

Glimcher, P. W. ([2014]). Value-based decision making. In P. W. Glimcher & E. Fehr (Eds.), *Neuroeconomics* (2nd ed., pp. 373–391). Academic Press.

Glimcher, P. W., Dorris, M. C., & Bayer, H. M. ([2005]). Physiological utility theory and the neuroeconomics of choice. *Games and Economic Behavior*, 52(2), 213–256. https://doi.org/10.1016/j.geb.2004.06.011

Glimcher, P. W., & Rustichini, A. ([2004]). Neuroeconomics: The consilience of brain and decision. *Science*. https://doi.org/10.1126/science.1102566

Godfrey-Smith, P. ([2014]). Sender-receiver systems within and between organisms. *Philosophy of Science*, 81(5), 866–878. https://doi.org/10.1086/677686

Grabenhorst, F. & Rolls, E. ([2011]). Value, pleasure and choice in the ventral prefrontal cortex.

Trends in Cognitive Sciences, 15, 56-67.

Hanks, P. ([2007]). The content-force distinction. *Philosophical Studies*, 134(2): 141-164.

Hanks. P. ([2015]). *Propositional Content*. Oxford University Press.

Juechems, K., & Summerfield, C. ([2019]). Where does value come from? *Trends in Cognitive Sciences*, 23(10), 836–850. https://doi.org/10.1016/j.tics.2019.07.012

Keramati, M., & Gutkin, B. ([2014]). Homeostatic reinforcement learning for integrating reward collection and physiological stability. *ELife*, 3, e04811. https://doi.org/10.7554/eLife.04811

Kim, J. ([1992]). Multiple realization and the metaphysics of reduction, *Philosophy and Phenomenological Research*, 52(1): 1–26. doi:10.2307/2107741

Klein, C. ([2015]). *What the Body commands: The Imperative Theory of Pain*. MIT Press.

Klein, C., & Martínez, M. ([2018]). Imperativism and pain intensity, in D. Bain, M. Brady, & J. Corns (Eds.), *Philosophy of Pain: Unpleasantness, Emotion, and Deviance* (pp. 13–26). Routledge.https://www.routledge.com/Philosophy-of-Pain-Unpleasantness-Emotion-and-Deviance-1st-Edition/Bain-Brady-Corns/p/book/9780815361640

Kringelbach, M. L. ([2004]). Food for thought: Hedonic experience beyond homeostasis in the human brain. *Neuroscience*, 126(4), 807–819. https://doi.org/10.1016/j.neuroscience.2004.04.035

Lewis, D. K. ([1988]). Desire as belief. *Mind*, 97, 323–332.

Lotze, H. (1852). *Medicinische Psychologie oder Physiologie der Seele*. Leipzig: Weidmannsche Buchhandlung.

Martínez, M. ([2011]). Imperative content and the painfulness of pain. *Phenomenology and the Cognitive Sciences*, 10(1), 67–90.

Martínez, M. ([2015a]). Disgusting smells and imperativism. *Journal of Consciousness Studies*, 22 (5–6): 191–200.

Martínez, M. ([2015b]). Pains as reasons. *Philosophical Studies*, 172(9), 2261–2274. https://doi.org/10.1007/s11098-014-0408-7

Martínez, M. ([2019]). Representations are rate-distortions sweet spots. *Philosophy of Science*, 86(12), 1214-1226.

Martínez, M. ([2022]). Imperative transparency. *Mind*, 131(522), 585–601.

Martínez, M., & Klein, C. ([2016]). Pain signals are predominantly imperative. *Biology & Philosophy*, 31(2), 283–298. https://doi.org/10.1007/s10539-015-9514-y

McClure, S. M., Laibson, D. I., Loewenstein, G., & Cohen, J. D. ([2004]). Separate neural systems value immediate and delayed monetary rewards. *Science*, 306(5695), 503–507. https://doi.org/10.1126/science.1100907

MacKay, D. J. ([2003]). *Information Theory, Inference and Learning Algorithms*. Cambridge University Press.

Millikan, R. G. ([1984]). *Language, Thought, and Other Biological Categories: New Foundations for Realism*. MIT Press.

Millikan, R. G. ([1995]). Pushmi-pullyu representations. *Philosophical Perspectives*, 9, 185-200.

Mower, G. D. ([1976]). Perceived intensity of peripheral thermal stimuli is independent of internal body temperature. *Journal of Comparative and Physiological Psychology*, 90(12), 1152–1155. https://doi.org/10.1037/h0077284

Olshausen, B. A., Anderson, C. H., & Essen, D. V. ([1993]). A neurobiological model of visual

attention and invariant pattern recognition based on dynamic routing of information. *Journal of Neuroscience*, 13(11), 4700–4719. https://doi.org/10.1523/JNEUROSCI.13-11-04700.1993

Rangel, A., Camerer, C., & Montague, P. R. ([2008]). A framework for studying the neurobiology of value-based decision making. *Nature Reviews Neuroscience*, 9(7), 545–556. https://doi.org/10.1038/nrn2357

Sanfey, A. G., Loewenstein, G., McClure, S. M., & Cohen, J. D. ([2006]). Neuroeconomics: Cross-currents in research on decision-making. *Trends in Cognitive Sciences*, 10(3), 108–116. https://doi.org/10.1016/j.tics.2006.01.009

Schlader, Z. J., Simmons, S. E., Stannard, S. R., & Mündel, T. ([2011]). The independent roles of temperature and thermal perception in the control of human thermoregulatory behavior. *Physiology & Behavior*, 103(2), 217–224. https://doi.org/10.1016/j.physbeh.2011.02.002

Schultz, W., Dayan, P., & Montague, P. R. ([1997]). A Neural substrate of prediction and reward. *Science*, 275(5306), 1593–1599. https://doi.org/10.1126/science.275.5306.1593

Shea, N. ([2018]). *Representation in Cognitive Science*. Oxford University Press.

Skyrms, B. ([2010]). *Signals: Evolution, Learning & Information*. Oxford University Press.

Spelke, E. S. ([1990]). Principles of object perception. *Cognitive Science*, 14(1), 29–56. https://doi.org/10.1207/s15516709cog1401_3

Textor, M. (manuscript). Sensory feeling, self and not-self: Lotze's account.

Weiss, B., & Laties, V. G. ([1961]). Behavioral thermoregulation. *Science*, 133(3461), 1338–1344. https://doi.org/10.1126/science.133.3461.1338

Zollman, K. ([2011]). Separating directives and assertions. *The Journal of Philosophy*, 11, 158–169.