

Luca Barlassina (University of Sheffield)

Valence: A reflection

Forthcoming in: *Emotion Researcher*, ISRE's Sourcebook for Research on Emotion and Affect

“He who laughs last has not yet heard the bad news.”
— Bertolt Brecht

There are many ways to refute a theory. Refutation by joke is the greatest of all. “Two behaviourists have just made love. One asks the other: ‘I know you liked it. Did *I* like it?’” The starting point of this short essay is another attempted refutation-by-joke. It concerns *valence*. What? Very simply, to say that an experience has valence is to say that it feels pleasant/good or unpleasant/bad. A number of experiences have valence: orgasms, happiness, and elation on the pleasant side; and pain, fear, and misery on the unpleasant one. Let’s call them ‘affective experiences.’

It goes without saying that affective experiences can be very different from one another—having a toothache is one thing, being afraid of flying is quite another. Still, some have argued that there is something common to all of them: affective experiences are all *(un)pleasant in the same way* (Bramble 2013). Hence the joke: “If you feel the same pleasure when you smell flowers and when you have an orgasm, either tell me where you buy flowers or please do something about your sex life.”

Did you laugh? If so, the joke might be on you. There are in fact good reasons to think that, despite appearances, valence might be a *natural kind* shared across the affective spectrum. Carruthers (2018) summarises them as follows. (1) Valence is underpinned by a single, domain-general, neurobiological network, which is activated by different stimuli (e.g., sex and money), across different modalities (e.g., vision, touch, and imagination). (2) The same interventions (e.g., paracetamol) modify valence across different types of affective experience—e.g., physical and social pain. (3) If different affective experiences are (un)pleasant in the same way, we can use their valence to choose among them. Valence can thus work as a *common currency* to decide between, say, gustatory and intellectual pleasures.

If you, like me, take the hypothesis that valence is a natural kind seriously, then you, like me, might want to explain what this natural kind *is*. In other words, given that affective experiences are so different from one another, what makes it the case that they all feel (un)pleasant? This is the question I tackle here.

Importantly, this question shouldn't be read *causally*. I am not asking what types of stimulus bring about affective experiences. My concern is a metaphysical one—it has to do with the *very nature* of valence. You are God, and here is an experience that lacks valence. What should you add to it to turn it into an (un)pleasant experience? That's the question, and there is nothing strange about it. If I ask you what water is, I am not asking you what brings about water. I am asking you what water consists of.

Sometimes, similar questions require very different answers. The correct answer to 'What is water?' details its *physical composition*: two atoms of hydrogen and one of oxygen, connected by covalent bonds. Affective experiences are physical too—in my world, there are no gods, ghosts, or immaterial minds. But I am not sure that the answer to 'What is valence?' should be pitched at the physical level. I am old school. Every time I encounter a mental phenomenon, I hope it can be explained at the *cognitive level*—that is, in terms of what information it carries, how this information is encoded and stored, how it interacts with other bits of information, and so on.

Luckily, I am not alone. In recent years, a number of philosophers have attempted to explain valence along these lines. The two most well-known approaches are world-directed evaluativism and world-directed imperativism. I want to convince you that they both fail. Should we then give up our search for a cognitive account of valence? Not at all. I have, in fact, a better alternative. It is called 'reflexive imperativism.'

A final note. The arguments that you see here against world-directed theories and in favour of reflexive imperativism—in fact, the idea of reflexive imperativism itself—have been developed together with my friend and colleague Max Khan Hayward. I honestly don't recall which ideas are mine and which are Max's. My heuristic is this: 'good ideas are mine; mistakes are his.' I have heard Max uttering the same sentence.

World-directed theories: In memoriam

Mental states have intentional, or semantic, content—they carry information; they “say” things. Some of them have *world-directed* content: they carry information about the non-mental world. My visual experience of an oak in the park is a case in point. Other mental states have instead *mind-directed* content. When I believe that you are happy, I am not representing the state of the park. I am representing the state *of your mind*.

World-directed theories propose that: (i) affective experiences have a special type of world-directed content; (ii) this world-directed content makes them (un)pleasant. What could this content be? One can find two main answers in the philosophical literature.

The first one goes under the name of *world-directed evaluativism* (Bain 2013; Carruthers 2018; Cutter and Tye 2011). According to it: (i*) affective experiences represent worldly objects as being *good/bad*; (ii*) it is in virtue of this *world-directed evaluative content* that they are pleasant/unpleasant. A couple of examples should help. Abe is having a backache. His pain experience represents the condition of his back *as bad*. This is why his experience is *unpleasant*. Zoe is eating a cookie. Her gustatory experience represents the cookie *as good*. This is why her experience feels *pleasant*.

Here are two arguments against this proposal.

The sadist and the do-gooder

Dr Do-Good is attending to a patient with a terrible injury on the neck. What a horrible sight! According to world-directed evaluativism, Dr Do-Good’s visual experience feels unpleasant because it represents the patient’s injury as bad. But this cannot be right. Dr Sadist is looking at the injury too, his visual experience also represents the injury as bad ... but it feels *so good*.

The Thinking Otherwise Problem

You have eaten a hot pepper and now your mouth is on fire. Your experience is so unpleasant that you run to the fridge and chug a litre of milk directly from the bottle. The interesting thing is that you do so even though you know full-well that there is no damage in your mouth. But then the unpleasantness of your experience cannot be

identical to the world-directed evaluation *There is a bad damage in my mouth*. If it were, knowing that this evaluation is false should have prevented you from acting like that.

Let's consider whether the second account—*world-directed imperativism* (Martínez 2011)—fares better. Again, the idea is pretty simple: (i**) affective experiences *command* their subjects to do something about some worldly object; (ii**) it is in virtue of this *world-directed imperative content* that they are pleasant/unpleasant. So, Abe's backache is unpleasant because it commands him *Less of this damage in the back!*, while eating a cookie gives Zoe pleasure since her experience has the content *More of this cookie!*

Yet again, the idea won't do:

Hunger

I am hungry. My hunger feels unpleasant. It feels so—world-directed imperativism says—because it commands me: *More food!* Fair enough, it is plausible that my hunger has such a content. But it cannot be the case that this content is what makes my hunger unpleasant. After all, if my hunger has that content, then *all* episodes of hunger have it ... but *not all* episodes of hunger are unpleasant.

A salty story

Bilateral damage to the central nucleus of the amygdala does *not* impact rats' hedonic responses to salt: their facial expressions indicate that they still experience *pleasure* in response to salty stimuli (Galaverna et al. 1993). This—world-directed imperativism says—means that these rats are tokening the command *More salt!* But this is false, since damage to the amygdala *does* abolish rats' motivation to eat salt: they reject solutions containing more than 0.2% NaCl (Flynn et al. 1991). If anything, these rats are tokening the command *Less salt!*

There are in fact two further arguments that show that *any* world-directed treatment of valence is doomed to fail.

World-undirected moods

You wake up one day and you feel miserable. Your misery doesn't command you to act upon the world a certain way, nor does it evaluate the state of world. In fact, it seems

that your experience isn't about the world at all. Still, your experience is very unpleasant. Thus, valence doesn't depend on world-directed content, be it imperative, evaluative, or whatever else.

Mind-directed motivations

You have twisted your right ankle and now you are in pain. The unpleasantness of your experience has *mind-directed motivational force*: it motivates you to get rid of this very experience—you don't want *to feel* like that. But then such an unpleasantness cannot reduce to contents like *There is a bad damage in my right ankle/Less of the damage in my right ankle!* Such contents can only motivate you against the state of your ankle, but they cannot motivate you against your pain experience. No world-directed content can do that.

Becoming reflexive

The attempt to explain valence in terms of world-directed content fails. But that's not the only type of content. Maybe affective experiences feel (un)pleasant in virtue of having a certain type of *mind-directed content*. This is the proposal I endorse. More precisely, I maintain that your affective experience has valence in virtue of *commanding* you to do something about *itself*. In particular, a pleasant experience P has reflexive imperative content *More of P!*, while an unpleasant experience U has reflexive imperative content *Less of U!* Hence the name of the account: *reflexive imperativism*. In a nutshell, it says that an experience feels pleasant/unpleasant in virtue of commanding you *More of me!/Less of me!* (Barlassina *under review*; Barlassina and Hayward 2019; Barlassina and Hayward *forthcoming*).

As always, examples should clarify the idea. Suppose that you are looking at a pen on your desk. Presumably, this visual experience is affectively neutral, that is, it feels neither good nor bad. This doesn't mean that it lacks world-directed content. It in fact represents the pen as being on your desk. What it lacks is reflexive imperative content. And this is why it lacks valence too. If it had reflexive imperative content *More of me!*, it would be a pleasant visual experience. If it had reflexive imperative content *Less of me!*, it would be an unpleasant one. This is exactly the difference between Dr Do-Good and Dr Sadist's visual experiences: even though they have the same world-directed content—they both represent the patient's injury as bad—the former

has reflexive imperative content *Less of me!*, while the latter has reflexive imperative content *More of me!* This is why one feels bad and the other good.

The fact that affective experiences command us to get more/less of themselves explains why they have *mind-directed motivational force*. When you twist your ankle, the unpleasantness of your pain commands you to get less of this experience, and so you don't want to *feel* like that. By the same token, when you eat a cookie, your pleasant experience commands you to have more of itself, and this is why you want to have more of that gustatory pleasure.

Oftentimes, these mind-directed motivations bring about *world-directed motivations*, i.e., motivations to change the state of the non-mental world. As we have just seen, the gustatory experience resulting from eating a cookie feels pleasant in virtue of having reflexive imperative content *More of me!* When this mind-directed command is sent to your decision-making system, the latter attempts to compute the best way to satisfy it. In this case, it is likely to hypothesise that the best course of action is to get another bite from the cookie. The corresponding world-directed motivation is thus produced, and you end up munching the cookie (all else being equal, of course). The "hot pepper scenario" above is another case in point. The sensation you get from eating the hot pepper is unpleasant because it commands you *Less of this sensation!* Upon receiving this mind-directed command as input, your decision-making system outputs the world-directed motivation to drink some milk, and you decide to act accordingly. Your decision is not discounted in the light of your knowledge that your mouth is not damaged, since this decision has nothing to do with how your mouth is faring. You intend to drink milk because you think that it will lessen the unpleasant *sensation* you are having. Your goal is to fulfil the reflexive imperative content *Less of this sensation!*

So far, I described the relation between world-directed motivations and mind-directed motivations as follows: one undergoes a pleasant/unpleasant experience; the experience's valence commands one *More of this experience!//Less of this experience!*, thus motivating one for, or against, this very experience; this experience-directed motivation is then transformed into a world-directed one by the decision-making system. Things, however, are slightly more complex than that, for at least three reasons.

First, on some occasions, the decision-making system doesn't know how to satisfy the mind-directed motivation it receives. You have been feeling miserable for months. You don't want

to feel like that, and unsurprisingly so, given that the unpleasantness of your misery keeps on commanding you *Less of this misery!* You tried a number of behavioural strategies to satisfy this mind-directed command, but none of them worked. Now, your decision-making system has, so to speak, given up. As a result, you still feel unpleasant misery, you don't want to feel like that, but this doesn't result in any world-directed motivation.

Second, the causal chain between mind-directed and world-directed motivations can be interrupted. This, I maintain, is what happened to the aforementioned amygdala-damaged rats. As you might remember, these rats find tasting salt pleasurable. According to reflexive imperativism, this means that they want more of that gustatory experience. This experience-directed motivation, however, doesn't bring about the world-directed motivation to ingest salt. Why so? Here's why: interventions to the central nucleus of the amygdala have set up the rat's decision-making system to constantly output the world-directed motivation not to eat any salt, *regardless of the experience-directed motivation it receives as input.*

Third, and more importantly, you should *not* think that reflexive imperativism claims that world-directed motivations are *typically causally dependent* upon the mind-directed motivations associated to affective experiences. Fair enough, reflexive imperativism has it that affective experiences (better: their reflexive imperative contents) normally bring about world-directed motivations. But this is not to say that the majority of world-directed motivations are brought about by such experience-directed motivations. Quite the contrary, world-directed motivations are by and large causally independent from affective experiences. (Compare with this: skiing normally makes one thirsty, but the great majority of thirst episodes are not due to skiing). In this moment, for example, I am typing these words on my laptop, so my brain is issuing a number of motor commands. These world-directed commands and their world-directed motivational force, however, have nothing to do with my current affective experience. The same applies to hunger. As said above, hunger always commands one *More food!*, but this world-directed command doesn't have to be generated by an affective experience. Some episodes of hunger feel neither pleasant nor unpleasant—but they have world-directed motivational force nonetheless.

This last point, of course, raises a very important question. If we are cognitive agents with a number of valence-independent, world-directed motivations, *what is the point of having valence-based, mind-directed motivations on the top of them?* In Barlassina and Hayward (2019, p. 1039) we

formulated the problem as follows: “what is the evolutionary advantage of [valence]? [...] Why do we need experiences that feel good or bad? Couldn’t nature just have endowed us with [world-directed] representations, desires, and affectless commands like the urge to defecate? Why did we need to *suffer* [or to feel *pleasure*]?” One year after, I am still puzzled by this question. So, I find myself in the following situation: I am pretty sure *that* affective experiences have valence in virtue of having reflexive imperative content *More of me!/Less of me!*, but I don’t know *why* that is the case. So, the joke might be on me after all.¹

References

- Bain, D. (2013). “What makes pains unpleasant?” *Philosophical Studies*, 166, 69-89.
- Barlassina, L. (under review). “Beyond good and bad: Reflexive imperativism, not evaluativism, explains valence”.
- Barlassina, L., & Hayward, M. K., (forthcoming). “Loopy regulations: The motivational profile of affective phenomenology”, *Philosophical Topics*.
- Barlassina, L., & Hayward, M. K., (2019). “More of me! Less of me! Reflexive imperativism about affective phenomenal character”, *Mind*, 128(512): 1013-1044.
- Bramble, B. (2013). “The distinctive feeling theory of pleasure”, *Philosophical Studies*, 62(2): 201-217.
- Carruthers, P. (2018). “Valence and value”, *Philosophy and Phenomenological Research*, 97(3): 658-680.
- Cutter, B., & Tye, M. (2011). “Tracking representationalism and the painfulness of pain”, *Philosophical Issues*, 21: 90-109.
- Flynn, F. W., Grill, H. J., Schulkin, J., & Norgren, R. (1991). “Central gustatory lesions: II. Effects on sodium appetite, taste aversion learning, and feeding behaviors”, *Behavioral Neuroscience*, 105(6), 944–954.

¹ I would like to thank Davide Bordini and James Turner for their suggestions.

Galaverna, O. G., Seeley, R. J., Berridge, K. C., Grill, H. J., Epstein, A. N., Schulkin, J. (1993).

“Lesions of the central nucleus of the amygdala I: Effects on taste reactivity, taste aversion learning and sodium appetite”, *Behavioural Brain Research* 59(1-2): 11-17.

Martínez, M. (2011). “Imperative content and the painfulness of pain”, *Phenomenology and the*

Cognitive Sciences, 10 (1):67-90, 2011.