PHILOSOPHICAL TOPICS
VOL. 47, NO. 2, FALL 2019

Loopy Regulations: The Motivational Profile of Affective Phenomenology

Luca Barlassina
University of Sheffield

Max Khan Hayward
University of Sheffield

ABSTRACT. Affective experiences such as pains, pleasures, and emotions have affective phenomenology: they feel (un)pleasant. This type of phenomenology has a *loopy regulatory profile*: it often motivates us to act a certain way, and these actions typically end up regulating our affective experiences back. For example, the pleasure you get by tasting your morning coffee motivates you to drink more of it, and this in turn results in you obtaining another pleasant gustatory experience. In this article, we argue that reflexive imperativism is the only intentionalist account of affective phenomenology—probably, the only account at all—that is able to make sense of its loopy regulatory profile.

1. BENTHAM'S DICTUM

In perhaps the most famous passage ever written by a philosopher on the nature of pleasant and unpleasant experiences, Jeremy Bentham claimed that:

> Nature has placed mankind under the governance of two sovereign masters, pain, and pleasure. It is for them alone to point out what we ought to do, as well as to determine what we shall do. (Bentham 1789/1970, 11)

Bentham's dictum is confronting, because it mixes an observation that is obviously correct with claims that are contestable at best. For it is undeniable that we undergo pleasant and unpleasant experiences and that their (un)pleasantness plays a role in governing our behavior. As we might say in more contemporary terminology, it plays a *regulatory* role. But it is dubious that it does anything so clean-cut as *determining* our behavior; more dubious still is the normative claim that pleasure and displeasure reliably indicate or ground facts about what we *ought to do*.

In fact, even the uncontroversial bit of Bentham's dictum raises difficult questions. Why *are* there such experiences? And *how* do they regulate what we do? These are the questions that will keep us busy in this article. So, it'd be better getting clear on them. Some examples will help.

1.1 THREE QUESTIONS ABOUT PLEASURE AND DISPLEASURE

Abe is eating a brownie. His gustatory experience feels pleasant. He thus keeps eating. And so he gets more gustatory pleasure. Zoe sprained her right ankle. Her experience feels unpleasant. So she takes a painkiller. Her unpleasant pain goes away.

These trivial stories highlight *three important psychological facts*. First, there are experiences that feel *good* or *pleasant*, as well as experiences that feel *bad* or *unpleasant*. Sensory pleasures and pains are cases in point. But the class of these experiences extends far beyond sensory pleasures and pains. It's easy to find examples: joy, happiness, elation on the pleasant side, and nausea, sadness, misery on the unpleasant one. Let us call these (un)pleasant experiences 'affective experiences' and call an experience's *feeling* pleasant/good or unpleasant/bad 'affective phenomenology'. Affective experiences thus contrast with affectively neutral experiences, which feel neither pleasant nor unpleasant. (Here is an instance of the latter category: the visual experience as of a pen next to your laptop. It does not feel good or bad. At least, *our* visual experience does not feel good or bad. You might be different. In that case, think of a different experience which does not feel good or bad. *That* is an affectively neutral experience.) Hence our *first question*: in virtue of what do some experiences have affective phenomenology? Or, which amounts to the same thing, what makes it the case that some experiences, but not others, feel pleasant or unpleasant?

In the grip of affective experiences, we typically *do* things. Abe kept eating his brownie; Zoe took a painkiller. It is not accidental that Abe and Zoe did this. It is exactly *because* their experiences had affective phenomenology—because they felt good or bad—that they were motivated to act like that. If eating the brownie hadn't been pleasant, Abe wouldn't have kept eating it (other things being equal). If the pain in Zoe's ankle hadn't been unpleasant, she would have probably left the painkillers in her purse. The affective phenomenology of our experiences has a

role in *regulating* our actions. This second psychological fact gives rise to our *second question*: in virtue of what does affective phenomenology motivate us to act a certain way?

And here is the third fact. The actions we perform guided by the (un)pleasantness of our experiences typically end up *affecting* these very affective experiences. Motivated by the pleasantness of his gustatory experience, Abe kept eating the brownie. Result: more gustatory pleasure. Motivated by the unpleasantness of her pain, Zoe took a painkiller. Result: less unpleasant pain. Just as much as our affective phenomenology plays a role in regulating our actions, so the actions that we are motivated to perform under the sway of our affective phenomenology *in turn* regulate our affective experiences. Affective phenomenology has a *loopy regulatory profile*: we undergo affective experiences; their affective phenomenology motivates us to perform actions; these actions either inhibit (for unpleasant experiences) or promote (for pleasant experiences) further affective experiences. What explains this loop? This is our third and final question.

So, here are our three questions in a nutshell:

(1) In virtue of what do affective experiences have affective phenomenology?—I.e., in virtue of what do they feel pleasant/good or unpleasant/bad?

(2) In virtue of what does affective phenomenology regulate us?—I.e., in virtue of what does it typically motivate us to act a certain way?

(3) In virtue of what does affective phenomenology have a loopy regulatory profile?—I.e., in virtue of what do the very actions motivated by the affective phenomenology of our experiences typically end up regulating our affective experiences back?

Any adequate theory of affective phenomenology should be able to answer (1)–(3). But, in this article, we focus our attention on a particular family of theories of affective phenomenology: *intentionalist theories*. You might want to know the reason for this focus. That's a fair request and we shall accommodate this shortly. Before doing that, however, a clarification is in order.

1.2 MIND VS. NORMATIVITY

Bentham, although not always given to philosophical subtlety, is careful to distinguish, in his *dictum*, between a descriptive claim ("Pain and pleasure determine what we *shall* do") and a normative one ("They determine what we *ought* to do"). Surprisingly enough, the contemporary literature on affective phenomenology often fails to draw this basic distinction. People start by saying that affective experiences *feel* good or bad and move seamlessly to the assertion that such experiences *are* good or bad for us. For example, David Bain claims that any theory of pain's unpleasantness must respect the following *Normative Condition*: "being in unpleasant pain could consist in being in state φ only if being in state φ is, in the relevant cases, noninstrumentally bad for its subject" (Bain 2019, 463).

We take it that the badness that Bain has in mind is *prudential* badness. Now, it *might* be true that unpleasant experiences are always noninstrumentally, prudentially bad for their subjects and that, *mutatis mutandis*, pleasant experiences are always noninstrumentally, prudentially good for their subjects. But this is a contested claim. Nietzsche (see Anderson 2017), as well as certain religious thinkers, appear to deny that unpleasantness is always noninstrumentally bad for us; and Moore (1903/1922, 209–10) argued that cruel or ugly pleasures are not even prudentially good. Be that as it may, the point we want to make here is that it is a job for ethics, not the philosophy of mind, to make any determination about the prudential value of pleasure and displeasure. And the present article is an article in the *philosophy of mind*. Thus, question (1) should be read as asking why affective experiences *feel* good/bad, not why (or whether) they *are* good/bad.

Similarly, you can find philosophers who, exactly like us, want to explain *how* affective phenomenology achieves the regulation of our actions—how it *motivates* us to act a certain way—but then end up asking why it *gives us reasons* to do this. Of course, if by 'reasons' one just means 'motivating reasons', then we have no problem with this. Indeed, there is an even further sense of 'reasons' such that one can interpret our question (2) in terms of reasons. Not only do pleasant and unpleasant experiences motivate action. It is also the case that the actions performed in the light of affective phenomenology seem to "make sense," or be intelligible, and so affective phenomenology may be said to provide reasons in a broadly Davidsonian sense (Davidson 2001). For example, it is entirely unsurprising that Abe kept on eating his brownie *since that gave him pleasure*.

But to say that actions performed in the light of pleasantness or unpleasantness are *rational*, in the sense of *being intelligible*, is not the same as to say that they are (pro tanto) *rationally required*, in the sense that we *ought* to perform them. Accordingly, we maintain that a theory of affective phenomenology *in the philosophy of mind* needs to explain how (un)pleasantness motivates behavior, and, moreover, why it is that these motivations appear to make sense; but we deny that such a theory needs to stray into the realm of full-blooded normativity. In this sense, question (2) asks why affective phenomenology provides *intelligible*, *motivating reasons*, not why it provides *normative reasons* (if it does at all).

1.3 UNINTELLIGIBLE DESIRES

'Intentionalism' (Byrne 2001; Dretske 1995; Tye 1995a) is the name for a family of theories that attempt to explain the phenomenology of an experience (including its affective phenomenology) in terms of the experience's intentional content—more on this later. This article attempts to establish which version of intentionalism (if any) best answers (1)–(3). "Why do you focus on intentionalism?" you asked us before. We can now answer your question.

The most prominent alternative to intentionalism is the desire theory (Heathwood 2007). On the face of it, this theory seems to offer an elegant and unified answer to (1)–(3). Why does Abe's gustatory experience feel pleasant? Because

Abe is desiring to have such an experience while he is having it. And why does such a pleasant experience motivate Abe to eat more of the brownie? Because Abe thinks that, by acting in that way, he will get more of the experience that he desires. And why will such an action end up generating more pleasure in Abe? Because Abe will then desire the experience he is having.

Nice story, isn't it? Not really, we say. As we stated above, even though a theory of affective phenomenology in the philosophy of mind is not required to explain why (or whether) affective phenomenology gives us *normative* reasons, it must explain why such a phenomenology gives us *intelligible*, *motivating* reasons. In this regard, the desire theory is a nonstarter. Why does Abe desire the gustatory experience he is having? The desire theorist cannot say that he desires it because it feels pleasant, since she claims that the reverse is true: it feels pleasant because Abe desires it. So, *why* does he desire it? The desire theory is in principle incapable of answering this question, since it takes this desire to be primitive: Abe doesn't desire his gustatory experience because it has this or that intrinsic feature; it *just so happens* that Abe desires it. End of story.

We find this account unsatisfactory. It doesn't seem true that Abe *merely happens* to desire his gustatory experience, or that Zoe *merely happens* to dislike her pain sensation—that they might just as easily have ended up reversing their attitudes. At least at the first-personal level, it appears that we have pro-/con- attitudes toward our experiences *because they feel pleasant/unpleasant*. But of course, we do not want to leave the story here either—our goal is to explain subjective experience in terms of something else. So that something else must be something which can also stand in this sort of rationalizing relationship to our attitudes.

The most likely candidate—maybe the *only* candidate—is affective experiences' intentional content. What else could in fact "rationalize" our response to such experiences? It might well be that affective experiences share certain neural properties, but it is hard to see how these properties can feature into a reason-giving explanation; such properties can cause our behaviors and attitudes, but they do not seem able to *make sense of* them. Other philosophers appeal to ineffable qualia (Bramble 2013). But it is unclear how qualia can play a reason-giving role either. To see why this is the case, it is helpful to consider perceptual cases. In response to the question "Why did you act as though that stick was straight?", it seems inapt to respond: "Because I was in brain state C", or: "Because I was experiencing quale No5." We make ourselves intelligible to our questioner only by responding: "Because it looked straight." This is because the explanation called for occurs at the *semantic-intentional level* (Fodor 1987).

We think the same considerations apply to affective states too. This is why, in this article, we investigate whether intentionalism can offer an adequate account of affective phenomenology. If intentionalism fails, it becomes unclear whether the role of affective phenomenology in our cognitive economy could be made *intelligible* at all.

2. INTENTIONALISM ABOUT AFFECTIVE PHENOMENOLOGY

Unless you have spent too much time in Oxford, you probably believe that many experiences have intentional content. For example, a visual experience as of a red car has an intentional content roughly like this: *There is a red car in front of me*. And unless you are Keith Frankish, you also believe that experiences have phenomenology—there is something it is like to have them.[1] For example, there is something it is like for you to have a visual experience as of a red car, and this something-it-is-like differs from the phenomenology of, say, a visual experience as of a blue car.

What's the relation between the phenomenology and the intentional content of an experience? Intentionalism says that the former depends on, or even reduces to, the latter. So, the phenomenology of your 'red car experience' differs from the phenomenology of your 'blue car experience' because these two experiences have different intentional contents.

In the last three decades, intentionalism has proven immensely popular among philosophers of mind. One reason for its success is that, as we said in the previous section, intentionalism promises to "rationalize" the role of phenomenology in cognition. Why did you form a belief with the content *There is a red car in front of me* in response to a visual experience with such-and-such a phenomenology? Because this phenomenology was nothing over-and-above the content *There is a car in front of me*. In addition, intentionalism paves the way to solving the hard-problem of consciousness (Chalmers 1995). It is not easy to find a place for phenomenology in the natural world. But if intentionalism is right and phenomenology is grounded in intentional content, then one only needs to naturalize the latter to naturalize the former.

But there's a catch. While intentionalists tend to agree on how an account of visual, auditory, or bodily phenomenology should work, disagreement about the proper treatment of affective phenomenology looms large. What is it that intentionalists disagree about?

2.1 EVALUATIVISM VS. IMPERATIVISM

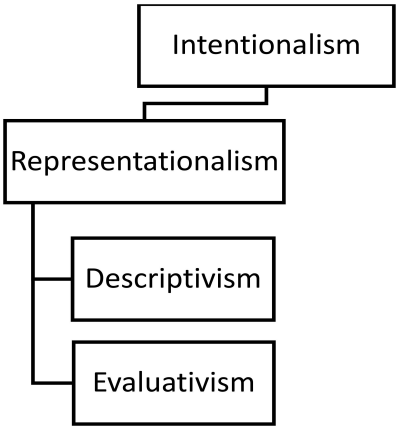
The standard story sees the greatest dividing line in the intentionalist camp as that which separates *evaluativism* from *imperativism*. Needless to say, both theories claim that experiences have affective phenomenology in virtue of their intentional content. Their disagreement concerns *what* intentional content grounds affective phenomenology. To understand the nature of this disagreement and how it came about, we need to go back to the oldest intentionalist account of affective phenomenology, namely Tye's *original* theory of the phenomenology of pain, which was *neither* evaluativist *nor* imperativist.

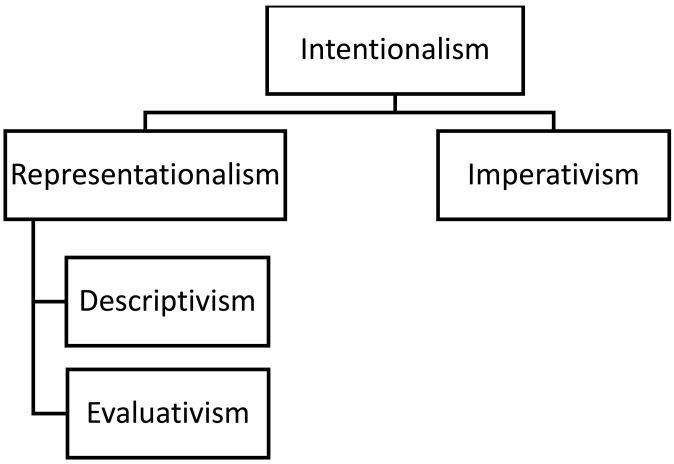
1. If you are Dan Dennett, you don't believe either of these claims. This is because you spent too much time in Oxford *and* with Keith Frankish.

According to Tye (1995b), pain feels the way it does in virtue of representing the presence of damage in one's body. In other words, the phenomenology of pain depends on pain's representational-descriptive content. This content is *representational* because it represents one's body as being a certain way; and it is *descriptive* because it doesn't evaluate such a bodily condition: it merely describes it. However, it soon became clear that this wasn't an account of the affective phenomenology of pain at all. At best, descriptive representationalism can explain why pain has *sensory* phenomenology, i.e., why, when in pain, you feel certain sensations as localized in your body. But why should your pain feel *unpleasant* simply in virtue of describing that your body is in such-and-such a state? (Aydede 2006).

Faced with this problem, intentionalists parted ways. *Evaluativists* (Bain 2013, 2019; Carruthers 2018) retained the representationalist component of Tye's proposal: experiences have affective phenomenology in virtue of representing things as being a certain way. But this representational content is *evaluative* rather than descriptive: it doesn't merely describe how things are; it evaluates them as *good* or *bad*. Accordingly, your backache doesn't feel *unpleasant* in virtue of *describing* that your back is in such-and-such a condition. Rather, it feels so because it *evaluates* such a condition *as bad*. The point generalizes. Why does your joy for having won the lottery feel *pleasant*? Because it evaluates your winning the lottery as *good*.

Imperativists (Barlassina and Hayward 2019; Klein 2015; Martínez 2011) adopted a different strategy. Representational content (whether descriptive or evaluative) is not the only type of intentional content. There is also a type of content that commands rather than represents: *imperative content*. As its name suggests, imperativism maintains that it is this nonrepresentational content that grounds the affective phenomenology of an experience. For example, Martínez (2011) proposes that your backache feels *unpleasant* because it commands you: *Get less of this bodily damage!*. By the same token, your joy for having won the lottery feels *pleasant* in virtue of commanding you: *Get more of these victories!*. We can also put





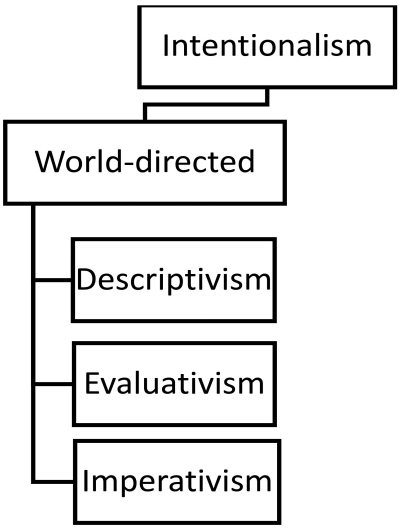
it like this: for imperativism, affective experiences feel (un)pleasant not because they tell us how things *are*, but because they tell us to *do* something.

2.2 WORLD-DIRECTED VS. EXPERIENCE-DIRECTED INTENTIONALISM

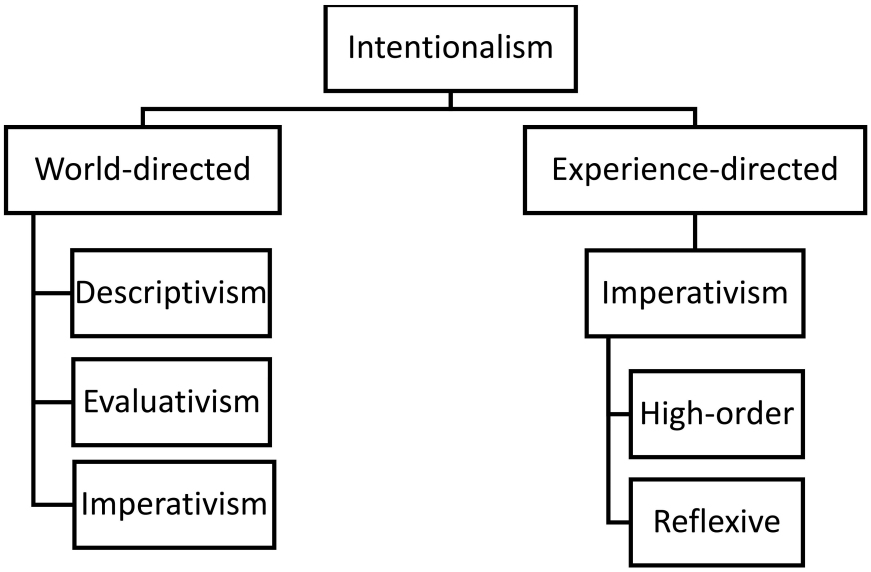
So far, we have presented the debate among intentionalists in terms of the disagreement between evaluativism and imperativism. But there is another way to slice the intentionalist pie. Tye's original descriptive representationalism is *world-directed*, or *first-order*: it maintains that the phenomenology of pain depends on pain representing the state of the *nonmental world*, in particular the condition of your body, as being such-and-such. Here again, evaluativists followed suit: affective experiences feel pleasant/unpleasant because they represent certain *nonmental*, *worldly objects* as being good/bad (Bain 2013; Carruthers 2018). Your backache feels bad because it represents the *condition of your back* as bad. Many imperativists would agree with this. Martínez is a case in point. As we have seen, he thinks that your backache feels *unpleasant* in virtue of commanding you: *Get less of this bodily condition!*.

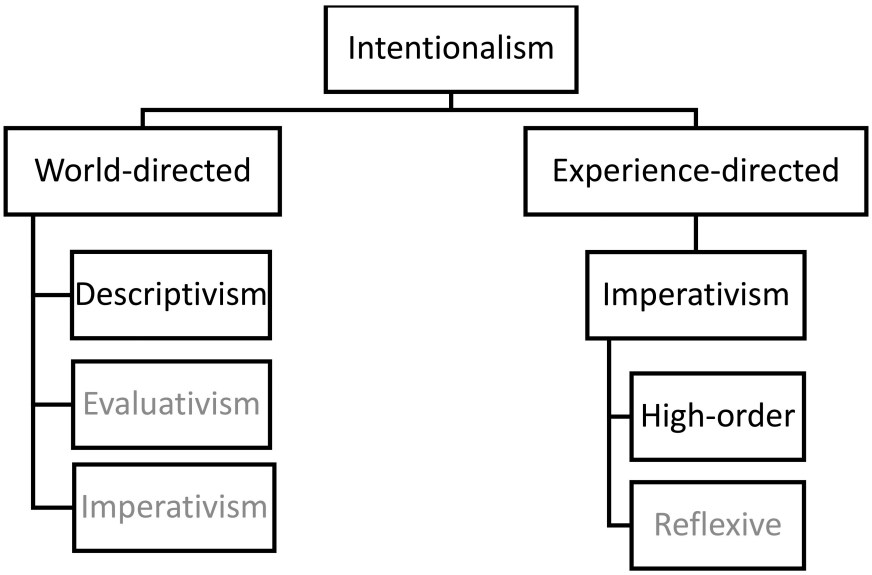
At least in principle, however, nothing prevents intentionalists from adopting an *experience-directed* view—that is, to propose that experiences have affective phenomenology in virtue of being directed at an *experience*. As far as we know, no evaluativist has ever taken this route.[2] But imperativists have. Klein (2015) sketched a form of *higher-order* imperativism; we developed a form of *reflexive*, or *same-order*, imperativism (Barlassina and Hayward 2019). The difference is as follows: for Klein, an experience E feels pleasant/unpleasant in virtue of command-

2. Bain (2013, 79–80) characterizes *dislike* theories of pain (according to which pains are sensations that we dislike) as experience-directed evaluativist views. We think these are much better understood as experience-directed *desire* views, and that they should be rejected for the reasons that we gave above for rejecting desire theories in general.



ing you: *Get more/less of H!*—where H is an experience numerically distinct from E. For reflexive imperativism, an experience instead feels pleasant, or unpleasant, in virtue of the fact that it commands its subject to get more, or less, *of itself*. In other words, affective phenomenology is grounded in *reflexive commands*: an experience E feels pleasant/unpleasant in virtue of commanding you: *Get more/less of E!*. For example, your backache feels bad because it has reflexive imperative content *Get less of this backache!*.





Accordingly, the intentionalist terrain can also be mapped as follows: *world-directed* theories, on the one hand, versus *experience-directed* theories, on the other. In this article, we discuss which of these two camps provides the best account of affective phenomenology. It goes without saying that we cannot discuss *all forms* of world-directed and experience-directed intentionalism about affective phenomenology. We shall rather compare and contrast two forms of world-directed intentionalism—i.e., world-directed evaluativism and world-directed imperativism—with one form of experience-directed intentionalism, namely, reflexive imperativism.

Our choice is well motivated. First, nobody endorses world-directed descriptive representationalism about affective phenomenology anymore—even Tye is now a world-directed evaluativist (Cutter and Tye 2011). In addition to world-directed evaluativism, the other main world-directed theory of affective phenomenology is world-directed imperativism. Hence our picks. As to experience-directed theories, we saw that there are two forms of experience-directed imperativism: higher-order imperativism and reflexive imperativism. We have argued elsewhere (Barlassina and Hayward 2019) that reflexive imperativism should be preferred to higher-order imperativism. This is why we focus on the former—well, the fact that it is *our* theory also played a role in *our* choice.

We have stated our aim already: we want to know whether there is a form of intentionalism that can provide an adequate theory of affective phenomenology, i.e., a theory that can satisfactorily answer questions (1)–(3). Now you also know how we are going to discover this: we will pit two varieties of world-directed intentionalism against one version of experience-directed intentionalism. You can anticipate our verdict: neither world-directed evaluativism nor world-directed

imperativism is up to the task. By contrast, reflexive imperativism offers a unified and principled account of (1)–(3). The only thing left us to do is to convince you that this is indeed the case.

3. THE WORLD IS NOT ENOUGH

In this section, we show that neither world-directed evaluativism (hereafter, WDE) nor world-directed imperativism (hereafter, WDI) offers convincing answers to (1)–(3). Since WDE and WDI are by far the best versions of world-directed intentionalism yet proposed, this suggests that any theory of affective phenomenology couched in terms of world-directed intentional content is unlikely to succeed.

3.1 WHY AFFECTIVE PHENOMENOLOGY?

Question (1) asks in virtue of what affective experiences have affective phenomenology. Intentionalists of any stripe will give an answer *of the form*: 'Experience E is pleasant/unpleasant in virtue of having intentional content *C*+/*C*-'. Depending on what one takes '*C*+' and '*C*-' to stand for, one will obtain different, *substantive* intentionalist theories of affective phenomenology.

How could one evaluate these theories? At the very least, they should be *extensionally adequate*: all pleasant/unpleasant experiences should have intentional content *C*+/*C*-; and all experiences with intentional content *C*+/*C*- should be pleasant/unpleasant. In what follows, we argue that neither WDE nor WDI satisfies this minimal requirement.

3.1.1 Against necessity

As you probably recall, WDE gives the following answer to (1): experience E is pleasant/unpleasant in virtue of representing some worldly object as *good/bad*. For example, Abe's gustatory experience feels pleasant in virtue of possessing world-directed evaluative content: *This brownie is good*. In the case of WDI, the idea is instead this: experience E is pleasant/unpleasant in virtue of commanding its subject to *get more/less* of some worldly object. Abe's gustatory experience, e.g., feels pleasant because it has world-directed imperative content: *Get more of that brownie!*.

There is a simple argument against the extensional adequacy of both WDE and WDI—in fact, against *any* form of world-directed intentionalism. Some experiences of *moods* do not appear to have *any* world-directed (i.e., first-order) content *at all*: they are *world-undirected* (Mendelovici 2013). But they do have affective phenomenology. Thus, no world-directed content—be it imperative, evaluative, or whatever you want—can be *necessary* for affective phenomenology. Take misery as an example (elation would do as well). You wake up one day and you feel blue. This is clearly an experience with affective phenomenology: it feels terribly unpleasant. However, there is no worldly object that your experience evaluates as bad. And

there is no worldly object that your experience commands you to act upon. This is because your experience isn't directed at any worldly object at all.[3]

This is instructive. World-directed intentionalism about affective phenomenology has been mainly developed to account for the unpleasantness of physical pain, with the affective phenomenology of other experiences left as an afterthought, on the assumption that a theory of pain's unpleasantness can be readily extended to accommodate them. Now, physical pain *does* typically have a worldly object—one's own body. But as soon as one tries to extend world-directed intentionalism to cases like world-undirected moods, it becomes clear that it lacks the resources to account for affective phenomenology across the affective spectrum.

3.1.2 Against sufficiency

So, some experiences have affective phenomenology but lack imperative, or evaluative, world-directed content. Conversely, it is also the case that some experiences *have* imperative, or evaluative, world-directed content, but lack affective phenomenology. Abe is hungry. His hunger feels unpleasant. It feels so—WDI says—because it commands Abe: *Get some food!*. Now, it is indeed plausible that Abe's hunger has such a world-directed imperative content. In fact, it is plausible that *all* episodes of hunger have such a content. However, *not all* episodes of hunger feel unpleasant. Hence, world-directed imperative content is *not sufficient* for affective phenomenology.

The same issue carries over to WDE. Dr. Expert and Dr. Novice are looking at a patient's severe injury. Dr. Novice's visual experience feels unpleasant. According to WDE, it feels so in virtue of possessing world-directed evaluative content: *That injury is bad*. If Dr. Novice's visual experience represents the patient's injury as bad, so does Dr. Expert's. But Dr. Expert is so used to these scenes that her visual experience doesn't feel unpleasant. For her, it is just another day at work. Conclusion: it is false that if an experience has world-directed evaluative content, then it has affective phenomenology.[4]

3.2 MOTIVATIONAL PROBLEMS

Recall our friend Abe: he was eating a brownie; his gustatory experience felt pleasant; this *motivated* him to get another bite; and so did he. Unfortunately, he is not always so lucky. The other day he opened a pack of shrimp. They were expired and smelled terrible. His disgust felt awfully unpleasant. This *motivated* him to throw the shrimp in the garbage. That's exactly what he did.

3. Just to be clear, we are *not* saying that all moods are world-undirected. In fact, we don't even think that all instances of misery/elation are that way. Our point is simply that *some* mood experiences—e.g., *some instances of misery*—lack world-directed content. This suffices to show that world-directed intentional content is *not necessary* for affective phenomenology.

4. Barlassina (under review) discusses a number of cases in which world-directed evaluative content and affective phenomenology come apart.

Question (2) asks in virtue of what affective phenomenology motivates us to act in these ways. WDI answers as follows. Affective phenomenology is grounded in world-directed imperative contents. These contents are *intrinsically motivational*: on their own, they motivate us to change the state of *the nonmental world*—they thus count as *world-directed motivations*. But why did Abe act in these *particular* ways? He ate more of the brownie because the pleasantness of his experience reduced to the content *Get more of this brownie!*—a world-directed *appetitive* motivation. And he threw the shrimps away because the unpleasantness of his disgust consisted in the content *Get less of these shrimps!*—this time, a world-directed *aversive* motivation.

WDE tells a similar story. Affective phenomenology is grounded in world-directed *evaluative* contents, and these contents are also intrinsically motivational. More precisely, pleasantness obtains in virtue of contents of the form *O is good!*, thus it has world-directed, *appetitive* motivational force; unpleasantness is instead grounded in contents of the form *O is bad!*, thus it has world-directed, *aversive* motivational force.

We don't find these stories the least convincing. In section 3.2.1, we raise a serious problem for the picture of motivation emerging from WDE; in 3.2.2, we show that WDI is immune to it; in 3.2.3, however, we argue that *both* WDE and WDI face a second, insurmountable difficulty.

3.2.1 The Thinking Otherwise Problem

Let us say from the outset that we are somewhat skeptical of the claim made by WDE that evaluative contents are *intrinsically* motivational. This is because we subscribe to the Humean thesis that no purely representational content—be it descriptive or, as in the case of WDE, evaluative—can motivate on its own. However, evaluativists are well aware of this worry (Bain 2013) and, at least in this article, we are prepared to grant them this point. But this concession is not going to help WDE much. In fact, it raises *two* other issues. In this section, we discuss the *first* one, which we call the *Thinking Otherwise Problem*.

Suppose that you have accidentally eaten one extremely hot habanero pepper—you mistook it for a boring sweet bell pepper—and now your mouth is on fire. So, you run to the fridge, pick a bottle of milk, and drink from it like there is no tomorrow. How does WDE explain your behavior? Very simply, your unpleasant sensation had world-directed evaluative content *There is a bad damage in my mouth* and this evaluation motivated you to fix this (represented) damage. But there's a complication. You are perfectly aware that the burning sensation resulting from eating hot food doesn't correlate with any bodily damage. So, why did you chug all that milk if you knew full well that the evaluation produced by your sensory system was *false*? (Of course, the problem need not be formulated in terms of *knowledge*. It is enough to imagine a case in which one has a belief—even a *false belief*—inconsistent with one's "affective evaluation.")

Bain (2013, 84) considers a similar objection and responds by invoking the *cognitive impenetrability* of experiences. Just as a stick in water continues to *visually*

appear bent even when the subject knows that it is straight, so the 'habanero pepper sensation' continues to represent badness even when the subject knows that nothing bad is occurring. This response, however, won't do. The Thinking Otherwise Problem doesn't in fact concern why your sensory system continued to generate the evaluative content *There is a bad damage in my mouth*. Bain is right: the impenetrability of your sensory system can easily explain this. The problem is why you—better: your decision-making system (hereafter, DMS)—kept on taking this evaluative content at face value and *acted* upon it, *irrespective of the fact that you knew it to be false*. Let us explain.

Your DMS is *not* cognitively impenetrable. Clearly, it can take the belief/knowledge that there is no damage in your mouth as input (this follows from the general thesis that your DMS can take *any* propositional attitude as input—it is an *isotropic* cognitive system, if there is one [Fodor 1983]). But then, in the light of this information, your DMS should have discounted the evaluation produced by your sensory system. Hence, the unpleasantness of your sensation shouldn't have resulted in any action. But it did result in action! This indicates that, *pace* WDE, it is not the evaluation *There is a bad damage in my mouth* that explains the motivation created by your unpleasant experience. If it did, your knowledge that there is no damage in your mouth should have prevented your unpleasant experience from steering your decision-making process.

We can also put it like this. Your chugging the milk is explained by the unpleasantness of your experience. However, it could *not* be explained by a world-directed evaluation. Hence, it seems that WDE is false: affective phenomenology and world-directed evaluations are not one and the same thing.

3.2.2 One cheer for world-directed imperativism

WDI doesn't face the Thinking Otherwise Problem. According to WDI, the unpleasantness of your 'habanero pepper sensation' depends upon the world-directed imperative content *Change the state of my mouth!*. In contrast to the evaluative content *There is a bad damage in my mouth*, this imperative content doesn't represent your mouth as being badly damaged. Therefore, it is consistent with the belief that there is no damage in your mouth and cannot be countered by it. This is why your decision-making system (DMS) output the intention to drink milk even though it received the 'no-damage belief' as input: your DMS simply didn't find any conflict between the content of this belief and the imperative content *Change the state of my mouth!*.

Moreover, even if there were some sort of friction between contents such as *Change the state of my mouth!* and *There is no damage in my mouth*, WDI explains how such a conflict (if it exists!) can be won by the imperative. The reason is simple. According to WDI, only imperative contents are intrinsically motivational; representational contents (either descriptive or evaluative) aren't. Thus, when the DMS has to decide whether to "obey" the imperative content *Change the state of my mouth!* or the content of the 'no-damage belief', there is no choice at all to be made:

only the imperative content exerts a motivational pull over the DMS. This is why you are likely to swallow a liter of milk directly from the bottle *no matter what your beliefs about your mouth's condition are*.

3.2.3 Hedo-motivational inversions

WDI has an advantage over WDE: it escapes the Thinking Otherwise Problem. But that's a trifling victory, since both theories face the same devastating objection.

You are drinking instant coffee. It tastes awful. According to WDI, the *unpleasantness* of your experience is grounded in the world-directed imperative content *Get less of this coffee!* and must then motivate you to stop drinking coffee—a world-directed *aversive* motivation. By the same token, had your experience been *pleasant*, it would have had content *Get more of this coffee!* and should thus have motivated you to keep on drinking coffee—this time, a world-directed *appetitive* motivation.

As we know, the picture emerging from WDE is basically the same. Pleasant experiences have world-directed evaluative content *This object is good*; such a content has intrinsic, world-directed, *appetitive* motivational force; hence, pleasant experiences cannot but motivate you to get more of their worldly objects. Unpleasant experiences have instead world-directed evaluative content *This object is bad*; such a content has intrinsic, world-directed, *aversive* motivational force; hence, unpleasant experiences cannot but motivate you to get less of their worldly objects.

Now, we are happy to grant DWI and DWE that there is a *reliable* connection between pleasantness/unpleasantness and world-directed appetitive/aversive motivations. However, this connection is not as intimate as these two theories predict. On many occasions, the connection breaks down and one either has (i) a *pleasant* experience accompanied by a world-directed *aversive* motivation for the experience's worldly object, or (ii) an *unpleasant* experience accompanied by a world-directed *appetitive* motivation for the experience's worldly object. How do we know that such *hedo-motivational inversions* occur?[5]

A RAT'S TALE

A first source of evidence comes from Kent Berridge and colleagues' celebrated work on the neuroscience of *liking* (i.e., affective phenomenology) vs. *wanting* (i.e., world-directed motivation)—Berridge (1996) is perhaps the *locus classicus*. Even

5. Please notice that even weaker dissociations would suffice to put these two theories in a lot of trouble: cases in which an intervention that increases/decreases the *level of pleasantness or unpleasantness* of an experience does *not* increase/decrease the *intensity of the experience's world-directed appetitive or aversive motivational force* (or vice versa). Even though one can find plenty of these weak dissociations in the neuroscientific literature (see Berridge et al. 2009 for a review), we shall not discuss them here, because they would force us to introduce a further level of complexity in our analysis of world-directed intentionalism, namely, the idea that both the level of (un)pleasantness of an affective experience E and the intensity of E's world-directed motivational force are grounded in the *strength* of E's world-directed imperative/evaluative content.

a couple of Berridge et al.'s findings should be enough to drive world-directed intentionalists to despair.

Facial expressions are a well-established measure of taste-elicited (un)pleasantness in many species (including, of course, human beings) (Berridge 2000). For example, rats respond to unpleasant stimuli with gapes and head shakes, and to pleasant ones with tongue protrusions. World-directed motivation can instead be assessed through the individual's object-oriented behavior: appetitive *behaviors* toward object O (e.g., approaching, ingesting, licking O) indicate appetitive *motivations* for O; O-aversive behaviors indicate aversive motivations.

Bilateral damage to the central nucleus of the amygdala abolishes salt intake in rats: even though these rats display normal water and food consumption, they reject solutions containing more than 0.2 percent NaCl, thus exhibiting a clear *aversive motivation* for salt (Flynn et al. 1991). However, their facial expressions strongly indicate that they experience *pleasure* in response to salty stimuli (Galaverna et al. 1993). This in an instance of hedo-motivational inversion (i): a *pleasant* experience accompanied by a world-directed *aversive* motivation for the experience's worldly object.

And here is an example of hedo-motivational inversion (ii)—(Berridge and Valenstein 1991). Electrical stimulation of the lateral hypothalamus (ESLH) considerably increases feeding behavior in rats and other animals, thus indicating the presence of a strong *appetitive* motivation for food. However, ESLH doesn't parallelly potentiate pleasantness in response to food. Quite the contrary, ESLH-bound rats eat increased quantities of food, even though they find it *unpleasant*.

JUNK FOOD

But one doesn't need to go into the lab to find examples of hedo-motivational inversions. One of the authors of this paper often finds himself in the following situation: he opens a pack of crisps; he starts munching them; the resulting experience is disgusting; but he keeps on munching at full speed. His gustatory experience is thus deeply *unpleasant*, yet, at the same time, it is accompanied by the *appetitive* motivation to keep on consuming the junk food in question. Therefore, it cannot be the case that his experience is unpleasant in virtue of possessing imperative content *Get less of that food!* or evaluative content *This food is bad*, if, as WDI and WDE say, these contents are intrinsically motivational.

3.3 OUT OF THE LOOP

Now, we come to question (3): in virtue of what does affective phenomenology have a loopy regulatory profile? Eating a brownie brings about a gustatory experience in Abe. Why does the *pleasantness* of this experience make Abe act in a way that results in *more gustatory pleasure*?

Here again, WDE and WDI answer in a similar fashion. Abe's gustatory experience has either world-directed evaluative content *This brownie is good*, or world-

directed imperative content *Get more of this brownie!*. Either content has intrinsic, world-directed, *appetitive* motivational force, hence motivates Abe to eat more of the brownie. This action, in turn, causes another *pleasant gustatory experience* in Abe, which then motivates him to get yet another bite, and so on. In other words, world-directed intentionalism sketches the following picture of affective phenomenology's *loopy motivational profile*:

(a) something (say, your drinking a coffee) causes a certain *affective experience* in a subject (e.g., you undergo a pleasant, or unpleasant, experience);

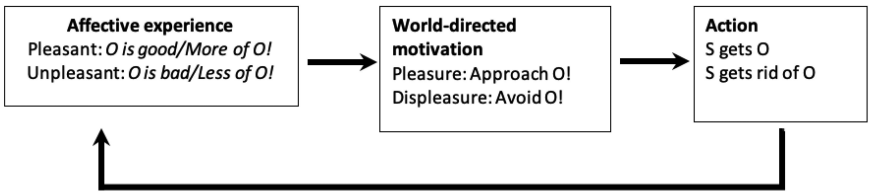
(b) in virtue of its world-directed imperative/evaluative content, this experience intrinsically motivates its subject to get more, or less, of its *worldly object* (e.g., it motivates you to drink again, or to stop drinking, your coffee);

(c) all else being equal, this motivation results in a corresponding action (e.g., you get another sip of your coffee, or you refrain from doing so);

(d) this action typically brings about a change in the subject's *affective experience* (e.g., you undergo some more gustatory pleasure, or you stop experiencing a gustatory displeasure).

We have a major misgiving about this model: it is silent about the numerous cases in which our affect-changing actions cannot be explained in terms of affective experiences' world-directed intentional contents. One such case is Zoe's, which we recounted at the very beginning of this article: she sprained her right ankle; her experience felt unpleasant; she thus took a painkiller and her unpleasant pain went away. It should be obvious why this case escapes the explanatory net cast by world-directed intentionalism. The latter has it that Zoe's pain is unpleasant in virtue of having world-directed content *There is a bad damage in my right ankle* / *Get rid of the damage in my right ankle!*. But unless Zoe literally knows nothing about the human body, such a content cannot motivate her to take a painkiller, since painkillers soothe one's unpleasant pain *without fixing the bodily damage*. If we want to explain Zoe's behavior, we should rather attribute her an *experience-directed motivation*: she took a painkiller because she wanted to get rid of her *unpleasant pain*. Our challenge for world-directed intentionalism is to explain where such a motivation comes from.

The challenge can also be introduced by again considering the case of *world-undirected moods* we discussed in section 3.1.1. Misery, we said, doesn't have any



world-directed intentional content, hence it lacks *world-directed* motivational force. But this doesn't mean that it is not associated to *any* motivation whatsoever. Quite the contrary, when one experiences this unpleasant mood, one doesn't want *to feel that way*. The same carries over, *mutatis mutandis*, to elation: this pleasant mood is associated to an experience-directed motivation—one wants to feel more of this *pleasant mood*—even though it doesn't have world-directed motivational force.[6]

In fact, it takes only a moment's reflection to realize that this is a general truth about affective experiences: pleasant experiences are such that, when we experience them, we want to have *more of them*; unpleasant experiences are such that, when we experience them, we want to have *less of them*. Hence our question for world-directed intentionalism: in virtue of what are affective experiences regularly associated with motivations pro, or against, them?

We now consider how proponents of WDI (section 3.3.1) and proponents of WDE (section 3.3.2) have attempted to answer this question, and we show that neither attempt works.

3.3.1 Spammy requests

Here is a natural answer to our question: it is the very nature of our affective experiences that makes us want having more/less of them. We are motivated against them because they feel *unpleasant*; and we are motivated toward them because they feel *pleasant*. The pleasantness, or unpleasantness, of our experiences *makes sense of* our desire to have, or to avoid, them. Unfortunately, this natural answer is not available to WDI. For this theory, the pleasantness/unpleasantness of an experience E is nothing over and above E's commanding us to have more/less of a certain *worldly object*. Clearly, this command cannot make sense of our attitudes *toward E itself*. For example, the fact that Zoe's unpleasant pain has world-directed imperative content *Get rid of the damage in my right ankle!* entirely fails to rationalize Zoe's negative attitude *toward her pain*—it can only rationalize Zoe's aversion for her *bodily damage*.

Martínez (2015) concedes this point and attempts to explain experience-directed motivations not in terms of affective experiences' *content*, but in terms of their *functional role*. Affective experiences are, *qua* world-directed imperatival states, demanding and distracting: they issue commands and, in this way, capture our attention and re-order our world-directed motivational preferences. For example, Zoe's unpleasant pain insists on telling her *Get rid of the damage in my right ankle!*, thus not letting Zoe attend to anything else. Now, this is perfectly functional *on many occasions*: if your right ankle is seriously damaged, then you'd better interrupt whatever it is that you are doing and focus on fixing your ankle. But *not all occasions* are like that. Suppose that Zoe has done everything in her power to obey

6. As a matter of fact, this counts as yet another type of dissociation between pleasantness/unpleasantness and world-directed appetitive/aversive motivational force. We don't elaborate on this point further because we have the impression that our take-down of world-directed intentionalism is already too cruel.

the command *Get rid of the damage in my ankle!*. At this point, such a command has become "spammy"—it has lost any beneficial function. Luckily, the human mind has the "general tendency to display avoidant reactions to insistent, unfulfillable, misguided, or otherwise inconvenient requests" (Martínez 2015, 2270). It is because of this *hard-wired response to spammy commands*, Martínez says, that Zoe forms the motivation to stop the pain experience by taking a painkiller.

Let's be honest: Martínez's proposal is utterly unbelievable. How could one say with a straight face that the *unpleasantness* of our pains has *nothing to do* with our decisions to soothe them through painkillers? But beyond its sheer implausibility, Martinez's account faces a dilemma as well. According to him, we desire not to have an unpleasant pain *only when* it has become spammy, and an experience is spammy *only when* there is no (further) action we can take to fulfill the world-directed command issued by it. The obvious problem with this is that we desire to end our unpleasant pains even when there are lots of actions open to us to alleviate our bodily damages. Consider Zoe. She has *just* injured her ankle, thus there are *many options open to her* to fix her body: she can massage her leg; go to the doctor; put ice on her ankle; and so forth. It follows that the command *Get rid of the damage in my right ankle!* issued by her unpleasant pain doesn't count as spammy according to Martínez's standards. Still, Zoe is motivated against her unpleasant pain: she doesn't want to feel like that!

In response to this objection, Martínez might redefine 'spamminess', so that *any insistent and distracting* command counts as spammy, whether or not it is unfulfillable or has gone on for a long time. That would make sense of the *immediacy* of our aversion to being in unpleasant pain. But this takes us to the other horn of the dilemma: it predicts that the more an experience is *pleasant*, the more we are motivated *against it*! You are having an orgasm. Your experience feels extremely pleasant. WDI explains such an affective phenomenology by saying that your experience has world-directed imperative content: *Get more of the stimulations of my genitals!*. Now, this is an insistent and distracting command if there is one—try to do anything else while orgasming!—but we want of course to have *more of such an experience*, not less of it. And why so? The answer is simple: because it feels *extremely pleasant*. But, as we know, this is not something that Martínez is allowed to say.

3.3.2 Background desires

In a sense, WDE is exactly in the same predicament as WDI: since unpleasantness/pleasantness is identical to world-directed evaluations, it cannot *per se* generate experience-directed motivations. For example, the unpleasantness of Zoe's pain is constituted by the evaluation *There is a bad damage in my right ankle*, and thus can only motivate her to fix the condition of her ankle. So why do we desire our pleasant/unpleasant experience to continue/end? Here is the answer given by Bain (2019, 485): we are motivated for/against our pleasant/unpleasant experiences because these experiences *are good/bad for us*, and we have *background desires* to have/not to have things occur which are good/bad for us.

What should we make of Bain's proposal? Now, even though we said earlier that a theory of affective phenomenology shouldn't rest on normative commitments of this kind, we can grant to Bain that we do have background desires for/against things that are good/bad for us. The question now becomes whether having a pleasant/unpleasant experience really is good/bad for us. Clearly, Bain owes us an argument to this effect. As it turns out, he has one, but we are going to show that it faces a dilemma: either it fails to support its conclusion, or it forces Bain to give up intentionalism.

Bain often gives the following *intuitive argument* for the claim that pleasant/unpleasant experiences are good/bad for us: they are so because *experiencing a pleasant/unpleasant feeling* is good/bad for us (Bain 2019, 482). Regardless of whether one finds this convincing or not (as a matter of fact, we don't), this argument is not available to Bain *qua intentionalist*. For a theory to count as genuinely intentionalist, its explanations should entirely take place at the semantic-intentional level. That is, intentionalists are supposed to reduce phenomenology to intentional content and then produce an explanation by exclusively adverting to the reducing content, with phenomenology playing no independent explanatory role. Bain's intuitive argument does not respect this stricture, since it rests entirely on how affective experiences *feel*, without any deeper explanation. Bain should rather show us that being in an intentional state that *evaluates things as good/bad* is good/bad for us. Could he come up with such an argument?

Sometimes, Bain seems to argue that merely tokening a mental state with positive/negative evaluative content is good/bad for somebody, and hence that we are motivated to token/not to token these states (Bain 2019). But this is clearly false. We believe that jumping off a cliff would be *terribly bad* for us, but we don't think that *this belief* is bad for us. Quite the contrary, we think that we are very lucky to have such a belief, and we are not at all motivated to get rid of it—God forbid!

In fact, even Bain seems to reluctantly accept this, since he is at pain to highlight that there is something *distinctive* about tokening evaluative *experiences*: when we have such experiences, the properties *goodness*/*badness* are "encountered" in some special way, a way that rationalizes our desires for/against these experiences (Bain 2019, 484–85). We are not sure we understand Bain's suggestion, but this is not a big problem, since it fails to count as an intentionalist explanation yet again. Remember: an intentionalist explanation should only be based on a mental state's content. However, the belief *This is bad for me* and the experience *This is bad for me* have exactly the same content. Thus, *given intentionalism*, it cannot be the case that the latter, but not the former, rationalizes our desires against it.

3.4 TAKING STOCK

In this section, we highlighted that the answers given to (1)–(3) by world-directed intentionalism generate the following problems:

(i) Neither WDE nor WDI is *extensionally adequate*: there are affective experiences lacking world-directed imperative/evaluative content,

and there are experiences possessing such a content but devoid of affective phenomenology.

(ii) WDE faces the *Thinking Otherwise Problem*: it cannot explain why affect-based decisions are not influenced by our knowledge and beliefs.

(iii) Neither WDE nor WDI has the resources to account for *hedo-motivational inversions*.

(iv) Neither WDE nor WDI can explain why we have *experience-directed motivations*: why we want more/less of our pleasant/unpleasant experiences.

In the next section, we argue that reflexive imperativism offers answers to (1)–(3) that avoid problems (i)–(iv).

4. THE IMPORTANCE OF BEING REFLEXIVE

The structure of this final section is as simple as it gets. We detail reflexive imperativism (Barlassina and Hayward 2019) by considering how it answers questions (1)–(3) in turn. By doing so, we both highlight reflexive imperativism's explanatory power as well as its capacity to evade problems (i)–(iv). The conclusion will not surprise you: we have to favor reflexive imperativism over world-directed intentionalism.

4.1 ON THE GROUNDS OF AFFECTIVE PHENOMENOLOGY

4.1.1 The basic idea

Question (1) asks in virtue of what an experience has affective phenomenology. Reflexive imperativism is a form of experience-directed intentionalism: it says that an experience has affective phenomenology in virtue of possessing *experience-directed*, rather than world-directed, intentional content. More precisely, reflexive *imperativism* has it that affective phenomenology depends on *imperative contents* that command the subject of the experience to do something about *some* experience. What experience exactly?

It is here that the *reflexive* aspect kicks in: an experience E has affective phenomenology in virtue of possessing a content that commands one to do something about *E itself*. In particular, an experience P feels pleasant in virtue of possessing reflexive imperative content (1), while an experience U feels unpleasant in virtue of possessing reflexive imperative content (2):[7]

(1) *More of P!*

(2) *Less of U!*

7. See Barlassina and Hayward (2019) for a discussion of the *syntactic structure* of affective experiences.

For example, Abe's gustatory experience G feels pleasant because it has reflexive imperative content *More of G!*, while Zoe's pain is unpleasant because it commands her *Less of this experience!*.

If you need to condense all this into a memorable sentence, here is one for you: affective experiences feel pleasant/unpleasant because they command us *More of me!* / *Less of me!* (bumper stickers are coming soon). This is how reflexive imperativism answers question (1).

This account of the grounds of affective phenomenology is, we maintain, *extensionally adequate*: all pleasant/unpleasant experiences E have intentional content *More of E!* / *Less of E!*; and all experiences E with intentional content *More of E!* / *Less of E!* are pleasant/unpleasant. But we are not so foolish as to try to prove these two universally quantified statements *by exhaustion*. In the next section, we shall rather content ourselves with showing that reflexive imperativism succeeds *vis-à-vis* extensional adequacy where world-directed intentionalism fails, and we leave to you, the reader, the task to devise counterexamples against it. Our bet is that you are not going to find any.

4.1.2 Extensional adequacy, or: hunger, doctors, and misery

World-directed imperativism (WDI) is committed to the claim that if an experience has world-directed imperative content, then it has affective phenomenology. Hunger, we said in section 3.1.2, shows this to be false. While it is plausible that *all* episodes of hunger command something like *Get some food!*, it is clearly *not* the case that all episodes of hunger have affective phenomenology—some of them feel in fact neither pleasant nor unpleasant.

Reflexive imperativism neatly solves the problem. Even though it might well be the case that all episodes of hunger have world-directed imperative content *Get some food!*, this content has *nothing to do* with affective phenomenology. Rather, affective phenomenology obtains in virtue of reflexive imperative content. Hence, those episodes of hunger that have world-directed imperative content *Get some food!*, but lack reflexive imperative content, will thereby lack affective phenomenology. On the contrary, if one has an episode of hunger that, in addition to commanding *Get some food!* also commands *Less of this command!*, one has an experience of unpleasant hunger.

In section 3.1.2, we raised a similar worry for world-directed evaluativism (WDE). Dr. Novice is looking at a patient's severe injury. What a terrible sight! According to WDE, this visual experience feels unpleasant because it has world-directed evaluative content *That injury is bad*. But this cannot be right. If Dr. Novice's visual experience represents the patient's injury as bad, so does Dr. Expert's; but the latter is not undergoing any unpleasant visual experience—Dr. Expert is just too used to this kind of stuff to be bothered.

Here comes reflexive imperativism with the right treatment. Dr. Novice and Dr. Expert's visual experiences have the same world-directed content. However, while Dr. Novice's experience also has reflexive imperative content *Less of this*

visual experience!, Dr. Expert's visual experience has lost this layer of content—probably due to desensitization. This is why Dr. Novice's experience feels unpleasant, while Dr. Expert's feels "affectively neutral."

Finally, in section 3.1.1, we argued that mood episodes like misery and elation put *both* WDI and WDE in trouble, since they have affective phenomenology, but need not have world-directed content. It is cases like these that make reflexive imperativism's explanatory power stand out.

Affective experiences—reflexive imperativism says—have affective phenomenology because they have reflexive imperative content. In the great majority of cases, affective experiences don't have this type of content *only*. For example, unpleasant hunger *also* has world-directed imperative content *Get some food!* and a pleasant taste experience of an ice cream *also* has world-directed descriptive content *This food is sugary*. The interesting thing is that, on some rare occasions, an affective experience can lack all nonreflexive content and be nothing more than a *bare* reflexive command. World-undirected moods are a case in point. When one experiences *pure misery*, one tokens a mental state whose *only* content is *Less of me!*. This explains why pure misery is experienced as *pure unpleasantness*: because unpleasantness is nothing over and above the reflexive command *Less of me!* and this command is *all that there is* when it comes to pure misery.

4.2 HOW TO GET MOTIVATED

4.2.1 On the relation between experience-directed and world-directed motivations

Let's now turn to question (2), which concerns the regulative power of affective phenomenology: in virtue of what does this phenomenology typically motivate us to act a certain way?

With regard to motivation, the crucial difference between world-directed intentionalism and reflexive imperativism is this: for the former, affective phenomenology has world-directed motivational force; for the latter, it has *experience-directed motivational force*—our affective experiences don't motivate us for, or against, some worldly object; they motivate us for, or against, *themselves*. Think again about Abe eating his brownie. According to world-directed intentionalism, his gustatory experience feels pleasant in virtue of telling him: *This brownie is good / Get more of this brownie!*, thus motivating him to get more *of the brownie*. For reflexive imperativism, it instead feels so because it commands Abe: *More of this experience!*. Thus, the pleasantness of Abe's experience motivates him to obtain more *of that experience*, rather than more of the brownie.

But then why does Abe, in response to this experience-directed motivation, get another bite from the brownie? Because his decision-making system *computes* that the best way to satisfy this experience-directed motivation is to keep on eating the brownie. Let us explain. Affective experiences don't come out of nowhere. Sometimes, we can generate/suppress/modify them by performing *mental actions*. For example, we can obtain *some* pleasure by *imagining* having a certain experience, or we can *somehow* reduce the unpleasantness of our pain by *focusing* on

something else. But these mental actions are of limited use. If you want to have more/less of an affective experience, the most effective strategy typically consists of performing a *nonmental action*, i.e., it consists of *acting upon the world* a certain way. Abe figured that out: eating the brownie brought about a pleasant experience P in him; P had imperative content *More of P!* and thus had *experience-directed* motivational force; Abe realized that the best way to satisfy this pro-P motivation was to eat more of the brownie; on this basis, he formed the *world-directed motivation* to eat the brownie and acted accordingly. If Abe is particularly smart and self-controlled, he will savor the brownie, eating it slowly in order to get the maximum quantity of pleasant gustatory experience from it.

Reflexive imperativism thus doesn't deny that affective phenomenology often *causes* world-directed motivations. The point is rather that *these* motivations are *not intrinsic* to affective phenomenology. It is instead *experience-directed motivation* that is intrinsic to affective phenomenology. When one undergoes a pleasant/unpleasant experience, one undergoes an experience issuing the command *More of me! / Less of me!*, hence one is motivated for/against one's affective experience. It is only through decision making that such an experience-directed motivation *might* bring about a world-directed motivation.

Why should you believe this story? Because it solves all the motivation-related problems faced by WDE and WDI.

4.2.2 Problem solving

For WDE, when you have a pleasant/unpleasant experience, you have an experience that evaluates its worldly object O as good/bad, and thus motivates you for/against O. As we have seen in section 3.2.1, one main issue with this proposal is that it predicts that your beliefs about the value of O should significantly influence your decision-making process, but this doesn't happen. For example, when you eat a hot habanero pepper, you might know full well that you are not undergoing any bad bodily damage; still, on the basis of the unpleasantness of your experience, you do things like chugging milk directly from the bottle. Conclusion: the unpleasantness of your experience is not grounded in the evaluation *There is a bad damage in my mouth*. If it were, your knowledge that this evaluation is false should have prevented you from acting like that.

This was the Thinking Otherwise Problem for WDE, and here is how reflexive imperativism (dis)solves it. Affective phenomenology obtains in virtue of reflexive imperative content. For example, the unpleasantness of your 'habanero pepper sensation' is due to the content *Less of this sensation!*. When this content is sent to your decision-making system, it is *not* discounted in the light of information concerning the condition of your mouth. The reason is simple: this content has nothing to do with how your mouth is faring, and thus information about that doesn't have any bearing on it. You drink milk because you know it will lessen the sensation, even though you don't believe that it will do anything to change the physical condition of your mouth.

More importantly, reflexive imperativism has the explanatory resources to account for *dissociations* between affective phenomenology and world-directed motivation. One such dissociation occurs in the case of a mood like misery, which is clearly unpleasant but need not have world-directed motivational force. This is a mystery for world-directed intentionalism. After all, if (i) affective phenomenology obtains in virtue of world-directed evaluative/imperative contents and (ii) such contents are intrinsically motivational, then the unpleasantness of misery *must* give rise to some world-directed motivation. But it often doesn't. When feeling blue, you are motivated *not to feel like that*, that's true. But you might not experience any world-directed motivation.

Here's how reflexive imperativism accounts for this phenomenon. First, misery—*qua* unpleasant affective experience—has reflexive imperative content *Less of me!* and so motivates you against itself. This explains the fact that when feeling miserable, you don't want to feel like that. Second, reflexive imperativism has it that an experience-directed motivation, ME, brings about a world-directed motivation, Mw, only when the decision-making system takes Mw to be a way of satisfying ME. E.g., if your decision-making system hadn't hypothesized that drinking milk would have stopped your 'habanero pepper sensation', your motivation against that sensation wouldn't have resulted in the motivation to drink milk.

But, of course, the decision-making system might *fail* to individuate a course of nonmental action that can satisfy the inputted experience-directed motivation. In such a case, the experience-directed motivation doesn't cause any world-directed motivation. This—or something close enough—is what happens with misery. The command *Less of this experience!* enters the decision-making system, but the latter is not sure how to accommodate this request. As is so often the case, it is unclear whether any nonmental action will lessen the misery. At the beginning, the decision-making system might initiate some experimental behavioral strategies, in order to see if they result in a reduction of the command *Less of this experience!*. After discovering that they are unsuccessful, the decision-making system gives up. Result: one feels miserable; one doesn't want to feel like that; but one doesn't have any world-directed motivation associated to one's affective experience.

Another relevant type of dissociation is what we called *hedo-motivational inversions* (section 3.2.3): cases where a *pleasant* experience is associated to a world-directed *aversive* motivation; or where an *unpleasant* experience is associated to a world-directed *appetitive* motivation. We have simply no idea how world-directed intentionalism can make sense of this phenomenon. If pleasure obtains in virtue of contents like *O is good / Get more of O!*, then it *must* have O-appetitive motivational force; and if displeasure obtains in virtue of contents like *O is bad / Get less of O!*, then it *must* have O-aversive motivational force.

How does reflexive imperativism fare in this regard? As you know, the theory proposes that experience-directed motivations and world-directed motivations are *normally* causally harnessed as follows: (i) one undergoes a pleasant/unpleasant experience, which possesses experience-directed motivational force; (ii) this

motivation is sent to the decision-making system, which attempts to compute the best way to satisfy it; (iii) if this computational process results in the individuation of a certain nonmental action, then a corresponding world-directed motivation obtains. The important thing about this causal structure is that it makes it possible to manipulate world-directed motivations *without manipulating experience-directed motivations*. This, we maintain, is exactly what happens in Berridge et al.'s experiments. We will show this by discussing the case of amygdala-damaged rats (Flynn et al. 1991; Galaverna et al. 1993).

As you might remember, those rats find salt pleasurable. According to reflexive imperativism, this means that they want more of that experience. Strikingly, however, this experience-directed motivation doesn't bring about the world-directed motivation to ingest salt: these rats in fact constantly refuse salty food. How's that possible? Very simply, interventions to the central nucleus of the amygdala *directly control* the rat's world-directed motivations, causing the production of the prepotent, world-directed motivational signal *Don't eat salt!*. Here is another way to put it. Berridge and colleagues disrupted the normal causal chain going from the experience-directed motivation *More of this experience!* to the world-directed motivation *Eat some salt!*. Typically, when the rat's decision-making system gets the first motivation as input, it produces the second motivation as output. Through experimental manipulation of the amygdala, however, the rat's decision-making system has been set up to constantly output the motivation not to eat any salt, *regardless of the experience-directed motivation it receives as input*.

In this section, we sketched some models that attempt to explain why some affective experiences, like misery, don't have *any* world-directed motivational force, and why some others feel *pleasant/unpleasant* but are associated to *world-aversive/world-appetitive* motivations—the so-called hedo-motivational inversions. Please notice that, given our present purposes, it doesn't matter much if our models get the *details* right—this is something that would require an article of its own. The central point is rather this. For reflexive imperativism, there is a considerable degree of independence between affective phenomenology and world-directed motivation, and this makes it possible that the typical *causal relation* between them goes awry. On the contrary, world-directed intentionalism ties the two so closely that it is almost impossible to see how they could come apart. And yet they do.

4.3 STAYING IN THE LOOP

Last, but not least, let's see how reflexive imperativism captures affective phenomenology's *loopy regulatory profile*, i.e., how it explains why the actions motivated by the affective phenomenology of our experiences end up regulating these very experiences back—this was our question (3). Again, the best way to introduce the answer given by reflexive imperativism is to contrast it with the one put forward by world-directed intentionalism.

For the latter, affective phenomenology's loopy regulatory profile does *not* depend on the nature of this phenomenology. In fact, such a phenomenology *only*

motivates one to get more/less of a certain *worldly object*. For example, the pleasantness of Abe's gustatory experience at time *t* obtains in virtue of world-directed content *The brownie is good / Get more of the brownie!*, thus *merely* motivates him to eat more *of the brownie* at time *t*+1. Eating a brownie, however, reliably brings about pleasant gustatory experiences. *It is in virtue of this reliable causal relation* that Abe's gustatory pleasure at time *t* results in Abe's getting more gustatory pleasure at time *t*+2.

In section 3.3, we argued that this proposal fails to account for those cases in which affect-changing actions are *not* grounded in world-directed motivations. Consider Zoe. She sprained her right ankle; she experiences unpleasant pain; she thus takes a painkiller; her unpleasant pain goes away. We *cannot* explain this regulatory loop in the following way: (i) Zoe's unpleasant pain has content *There is a bad bodily damage in my right ankle / Get rid of the bodily damage in my right ankle!*, thus motivates Zoe to fix the bodily damage in her right ankle; (ii) upon receiving this world-directed motivation as input, Zoe's decision-making system generates the motivation to take a painkiller as output; (iii) by taking the painkiller, Zoe gets rid of her unpleasant pain.

Clearly, there is something wrong with (ii): unless Zoe is really in the dark about how painkillers work, she will *not* form the motivation to take a painkiller on the basis of the world-directed motivation to fix her right ankle. The only way to explain Zoe's decision to take a painkiller is to credit her with an experience-directed motivation: Zoe decided to take a painkiller *because she didn't want to feel unpleasant pain*. But where did this experience-directed motivation come from? We argued above that world-directed intentionalism fails to answer this crucial question.

Reflexive imperativism has instead a very convincing answer: Zoe doesn't want to feel unpleasant pain because her unpleasant pain experience has reflexive content *Less of this experience!*. This also explains why Zoe's affective phenomenology has a loopy regulatory profile. This phenomenology commands *Less of this experience!*. Upon receiving this experience-directed command as input, Zoe's decision-making system generates the motivation to take a painkiller as output. Zoe then takes the painkiller and her unpleasant pain goes away.

The same type of explanation applies to *all instances* of affect-based regulatory loops. Let's go back to Abe. The pleasantness of his gustatory experience has content *More of this experience!*. Abe believes that, by eating more of the brownie, he will satisfy such an experience-directed motivation. Thus, he acts accordingly and gets more gustatory pleasure.

The general structure is always the same: (A) S has a pleasant/unpleasant experience E; (B) E's affective phenomenology motivates S for/against E itself; (C) S's decision-making attempts to compute the best course of nonmental action to satisfy this experience-directed motivation; (D) *if* an appropriate course of nonmental action is individuated, S *ceteris paribus* acts that way and gets more/less of the relevant experience. Hence the regulatory loop. It is thus the *very nature* of affective phenomenology that makes its regulatory profile loopy.

CONCLUSION

So how does affective phenomenology regulate? According to world-directed intentionalism, affective phenomenology motivates us to approach or avoid certain *worldly objects*, and these *world-directed motivations* happen to bring about more or less of our affective experiences as a side effect.

Reflexive imperativism understands the loopy regulatory profile of affective phenomenology in a precisely inverse manner. What affective experiences fundamentally tell us to do is to regulate *them*. A pleasant experience P has reflexive imperative content *More of P!* and thus motivates us to have more of itself; an unpleasant experience U has reflexive imperative content *Less of U!* and thus motivates us to have less of itself. Since most often the best way to satisfy an *experience-directed motivation* is to perform a nonmental action (to eat a brownie, take a painkiller, or drink soothing milk), affective phenomenology typically brings about world-directed motivations as well. But these world-directed motivations are ultimately in the service of the experience-directed motivations that caused them. When we act under the sway of affective phenomenology, our end goal is that of getting more/less of our affective experience.

In this article, we argued that there are many reasons to prefer reflexive imperativism to world-directed intentionalism. The former, but not the latter, is extensionally adequate, solves the Thinking Otherwise Problem, has the resources to deal with world-undirected moods, and accounts for hedo-motivational inversions. Reflexive imperativism is the best *intentionalist* account of affective phenomenology—probably, the best account *tout court*. "More of it!", we say.

REFERENCES

Anderson, R. L. 2017. "Friedrich Nietzsche." In *The Stanford Encyclopedia of Philosophy*, edited by E. N. Zalta. https://plato.stanford.edu/archives/sum2017/entries/nietzsche/.

Aydede, M. 2006. "The Main Difficulty with Pain." In *Pain: New Essays on the Nature of Pain and the Methodology of Its Study*. Cambridge, MA: MIT Press.

Bain, D. 2013. "What Makes Pains Unpleasant?" *Philosophical Studies* 166:69–89.

Bain, D. 2019. "Why Take Painkillers?" *Noûs* 53:462–90.

Barlassina, L. Under review. "Beyond Good and Bad: Reflexive Imperativism, not Evaluativism, Explains Valence."

Barlassina, L., and M. K. Hayward. 2019. "More of Me! Less of Me! Reflexive Imperativism about Affective Phenomenal Character." *Mind* 128(512): 1013–44.

Bentham, J. 1789/1970. *An Introduction to the Principles of Morals and Legislation*. Edited by J. H. Burns and H. L. A. Hart. Oxford: Oxford University Press.

Berridge, K. C. 1996. "Food Reward: Brain Substrates of Wanting and Liking." *Neuroscience & Biobehavioral Reviews* 20(1): 1–25.

Berridge, K. C. 2000. "Measuring Hedonic Impact in Animals and Infants." *Neuroscience and Biobehavioral Reviews* 24:173–98.

Berridge, K. C., and E. S. Valenstein. 1991. "What Psychological Process Mediates Feeding Evoked by Electrical Stimulation of the Lateral Hypothalamus?" *Behavioral Neuroscience* 105(1): 3–14.

Berridge, K. C., T. E. Robinson, and J. W. Aldridge. 2009. "Dissecting Components of Reward: 'Liking', 'Wanting', and Learning." *Current Opinion in Pharmacology* 9(1): 65–73.

Bramble, B. 2013. "The Distinctive Feeling Theory of Pleasure." *Philosophical Studies* 62(2): 201–17.
Byrne, A. 2001. "Intentionalism Defended." *Philosophical Review* 110:199–240.
Carruthers, P. 2018. "Valence and Value." *Philosophy and Phenomenological Research* 97(3): 658–80.
Chalmers, D. 1995. "Facing up to the Problem of Consciousness." *Journal of Consciousness Studies* 2:200–19.
Cutter, B., and M. Tye. 2011. "Tracking Representationalism and the Painfulness of Pain." *Philosophical Issues* 21:90–109.
Davidson, D. 2001. *Essays on Actions and Events*. Oxford: Clarendon Press, 2nd ed.
Dretske, F. 1995. *Naturalising the Mind*. Cambridge, MA: MIT Press.
Flynn, F. W., H. J. Grill, J. Schulkin, and R. Norgren. 1991. "Central Gustatory Lesions: II. Effects on Sodium Appetite, Taste Aversion Learning, and Feeding Behaviors." *Behavioral Neuroscience* 105(6): 944–54.
Fodor, J. 1983. *The Modularity of Mind*. Cambridge, MA: MIT Press.
Fodor, J. 1987. *Psychosemantics*. Cambridge, MA: MIT Press.
Galaverna, O. G., R. J. Seeley, K. C. Berridge, H. J. Grill, A. N. Epstein, and J. Schulkin. 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.
Heathwood, C. 2007. "The Reduction of Sensory Pleasure to Desire." *Philosophical Studies* 133:23–44.
Klein, C. 2015. *What the Body Commands*. Cambridge, MA: MIT Press.
Martínez, M. 2011. "Imperative Content and the Painfulness of Pain." *Phenomenology and the Cognitive Sciences* 10(1): 67–90.
Martínez, M. 2015. "Pains as Reasons." *Philosophical Studies* 172(9): 2261–74.
Mendelovici, A. 2013. "Intentionalism about Moods." *Thought: A Journal of Philosophy* 2(1): 126–36.
Moore, G. E. 1903/1922. *Principia Ethica*. Cambridge: Cambridge University Press.
Tye, M. 1995a. *Ten Problems of Consciousness*. Cambridge, MA: MIT Press.
Tye, M. 1995b. "A Representational Theory of Pains and Their Phenomenal Character." *Philosophical Perspectives* 9:223–39.