Article:
Barlassina, L. and Khan Hayward, M. (2019) More of me! Less of me! Reflexive imperativism about affective phenomenal character. Mind, 128 (512). pp. 1013-1044. ISSN: 0026-4423

https://doi.org/10.1093/mind/fzz035

More of me! Less of me!

Reflexive imperativism about affective phenomenal character

Forthcoming in *Mind*

LUCA BARLASSINA
University of Sheffield
l.barlassina@sheffield.ac.uk

MAX KHAN HAYWARD
University of Sheffield
mh3173@columbia.edu

Abstract

Experiences like pains, pleasures, and emotions have *affective phenomenal character*: they feel pleasant or unpleasant. Imperativism proposes to explain affective phenomenal character by appeal to *imperative content*, a kind of intentional content that directs rather than describes. We argue that imperativism is on the right track, but has been developed in the wrong way. There are two varieties of imperativism on the market: first-order and higher-order. We show that neither is successful, and offer in their place a new theory: *reflexive imperativism*. Our proposal is that an experience P feels pleasant in virtue of being (at least partly) constituted by a Command with reflexive imperative content (1), while an experience U feels unpleasant in virtue of being (at least partly) constituted by a Command with reflexive imperative content (2):

(1) More of P!
(2) Less of U!

If you need a slogan: experiences have affective phenomenal character in virtue of commanding us *Get more of me!, Get less of me!*

Introduction

Experiences like pains, pleasures, and emotions have *affective phenomenal character*: they feel good or bad, pleasant or unpleasant. Imperativism says that we can explain affective phenomenal character by appeal to *imperative content*, a kind of intentional content that directs rather than describes. In this paper, we argue that imperativism is on the right track, but has been developed in the wrong way. There are two varieties of imperativism currently on the market: first-order imperativism (Martínez 2011, 2015*a*,

2015*b*) and higher-order imperativism (Klein 2015). We shall show that neither is successful, and offer in their place a new theory: *reflexive imperativism*.

Our proposal is that an experience has affective phenomenal character in virtue of possessing *reflexive imperative content.*[1] More precisely, an experience P feels pleasant in virtue of being (at least partly) constituted by a Command with reflexive imperative content (1), while an experience U feels unpleasant in virtue of being (at least partly) constituted by a Command with reflexive imperative content (2):

(1) More of P!
(2) Less of U!

If you need a slogan: experiences feel pleasant or unpleasant in virtue of commanding us *Get more of me!, Get less of me!*[2]

Imperativism is one approach to affective phenomenal character. Other prominent accounts include: evaluativism (Bain 2013; Carruthers 2018; Cutter and Tye 2011; Tye 2005), psycho-functionalism (Aydede *forthcoming*), and the desire theory (Brady *forthcoming*; Heathwood 2007). We think that our proposal is not just the best formulation of imperativism, but the best theory of affective phenomenal character *altogether*. But we shall restrict our arguments to the former, more local claim. Still, we will show that reflexive imperativism has remarkable explanatory power. For this reason, it should be considered as one of the most promising out of all the candidate accounts.

The paper is structured as follows. In section 1, we further clarify our explanandum: affective phenomenal character. In section 2, we introduce the common

[1] The relation *in-virtue-of* will play a big role in this article. But we will avoid giving a precise characterisation. We take it to express *some* relation of dependence/determination, on a scale that goes from supervenience to identity. For our aims, it is unnecessary to pick a precise point on the scale.

[2] In English, there might be subtle differences in meaning among 'More of me!', 'Get more of me!', 'Have more of me!', etc. We doubt that the imperative content of affective experiences is so fine-grained to express these differences, so we can safely ignore them. The crucial point is simply this: an experience feels good/bad in virtue of commanding *the subject of the experience* (or a cognitive sub-system of the subject) to get more/less of this very experience, or something near enough. We thank the editors of this journal for urging us to clarify this.

core of imperativism. After this, we illustrate the different versions of imperativism and assess their respective merits. We start in section 3 with first-order imperativism, and argue that it won't do. In section 4, we do the same with respect to higher-order imperativism. We then present, in section 5, our variety of imperativism about affective phenomenal character—reflexive imperativism—and show how it can solve the difficulties that beset first-order and higher-order imperativism. We conclude, in section 6, by highlighting a further benefit of reflexive imperativism: it paves the way to accounting for the evolutionary function of affective phenomenal character and, by doing so, it illuminates the relations among affective phenomenal character, motivation, and learning.

1. The explanandum

Some mental states are such that there is something it is like for their subjects to be in them. If your visual system is properly working and you stare at a banana in good light, there is something it is like for you to see that banana. Call these mental states 'experiences', and call what it is like to have them their 'phenomenal character'.

'Affective experiences' have 'affective phenomenal character': they feel good (pleasant) or bad (unpleasant). An experience has *positive* affective phenomenal character when it feels good, and *negative* affective phenomenal character when it feels bad. It is controversial exactly which experiences count as affective experiences, but here are some paradigmatic examples: pain experiences[3] (e.g., headaches and cramps), pleasant and unpleasant bodily sensations (e.g., orgasms and intense itches), felt emotions (e.g., happiness and shame), and felt moods (e.g., elation and misery). Affective phenomenal character is our target explanandum.

A theory of affective phenomenal character has to answer the following questions:

[3] Subjects suffering from pain asymbolia report that they experience pain, but that it does not feel unpleasant (Berthier et al. 1990). Lacking affective phenomenal character, asymbolic pain does not count as an affective experience. Does it count as pain at all? This question is beyond the remit of this article.

(Q1) In virtue of what does an experience have affective phenomenal character?

(Q2) In virtue of what does an experience have positive, rather than negative, affective phenomenal character (or vice versa)?

To answer these questions, we adopt the well-known *intentionalist* strategy, which attempts to explain phenomenal character by reference to *intentional* content (more on this in the next section). Philosophers working on consciousness are often puzzled by the following question: how is it *possible* that there is any such thing as phenomenal character? This puzzlement involves the thought that phenomenal character is radically different from anything else in the world. This leaves the existence of phenomenal character in the natural world seeming mysterious. We think that the intentionalist strategy helps us to see how phenomenal character is possible, for the following reason.

We often appeal to phenomenal character in providing psychological explanations. We might ask: 'Why do you *think* that fruit is a mandarin, not a lime?' And you might answer: 'Because it *looks* orange!' But you might *also* explain this by talking solely of the intentional content of your experience, without mentioning phenomenal character: 'Because it *is* orange', you can say. So intentional content can play the same role in psychological explanations that phenomenal character does.

This suggests that if we can work out how mental states can have intentional content, it should become less mysterious how they can have phenomenal character. Phenomenal character is not radically different from everything else, but could rather be explained by—or even identified with—intentional content. This is enticing for philosophers searching for a naturalistic account of the mind. If phenomenal character can be explained by appeal to intentional content, then we 'only' need to naturalise the latter in order to naturalise the former. Intentionalism thus paves the way to crack the hard problem of consciousness. This is why it has proven such an attractive position to philosophers in the last three decades (Hellie 2009).

What is the distinctive role that affective phenomenal character plays in psychological explanations? It explains *motivations*. 'Why don't you *want* to feel pain any longer?', we ask. 'Because it feels *bad*!', you answer. Any theory of affective

phenomenal character must then explain its connection with motivation. But affective phenomenal character does not just happen to be typically motivating; it has *intrinsic motivational force* (Bain 2013; Jacobson 2013). Suppose that you are having a visual experience as of a red apple. The visual phenomenal character of your experience can motivate you to reach out and grab the apple … but only if you have a background desire for apples. Affective phenomenal character is different: it can motivate you *independently of any other conative state*. Once you have told us that your experience feels bad, you have *fully* explained why you are motivated to get rid of it (but see Corns 2014).[4]

There are things other than affective phenomenal character which are intrinsically motivating. A desire can be intrinsically motivating without being pleasant or unpleasant. Hunger is intrinsically motivating, but does not always feel bad (or good). What is special about affective phenomenal character's intrinsic motivational force is that it is self-directed, or *reflexive*. Reflexive motivational force is a particular form of *mind-directed* motivational force. It motivates us to do something about our own mental states. But it doesn't just motivate us to get into or out of some mental state or another. The mental state which affective phenomenal character motivates us for or against is the *very same* mental state that has affective phenomenal character. When we are having a pleasant/unpleasant experience, it is the very pleasant/unpleasant experience that we want/not want to have: we want to avoid pain, misery, shame, and unpleasant itches; we want happiness, orgasm, and elation. Again, in explaining our motivations for or against affective experiences, we need do no more than appeal to their affective phenomenal character: why does misery motivate you to have no more misery? Because it feels bad! Thus, an adequate theory of affective phenomenal character has to answer (Q3) as well:

[4] 'Intrinsic motivational force' is a dispositional notion. Thus, strictly speaking, to say that affective phenomenal character has intrinsic motivational force is to say that it has the *capacity*, or the *power*, to motivate independently of any other conative state. This capacity might fail to manifest itself. For example, your attention may be diverted from an unpleasant experience by a sudden distraction. Furthermore, even when motivations are produced, this does not imply that they are efficacious in bringing about intentional action. The subject might in fact have stronger, countervailing motivations—like when a marathon runner endures her unpleasant pain, because the thing she wants the most is to reach the finish line. Since none of our arguments hinge upon intrinsic motivational force's dispositional and *pro tanto* nature, we will omit the qualification in what follows, so to keep things as simple as possible.

(Q3) In virtue of what does affective phenomenal character have intrinsic and reflexive motivational force?

2. Imperativism

We adopt the *intentionalist strategy*: we propose to explain affective phenomenal character by appeal to intentional content. But while standard, or 'representationalist', intentionalism attempts to explain affective phenomenal character in terms of *indicative content*, we shall instead resort to imperative content. Our theory is a form of imperativism.

2.1 Indicative and imperative contents

Consider the following two sentences:

(3) Ben is closing the door
(4) Ben, close the door!

Recent work in the philosophy of language and natural language semantics suggests that an imperative sentence like (4) has a *different type* of content than a declarative sentence like (3) (see Charlow 2014 and Portner 2016 for an overview). (3) has indicative content, (4) has imperative content.[5] Here is a characterisation of the distinction:

[5] Why not stick with Frege's idea that there is one type of content only, but different illocutionary forces? For example, why not think that utterances (A) and (B) have the same content, namely, *that Ben will buy ice-cream tomorrow*, but assertoric and directive force, respectively?

A. Ben will buy ice-cream tomorrow
B. Buy ice-cream tomorrow! (said to Ben)

The issue is complex and we cannot do it justice in this article. For discussion, see Hanks 2007. In any case, we suspect that Fregeans will be able to accept much of our substantive view by formulating their own style of imperativist theory. We thank the editors of *Mind* for pointing out this possibility to us.

INDICATIVE CONTENT:

(i) Has the function of carrying information, e.g., that P is the case.

(ii) Has truth conditions: it is true if P is the case and false if not.

(iii) The audience correctly uptakes it by forming a belief.

IMPERATIVE CONTENT:

(iv) Has the function to direct its addressee to do something, e.g., to φ.

(v) Has satisfaction conditions: it is satisfied if and only if the addressee φs.

(vi) The audience correctly uptakes it by forming a motivation.

The content of (3) carries the information that Ben is closing the door and is true if and only if Ben is closing the door. The content of (4) does not have information-carrying function and cannot be evaluated as true or false. Rather, it has the function of directing Ben to close the door and is satisfied if and only if Ben does so. If Ben is receptive to the speaker, he will uptake the content of (3) by forming a belief, and uptake the content of (4) by forming a motivation.[6]

The distinction between indicative and imperative content is also relevant for the philosophy of mind and cognitive science (Shea 2013 is a good place to start). Some mental states have indicative content (call these states '*Indicators*'). Other mental states have imperative content (call them '*Commands*'). This distinction, we maintain, paves the way to account for the difference between affective phenomenal character and other types of phenomenal character.

[6] According to this picture, the features possessed by each type of content are deeply intertwined. Take imperative content. Its *function* is that of directing its addressee to φ. This is why it is *satisfied* only if the addressee φs. But for the addressee to φ, the addressee needs to be *motivated* to φ. Therefore, the correct, or successful, uptake of an imperative content cannot simply involve the addressee forming the belief *I am commanded to φ*. It also requires that the addressee forms the motivation to φ. This understanding of imperative content is somewhat controversial. Some accounts distinguish between *uptake* and *acceptance*—I uptake the imperative content by understanding that I am being told to φ, and I accept it by forming the motivation to φ. But that more complex model is equally congenial for our purposes—the function of the imperative content is to get the audience (say, Ben) to φ, and if Ben is receptive to the speaker, he will *both* uptake the imperative content by understanding he is being told to φ *and* accept it by forming a motivation to φ.

2.2 Imperativism about affective phenomenal character

Mary is having a visual experience as of a red tomato. Intentionalists want to explain the phenomenal character of Mary's experience by reference to intentional content. Some say the relevant content is first-order content (5) (Dretske 1995; Tye 1995), others point to higher-order content (6) (Rosenthal 2005):

(5) *There is a red tomato*

(6) *I am having a visual experience as of a red tomato*

Both (5) and (6) are *indicative* contents. In spite of the countless divergences among them, until recently all intentionalists assumed that indicative content was the only type of content to be invoked in an explanation of phenomenal character.

Imperativism marks a reaction against this stricture. Imperativists maintain that imperative content is a central explanatory tool too. Perhaps indicative content suffices to explain *certain types* of phenomenal character (e.g., visual phenomenal character). But there are *other types* of phenomenal character that are bound to resist a 'purely indicativist' treatment. This is the central tenet of imperativism.

Imperative content was first used to explain the *sensory* phenomenal character of bodily sensations like itches (Hall 2008) and pains (Klein 2007). Subsequently, it was deployed to deal with pain experiences' affective phenomenal character (Martínez 2011). Finally, it has been applied to affective phenomenal character *in general* (Martínez 2015*a*; Klein 2015). This article belongs to the latter project. It is easy to grasp the intuitive motivation to account for affective phenomenal character through imperative content. You can explain coming to *believe* that there is a red apple on the table by appealing to the *visual* phenomenal character of your experience. Why did you believe this? Because that's how it *looked!* This explanation makes sense if visual phenomenal character depends on *indicative* content—forming that belief was in fact the condition of correctly uptaking this type of content. But visual phenomenal character is motivationally inert, unlike, for example, the affective phenomenal character of a toothache. Imperativism proposes that affective phenomenal character depends on imperative content. Forming a motivation is the condition of correctly uptaking this type of content. Imperative content thus has intrinsic motivational force—

its function is to direct, not to describe. Therefore, we begin to understand why affective phenomenal character has this motivational force too.

We begin to understand, of course. We are not suggesting that what we have just told you counts as an explanation of affective phenomenal character. For while all imperativist theories correctly appeal to the intrinsically motivational nature of imperative content, the extant theories do not, we claim, successfully explain affective phenomenal character.

3. First-order imperativism

3.1 The theory introduced

The central tenet of *first-order imperativism* (Martínez 2011, 2015*a*, 2015*b*) can be expressed as follows:

> **First-order imperativism**: an experience has affective phenomenal character in virtue of having first-order imperative content.

The content of a mental state is a *first-order content* if and only if it does not feature any mental state. Accordingly, the content of a mental state is a *first-order imperative content* if and only if it directs the subject to do something about the non-mental world only.[7] It is in virtue of having this type of imperative content that an experience has affective phenomenal character—Martínez says.

George is feeling pain because his right hand is under scalding water. George's experience has *sensory* phenomenal character (George experiences certain sensory qualities as instantiated in his hand) and affective phenomenal character (George's experience is unpleasant). Martínez proposes that George's experience has sensory phenomenal character in virtue of having first-order *indicative* content:

(7) *There is a burning disturbance in your right hand*

[7] By the same token, a *first-order indicative content* is a content that represents the non-mental world only.

On the other hand, his experience possesses affective phenomenal character in virtue of having first-order *imperative* content:

(8) *See to it that the disturbance in your right hand does not exist!*

Thus, George's pain experience is a compound mental state, made up by an Indicator *conjoined with* a Command—the content of the first determining the burning sensation, the content of the second determining the feeling of unpleasantness.

What about affective phenomenal character's intrinsic motivational force? Imperative contents have intrinsic motivational force, in that they direct their addressees to do something. Therefore, if the affective phenomenal character of George's experience depends on content (8), it is not hard to see why it has intrinsically motivational force.

First-order imperativism is an elegant and simple theory. These virtues, however, come at a high price: the theory is false. So, at least, we argue in the next section.

3.2 Four problems

First-order imperative content without affective phenomenal character

Imperativism was first introduced as an explanation of the *sensory* phenomenal character of bodily sensations like itches and hunger (Hall 2008). These experiences can motivate one to scratch or eat in the absence of any other conative state. Thus, we agree that it is plausible to think that they have first-order imperative content. An itch commands *Scratch!*, while hunger says *Eat!* But here is the rub: itches don't always feel bad, nor does hunger. And when these experiences don't feel bad, they don't necessarily feel good either. Sometimes, they feel neither good nor bad. That's to say, sometimes they have no affective phenomenal character. But these experiences have intrinsic, world-directed motivational force, and thus have first-order imperative content, nonetheless. So, *pace* Martínez's, first-order imperative content is *not sufficient* for affective phenomenal character.

Affective phenomenal character without first-order imperative content

Alice has been suffering from depression. Today, she is feeling miserable. Her experience has negative affective phenomenal character—if there is an unpleasant mood, misery is one. But it does not seem to have any first-order imperative content, since it elicits no world-directed motivations. Alice, like many who suffer from depression, spent the entire day in bed, completely still. To be sure, she wanted not to *feel miserable*, but she didn't want to do anything else. But if an experience can have affective phenomenal character *without* having first-order imperative content, then the former does not obtain *in virtue of* the latter.

Martínez could reply as follows. It is true that, when feeling miserable, one ends up doing nothing. But this is because misery has first-order imperative content:

(9) *Don't do anything!*

In other words, misery does have world-directed motivational force. It is just that this force is *negative*: it motivates one *not to act* upon the non-mental world.

This reply is not very convincing. We do not know of any model of misery/depression according to which this condition is due to the presence of a global, negative, world-directed motivational signal. In fact, the consensus is that misery/depression concerns a global *loss* of world-directed motivation. When one feels miserable, one does not experience one's ordinary urges *together with* the stronger urge not to act upon the non-mental world. Rather, one feels as if one's world-directed urges have disappeared.

Misery is another counter-example to first-order imperativism: an experience with affective phenomenal character, but lacking any first-order imperative content. First-order imperative content is *not necessary* for affective phenomenal character either.

Positives and negatives

Martínez offers no explicit answer to why an experience feels good rather than bad (or vice versa). But it is possible to reconstruct what he has in mind from the examples he

considers. He suggests that the *negative* affective phenomenal character of fear and disgust depend on imperative contents like (10) and (11):

FEAR: (10) *Stay away from that danger!*
DISGUST: (11) *Stay away from those pathogens!*

And he proposes that the *positive* affective phenomenal character of an orgasm or of tasting chocolate depend on the imperative contents such as:

ORGASM: (12) *Get more stimulation of the genitals!*
TASTE: (13) *Get more chocolate!*

Hence, it seems fair to say that, according to Martínez:

- An experience has *negative* affective phenomenal character in virtue of having first-order *aversive* imperative content (i.e., an imperative content that directs its addressee to stay clear from, or to avoid, or to get less of, something in the non-mental world).
- An experience has *positive* affective phenomenal character in virtue of having first-order *appetitive* imperative content (i.e., an imperative content that directs its addressee to approach, or to get more of, something in the non-mental world).

This proposal does not strike us as tenable. Consider *agonising* hunger. It has first-order *appetitive* imperative content: *Eat something!*, or *Put something in your stomach!* But it is unpleasant. Martínez might respond that *agonising* hunger has in fact first-order *aversive* imperative content: *Stop having an empty stomach!* But such a move highlights another problem: there does not seem to be a principled distinction between appetitive and aversive first-order imperative contents. What is the difference between *Stop having an empty stomach!* and *Put something in your stomach!*? The choice of whether to characterise such contents as appetitive or as aversive is arbitrary. But whether or not an experience is pleasant or unpleasant is not arbitrary. Martínez is in trouble again.

Reflexive and intrinsic motivational force

Try to remember the last time you had an excruciating pain. The unpleasantness of your pain motivated you to *get rid* of the pain, didn't it? We bet that you did all sorts of things you could to silence your pain experience: you took painkillers, directed your attention elsewhere, you even smoked all of Bob's marijuana. This is why we said that affective phenomenal character is *intrinsically* and *reflexively* motivational: all by itself, the affective phenomenal character of an experience E motivates us for or against E. This is not something that Martínez can explain. If the unpleasantness of a toothache depends on first-order imperative content:

(14) *See to it that the cavity in your tooth does not exist!*

then it will motivate you to get rid of the cavity in your tooth. But it's unclear why it would motivate you to stop feeling pain.

How could Martínez respond? He could accept that affective phenomenal character is *typically associated* with *reflexive* motivational force, but deny that such force is *intrinsic* to it. It is only in virtue of other, background desires that we are motivated to have more or get rid of affective experiences. This, in effect, is exactly what Martínez says about pain:

> No pain … is directing us to do something *about itself*. … Let's assume that [Iris's toothache] has being going on *for hours*. … Iris has already made an appointment with a dentist early the next morning. There is nothing more she can do now to follow the toothache command. In such a situation, the toothache … is just spam. If Iris is able to limit the impact of such unhelpful advice …, she should do so. (Martínez, 2015*b*, 2269-70, emphasis added)

Martínez would have us believe that there is no constitutive connection between the *unpleasantness* of pain and Iris's motivation to get rid of it. Rather, unpleasantness motivates Iris to do something *other* than get rid of her pain—in the case, to call the doctor. It is only accidentally that Iris wants to get rid of the pain—because there is no further action she can take to improve her bodily state, and the human mind has the general tendency to avoid 'insistent and unfulfillable requests' (Martínez, 2015*b*, 2270). This theory predicts that any action Iris takes to get rid of *the pain* (for example to take

a painkiller, or try to distract herself) would only arise after she has realised that there is nothing further she can do to fix her cavity.

We are not persuaded. We may be atypical, but when we experience pain, our *first* motivation is to get rid of the pain. And we are motivated to do that not because pain is giving us some *other* motivation that we cannot currently act on, but simply because pain feels unpleasant. In fact, very often we only take steps to protect our bodies *because* these seem like the best way to avoid pain—perhaps irresponsibly, it's only when the toothache is persistent that we tend to call the dentist. This indicates that when we are motivated for or against affective experiences, our motivation is *intrinsic* to them, not extrinsic.

3.3 Taking stock

Something went badly wrong with Martínez's first-order imperativism. What exactly? One might say that the problem lies with *imperative* contents. But this cannot be right, since the same kind of issue arises for any *first-order* account of affective phenomenal character. Consider *evaluativism* (Bain 2013; Cutter and Tye 2011). The idea here is that an experience has affective phenomenal character in virtue of having first-order *evaluative* content—i.e., in virtue of representing a certain worldly object as good or as bad (evaluative contents are thus *indicative*, rather than imperative). For example, Joe's back pain is said to be unpleasant in virtue of representing a certain bodily damage in Joe's back *as bad* (for Joe). Assuming that such a first-order content has intrinsic motivational force *at all*, it appears only to motivate Joe to take care of his *body*, hence failing to explain the fact that the unpleasantness of Joe's experience motivates him to get rid of his very *experience*.[8] By the same token, first-order evaluative contents are also ill-suited to explain the negative affective phenomenal character possessed by misery: what is the first-order target that an experience of misery evaluates as bad? Nothing comes to mind.

Accordingly, the problem with the contents chosen by Martínez is not that they are imperative. It is that they are first-order. Imperativists should better resort to a type

[8] To be fair, evaluativists are aware of this problem and have tried to deal with it on multiple occasions (Bain 2013, *forthcoming*; Cutter and Tye 2014). We are not convinced by their responses, but we leave this to another paper.

of imperative content that rather directs one to do something about *one's mental states*. The natural way to implement such suggestion is to go higher-order. This is exactly what Colin Klein (2015) did. Let us then see whether higher-order imperativism fares better than first-order imperativism.

4. Higher-order imperativism

Let's say that the content C of a mental state M is a higher-order content if and only if C features some mental state M* different from M. For example, Mary's belief that she *likes* ice-cream has higher-order (indicative) content:

(15) *I like ice-cream*

Mutatis mutandis, the content C of a mental state M of a subject S is a *higher-order imperative content* if and only if C directs S to do something about some mental state M* different from M.

The general idea behind higher-order imperativism is that affective phenomenal character depends on higher-order imperative content. Here is how Klein puts it:

> Pleasantness [and unpleasantness] are higher-order mental states. [Unpleasantness] is an attitude taken towards pain. That attitude could also be taken towards a variety of other sensations. Hence, it is possible to be [unpleasantly] hungry, tired, or lonely. That makes [unpleasantness] a higher-order mental state: … a state that is … directed towards *some other mental state*. … It is a second-order imperative directed towards a first-order sensation. (Klein, 2015, 183-6)[9]

On Klein's view, there are two distinct states: one that commands the subject to get less/ more of the other. There are two ways of interpreting this proposal, each corresponding to different answers to the question 'which of the two states is the one that feels bad/good?' Under either interpretation, higher-order imperativism does a better job in accounting for affective phenomenal character than first-order imperativism. However, both versions face serious difficulties.

[9] Note that Klein here is using the term 'attitude' simply as a catch-all for various kinds of higher-order intentional state. As he explains, the particular kind of attitude he has in mind is a higher-order imperative—a higher-order Command, in our terminology.

4.1 Higher-order imperativism: First formulation

Here is the most straightforward reading of the central claim made by higher-order imperativism:

> (**HO_1**): An experience E of a subject S has affective phenomenal character in virtue of having higher-order imperative content directing S to do something about some of S's mental states distinct from E.

On this view, affective experiences have higher-order imperative content. It is in virtue of this content that they feel pleasant or unpleasant. More precisely, an unpleasant experience U is a higher-order Command with higher-order imperative content (16), while a pleasant experience P is a higher-order Command with higher-order imperative content (17):

> (16) *Less of M!*
> (17) *More of M!*

where 'M' picks a mental state *different* from U and P.

Remember George, who is feeling an unpleasant pain because his right hand is under scalding water? According to HO_1, George is in fact having *two numerically distinct* experiences at the same time:

(i) A sensory experience S (i.e., pain), which has sensory phenomenal character, but lacks affective phenomenal character.
(ii) An affective experience U, which has negative affective phenomenal character, but lacks sensory phenomenal character.

For HO_1, U has negative affective phenomenal character in virtue of possessing higher-order imperative content:

> (18) *Less of S!*

In other words, it is not the pain that feels bad. It is the distinct experience that tells us not to have pain that feels bad.

Analogously, a pleasant gustatory experience consists in the *co-occurrence* of two separate experiences: a first-order gustatory experience G, which has sensory phenomenal character, but lacks affective phenomenal character; and a higher-order experience P, possessing affective phenomenal character, but devoid of any sensory phenomenal character. P has positive affective phenomenal character in virtue of having higher-order imperative content:

(19) *More of G!*

Again, it is not the gustatory experience that feels good, but the distinct experience that tells us to have more of the gustatory experience.

4.2 The good and the bad of HO_1

HO_1 has some advantages over first-order imperativism. First, it fares better taxonomically. It needn't count as having affective phenomenal character experiences which are neither pleasant nor unpleasant, like mild hunger or minor itches. HO_1 can allow that these experiences are first-order Commands (they have first-order imperative contents like *Eat something!* and *Scratch there!*), and simply deny that they have higher-order imperative content. Second, it is not saddled with trying to find a world-directed motivation to associate with misery. Misery consists in the co-occurrence of two experiences: a first-order experience, S, and a higher-order one, H. S does not have imperative content at all. H does, but it is higher-order imperative content *Less of S!* This is why misery does not have world-directed motivational force.

Apparently, HO_1 possesses a further virtue. Affective phenomenal character has intrinsic, mind-directed motivational force: it motivates us for or against our own mental states. HO_1 makes it clear why this is the case—after all, imperative contents are intrinsically motivational, and higher-order contents are mind-directed by definition. The problem for HO_1 is that it misidentifies *which* state it is that we are motivated to get less of (or more of). Suppose that you are having an unpleasant pain experience. How does HO_1 describe this case? It says that you are having two distinct experiences: an

affectively neutral, sensory experience, S, and an unpleasant affective experience, U. The latter is unpleasant in virtue of having higher-order imperative content:

(18) *Less of S!*

It is true that (18) has mind-directed motivational force, but it motivates you to get rid of experience S, which is *not unpleasant at all*! This is clearly absurd. It is unpleasant states *themselves* (in this case U) that we wish to be rid of, pleasant states *themselves* that we are motivated to get. This is why we have said that the affective phenomenal character of an experience E has a particular type of mind-directed motivational force, namely, *reflexive* motivational force. It motivates us for or against E *itself*. HO_1 cannot account for this.

4.3 Higher-order imperativism: Second formulation

HO_1 is the most natural way to read Klein's view. But, faced with the argument above, Klein might argue that a revision is needed:

> (**HO_2**): An experience E of a subject S has affective phenomenal character in virtue of being *targeted* by a mental state H (distinct from E) whose higher-order imperative content directs S to do something about E.

Exactly as HO_1, HO_2 proposes that affective phenomenal character has to do with the co-occurrence of two numerically distinct mental states, E and H, where the latter is a higher-order Command directing its subject to do something about E. The crucial difference is that while HO_1 proposes that it is H that has affective phenomenal character, HO_2 maintains that it is E that has it. Higher-order Commands *do not have* affective phenomenal character; rather, they *confer* it to the states that they target. On this view, an unpleasant experience E is an experience targeted by a Command with higher-order imperative content (20), while a pleasant experience E* is an experience targeted by a Command with higher-order imperative content (21):

(20) *Less of E!*

(21) *More of E*!*

Take this case. Louise is feeling her right index finger touching a piece of wood. Let's call this sensation 'S'. S does not have any affective phenomenal character. What would it take for S to be pleasant or unpleasant? According to HO_2, S should be targeted by another mental state, call it 'H', with higher-order imperative content. If H's content is *More of S!*, then S will feel good; if it is *Less of S!*, then S will feel bad.

HO_1 explained how affective phenomenal character has *intrinsic*, *mind-directed* motivational force. But it mischaracterised the *object* of that motivation, or, which is the same, failed to account for its *reflexivity*. HO_2 fares better in this respect. If the affective phenomenal character of an experience E depends on a higher-order Command that says *More of E!/Less of E!*, then it is the E *itself* that E's affective phenomenal character motivates us to have/not to have. Nevertheless, we have two misgivings about HO_2.

The good, the bad, and the neutral

Mental states with imperative content are often referred to in folk psychology as *urges*. For example, the urge to scratch, defecate or eat seem to be aptly described as states with first-order imperative content. But we also have urges to feel/not feel experiences—like when a smoker has a strong urge to *feel the sensation* of smoke rushing down her throat. These *experiential urges* are best understood as higher-order Commands directing us to have more or less of a certain experience. Accordingly, HO_2 predicts that if you are having the experiential urge to feel a sensation *while you are in fact feeling it*, the sensation will be pleasurable. Unfortunately for HO_2, it is very easy to disconfirm this prediction.

As any smoker knows, the following situation often happens. You are feeling the sensation of smoke hitting the back of your throat—call this sensation 'S'. *Exactly at the time in which you are having S*, you have the urge to feel S (that is, you are tokening a Command with higher-order imperative content *More of S!*). Still, S fails to be pleasurable. It might be entirely neutral. It might even be unpleasant. Either way, HO_2 runs into troubles.

This should not come as a surprise. HO_2 is, after all, rather like the desire theory (Heathwood 2007). Desire theorists say that a sensation is pleasant if it is targeted by

the intrinsic desire to have it, and it is unpleasant if it is targeted by the intrinsic desire not to have it. But this cannot be right. A celibate ascetic may feel sexual arousal arising unbidden, and desire it to end. The ascetic's desire is an instantiation of his deontic commitment to renounce the pleasures of the flesh, and so is paradigmatically intrinsic.[10] Perhaps the ascetic will feel some other unpleasant emotion of guilt or shame as a result. But the sensation of arousal itself may still be thoroughly pleasant. HO_2 differs from the desire theory in positing a higher-order *urge* in place of a higher-order *desire*. Still, the difficulty that they face is the same. One can have an urge or desire not to have an experience which is, in fact, thoroughly pleasant.

Pure affect

According to HO_2, it is *lower-order* experiences that have affective phenomenal character. In the case of pain, a first-order sensory experience *becomes* unpleasant in virtue of being targeted by a higher-order Command. But it is not clear that we *can* always characterise affective experiences in terms of an independent lower-order experience. What is the lower-order experience in the case of misery and depression that gets targeted by a higher-order Command? We're tempted to say—nothing at all. True enough, when one feels depressed, often all sorts of experiences become unpleasant. But depression and misery are also often experienced as a feeling of *pure unpleasantness*—some patients call it a 'black feeling'. They feel bad. And that's all there is to them.

[10] Heathwood might reject the claim that this is an *intrinsic* desire. But this is because Heathwood has a rather idiosyncratic account of intrinsicality. While the standard view has it that intrinsic desires are 'desires … for states of affairs that are wanted for themselves', (Schroeder 2017, section 2.2), Heathwood says that there must be '*no reason* you can give for wanting' (Heathwood 2007, 30, emphasis added) whatever you desire intrinsically—even if that reason points to an intrinsic property of the thing desired. He adds (*ibid*., 30, footnote 13) that if one desires something because it *exemplifies* a broader class that one has desires towards, then that desire counts as extrinsic. This leads him to the incredible conclusion that if we desire some sensation 'because it is pleasant' (*ibid.,* 38), that would be a case of *extrinsic* desire! Clearly Heathwood's account should be rejected. In any case, Heathwood's response can hardly help the higher-order *imperative* theorist—urges are never had for a reason, and so would *always* count as intrinsic for him. But it is possible to have an urge not to have a state and still not find that state unpleasant.

You may disagree with our portrayal of depression and misery. There is, nevertheless, good empirical evidence that pure affect occurs. The most famous case is probably the one described in Ploner et al. (1999):

> A 57 year-old male … suffered from a stroke … [resulting in] a lesion … comprising primary and secondary somatosensory cortices. … Thermonociceptive stimuli were applied by means of cutaneous laser stimulation. Pain thresholds were 200 mJ for right hand. Evoked pain sensations were characterized as 'pinprick-like' and were well localized. For left hand, up to an intensity of 600 mJ, no pain sensation could be elicited. However, at intensities of 350 mJ and more, the patient spontaneously described a 'clearly unpleasant' intensity … that he wanted to avoid. The patient … was completely unable to describe quality, localization and intensity of the stimulus. Suggestions from a word list containing 'warm', 'hot', 'cold', 'touch', 'burning', 'pinprick-like', 'slight-pain', 'moderate pain', and 'intense pain' were denied. Our results demonstrate … loss of pain sensation with preserved pain affect. (Ploner et al. 1999, 212-13)

The patient reported the occurrence of an experience with negative affective phenomenal character. However, no sensory phenomenal character whatsoever was reported. This is thus a case of pure affect, of an affective experience occurring in the absence of any lower-order sensory experience. HO_2 cannot deal with a case like this.

Wait a second! Klein might reply that even though no lower-order *experience* obtained here, the patient tokened some lower-order *state* nonetheless—an *unconscious somatosensory state*. This strikes us as *ad hoc* and implausible. There is no behavioural evidence in support of this hypothesis. Even worse, there is neural evidence *against* it: the patient's somatosensory cortex is severely damaged, so we should expect that he cannot token any somatosensory state, either conscious or unconscious. We conclude that HO_2 fails to capture the phenomenon of pure affect and is thus inadequate.

4.5 Taking stock

Higher-order imperativism fails. However, it does a better job than first-order imperativism in accounting for affective phenomenal character. This suggests that there is something to the idea that affective phenomenal character should be accounted for in terms of imperative contents directing their addressees to do something about their own mental states. Klein cashed out this idea in terms of higher-order imperative contents.

Our proposal is instead that affective phenomenal character has to be explained in terms of reflexive imperative content.

5. Reflexive imperativism

5.1 The theory introduced

Intentional states have objects. Indicative states represent their objects as being a certain way. Imperative states direct their subjects to do something about their objects. First-order intentional states have non-mental objects, and higher-order intentional states have mental objects *which are distinct from those self-same states*. This taxonomy leaves space for another kind of intentional state: intentional states whose objects are (at least in part) *themselves*—we can call these 'reflexive', or 'same-order' (Kriegel 2006), states.

The intuitive idea is this. A reflexive Indicator represents itself; a reflexive Command directs us to do something about itself. But the intuitive idea only gets you so far. Here is a more precise formulation of the notion of reflexive Command (given our aims, we do not need to dwell on reflexive Indicators): a reflexive Command K is a mental state with *reflexive imperative content*, i.e., a content directing its addressee to do something about the mental state M of which K is a *constitutive part*. Since everything is part of itself, it goes without saying that if M has no constituent other than K (i.e., if M = K), then K's reflexive imperative content will direct its addressee to do something about K itself only.

Reflexive imperative content, we maintain, allows the imperativist to adequately explain affective phenomenal character. Our proposal is as follows:

> **Reflexive imperativism**: An experience E of a subject S has affective phenomenal character in virtue of being (at least partly) constituted by a Command K with reflexive imperative content (i.e., a content directing S to do something about the mental state of which K is a constitutive part—thus, about E).

In particular, an experience P is pleasant in virtue of being (at least partly) constituted by a Command K+ with reflexive imperative content (22), and experience U is unpleasant in virtue of being (at least partly) constituted by a Command K- with reflexive imperative content (23):

(22) *More of the experience of which K+ is a constitutive part!* (that is, (1) *More of P!*)

(23) *Less of the experience of which K- is a constitutive part!* (that is, (2) *Less of U!*)

In a nutshell, an experience is pleasant/unpleasant in virtue of commanding us: *More/less of me!*

After all these pages, George has still his right hand under scalding water and is feeling an unpleasant pain because of this. According to higher-order imperativism, George is having two numerically distinct mental states at the same time: a first-order sensory experience, S, accompanied by a higher-order Command targeting S. Our proposal is instead that George is having *one compound* experience—call this experience 'U'. One of U's constituents is an Indicator F with first-order indicative content:[11]

(7) *There is a burning disturbance in your right hand*

But U has another constitutive part, namely Command K-, with reflexive imperative content (23):

(23) *Less of the experience of which K- is a constitutive part!* (that is, (2) *Less of U!*)

[11] We are open to the option that F might be a first-order Command. It might even be a compound state in itself, made up by a first-order Indicator *plus* a first-order Command.

Thus, George's experience (namely, U) is a compound mental state made up by a first-order Indicator F *conjoined with* a reflexive Command K- (that is U = F + K-). The sensory phenomenal character of U depends on F's indicative content (7), while U's affective phenomenal character is determined by K-'s reflexive imperative content (23).

The points above apply to pleasant experiences. What is it for you to have a pleasant bodily sensation in the neck (call it 'P')? It is for you to instantiate a compound experience made up by a first-order indicator with indicative content (24) conjoined to a Command K+ with reflexive imperative content (22):

(24) *Your neck is in such and such condition*

(22) *More of the experience of which K+ is a constitutive part!* (that is, (1) *More of P!*)

(24) determines P's sensory phenomenal character, while (22) determines its affective phenomenal character.

Now that you have an idea of what reflexive imperativism is, we can move to a more pressing question: Why should you believe it? Well, to begin with, it solves all the difficulties faced by first-order and higher-order imperativism.[12]

5.2 Problem solving

Reflexive imperativism has the capacity to provide the right taxonomy of affective experiences. Like higher-order imperativism, it needn't suppose that all experiences with first-order imperative content—like mild hunger or slight itches—have affective phenomenal character. These first-order Commands simply are not conjoined with reflexive Commands. Also like higher-order imperativism, reflexive imperativism

[12] One might say that it is incorrect to characterise reflexive, or same-order, imperativism as an alternative to higher-order imperativism. What we call 'reflexive imperative contents' are in fact higher-order contents: they direct one to do something about one's own *mental states*. Hence, it would be more appropriate to conceive of our proposal as a version of higher-order imperativism, and label it, say, 'reflexive higher-order imperativism'. The point is moot. If it is true that one can see same-order theories as variants of higher-order ones, it is also true that everybody agrees that there is an important distinction between them: *standard* higher-order theories posit two distinct mental states, one *targeting the other*; reflexive theories instead maintain that there is one mental state *targeting itself* (Kriegel 2006). As we show in section 5.2, this apparently small distinction makes a huge *explanatory difference*.

allows that any *kind* of experience could, in principle, have affective phenomenal character. Since all it takes for an experience to have affective phenomenal character is for the experience to be (at least partly) constituted by a reflexive Command, the theory has the ability to accommodate as extensive array of affective experiences as you please.

Unlike higher-order imperativism, reflexive imperativism does a good job of explaining pure affect cases. Normally, a pain experience is a conjunction of a first-order sensory experience and a reflexive Command. These constituents can doubly dissociate. In pain asymbolia, the reflexive Command is missing. This is why asymbolic pain has sensory phenomenal character, but not affective phenomenal character (Berthier et al. 1990). In the case of Ploner et al.'s patient, it is the other constituent that is missing: the patient's experience has *no constituent other than* a reflexive Command. Such experience says nothing more than *Get less of me!* This is why it has affective phenomenal character only. Analogously, if we think that certain forms of misery and depression have no phenomenal character beyond feeling awful, this allows us to account for them too.

Reflexive imperativism also manages to deal with a phenomenon which we have not as yet dwelt upon much. In one sense, all pleasant experiences have something in common, and diametrically opposed to the common feature of all unpleasant experiences: we describe them as having a common phenomenology when we say that they feel good, or feel pleasant. At same time, pleasant experiences are diverse, heterogeneous. Heathwood (2007, 25) captures the latter point nicely: 'pleasure is a diverse phenomenon. There are bodily pleasures, like those had from sunbathing or from sexual activities. There are gustatory pleasures, etc. … There doesn't seem to be any one feeling common to all occasions on which we experience pleasure.' Our theory captures *both* these commonalities and differences. Much as all pleasant experiences are partly constituted by a 'positive' reflexive Command, they can differ in terms of their other constituents. It is in virtue of their shared reflexive Command that orgasms, happiness, and elation all count as pleasant experiences.[13] And it is in virtue of the

[13] Margot says: 'I am French'; Charlotte says: 'I am French'. Have they said the same thing? Yes, each of them has said of herself that she is French … and no—Margot has said that Margot is French; Charlotte has said that Charlotte is French. Something similar applies to reflexive

differences in the states which are *conjoined* to the Command that these experiences are so different. Thus, pleasure may be mental or sensational, simple or complex, precisely located or diffuse, informative or purely affective. The same applies to unpleasant experiences.[14]

Finally, reflexive imperativism explains why affective phenomenal character has intrinsic and reflexive motivational force: it explains why the affective phenomenal character of an experience E motivates us, all by itself, to have more/less of E. Consider an unpleasant experience U. Its unpleasantness depends on U's content *Less of U!* This content, being imperative, has *intrinsic* motivational force. Thus, we do not need to appeal to any other mental state to explain U's unpleasantness motivational force. Moreover, this content is *reflexive*: it commands us to get less of U itself. This is why U's unpleasantness has reflexive motivational force.

Commands. Token experience E_1 has token Command $K+_1$ as constituent, while token experience E_2 has Command $K+_2$. In a sense, these two Commands 'say' something different: $K+_1$ has imperative content *More of E_1!*; $K+_2$ has imperative content *More of E_2!* In another sense, these two Commands issue the same order: they both direct their addressees to produce more of the token experience of which they are a constitutive part. This is what is common to all 'positive' reflexive Commands and, according to us, what explains the phenomenological commonalities among all pleasant experiences.

[14] Some will be outraged. How could we say with a straight face that the only phenomenal difference between, say, an orgasm and tasting white wine consists in the different *non-affective* qualities making up the phenomenal character of these two experiences? Isn't it obvious that, in addition to those differences, it is also the case that an orgasm and tasting white wine *feel pleasant in a different way*, and thus differ with regard to their affective phenomenal character? As a matter of fact, it is not obvious at all. In fact, this strong claim about affective heterogeneity boils down to an un-argued intuition. By contrast, there are at least three good reasons to maintain that (un)pleasantness is phenomenally homogenous.

First, the neural correlates of (un)pleasantness are the same irrespectively of the type of affective experience (Berridge and Kringelbach 2015; Leknes and Tracey 2008).

Second, we often take decisions ('Should I go for A or for B?') based on calculating which outcome will give us the greatest pleasure/the less dis-pleasure (Gilbert and Wilson 2005). For this calculation to be possible, the pleasantness/unpleasantness of different experiences has to be *commensurable*. As it is sometime put, there should be a *common currency* circulating in our affective life (Levy and Glimcher 2012). This is explained by hypothesizing that affective experiences feel (un)pleasant *in the same way*.

Third, when we try to *describe* the difference between diverse (un)pleasant states, we only talk about their non-affective differences—we point to the difference between sensory and mental, or between distinct sense modalities, or concerning location and complexity, etc. This is exactly what our theory predicts.

6. But how? And why?

First-order imperativism proposes that an experience has affective phenomenal character in virtue of having first-order imperative content. Higher-order imperativism says that affective phenomenal character depends on the co-occurrence of two distinct mental states, one of them being a higher-order Command. Both views, we argued, face significant problems. We showed that all these problems can be solved at once by hypothesising that an experience has affective phenomenal character in virtue of being (at least partly) constituted by a Command with reflexive imperative content. This was, in a nutshell, our argument for reflexive imperativism.

Despite its explanatory virtues, there are two families of worries that our view is bound to attract. The first concerns its underlying metaphysics; the second has to do with the place of reflexive imperative content in the natural world. We consider them in turn.

6.1 Get it together!

You are feeling an unpleasant pain in your right foot. Both reflexive and higher-order imperativism say that you are tokening a first-order sensory experience *and* a Command. However, while Klein interprets the 'and' roughly as 'at the same time as' (so that your unpleasant pain consists of two co-occurring, but *numerically distinct*, mental states), we read it as 'conjoined with': your unpleasant pain is a *single*, but composite, mental state.

The semantic implications of this distinction are clear. Whereas Klein's view posits that the content of the Command is to get more/less of a state that is *distinct from* itself, according to our theory the Command tells one to get more/less of the experience of which it is a *part*, thereby targeting itself. And this is why our theory answers our third overarching question in a way that neither formulation of Klein's can:

> (Q3) In virtue of what does affective phenomenal character have intrinsic and reflexive motivational force?

According to HO_1, it is the Command that has affective phenomenal character. In that case, the motivational force of affective phenomenal character is not *reflexive*—it

motivates one to get more/less of a state that is *different from* the one that is pleasant/unpleasant. According to HO_2, it is the sensory state targeted by the Command that has affective phenomenal character. In this case, the motivational force of affective phenomenal character is not *intrinsic* to the affective experience—it entirely depends on a separate conative state, namely, an affectless Command. By contrast, if the affective phenomenal character of an experience E depends, as we argue, on E being (at least partly) constituted by a reflexive Command, then affective phenomenal character can be both reflexive (it motivates one pro/against E) and intrinsic to E (it does not depend on any conative state other than E itself).

This invites the following question: what is the difference between there being one single, complex state of this type, rather than two distinct, but co-occurring, simpler mental states? Since cognitive systems can be described at different levels, there is more than one way to address this question. For reasons of space, we confine ourselves to discuss it at the *syntactic* level (Pylyshyn 1984).

What is the difference between, on the one hand, believing that the sun is shining and believing that the sky is blue, and, on the other hand, believing that: the sun is shining *and* the sky is blue? In the first case, one is tokening two distinct mental representations in one's mind (say, #SUN-SHINE# and #SKY-BLUE #), while in the second case one is tokening a single, but more complex, mental representation (#SUN-SHINE & SKY-BLUE#). This is not loose talk. We take something along these lines to be *literally true* of human minds, as something along these lines is literally true of computers. In fact, something along these lines, we claim, not only is literally true of beliefs, desires, and the like, but of experiences as well, including affective experiences.

Accordingly, one way to articulate the disagreement between reflexive and higher-order imperativism is as follows: while the latter maintains that to have, say, an unpleasant pain is *just* to token two distinct mental representations *at the same time*, reflexive imperativism proposes that unpleasant pain is constituted by a single, complex, *conjunctive* representation with the syntactic form #F & K-#—where, as you already know, #K-# is a reflexive Command and #F# a first-order Indicator.

A difficulty still stands in the way. In order for you to believe *that the sun is shining and the sky is blue*, your mind needs to conjoin two mental representations of the same type—your mind has to put together two beliefs to generate another belief.

However, according to our view, for you to feel an unpleasant pain, your mind has to conjoin two different types of representation, namely, a Command and an Indicator. How do we know that this is possible?[15] Three quick and interconnected answers.

First, we follow our explanation where it leads. If the best account of affective experiences and their phenomenal character commits us to mixed indicative-imperative conjoined representations, then that is a commitment we are happy to make. In the absence of a demonstration that such representations are not possible, the explanatory power we gain justifies the commitment.

Second, these mixed representations do appear to be possible in a functionalist framework. An Indicator has the function of representing how things are; thus, it will be consumed by the belief system. A Command is poised to make an impact on one's motivational system. Accordingly, a mixed indicative-imperative representation is such that, all else being equal, it will have an impact on both one's beliefs and one's motivations. This is why, as a result of tokening an unpleasant pain, one typically ends up believing that there is something going wrong in one's body *and* in being motivated to get rid of the unpleasant experience.

Our third and final point is that similar mixed representations have already been introduced in the philosophical and psychological literature. Even leaving aside Millikan's (1995) pushmi-pullyu representations, a prominent view in the cognitive science of emotions is that the latter are complex states made up by a variety of indicative representations (appraisals, bodily perceptions, etc.) and a variety of imperative representations (motor commands, action tendencies, etc.) *all bound together* in a single, complex mental representation (see Seth and Friston 2016 for a review). In fact, we are not the first *imperativists* to have proposed the existence of mixed indicative-imperative representations. As we have seen in section 3, Martínez's first-order imperativism proposes that prototypical affective experiences are compound mental states made by an Indicator conjoined with a Command. The fundamental difference, of course, is that, *contra* Martínez, we argue that the Command has reflexive imperative content. Maybe, it is here where the real difficulty lies. Maybe, there is

[15] We thank the editors of *Mind* for pressing us on this point.

something particularly problematic about such reflexive Commands. It is to this issue that we now turn.

6.2 Reflexive imperative content naturalised

Recall that one of the attractive features of intentionalism is that it paves the way to naturalising phenomenal character. However, one might think that by attempting to explain affective phenomenal character in terms of reflexive imperative content we have traded one mystery for another. From a naturalistic point view, reflexive imperative content might appear just as puzzling as phenomenal character. '*How can* a mental representation have such a content?', we hear you asking.

Since this is a 'how-possible' question, we give a 'how-possible' answer. That is, we show that there are no obstacles *in principle* to a naturalistic psychosemantics for reflexive imperative content. To do this, we sketch a *toy* teleo-semantics. Importantly, we are *not* committed to it, or to teleo-semantics more generally. In fact, we intend our theory of affective phenomenal character to be as neutral about content-determination as possible. Our aim here is just to show that there is nothing mysterious or spooky about reflexive Commands. Here is a teleo-semantics for imperative content *in general*: Command K has imperative content *C* if and only if K has the biological function to make it the case that C. The passage to reflexive imperative content is straightforward: Command K- has the content *Less of the experience of which K- is a constitutive part!* if and only if K- has the biological function of producing less of the experience of which it is a constitutive part. The same applies, *mutatis mutandis*, to 'positive' reflexive imperative content.

This gives rise to a further question: *why* is there something with no function except to get more or less of itself (better, of the state of which it is a constituent)? What is the evolutionary advantage of reflexive Commands? Since we maintain that affective phenomenal character depends on these Commands, an answer to *that* question will also be an answer to the following question: what is the evolutionary advantage of affective phenomenal character? Or, if you prefer, why do we need experiences that feel good or bad? Couldn't nature just have endowed us with representations, desires, and affectless first-order imperatives like the urge to defecate? Why did we need to *suffer?* These are deep problems, and we shall not answer in detail here—not just because there is not

enough space, but also because nobody knows the exact details. Still, we can tell you a nice story.

Once upon a time, planet Earth was populated with quite simple-minded creatures—spiders, scorpions, flies, snails, etc. Each of them faced the following formidable tasks: get food, avoid predators, reproduce, and so forth. Still, given the *relative* non-flexibility of their behaviours, we might suppose that they performed these tasks in the absence of *complex* decision-making activities. Presumably, they could go by on the basis of a more or less fixed set of pre-programmed responses. These responses were more useful the more specific they were. Otherwise, they would have offered the organism little guidance.

As more complex creatures came into being, such specific responses ceased to be sufficient. Complex creatures had a greater number of goals than their simpler kin. They also had far more varied means at their disposal for achieving them. This gave them a vast array of sub-goals. And in many cases even basic goals had become so complex as to make it impossible to pre-program these creatures with instructions for dealing with the manifold challenges they faced. For example, their bodies were now capable of undergoing a nearly endless variety of damage, each calling for quite different courses of action in order to promote recovery. These creatures thus needed the capacity to *learn* what the best means to achieve a certain goal is. This is when affective phenomenal character (or, which is the same, reflexive Commands) kicked in.[16]

Affective phenomenal character works as a system of *reward* and *punishment*. A complex creature, call it 'CC', is in a certain predicament—say, its body has been damaged, or it is looking for a mate. CC does not possess in advance a solution to these problems: it has not been hardwired with a comprehensive set of instructions for how to

[16] Our theory thus ascribes experiences with affective phenomenal character only to creatures of *relative* cognitive complexity. This seems to generate a problem. Surely human infants can experience *at least some* affective experiences. But aren't they too cognitively unsophisticated to token mental states with reflexive imperative content? Quite the contrary. A large body of evidence indicates that the capacity to represent others' mental states is functional in human infants as young as 6 or 8 months (see Carruthers 2013 for a review). Since the meta-representations involved in mental states attributions are more complex than the kind of reflexive content we have introduced to explain affective phenomenal character, we maintain that there is nothing implausible in the claim that human infants can token reflexive Commands *from the very beginning.*

fix the damage, or for how to perform courtship behaviour. It is up to its thinking brain to work out what to do. CC needs to learn. This often takes trial and error.

CC has injured its ankle. It starts by trying to use the leg normally—*ouch!* An unpleasant experience obtains. It acts as a *punishment*. Its unpleasantness commands: *Less of this experience!* CC has *learned* that this is not the right way to go and looks for another strategy. It experiments with changing its gait as it walks, and adopts whichever solution brings about least unpleasantness. Over time it learns more sophisticated behaviours. It realises that sleeping in a certain position decreases unpleasantness in the morning, and accordingly does that, or that applying ice to the site of the injury makes things feel better. When, in the future, it has another injury of this apparent type, CC may re-enact the successful strategy and, *if the injury* is *indeed similar*, it will likely be successful again. If all else is *not* similar, a new learning process will begin.

Now CC needs a mate. It tries a strategy to attract one. This one doesn't work—so it tries something else. Hopefully, CC will eventually do the right thing. *Bingo!* It gets lucky, and CC is *rewarded* with a pleasant experience. The pleasant experience commands: *More of me!* CC will then re-enact the successful strategy in the future and, *all else being equal*, it will be successful again. If all else is *not* equal, a new learning process will begin.

CC has learned complex strategies for dealing with difficult predicaments, guided by pleasantness and unpleasantness. It is for these 'two sovereign masters … alone to point out what we ought to do, as well to determine what we shall do' (Bentham, 1789/1970, 11).

You might be still unconvinced. CC has twisted an ankle, and does not know how to take care of this. When CC puts weight on the ankle, it experiences an unpleasant pain U. In this way, CC learns not to *behave like that*. Isn't it thus natural to say that the *function* of U's unpleasantness is that of preventing CC from putting weight on the ankle? Accordingly, given that in teleo-semantics content is fixed by function, should not we conclude that U's unpleasantness depends on the first-order content *Don't put weight on your ankle!* rather than on a reflexive content? We should not.

As soon as we consider how many affective experiences are there, it becomes obvious that there is *no one* world-directed behaviour that affective phenomenal character has the function to produce. Even in such a simple case as the unpleasant

injured ankle, there are a *vast array* of appropriate behaviours that CC is liable to adopt. Thus, even though there is no doubt that affective phenomenal character *often* (but not *always*, see 3.2) brings about *one or another* world-directed behaviour, the causal relation between the two cannot be explained in terms of first-order imperative content (unless one wanted to cash out affective phenomenal character in terms of a very, very long disjunction of first-order contents—and, trust us, you do not want that!)

The right thing to say is instead that an experience's affective phenomenal character tells you *More/Less of this experience!* and leaves to you the task to figure out which behaviour, *if any*, can satisfy this request. Sometimes, this process leads to a world-directed behaviour. After all, getting rid of a cavity is a good way of getting rid of an unpleasant toothache. Other times, this process engenders a mind-directed behaviour, like taking a painkiller. In fact, *any behaviour* might arise, insofar as it culminates in getting more/less of your affective experience. This is the *only* thing affective phenomenal character cares about. It is a very self-centred character indeed.[17]

References

Aydede, M. (forthcoming). "A Contemporary Account of Sensory Pleasure," in Shapiro, L. (ed.), *Pleasure: A History*. Oxford University Press.

Bain, D. (2013). "What Makes Pains Unpleasant?" *Philosophical Studies*, 166(1): 69-89.

Bain, D. (forthcoming). "Why Take Painkillers?" *Noûs*, DOI: 10.1111/nous.12228

Bentham, J. (1789/1970). *An Introduction to the Principles of Morals and Legislation*, ed. J. H. Burns and H. L. A. Hart. Oxford: Clarendon Press.

[17] Earlier versions of this article were presented at the universities of Cambridge, Edinburgh, Glasgow, Helsinki, Milan, Sheffield, Stirling, and at the 2016 meeting of the *European Society for Philosophy and Psychology*. We are grateful to our audiences for their criticisms and suggestions. In addition to two anonymous referees for *Mind*, we would like to thank the following people for helping us to improve our ideas about affective phenomenal character: Francesco Antilici, Murat Aydede, David Bain, Simon Blackburn, Davide Bordini, Clotilde Calabi, Tom Cochrane, Niall Connolly, Jennifer Corns, Tim Crane, Fabio Del Prete, Jeremy Dunham, Jihi Huang, Rosanna Keefe, Colin Klein, Damiano La Manna, Jess Leech, Stéphane Lemaire, Jimmy Lenman, Jussi Palomaki, Christopher Peacocke, Patrizia Pecl, Kevin Reuter, Komarine Romdenh-Romluc, Jacopo Tarantino, Mark Textor, and Sandro Zucchi. We owe a special debt of gratitude to Steve Laurence, who read multiple drafts of the article and provided an incredible amount of insightful feedback.

Berridge, K. C., and Kringelbach, M: L. (2015). “Pleasure Systems in the Brain.” *Neuron*, 86(3): 646-64.

Berthier, M. L., Starkstein, S., Nogues, M. A., and Robinson, R. (1990). “Bilateral Sensory Seizures in a Patient with Pain Asymbolia,” *Annals of Neurology*, 27(1): 109.

Brady, M. (forthcoming). “Painfulness, Desire, and the Euthyphro Dilemma.” *American Philosophical Quarterly*.

Carruthers, P. (2013). “Mindreading in Infancy”, *Mind and Language*, 28(2): 141–172.

Carruthers, P. (2018). “Valence and Value.” *Philosophy and Phenomenological Research*, 97(3): 658-680.

Charlow, N. (2014). “The Meaning of Imperatives.” *Philosophy Compass*, 9: 540–555.

Corns, J. (2014). “Unpleasantness, Motivational Oomph, and Painfulness.” *Mind and Language*, 29 (2):238-254.

Cutter, B., and Tye, M. (2011). “Tracking Representationalism and the Painfulness of Pain.” *Philosophical Issues*, 21(1):90-109.

Cutter, B., and Tye, M. (2014). “Pain as Reasons: Why it is Rational to Kill the Messenger.” *Philosophical Quarterly*, 64(256): 423-433.

Dretske, F. (1995). *Naturalizing the Mind*. Cambridge, MA: Bradford Books/MIT Press.

Gilbert, D. & Wilson, T. (2005). “Affective Forecasting: Knowing What to Want”, *Current Directions in Psychological Science*, 14, 131–13

Hall, R. J. (2008). “If it Itches, Scratch!”, *Australasian Journal of Philosophy*, 86(4): 525–535.

Hanks, P. W. (2007). “The Content-Force Distinction”, *Philosophical Studies,* 134(2): 141-164.

Heathwood, C. (2007). “The Reduction of Sensory Pleasure to Desire.” *Philosophical Studies*, 133(1):23-44.

Hellie, B. (2009). “Representational Theories of Consciousness,” in T. Bayne et al. (Eds.) *Oxford Companion to Consciousness*. Oxford: Oxford University Press.

Klein, C. (2007). “An Imperative Theory of Pain.” *Journal of Philosophy*, 104(10): 517–532.

Klein, C. (2015). *What the Body Commands: The Imperative Theory of Pain*. Cambridge: MIT Press.

Kriegel, U. (2006). "The Same-Order Monitoring Theory of Consciousness," in U. Kriegel and K. Williford (eds.), *Self-Representational Approaches to Consciousness* (pp. 143-170). Cambridge MA: MIT Press.

Leknes, S. & Tracey, I. (2008). "A Common Neurobiology for Pain and Pleasure." *Nature Reviews Neuroscience*, 9, 314–320.

Levy, D. & Glimcher, P. (2012). "The Root of All Value: A Neural Common Currency for Choice", *Current Opinion in Neurobiology*, 22, 1027–1038.

Martínez, M. (2011) "Imperative Content and the Painfulness of Pain." *Phenomenology and the Cognitive Sciences*, 10:67-9

Martínez, M. (2015*a*) "Disgusting Smells and Imperativism." *Journal of Consciousness Studies*, 22(5-6):191-200

Martínez, M. (2015*b*) "Pains as Reasons." *Philosophical Studies*, 172(9):2261-2274.

Millikan, R. G. (1995). "Pushmi-Pullyu Representations.", *Philosophical Perspectives*, 9: 185-200.

Ploner, M., Freund, H. J., and Schnitzler, A. (1999). "Pain Affect without Pain Sensation in a Patient with a Postcentral Lesion." *Pain*, 81(1/2): 211–214.

Portner, P. (2007). "Imperatives and modals." *Natural Language Semantics*, 15(4):351–383.

Portner, P. (2016). "Imperatives", in M. Aloni and R. van Rooij, (eds). *The Cambridge Handbook of Formal Semantics*. Cambridge: Cambridge University Press.

Pylyshyn, Z. (1984). *Computation and Cognition: Toward a Foundation for Cognitive Science*. Cambridge, MA: MIT Press.

Rosenthal, D. (2005). *Consciousness and Mind*. Oxford University Press.

Seth, A.K., and Friston, K.J. (2016). "Active Interoceptive Inference and the Emotional Brain." *Philosophical Transactions of the Royal Society B: Biological Sciences*, 371(1708).

Schroeder, T. (2017) "Desire", in *The Stanford Encyclopedia of Philosophy (Summer 2017 Edition),* Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/sum2017/entries/desire/>

Shea, N. (2013). "Naturalising Representational Content", *Philosophy Compass*, 8(5):496-509.

Tye, M. (1995). *Ten Problems of Consciousness*. Cambridge, MA: Bradford Books/MIT Press.

Tye, M. (2005) "Another Look at Representationalism about Pain," in Murat Aydede (ed.), *Pain: New Essays on its Nature and the Methodology of its Study*. MIT Press.