The Intentional Content of Homeostatic Drives:

A Philosophical Analysis

Viktoriia Kononova

A thesis submitted in partial fulfilment of the requirements for the degree of

Doctor of Philosophy

The University of Sheffield

Faculty of Arts and Humanities

School of History, Philosophy and Digital Humanities

February 2025

Declaration

I, Viktoriia Kononova, confirm that the Thesis is my own work. I am aware of the University's Guidance on the Use of Unfair Means (www.sheffield.ac.uk/ssid/unfair-means). This work has not been previously been presented for an award at this, or any other, university.

Acknowledgments

The value of a PhD is difficult to condense into a few sentences. I have done a lot of research and had some good ideas. I have spent hundreds of hours discussing highly advanced subjects at seminars, conferences, and in the pub. I have vastly improved as a philosopher and a teacher. Yet, reflecting on it all, what stands out to me is how much I have grown *as a person*.

The most important lessons of my PhD were the ones I did not know I needed. Perseverance – in spite of the circumstances, but also in spite of myself. Kindness – to others and to myself. Courage – to ask the necessary questions, to be clear and honest, to expose myself to criticism, and not to take myself too seriously. To accept, in other words, that perfection does not exist, that seeking it distracts us from what is actually important, and that the absence of doubt would signal the death of real inquiry. My PhD journey has been a rollercoaster of emotions, equal parts joy and despair. Now that it is at an end, my main feeling is that of gratitude.

My deepest thanks go to my brilliant supervisors, Luca Barlassina and Dominic Gregory. Luca has an extraordinary talent for getting to the heart of any problem, and his easy humour, unwavering focus on the bigger picture, and no-nonsense approach to philosophy and life have both grounded and guided me. Dominic's astute suggestions and gentle encouragement have likewise kept me going. At different stages of my PhD, I also benefited greatly from the kind advice and mentorship of Rosanna Keefe, Jennifer Saul, and Gerardo Viera. And when the time finally came for my viva, I was fortunate to receive immensely helpful feedback from my examiners, Stephen Laurence and Manolo Martínez. Their insightful questions not only gave me a better perspective on my work, but they also inspired me to continue it.

It would be impossible to mention everyone who has made my time in Sheffield so special. Since my first day in the Philosophy Department, I have had a deep feeling of peace, belonging, and kinship. The research community there is warm, caring, fun, and driven — the perfect soil in which to thrive. The

administrative staff, too, have been more supportive than I can describe – thank you, Patrizia Baldi and Ed Matthews. Wherever I go, I will carry a piece of Sheffield in my heart. This is where I fell in love with philosophy and with my life partner (in that order); this is where I made friends from all sorts of backgrounds and found a home away from home. And I will never forget Mark Wilson of the University Health Service, who has helped me through some of the most challenging parts of my PhD.

Last but not least, I wish to express heartfelt gratitude to my family and especially to my parents, Dmitriy Kononov and Irina Kononova. They have done everything in their power to help me get where I am now. They read me books, and listened to me, and accepted me, and nurtured me, and let me do things my own way. They even asked about my research. Thank you, a thousand times.

And I thank my life partner, Jack Herbert, who has been by my side this entire time, has believed in me, has debated with me late at night, and has made me hundreds of excellent cups of coffee. I would not have been able to do this without you (or the coffee).

Abstract

This thesis explores the intentional content of *homeostatic drives (HDs),* a group of mental states that include, among others, pain, itch, hunger, thirst, nausea, breathlessness, fatigue, and temperature sensations. HDs, I argue, constitute a philosophically interesting cluster of bodily experiences united by common phenomenology and function; yet with the exception of pain, they have rarely been the subject of philosophical analysis and have not been seriously investigated as a group. I seek to rectify this by expanding the ongoing debate on the representational content of pain. My goal is to determine whether this debate has already yielded an explanation for *the entire class of HDs* – and, if not, to identify the direction in which the inquiry should proceed. To this end, I consider, in turn, whether the attribution of *purely indicative* or *purely imperative* content to HDs might account for the key features of these mental states. I show that although each theory gets something right, neither pure indicativism nor pure imperativism fully explains HDs. I then show that the most obvious solution, which is to treat HDs as *pushmi-pullyu* mental states whose content is at once indicative and imperative, is equally inadequate, since it inherits the problems of the pure theories but does not combine their strengths. Ultimately, I suggest that the pure theories and the pushmi-pullyu theory fail for the same reason: they assume that HDs are *simple* mental states with a single layer of intentional content. I recommend an *impure* and *complex* approach to HDs and offer a sketch of a model that satisfies these desiderata.

Table of Contents

Introduction

Our bodies are vulnerable in many ways. Bones can fracture, and muscles can tear. Skin can be pierced by sharp objects, scraped by rough surfaces, or burned by the sun. Wounds, in turn, pose the risk of infection and irreversible blood loss. Our energy reserves, however ample, are still finite, and the functioning of most bodily systems relies on a steady supply of nutrients. A lack of water can kill us in a few days; a lack of oxygen can do the job in a few minutes. Significant fluctuations in body temperature and blood glucose can be deadly, too. The list goes on.

Our brains address these vulnerabilities by constantly monitoring and adjusting physiological parameters in line with varying demands on them. This process of maintaining the optimal physiological condition of the body is known as *homeostatic regulation* (Berridge 2004; Billman 2020; Craig 2003; Critchley and Harrison 2013; Woods and Ramsay 2007). Much of it occurs outside our awareness and cognitive control: for example, hormone release, heart rate modulation, and the conversion of stored fat into energy. Yet homeostatic regulation also requires deliberate action on the part of the agent: for example, eating, drinking, and urinating (Pezzulo, Rigoli and Friston 2015). How do we know *when* to act, though? The answer is: through our *bodily experiences*.

Historically, bodily experiences have lain at the margins of philosophical inquiry. Recently, however, the alleged problem status of *pain* for an influential position in the philosophy of mind has thrust it into the spotlight. It is generally accepted that pain has phenomenal character, that it *feels a certain way*. At the same time, it has been repeatedly asserted – for example, by Colin McGinn (1996: 8) – that the sensation of pain has no intentional object, that it is not *about* anything.[1] If this assertion is correct, then pain constitutes a serious challenge to *representationalism*, the view that the phenomenal character of experiences is given by their representational (intentional) content.[2]

[1] For discussions of anti-representationalist claims, see Aydede (2001, 2009, 2023), Block (2006), Byrne (2001), Cutter (2017), Martínez (2011), and Tye (2006).

[2] For expositions of representationalism, see Byrne (2001), Dretske (1995, 2003), Lycan (1996), and Tye (1995b).

Quite a few sophisticated attempts have now been made to defend representationalism by showing that pain can, in fact, be explained in representational terms.[3] What remains to be determined, however, is whether any of the theories that have been put forward are superior to the others – and whether any of these theories can be successfully applied to *an entire class of pain-like mental states*. My thesis addresses these unresolved questions. Pain, I argue, shares its key characteristics with other *homeostatic drives (HDs)*: thirst, hunger, satiety, itch, breathlessness, nausea, temperature sensations, toilet urges, fatigue, and so forth. A unified theory of HDs is therefore warranted, and the first place to look for it is among the existing representational theories of pain.

The structure of the thesis is as follows. In Chapter 1, I introduce HDs as a major variant of *interoceptive experiences*. I begin by defining interoceptive experiences as those that relate to physiological, or homeostatic, parameters, and go on to show that while all interoceptive experiences exhibit bodily, or sensory, phenomenology, only some members of this class, or HDs, additionally exhibit affective and motivational phenomenology. The *distinctive phenomenology* of HDs, I further show, is likely dictated by their *distinctive function* of promoting, or driving, behaviours that restore homeostatic parameters to acceptable values. I conclude that HDs should be treated as an *investigative kind* and propose to explore them from the representationalist standpoint.

In Chapter 2, I evaluate *pure indicativism*. On this view, HDs represent *homeostatic events*, or the relevant changes in homeostatic parameters. After outlining the advantages of pure indicativism, I examine and defuse the common objections to this view. Following this, however, I argue that pure indicativism fails to account for the *motivational aspect* of HDs. Based on this, I conclude that HDs may have *some* indicative content, but their content is not purely indicative.

In Chapter 3, I consider *pure imperativism* as an alternative to pure indicativism. On this view, HDs do not inform us about homeostatic events. Instead, they *command* us to carry out homeostatic behaviours. I

[3] For recent representational accounts of pain, see Bain (2003, 2007, 2011), Cutter and Tye (2011, 2014), Dretske (1995), Grahek (2007), Klein (2007, 2012, 2015), Martínez (2011, 2015), Martínez and Klein (2016), and Tye (1995a, 1995b, 2006).

argue that the foundations of pure imperativism are flawed and that on top of other issues, it particularly struggles to accommodate the bodily (sensory) phenomenology of HDs. After addressing two possible rejoinders, I conclude that the content of HDs is not purely imperative.

In Chapter 4, I move on to *impure accounts* and ask if HDs are *pushmi-pullyu representations*, i.e. if they *simultaneously* inform us about homeostatic events and command us to engage in appropriate homeostatic behaviours. After formulating this view, which has been repeatedly described as promising without having been explicitly defended in the philosophical literature, I answer in the negative. The pushmi-pullyu account, I argue, does not only inherit the problems of pure indicativism and pure imperativism, but also suffers from problems unique to it. The lesson here, I suggest, is that we should stop treating HDs as representations with a single layer of content (indicative, imperative, or pushmi-pullyu). Instead, we should go both *impure* and *complex*.

In the extended conclusion, I lay out avenues for future research and offer a sketch of such an impure and complex account of HDs. I suggest that, having cleared the field and analysed the mistakes of the simple theories, we have improved the prospects for understanding the nature of HDs.

Chapter 1:

Homeostatic Drives as an Investigative Kind

Abstract:

In this chapter, I make a case for seeking a unified explanation for a range of bodily experiences. I begin by highlighting the inconsistent and insufficient engagement with this subject in contemporary philosophy. In particular, I note that although the puzzling features of pain have been widely discussed in recent years, *pain-like mental states* such as hunger, satiety, thirst, itch, breathlessness, fatigue, nausea, toilet urges, and temperature sensations have rarely entered these discussions. I submit that this has been a mistake: these phenomena should be addressed together as a distinct and philosophically interesting subset of *interoceptive experiences*. Defining the latter as experiences relating to the physiological state of the body, I argue that all of them are felt as sensations in the body, but some are also felt as pleasant or unpleasant, and as motivating body-directed behaviours. This distinctive phenomenological profile of some interoceptive experiences, I further argue, is dictated by their distinctive function of promoting, or *driving*, one's physiological, or *homeostatic,* needs. I suggest we refer to them as *homeostatic drives, or HDs*, and analyse them in representational terms.

1. Framing the question

Humans undergo a wide range of experiences that may be classified as *bodily*. We feel the position of our body and its parts in space and relative to each other. We perceive vibration, pressure, temperature, the distension of internal organs, and the stretching of muscles and skin. The sensations of breathing and heartbeat are our constant – if mostly quiet – companions, and each day brings episodes of thirst, hunger, toilet urges, fatigue, or sleepiness. On bad days, we may feel faint, short of breath, nauseous, feverish, sore, achy, or stiff. On good days, we enjoy pleasant touch and orgasms. On top of this, there are

miscellaneous tickling, tingling, crawling, and 'pins and needles' sensations, as well as sensations of skin, eye, and mouth dryness. Some bodily experiences are rare (e.g. floating) or difficult to put into words. And others, such as feelings of 'enlargement', 'shrinking', or 'splitting' of body parts, arise only in the context of neurological or psychiatric disorders (Röhricht and Priebe 2002; Todd 1955).

The diversity of bodily experiences is sharply contrasted by the paucity of philosophical inquiry into them. The only real exception here is *pain*, which has recently emerged as the subject of spirited debates within the philosophy of mind.[4] Given that it presents us with several puzzles, it is not surprising that pain should have attracted as much attention as it has. We want to know why pain has a bodily location, why it is unpleasant, and why it motivates us to act. We want to understand why pain seems to be incorrigible and objective, yet medical professionals often struggle to find adequate explanations for its intensity and persistence. Above all, we want to settle the issue of the *representational content* of pain.[5] What *is* surprising, however, is the lack of equally vigorous theorising about other bodily experiences that, I argue, strongly resemble pain in all significant aspects. For reasons that will be specified in due course, I call these bodily experiences ***homeostatic drives (HDs)***. In addition to pain, examples of HDs include, but are not limited to, itch, hunger, satiety, thirst, nausea, breathlessness (dyspnoea, 'air hunger', or shortness of breath), fatigue, toilet (bladder and bowel) urges, and temperature sensations. [6]

In this thesis, I aim to advance our understanding of previously neglected bodily experiences by working on a *unified* account of HDs. In this chapter, I show that such a project is justified: HDs share key

---

[4] While it had not been entirely neglected prior to that (see Armstrong 1962; Cornman 1977; Graham and Stephens 1985; Pitcher 1970), the past three decades have witnessed a veritable explosion of philosophical engagement with pain (see Aydede 2001, 2009, 2023; Aydede and Fulkerson 2014; Bain 2003, 2007, 2011, 2013, 2014; Barlassina and Hayward 2019; Casser 2021; Corns 2014a, 2014b, 2016; Cutter and Tye 2011, 2014; Dretske 1995; Grahek 2007; Klein 2007, 2012, 2015; Martínez 2011, 2015; Martínez and Klein 2016; Tye 1995a, 1995b, 2006; De Vignemont 2017).

[5] Representationalism is the view that 'the phenomenal character of an experience supervenes on its representational content' (Cutter and Tye 2011: 90). While many authors differ only in their specific take on the representational content of pain (Armstrong 1962; Bain 2003, 2007; Cutter and Tye 2011; Dretske 1995; Grahek 2007; Klein 2007, 2012, 2015; Martínez 2015; Martínez and Klein 2016; Pitcher 1970; Tye 1995a, 1995b, 2006), some have denied that pain can be explained (or that it can be fully explained) in representational terms (see Aydede 2001, 2009, 2023; Block 2006; Maund 2006; McGinn 1996).

[6] While these experiences have been grouped together before (Fulkerson 2023; Klein 2015), they remain poorly understood and rarely discussed, especially *as a group*. Hall (2008) argues that itch has imperative content. Fulkerson (2021) focuses on thirst in his analysis of *'sensory motivations'*. Armstrong (1962) and Klein (2015) defend theories intended to be applicable to many types of HDs, but both are mainly concerned with explaining pain.

characteristics and thus warrant unified investigation. The position I defend is that HDs fall under the umbrella of ***interoceptive experiences***, of which they make up *a major variant*. I begin by introducing the notion of interoceptive experiences. Then, I outline the reasons for regarding them as products of the sensory modality that tracks physiological (homeostatic) parameters such as fluid and oxygen levels. Next, I argue that while all interoceptive experiences relate to homeostatic parameters and have bodily (sensory) phenomenology, those that I call *homeostatic drives (HDs)* have additional characteristics that set them apart from *non-HDs*. First, HDs are not only felt as (i) sensations in the body, but also as (ii) pleasant or unpleasant and (iii) motivating us to engage in specific body-directed behaviours. That is to say, HDs have (i) sensory, (ii) affective, and (iii) motivational phenomenology. Second, HDs have the same *distinctive function*, which is to promote homeostatic behaviours, i.e. behaviours that contribute to the stability of homeostatic parameters. I argue that the distinctive phenomenology of HDs is dictated by their distinctive function, and that HDs are therefore worth investigating together, as a group. I wrap up the chapter by outlining the aims and strategy I adopt in my investigation.

2. The definition of interoception

As noted at the outset, bodily experiences are a mixed bag – and a full bag, at that. Yet their varied phenomenology and sheer number do not exhaust the reasons why *bodily senses* have been relatively neglected by philosophers compared to *worldly senses* like vision (de Vignemont 2018: 66). What serves as a further challenge in the study of bodily experiences is that the sensory data on the basis of which they are formed appears to be (1) transduced by a multitude of receptor types (including non-specialised) scattered throughout the body and (2) conveyed to the brain via multiple central and peripheral nervous system and humoral pathways (Aziz and Ruffle 2019; Feldman, Bliss-Moreau and Lindquist 2024; Ritchie and Carruthers 2015; Wang and Chang 2023). In light of such complexity of processing, classifying bodily experiences cannot be a straightforward task – especially if we take into account the prevalence of non-specialised receptors and the ubiquity of multisensory integration (Köteles, 2021, pp. 44–45).[7]

[7] See Fulkerson (2014) for an argument in favour of *'sensory pluralism'* and for an analysis of the challenges involved in classifying the temperature sense.

The situation, however, is not too dire. While bodily experiences are not easy to sort into neat clusters in a way that would satisfy everyone, considerable progress has been made in the field in recent years, with selected categories of bodily experiences securing the endorsement of scientists and philosophers alike (Carvalho and Damasio 2021; de Vignemont 2018: 5; Garfinkel, Critchley and Pollatos 2017; Macpherson 2011; Ritchie and Carruthers 2015). First, there are *the proprioceptive experiences*, such as the experience of crossing one's arms. These concern the position of the body and its parts relative to each other and their surroundings and come to the forefront of our awareness when we walk in the dark (Macpherson 2011: 15–16). Second, there are *the equilibrioceptive experiences*, such as the experience of rotational movement. These have to do with the orientation of the head in relation to the gravitational field and become pronounced when we step off a merry-go-round or a boat (Macpherson 2011: 16–17).[8] Finally, there are *the interoceptive experiences;* what are they?

Had this question been posed two or three decades ago, it would not have elicited the answer which will be presented here. Indeed, until then, the term '*interoception*' had merely referred to *visceroception*, or the processing of the sensory input from the internal organs by the brain (Cameron 2002: 4; Köteles 2021: §1.2). On this narrow view of interoception, only *visceral experiences,* or the experiences felt inside the body, below the surface of the skin (e.g. gastric, cardiac, respiratory sensations) had been designated as interoceptive, while *cutaneous experiences*, i.e. any experiences felt on the surface of the skin (e.g. sensations of touch, pressure, temperature, itches, cutaneous pains), had been automatically excluded from this group (Ceunen, Vlaeyen and Van Diest 2016; Craig 2015: 3). This had entailed, inter alia, treating (visceral) stomach or kidney pains and (cutaneous) pains from a scratch or a burn as products of *separate sensory modalities*, an interoceptive and a non-interoceptive one (Ritchie and Carruthers 2015: 353).

In recent years, however, there has been a major shift in perspective. Following the wide dissemination of the groundbreaking neuroanatomical discoveries by Arthur D. (Bud) Craig (1996, 2002, 2003, 2015) and

[8] See Wong (2017) for the suggestion that equilibrioception is a separate sense which makes unique contributions towards anchoring the *self* to the *body*.

the surge of research and discussion they had inspired, interoception has acquired a more nuanced definition (Ceunen, Vlaeyen and Van Diest 2016; Chen and others 2021; Garfinkel, Critchley and Pollatos 2017; Tsakiris and Critchley 2016). In today's usage, this term typically encompasses the processing of *all* of the sensory data relating to *homeostatic parameters*, that is, the physiological parameters which must be kept within certain ranges to guarantee survival (Cameron 2002; Carvalho and Damasio 2021; Khalsa and others 2018; Seth 2013). This is because, unlike the narrow view outlined above, the modern *inclusive view* of interoception puts the emphasis not on the *source* of the sensory input (viscera or skin), but on its use by the brain – that is, on the *function* for the implementation of which the input is gathered. Since interoception is thought to underpin *the regulation of homeostatic parameters* – a task which requires accurate knowledge of homeostatic goings-on – it can involve the processing of any type of bodily data relevant to this task, regardless of whether the data is obtained via the sensors located in the internal organs, skin, muscles, or joints (Ceunen, Vlaeyen and Van Diest 2016: 7; Chen and others 2021: 4; Craig 2008). In Craig's words (2015: 33):

> *The ANS [Autonomic Nervous System] is a complicated system with a very well-defined anatomical organization, and it controls a broad variety of peripheral organs efficiently and reliably. […] Such a highly evolved control system certainly requires ongoing sensory information about the somatic tissues and organs it controls.*

The modern understanding of interoception as the perception of the body's physiological state provides us with an alternative way to categorise bodily experiences – one which does not entail treating superficial (cutaneous) and deep-in-the-body (visceral) pains as products of different modalities. From the inclusive standpoint, the dissimilarities between cutaneous and visceral experiences do not reach the level that would warrant the construction of *separate kinds*. Instead, *all* the experiences associated with the physiological processes taking place in a human body may be taken to form *a unified interoceptive kind*, including regular breathing sensations and the feeling of breathlessness, itches, pains and aches, heartbeat sensations, thirst and hunger, the sensations which accompany swallowing and the movement

of food and fluids along the digestive tract, toilet urges, fatigue, nausea, temperature sensations, and so on (Craig 2002, 2008, 2015; Feldman, Bliss-Moreau and Lindquist 2024; Garfinkel, Critchley and Pollatos 2017; Köteles 2021; Strigo and Craig 2016; Wang and Chang 2023).

In this thesis, I adopt the inclusive understanding of interoception. Accordingly, whenever I say *'interoceptive experiences'*, I am referring to both cutaneous and visceral experiences that may be reasonably construed as *homeostatic*, i.e. related to homeostatic parameters, arising usually as a result of *fluctuations* in these parameters (Ceunen, Vlaeyen and Van Diest 2016: 8; Köteles 2021: 48).[9] Later in this section, I discuss this in more detail, but a good rule of thumb for identifying such experiences is to think, like in the paragraph above, about the processes involved in the daily maintenance of our bodies, from digestion and excretion to the contractions of the heart.

At the same time, I propose that the set of interoceptive experiences can be individuated *further*. Specifically, I offer homeostatic drives (HDs) as a *distinct variant* of interoceptive experiences that support homeostatic regulation by *promoting homeostatic behaviours*. HDs, I propose, are interoceptive experiences that have evolved to fulfil a specific homeostatic function, and this is why, in addition to the phenomenal features shared by all the members of the interoceptive kind, they have additional phenomenal features not shared by other members of this kind. The arguments for treating HDs as a subkind of interoceptive experiences are laid out in Sections 4–5 of this chapter. Before I turn to them, however, I defend, in Section 3, the cornerstone of my proposal – namely, the idea that interoceptive (homeostatic) experiences do form a unified kind in virtue of belonging to the same interoceptive modality and sharing joint phenomenology. So far, I have only introduced this idea and pointed out that many contemporary neuroscientists would agree with it. In what follows, I expound on the evidence behind it and, in doing so, build the foundation for my overall project of launching a joint investigation into the nature of HDs.

[9] To illustrate, although the 'normal' fluid levels are not felt, dehydration produces the feeling of thirst.

3. Interoceptive experiences as a unified kind

In this section, I justify my decision to treat bodily experiences relating to homeostatic parameters as a unified interoceptive kind. I begin by presenting (1) *the argument from joint processing*. According to this argument, the sensory data used in the generation of the experiences labelled here as interoceptive is processed in the same way, and the best explanation for such uniformity of processing is the generation of all these experiences by a specialised sensory system tracking the state of homeostatic parameters. Next, I present (2) *the argument from joint phenomenology*. According to this argument, the experiences labelled here as interoceptive share the same bodily phenomenology: they are all felt as sensations in the body. Given that this was famously contested by David Armstrong (1962), I devote a considerable part of this section to countering Armstrong's points. At the end of it, I conclude that there are good reasons to accept the existence of a unified interoceptive kind, though I also suggest that interoceptive experiences come in two kinds, one of which has the specific function of promoting homeostatic behaviours and would be of particular interest to philosophers of mind.

3.1. The argument from joint processing

As explained earlier, the redefinition of interoception as the perception of the body's physiological state has been mainly driven by the developments in the neurosciences. Let us now take a closer look at these developments and the interpretations they have received. For a long time, there was general agreement among neuroscientists that the sensory signals coming from the skin are processed together and in a manner *distinct* from the processing of the sensory signals coming from the viscera (Ceunen, Vlaeyen and Van Diest 2016; Craig 2015: 23–24).

Yet neuroanatomical findings by Craig and other researchers have shown this assumption to have been incorrect (Berntson and Khalsa 2021; Carvalho and Damasio 2021; Strigo and Craig 2016). Instead, it has been determined that the sensory data that forms the basis for discriminative touch and proprioceptive

experiences travels to the brain via the *dorsal column medial lemniscal pathway*, while the sensory data that forms the basis for the cutaneous experiences of pain, temperature, itch, and sensual touch, as well as the experiences of muscle ache travels to the brain via the morphologically distinct *spinothalamic pathway* (Andrew and Craig 2001; Craig 2015: 3–4, 23–24, 302).[10] In the brain, the dorsal column medial lemniscal pathway terminates in the somatosensory cortex, but the spinothalamic pathway ascends instead to the *insular cortex,* the region which also receives visceral sensory input from the various tissues and organs of the body via *non-spinal* routes, such as the vagus nerve and the direct, in-the-brain sensing of any chemical substances that can permeate the blood-brain barrier (Aziz and Ruffle 2019; Craig 2015: ch. 5; Köteles 2021: 46–47; Tsakiris and Critchley 2016; Wang and Chang 2023: 17.4–17.5).[11] While the insular cortex has been implicated in a whole range of cognitive functions, one of its major roles is believed to be the integration of interoceptive information in conscious interoceptive experiences, as confirmed by the sustained activity of this region during pain and itch, non-painful distension of the stomach and the gut, focusing one's attention on heartbeat sensations, breathlessness, exercise, temperature changes, erotic sensations, hunger, thirst, and nausea (Chen and others 2021: 156–7; Garfinkel, Critchley and Pollatos 2017: 435; Craig 2015: 177, ch.6). In line with this, the insular cortex is commonly described as *'the interoceptive cortex'* and one of the key structures for *'interoceptive awareness'* (Craig 2002, 2009, 2015; Berntson and Khalsa 2021; Feldman, Bliss-Moreau and Lindquist 2024; Livneh and Andermann 2021).

For Craig and many others, the broadcasting of various types of bodily data from all the tissues and organs of the body to a dedicated cortical region via dedicated sensory pathways has served as evidence for the *common homeostatic nature* of this data (Berntson and Khalsa 2021; Ceunen, Vlaeyen and Van Diest 2016; Craig 2002, 2009, 2015; Ritchie and Carruthers 2015; Quigley and others 2021). Call it *the argument from joint processing*. According to this argument, the joint handling of some sensory input by what appears to be *a specialised sensory system* must be the result of evolutionary design, and evolutionary design, in turn, must be necessitated by the common use of the input in the service of the

[10] This was confirmed by lesion studies: *'surgical interruption of the spinothalamic pathway […] can reduce or eliminate contralateral pain and temperature sensations yet leave discriminative touch and the sense of limb position intact'* (Craig 2015: 32).

[11] While there are many brain structures implicated in the processing of the sensory data consolidated in the insular cortex, given the aims of this section, it is not necessary to list all these structures and their interactions here.

same *function* (e.g. olfaction, vision). When it comes to the diverse sensory inputs on the basis of which the experiences of itch, temperature, breathing, cutaneous and visceral pain, muscle ache, hunger, nausea, and heartbeat sensations are generated, the best candidate for the reason why a sensory system should gather all this input is *the homeostatic function*, i.e. the tracking of homeostatic parameters. To use an analogy, if a group of students consistently attends only philosophy lectures, we may reasonably conclude that they are philosophy students. By the same token, if a group of sensory data is processed in the same way, and each type of data relates to the physiological body state, we may reasonably conclude that this data is homeostatic, i.e. that the receptors transducing it have evolved to respond to the stimuli from which the data about homeostatic fluctuations may be successfully extracted.

While the argument from joint processing has solid neuroanatomical foundations, it faces a potential objection with respect to its *interpretation* of the empirical findings. Specifically, the treatment of bodily data processed together as homeostatic could be contested on the grounds that visceral experiences arise in response to changes in the *internal milieu*, whereas cutaneous experiences are produced in response to *external stimulation* (Ceunen, Vlaeyen and Van Diest 2016). A sceptic could point out that while heartbeat sensations are caused by heartbeats and nausea is caused by indigestion, temperature sensations are triggered by changes in the external environment, cutaneous pains are triggered by contact with noxious stimuli (such as excessive heat or sharp objects), and itches are triggered by contact with irritants. Based on this, a sceptic could surmise that temperature sensations, cutaneous pains, and itches are the products of cutaneous senses tracking the properties of the external environment as opposed to the products of the interoceptive sense tracking the homeostatic properties of the body.

This objection, however, does not hold. While cutaneous pains are more sharply localised than visceral pains due to the skin being innervated more densely than the viscera (Garfinkel, Critchley and Pollatos 2017; Strigo and others 2002), *all* pains – from pinprick to muscle cramps and stomach aches – share a common felt quality of *painfulness* which enables us to distinguish them from other bodily experiences. With this in mind, it would be best to treat cutaneous and visceral pains as products of the same sensory

system tracking similar homeostatic properties. Otherwise, we would have to accept the implausible conclusion that sensory systems tracking different (external and internal) properties nonetheless produce experiences with the same phenomenal quality.

Second, the skin is not only the point of contact between our bodies and the external world, but also an important bodily organ, damage to which can result in infection and death. We may reasonably expect, therefore, that the brain would have evolved to gather information about the physiological condition of the skin – and the sensory information which gives rise to cutaneous pains, temperature sensations, and itches certainly qualifies as such. Indeed, cutaneous pains follow skin damage from scratches and burns, temperature sensations follow changes in skin temperature (which precede the changes in the internal body temperature and thus provide early warning), and itches follow skin irritation from insect bites or poisonous plants.

Third, our bodies do not exist in a vacuum but are in constant contact with the world. To illustrate, homeostatic events such as tissue damage and suffocation are often causally dependent on worldly objects or events, such as sharp objects and the lack of oxygen in the air as a result of a fire. This means that homeostatic data gathered by a sensory system tracking homeostatic parameters could also contain *inferable* information about the environment – information that other cognitive systems could exploit.[12]

I conclude, therefore, that the convergence of sensory data from all the tissues and organs of the body in the insular cortex suggests the existence of a sensory system optimised for tracking homeostatic parameters. From now on, I will refer to this system as *the interoceptive system*. While it is likely constituted by a collection of systems, the precise cognitive architecture of interoception remains to be elucidated (Ritchie and Carruthers 2015). What hinders progress in this line of research is that, owing to the nature of the work it has evolved to undertake, the organisation and function of the putative

[12] Another weakness of this objection is that it can be easily turned on its head. Indeed, at least some visceral experiences – not *only* cutaneous experiences – may be interpreted as non-homeostatic, formed on the basis of sensory data signalling external rather than bodily events. For example, the sensation of stomach fullness may be seen as being formed on the basis of the data about the amount of food in the stomach – and food is external to the body (see Köteles 2021: 3–4).

interoceptive modality do not appear to be very similar to those of the more-studied exteroceptive modalities (Carvalho and Damasio 2021). For one thing, visual or auditory information does not have to travel far to reach the brain – but interoceptive information does. Under selective pressure to reduce the metabolic costs of long-distance data transmission, the axons of interoceptive neurons have become thinner, which may be more adaptive overall, but also necessitates *'lossy'* data compression, thus imposing constraints on the speed of transmission (Feldman, Bliss-Moreau and Lindquist 2024: 651–52). And with its sensory input somewhat less reliable and not updated as promptly as the input received by, say, the visual system, the interoceptive system has been theorised by some to rely more heavily on its own *predictions* about the upcoming changes in the input than any exteroceptive system (Barrett and Simmons 2015; Feldman, Bliss-Moreau and Lindquist 2024; Seth 2013).

The starkest feature of the interoceptive system, however, is its involvement in the *autonomic regulation* of homeostatic parameters. While we are not aware of this, our brains exert constant and powerful influences on virtually every aspect of the bodily state, from breathing to energy metabolism and immune responses (Berntson and Khalsa 2021; Carvalho and Damasio 2021; Chen and others 2021; Pezzulo, Rigoli and Friston 2015; Quigley and others 2021). Far from being 'rigid' and 'preprogrammed', even the most basic of our homeostatic parameters such as the heart rate are subject to flexible and complex modulation by a network of brain regions including the insular cortex, as demonstrated by the fact that dysfunction of this cortical region (e.g. during a seizure) may induce fatal cardiac arrhythmia (Köteles 2021: 4; Oppenheimer and Cechetto 2016: 1081). Accordingly, given that effective regulation of fluctuating variables is unattainable without adequate sensory feedback about their current status, the interoceptive system must play a fundamental role in maintaining homeostatic parameters within acceptable ranges (Berridge 2004; Khalsa and others 2018). The bottom line is that whether or not there is a sharp division of labour between the systems that *sense* and the systems that *regulate* the physiological condition of the body – and, as Chen and others note (2021: 10) this question remains open – these systems must work closely together to guarantee our survival:

> *Much remains to be understood about the many brain regions involved in interoception, especially whether specific neuronal populations in these regions function as interpreters, integrators, or regulators of interoceptive information*.

In having the capacity to *control its own input*, the interoceptive system clearly stands out among other perceptual systems (Carvalho and Damasio 2021; Taggart and others 2016). To illustrate, there is nothing the olfactory system can do about a bad smell. Meanwhile, the interoceptive system (or a system working in tandem with it) can initiate shivering in response to decreasing body temperature, lower blood glucose in anticipation of a meal, and so forth (Woods and Ramsay 2007).

This difference is highly relevant to understanding interoceptive experiences, because if the interoceptive system is not set up in the same way as the more 'familiar' perceptual systems, then we should not assume that the experiences it generates are 'typical' perceptual experiences that represent some worldly or bodily property. That is to say, while it is *plausible* that interoceptive experiences are mere representations of homeostatic fluctuations, in light of philosophical claims to the contrary (see Klein 2015) and the peculiarities of interoception, it would not do to rush to this conclusion. For the purposes of this chapter, it is sufficient to show that experiences of hunger, thirst, itch, pain, heartbeats, breathing, stomach fullness, gut distension, nausea, and so on constitute *a family of experience*s insofar as they are all produced on the basis of computations carried out within the same sensory system. The precise nature of these experiences — or, more specifically, of HDs as their major variant — will be discussed at length in the chapters to come.

A final caveat here. A sceptic might argue that joint processing alone does not justify treating interoceptive experiences as a unified kind. After all, exteroceptive modalities are generally regarded as independent from each other and are discussed separately (Macpherson 2011). So why assume that interoceptive modalities are sufficiently homogenous to justify a different approach?

My answer is that interoceptive modalities do not appear to be as completely segregated or as well-defined as vision and hearing are. First, as I have already indicated, interoceptive modalities share the same interoceptive cortex. By contrast, visual and auditory cortices are located far apart. Second, the receptors supplying the interoceptive system with sensory data – i.e. *interoceptors* – may respond to more than one type of stimulation: thermal, mechanical, or chemical (Dubin and Patapoutian 2010; Köteles 2021: 44; Lumpkin and Caterina 2007). Thus, these receptors are difficult to classify as, say, hunger or pain receptors. Third, interoceptive experiences are often formed on the basis of the information supplied by different types of receptors and subsequently consolidated. Heartbeat sensations are a case in point: despite their simplicity, they necessitate the integration of sensory data from pressure receptors in the blood vessels, mechanoreceptors in the chest wall, and other types of receptors (Khalsa and others 2009). Pains and itches provide another striking example of interoceptive integration, since both can have *a burning quality*, a quality of temperature sensations. Taken together, these features of interoceptive processing imply a significant degree of cross-talk and shared processing within it, which is not consistent with a set of modalities working in isolation from each other. And this is not to mention a further reason to see interoceptive experiences as a kind. As I argue next, they share *bodily (sensory) phenomenology*.

3.2. The argument from joint phenomenology

Unlike the previous argument, which required a lengthy exposition, this argument is short. Here it is: all interoceptive experiences have bodily (sensory) phenomenology, i.e. they are all felt as sensations in the body. Heartbeat sensations, for example, are felt as pulsations on the left side of the ribcage. Breathlessness is felt as tightness in the chest and upper airways. Pain is felt in various organs and tissues of the body and may be described as sharp or dull, throbbing or shooting. And so on.

The problem here is that David Armstrong (1962, 1993) famously claimed that the experiences which I call interoceptive do not form a kind, precisely because they do not share bodily phenomenology. According to Armstrong, these experiences belong to two different kinds: namely, *'bodily sensations'* (e.g. pains,

itches, tickles, 'erotic sensations', heartbeat sensations, sensations of motion and distension of bodily organs) and *'bodily feelings'* (e.g. fatigue, hunger, thirst). Bodily sensations have a felt bodily location, whereas bodily feelings do not.[13] If Armstrong is right, then I am doubly wrong: not only do interoceptive experiences fail to form a kind, but neither do homeostatic drives (HDs), which I introduce as a variant of interoceptive experiences. Fortunately, it is Armstrong who is wrong.

At a glance, Armstrong's position is appealing enough. Experiences of pain, itch, and internal organ distension are clearly *localised*: take pinprick pains, mosquito-bite itches, or bladder fulness. These experiences direct our attention to specific body parts, prompting us to act on them: for example, to rub where it hurts, scratch where it itches, or empty the bladder. Conversely, an experience like fatigue may be described as *generalised*. Yet, to say that fatigue is generalised does not mean that it is *not located anywhere*. Rather, it means that fatigue is *diffuse*: poorly localised, often spread across several parts of the body. After all, when one is very fatigued, does one not feel weakness and heaviness in every part of the body that one tries to move or even thinks about moving? In fact, fatigue is not even *that* widespread, since it is confined to a particular organ system: the muscular system. As such, fatigue is far from being nowhere in particular: it is in the muscles, sometimes in many, sometimes in just a few.[14]

At this stage, one might still think that Armstrong's categorisation is useful on the grounds that we ought not to lump together *sharply localised* bodily experiences and those that are more *diffusely localised*. Even such a diluted version of Armstrong's view, however, would be hard to defend. On the one hand, what Armstrong calls 'bodily feelings' may be fairly sharply localised. Hunger, for example, may be felt as an emptiness in the stomach. On the other hand, what Armstrong calls 'bodily sensations' are not always sharply localised. Orgasms, for example, may be felt throughout the body (Vance and Wagner 1976).

---

[13] Armstrong (1962, 1993) also distinguishes between *transitive* and *intransitive* bodily sensations. The former (e.g. temperature sensations) are experiences of something other than themselves – in this case, the temperature of a body part – while the latter (e.g. pains, itches) are not.

[14] What about mental fatigue? I find this concept somewhat unclear, so I choose to omit it from the later discussion. We are certainly able to *infer* that we need a break from thinking (e.g. by judging how hard it is to focus). The question is whether mental fatigue possesses bodily phenomenology. I think it might, and its bodily phenomenology resembles that of sleepiness. Indeed, both experiences seem to be mainly located in the head. If they are generated on the basis of roughly the same computations, they might even constitute a single phenomenon.

Likewise, pains may be spread over ill-defined regions. Menstrual pains are a good example of this, as they have only an approximate location and may radiate into the lower back and down into the legs.[15] In fact, some pains are felt all over the body, as in the widespread aching associated with viral infections, hangovers, or conditions such as fibromyalgia (Clauw 2015). Itches, too, can be highly diffuse, as seen, for example, in drug withdrawal syndromes (Reich, Ständer and Szepietowski 2009).

To this, Armstrong could respond that I am missing the point. When he says that bodily feelings do not have a felt bodily location, he does not imply that no bodily location is involved *at all*. Rather, he is trying to convey that bodily location is not an essential part of these experiences. Take hunger and thirst. Armstrong acknowledges that the former is associated with the sensation typically described as stomach pangs and that the latter is associated with the sensation of dryness in the throat and mouth (1962: 113–114). Yet Armstrong regards these sensations as merely *incidental* — that is, as not actually constitutive of hunger and thirst (1993: 320). Hunger and thirst, Armstrong holds, are *desires* for food and drink, nothing more (1962: 113–114). And this, for Armstrong, is why bodily sensations and bodily feelings form different kinds. Bodily sensations are what Armstrong calls *portmanteau concepts*: sensations plus characteristic reactions (such as desires or dislikes) to those sensations (1993: 312). Itch, for instance, is a sensation in the body plus the desire to scratch. As Armstrong (1962: 112) writes:

> *To have an itch in a certain place is to feel a disturbance of our normal bodily state at that place, together with an immediate and interested dislike of that feeling, and an impulse to try to remove the disturbance by rubbing or scratching*.

Meanwhile, bodily feelings, according to Armstrong, have a fundamentally different structure: they are not sensations plus desires, but desires which may or may not be *accompanied* by sensations (1962: 113–114). The reason why Armstrong is forced to take this view is that bodily feelings can manifest as a variety of sensations. Hunger, for example, can manifest as a hollow sensation in the stomach, a feeling of

[15] This may be one of the reasons why menstrual pain is often pathologised in medical settings, despite its high prevalence among those who menstruate. See Serrahima and Martínez (2023) for a detailed discussion.

lightheadedness or weakness, or even all of these sensations at the same time. In other words, hunger does not have a *bodily signature*; it is elusive in that respect. But if there is no single sensation typical of hunger, then hunger cannot be construed as a sensation plus a desire provoked by that sensation. Thus, hunger ends up as *just* the desire to eat.[16]

This conclusion, however, is not one I am prepared to accept. We may, after all, desire to eat even if we are not hungry. For example, we may desire to eat to pass the time, to gain muscle mass, or to avoid offending the person who has prepared a meal for us. Of course, we also eat *because* we are hungry, but this only indicates that hunger and the desire for food are distinct. Hunger is a set of bodily sensations. It might not be identical to a specific sensation, but it is a sensory phenomenon nonetheless. As such, it has more in common with pains, itches, and breathing-related sensations than Armstrong admits. [17]

Let us take stock. A considerable number of our bodily experiences may be construed as relating to the physiological state of the body: hunger and thirst, stomach fullness and nausea, fatigue and sleepiness, the sensations which accompany breathing and swallowing, heartbeat sensations, temperature sensations, toilet urges, pains, itches, tickles, and pins-and-needles sensations. These experiences, I have argued, are all generated on the basis of the activity of the interoceptive system, which has evolved to process and integrate sensory data about homeostatic parameters.[18] Given that these experiences also share the phenomenal quality of being felt as sensations in the body, I conclude that we are justified in looking at them together. That said, I think Armstrong (1962) was right in seeking to further differentiate these phenomena. In what follows, I argue that the most philosophically interesting examples of interoceptive experiences constitute their own kind by virtue of their distinctive phenomenology and function.

---

[16] Armstrong also mentions 'feeling sick' as an example of an interoceptive feeling (1962: 33). I omit this experience from future discussion since I believe that it is best viewed as a *combination* of interoceptive experiences which coincide in times of sickness (severe fatigue, diffuse aching, sore throat, changes in felt body temperature, nausea, and so on). Note that in having varied bodily phenomenology, 'feeling sick' – like hunger – is a poor fit for the portmanteau view; but neither can it be defined as a kind of desire, given that it involves a number of desires.

[17] See Hall (2008) for an extended critique of Armstrong's (1962) portmanteau treatment of pain and itch.

[18] My claim that all interoceptive experiences are generated by the interoceptive system should be understood as *with participation* of the interoceptive system. Indeed, there may be several systems involved in generating HDs, and I am not wedded to a particular version of the cognitive architecture of HDs.

4. The distinctive phenomenology of homeostatic drives

While introducing the notion of interoceptive experiences, I focused on what these phenomena have in common. Now, I turn my attention to the differences between them. Specifically, I argue that a set of homeostatic experiences, which I *call homeostatic drives (HDs)*, typically exhibit *affective phenomenology* and *motivational phenomenology* – something of which the rest of them, which I call *non-HDs* for the sake of simplicity, are devoid. Examples of HDs include pain, itch, hunger, thirst, breathlessness, nausea, fatigue, temperature sensations, and toilet urges (bladder and bowel). Examples of non-HDs include heartbeat sensations, regular breathing sensations, and sensations which accompany swallowing and the movement of food along the digestive tract. I will further refine and detail this categorisation as I go.

4.1. The affective phenomenology of homeostatic drives

In addition to bodily (sensory) phenomenology, the experiences I call HDs also exhibit *affective phenomenology*, or *valence*. On this point, there is now sufficient agreement among philosophers (Aydede and Fulkerson 2014; Barlassina and Hayward 2019; Carruthers 2017; 2023; Grahek 2007; Martínez 2011; Klein 2015) as well as neuroscientists (Berridge and Kringelbach 2015; Levy and Glimcher 2012; Tye 2018).

Some HDs have a negative valence: they are *unpleasant*. Pain is a patently unpleasant experience, but so are itches, thirst, hunger, nausea, breathlessness, fatigue, toilet urges, and tickles in the throat or nose. Other HDs have a positive valence: they are *pleasant*. Erotic sensations provide the clearest example of this. Finally, there are HDs, such as temperature sensations, which have no fixed valence: that is, they can be pleasant or unpleasant. Indeed, one can enjoy the feeling of warmth when sitting in the sun, but one can also feel uncomfortably warm after chasing a bus in a thick jacket.

It is important to remember, however, that valence is not an essential characteristic of interoceptive experiences. Indeed, the interoceptive experiences I call 'non-HDs' are *affectively neutral*. In everyday life, we give little thought to these interoceptive experiences, precisely *because* they are neither pleasant nor unpleasant. Unless disrupted, they tend to remain in the background of awareness. Take, for example, heartbeat sensations. These sensations only have bodily phenomenology: they are nothing but pulsations in the chest which vary in their intensity and rhythm. A significant change in any of these two characteristics is likely to capture one's attention and provoke an emotional response, but the sensation itself would remain neutral. The same goes for the sensations which accompany regular breathing. While breathlessness, defined as *'a subjective experience of breathing discomfort that consists of qualitatively distinct sensations that vary in intensity'*, involves a highly unpleasant and urgent sensation of tightness in the chest, the sensations that accompany unhindered breathing have no negative valence (Ambrosino and Serradori 2006).

Now, it may be objected here that occasionally, non-HDs appear affectively coloured. Say that Valentina has a family history of serious heart disease (and knows it). She herself has never had any tests done to confirm whether or not she is affected by it. One day, though, Valentina notices that her heartbeat is irregular for no apparent reason. Immediately, she feels scared and upset. She resents the experience of irregular heartbeats and wishes they would stop. Can we not say, then, that in this case the heartbeat sensation is unpleasant? We cannot – because it is not the heartbeat sensation itself that is unpleasant, but *the negative emotions* caused by the negative appraisal of the heartbeat sensation. Valence is a general characteristic of affective states, the category which includes HDs, but also emotions and moods (Carruthers 2023). So, while Valentina has an unpleasant experience as a result of her palpitations, it is her fear and sadness that feel bad, not the palpitations themselves.

Consequently, interoceptive experiences can be divided into two groups: affectively neutral (non-HDs) and affectively charged (HDs). The second group is larger, but this is hardly surprising. Indeed, if the evolutionary role of valence is to draw attention to what is important, then the majority of interoceptive

experiences *should* be valenced, given the relevance of changes in the physiological state of the body for survival.[19]

Furthermore, HDs vary in the extent of their (un)pleasantness. Migraines, severe thirst, and relentless allergic itching feel terrible, but the slight sensation of bladder fullness or very light hunger is so easy to ignore that some have even described them as affectively neutral (Klein 2015: 48–49). Of course, these experiences may just as well be viewed as only *slightly* unpleasant; it is not obvious how to adjudicate between the two options. Even so, there do seem to be some cases of HDs which lack valence. Some such cases will be presented later in this section, since they also involve a possible absence of the normal motivational character of HDs. I will, however, discuss one of them here: namely, *depersonalisation*.

Depersonalisation is a rare condition characterised by strong and persistent feelings of unreality. There is a bodily dimension to this condition, too: those who suffer from it have sometimes been described in the literature as having ceased to experience pain, hunger, thirst, fatigue, and so on (Simeon and Abugel 2006: 69). This, however, is an overstatement, or perhaps an unfortunate figure of speech: the depersonalised subjects rather report a pronounced *indifference* to pain, hunger, and so forth (Sierra and Berrios 1998).[20] Sierra (2009), one of the leading experts on this disorder, notes this indifference and attributes it to a general top-down suppression of affective processing, which, among other things, strips HDs of their normal affective character. Consequently, if this theory is correct, then there is an example of HDs across the board losing valence. In light of the possible existence of cases of HDs lacking valence, I refrain from claiming that HDs are *necessarily* valenced. Instead, I conclude that while all interoceptive experiences exhibit bodily phenomenology, only HDs *typically* exhibit affective phenomenology.[21]

---

[19] See Carruthers (2023) for an overview of the possible functions of valence.

[20] Besides, if HDs were truly absent in depersonalisation, many individuals who suffer from it would be hospitalised with dehydration, anorexia, or bladder problems. This does not happen, though: depersonalisation has remarkably little effect on everyday functioning (Sierra 2009). It would seem, therefore, that HDs are *altered* rather than *absent* in depersonalisation.

[21] There have been some reports of the so-called congenital *indifference* to pain, which must be distinguished from congenital *insensitivity* to pain because it is not characterised by the absence of pain sensation, merely by indifference to it. Landrieu, Said and Allaire (1990), for example, describe two patients capable of registering pinpricks and exhibiting normal withdrawal reflexes and vocal responses to pain, who nevertheless reported no associated unpleasantness. One patient, in fact, had sustained several painless limb fractures. It is, however, not clear from the description whether the patients in question actually felt pain as such.

4.2. The motivational phenomenology of homeostatic drives

Another aspect of HDs that distinguishes them from non-HDs is *motivational phenomenology*. Not only do HDs typically feel pleasant or unpleasant, but they also typically feel as motivations to perform specific types of body-directed behaviours (Hall 2008; Klein 2015). Thirst, for example, is felt as a motivation to drink. Hunger is felt as a motivation to eat. Itch is felt as a motivation to scratch. A tickle in the throat is felt as an urge to cough. Bladder fullness is felt as a motivation to void the bladder. And so on.

The term 'behaviour' is used here in the broad sense. When I say that HDs motivate specific types of body-directed behaviours, I mean both *engaging* and *refraining* from actions, since HDs are able to motivate both. Nausea, for example, is associated with a motivation to vomit and with a motivation not to consume food (even the idea of food is very unwelcome to a person experiencing nausea). Similarly, pains and aches may be felt as motivations to move away from very hot or very sharp objects, but also as motivations to limit certain activities (e.g. not to chew with an aching tooth). And while fatigue may be interpreted as motivating us to rest, it may also be interpreted as motivating us to stop whatever we are doing (e.g. stop running). We may say, then, that HDs have the capacity to engender both *positive* and *negative* motivations. Some HDs (thirst, hunger, bladder fullness, and so on) tend to issue mainly positive motivations, some HDs tend to issue mainly negative motivations (fatigue), and some HDs (nausea and pain) may issue both positive and negative motivations.

The motivational phenomenology of HDs may be further specified in three ways. First, HDs motivate *directly*, simply in virtue of being felt. Indeed, hunger motivates us to eat whether or not we judge that now is a good time to eat. Itch motivates us to scratch whether or not we know that it would make the itching worse. Feeling out of breath motivates us to stop running whether or not we take this feeling to be psychosomatic in origin. Bladder fullness motivates us to go to the toilet whether or not we believe that the bladder is, in fact, almost empty. And so on.

Second, HDs motivate *efficiently*. Admittedly, it is unclear – and rarely discussed – whether our ability to experience HDs as motivating specific behaviours is *learnt* or *innate*. Some psychologists belong firmly to the 'learning camp'. For example, Bruch (1973), who both studied and successfully treated eating disorders, was of the opinion that these disorders stem from the impaired 'matching' of bodily sensations with the motivation to consume food.[22] A number of contemporary researchers share this view (Stevenson 2024), while philosophers, as far as I can tell, have yet to address the issue. Still, regardless of *how* we acquire the ability to do so, as adults, we seem to have an immediate grasp of the motivational meaning of HDs (Klein 2015: 17). That is, we do not, as a rule, experience HDs as motivating us to do *something* and then *infer* what needs to be done. Rather, we tend to experience them as stipulating specific types of actions. Thirst, for example, involves an instantly recognisable motivation to drink. Nausea includes an instantly recognisable motivation to vomit and abstain from eating. And so on. In fact, some HDs – itches and toilet urges come to mind – may be impossible to define without reference to the behaviours which they engender (Hall 2008). This is why I say that HDs motivate efficiently: because they do not need to be decoded by the agent.

Now, it has been argued that some pains, both common and relatively uncommon, do not feel as motivations for specific actions (Tumulty 2009). Such *apparently purposeless pains* include headaches, menstrual pains, and kidney stone pains. Despite influencing our behaviours – getting us to stay in bed, take painkillers, cry, rub the aching body part, and search online for remedies – purposeless pains do not direct us to do all these things in the same targeted way that itch directs us to scratch and bladder fullness directs us to urinate. Rather, we seem to do all these things *in the hope* of getting rid of pain.

Klein (2015), whose position is analysed in greater detail in Chapter 3, disagrees with Tumulty's assessment. According to Klein, all pains motivate us to protect our body – it is simply that sometimes, there is not much to be done. I am not convinced, however, that purposeless pains involve an urge to

[22] As cited in Harshaw (2009), in an overview of the research on the developmental aspects of hunger, satiety, and thirst.

protect the body. Indeed, there is nothing to show that they do. Accordingly, I will assume, for the purposes of this thesis, that while HDs tend to motivate efficiently, they may occasionally fail to do so.

Third, HDs motivate *reliably*. Of course, the motivational import of HDs is neither immediate nor unchecked. Indeed, most people manage to suppress the urge to urinate for a few hours, and those of us who live in mosquito-rich areas learn to abstain from scratching mosquito bites, as it invariably makes them worse. The motivations engendered by HDs are also not immune to cultural and other influences. For example, hunger motivates one to eat, but one's personal preferences shape this motivation into a more specific motivation to consume particular kinds of food in particular settings.

Whether or not HDs can lack motivational force is not entirely clear, though there is some evidence in support of this possibility – though mainly with respect to pain. Indeed, changes in the experience of pain observed in morphine use and brain resection (lobotomy, cingulotomy) are typically interpreted as changes in the affective and motivational aspects of pain which leave its sensory aspect intact (Carruthers 2017: 665; Corns 2014a: 360–65; Grahek 2007: ch. 3). Of note, certain types of brain lesions and surgeries appear to produce pains which combine *reduced* unpleasantness with *heightened* pain behaviours such as withdrawal (Talbot and others 2009). This suggests that the affective and the motivational aspects of pain may dissociate not only from the sensory aspect of pain, but also from each other.

> *During the POST-OP session, however, the patient simply withdrew his hand abruptly after 10-20 sec; he appeared somewhat perplexed and immediately replaced it in the water, only to withdraw it again after an even shorter period of time […] although the water seemed "less intense," he was unable to keep his hand in the bath as long (Talbot and others 2009: 123).*

The so-called *pain asymbolia* has also been offered as a case of pain devoid of either motivational (Klein 2015) or both affective and motivational character (Bain 2014; Grahek 2007; de Vignemont 2015). Those who develop this syndrome following a brain injury appear to retain the ability to detect pain, identify it as

such, and describe its bodily features, but no longer act appropriately with respect to it, showing no withdrawal response and voicing no complaints when pricked with a needle (Grahek 2007: 47). The problem with using pain asymbolia to inform pain theories, however, is that this condition is extremely rare and has not been rigorously studied. Among other things, patients may have disordered speech, engage in acts of compulsive self-harm, and show broader disturbances of affect, action organisation, and attention (Schilder and Stenghel 1930; Weinstein, Kahn and Slote 1955). These factors, exemplified below, make it difficult to determine the precise nature and causes of their abnormal pain behaviours:

> *We observed a patient with sensory aphasia who would have hurt herself severely if left alone. She pushed everything that came into her hand against her eyes, heedless of the pain […] sometimes she took a needle and stuck herself deeply with it. The patient perseverated in actions once undertaken* (Schilder and Stenghel 1930: 598).

> *Even on the occasions when a patient might act as if a stimulus was painful, on being questioned, he would deny that it hurt. […] Pain asymbolia is brought out under certain conditions of examination, while in other situations the patient may seem to feel, and even complain of, pain* (Weinstein, Kahn and Slote 1955: 240, 246–47).

What about HDs other than pain? Depersonalisation, as previously discussed, may be characterised as a disorder of valence that does not impact the motivational dimension of HDs. An argument might also be made that depressed individuals are not motivated to eat by their hunger, though it is doubtful that similar arguments could be advanced for experiences of itch and bladder fullness. Indeed, what would it feel like to have an itch devoid of motivational character? I cannot imagine it; but then I am also unable to simulate the experience of pain asymbolia, so intuitions may carry little evidential weight here. In sum, it appears that – pathological cases aside – HDs generally involve a motivation to act in a body-directed way. This motivation may be relatively weak and easy to ignore, as in the case of slight hunger, or intense and urgent, as in the case of severe breathlessness, where suppression is virtually impossible.

And here we get to the heart of the matter. HDs, I have suggested, motivate directly, efficiently, and reliably. Putting it all together, I suggest that HDs feel like *urges* to act in a certain way. Here is how Richard Hall describes the distinction between the cognitive origin of urges and desires: *'Desires, along with beliefs and other propositional attitudes, are central. Your senses don't have them. […] You usually believe what your eyes tell you. But your sense of sight doesn't itself have beliefs. And your bodily sensory systems don't have desires'* (2008: 532).

If desires are produced centrally while urges are produced by sensory or other low-level modular systems, it becomes clear how we can feel *compelled* to act in ways we neither want nor judge to be necessary or appropriate. Those who suffer from the obsessive–compulsive disorder (OCD), for example, may feel compelled to fold their socks in a particular way or arrange objects symmetrically, despite recognising such acts as pointless and deriving no pleasure from them.[23]

HDs, I argue, also involve urges, which is why they can be hard to resist (consider the urge to breathe in when holding your breath). I do not deny, of course, that HDs can be, and often are, accompanied by desires with the same satisfaction conditions as the associated urges – i.e. by *congruent desires*. An urge to eat when hungry, for example, is typically accompanied by a congruent desire to eat, while an urge to rest is typically accompanied by a congruent desire to rest. This accords with the modern neuroscientific understanding of urges routinely *translating* into desires (Davenport 2008; Jackson and others 2011). What I propose, then, is that HDs are typically experienced as urges whether or not congruent desires are also experienced alongside them. When congruent desires *are not* present, the phenomenology of urges becomes more pronounced and the agent may feel *coerced* into acting (e.g. scratching an itch or eating). When congruent desires *are* present, their phenomenology may blend with, or even overshadow the phenomenology of urges, with the agent feeling more excited about eating than coerced into it.

[23] OCD has been found to be associated with a decreased ability to resist sensory urges, such as the urge to blink (Bragdon and others 2023). This suggests the presence of a *general capacity* to resist urges, one that may be disrupted in certain conditions.

Now that I have broken down the motivational phenomenology of HDs, I am in a position to clarify why I consider non-HDs to be devoid of it. While I doubt that anyone has ever been spurred into action by the sensations that accompany swallowing, sensations of heartbeat and breathing seem perfectly able to do that. For example, if I notice that my heart is beating faster than usual, I may decline another cup of coffee. If I am in a yoga class and I notice that my breathing is becoming shallow and uneven, I may try to relax or adjust my pose. Yet, although heartbeat and breathing sensations do motivate me to act in these examples, they do not have the motivational phenomenology which is typical of HDs. These sensations do not motivate me directly, efficiently, or reliably; they do not feel like urges. Instead, it is my desires – to be healthy, to improve at yoga – doing the motivating job here.

I conclude, then, that there exists a distinctive class of interoceptive experiences that tend to exhibit, in addition to bodily (sensory) phenomenology, affective and motivational phenomenology. These experiences are felt as sensations in the body, but they are also felt as *(un)pleasant* and *urging body-directed behaviours*. I call them homeostatic drives (HDs). These experiences are interoceptive in that they are generated *from* the homeostatic data processed by the interoceptive system, but the exact number of systems involved in their generation remains an open question. Valence, as a *general characteristic of affective states* (interoceptive experiences, emotions, moods, gustatory pleasures), is thought to be generated *a domain-general system* rather than by the interoceptive system (Barlassina and Hayward 2019; Berridge and Kringelbach 2015; Carruthers 2017; 2023; Levy and Glimcher 2012; Tye 2018). Felt urgency, too, is likely generated by a dedicated system (or systems): given the adaptive value of having one's urges reflect the *relative* importance of engaging in certain actions, there is an argument to be made in favour of urgency computations for different types of mental states being performed in the same place. In this thesis, I do not aim to provide a cognitive architecture of HDs; although I discuss in several places how it could and could not work, I do not address it in sufficient detail since my focus lies elsewhere. In the next chapter, I complete my presentation of HDs as a specific variant of interoceptive experiences by arguing that their phenomenological profile is dictated by their *homeostatic* function.

5. The distinctive function of homeostatic drives

Our bodies, as I noted in the introduction, are fragile and thus require constant *maintenance* and *protection*. Fluid levels, oxygen levels, body temperature – all these physiological parameters must remain within fairly narrow margins if we are to survive. In addition to that, nutrients must be consumed, metabolic waste and poisons must be evacuated, injuries must be avoided, and viruses and bacteria must be neutralised by the immune system. And so on.

In having access to homeostatic data, the interoceptive system is uniquely positioned to ensure the stability of our homeostatic variables. How exactly does it achieve this goal, though? As mentioned previously, much of the interoceptive activity occurs outside awareness and cognitive control, requiring no goal-directed behaviour on the part of the agent (Berntson and Khalsa 2021; Billman 2020; Chen and others 2021). Consider, for example, the conversion of stored fat into energy or the immune response to infection. These vital processes are initiated and carried out *automatically*, without any input from us; indeed, we could not control them even if we wanted to.

Often, however, the task of body maintenance and protection makes it necessary for the agent to *act in particular ways*. Consider, for example, common behaviours like eating, drinking, or wearing a warm sweater on a cold day. If we do not engage in these actions, we put our bodies at risk – and yet, we have a great deal of control over their implementation, being able to delay their execution (e.g. postponing eating or drinking for a few hours) or even suppress them entirely (e.g. resisting the urge to scratch a mosquito bite). This control, however, is not absolute: vomiting, coughing, or withdrawing from a painful stimulus cannot *always* be suppressed. There is a limit to how long that you can hold a hot cup before dropping it, however much you may wish to avoid it. As such, the behaviours necessary for body maintenance and protection seem to fall somewhere between voluntary and involuntary. For this reason, I prefer to call these behaviours *controlled* rather than *intentional*.

Let us put this in more precise terms. There are a number of processes – automatic and controlled – that serve to secure *homeostasis*, i.e. the stability of physiological variables such as oxygen levels, fluid balance, and body temperature (Berntson and Khalsa 2021; Billman 2020; Chen and others 2021; Modell and others 2015). The controlled behaviours may be termed '*homeostatic behaviours*' (Klein 2015). Sometimes, their importance is culturally derived. For example, one often hears that one should drink two litres of water a day. In most cases, however, we engage in homeostatic behaviours because we experience the drive – because we *feel like it*. Indeed, we do not typically drink water because we know how important it is to do so (though we are aware of this). Rather, we drink water because we are thirsty, and drinking quenches it. Likewise, although we know we ought not to touch a skin graze with dirty fingers, we would in any case avoid doing so, since contact with broken skin tends to elicit pain.[24]

I propose that HDs have acquired affective and motivational features because these features equip them *to promote homeostatic behaviours*. To be sure, all interoceptive experiences, HDs and non-HDs, can in principle lead us to modify our behaviour in ways that are broadly beneficial to our bodies. For example, heightened heartbeat sensations may prompt us to slow down after exertion. I maintain, however, that only HDs have the *function* of getting us to satisfy homeostatic needs. Indeed, as argued in the previous section, heartbeat sensations and breathing sensations can serve as reasons to act, but only in the presence of *additional* motivating factors such as the recognition of medical risk. Since they do not typically urge us to act, non-HDs cannot be said to have specific *homeostatic jobs*.

Conversely, each member of the HDs group may be viewed as a targeted solution to a specific homeostatic 'problem', that is, an efficient mechanism for ensuring the timely satisfaction of a specific homeostatic need. We need to eat in order to replenish our energy resources – so we feel hunger. We need to have the waste products of the digestive process eliminated from the body – so we feel toilet urges. We need our body temperature to remain within certain limits – so we feel warm or cold when our

[24] Many homeostatic variables are regulated both automatically and behaviourally. The two strategies may even be employed simultaneously. For example, optimal fluid levels are maintained by (1) triggering thirst and (2) reducing the production of saliva, which produces the characteristic sensation of dry mouth. While the sensation of dry mouth alone can motivate us to drink, research shows that thirst reliably occurs even when saliva production is artificially kept high (Inenaga and Ono 2010).

body temperature fluctuates. And so forth. The homeostatic significance of many HDs is evident and does not require a lengthy defence. Indeed, it is not difficult to see how nausea, thirst, breathlessness, fatigue, and sleepiness all contribute to keeping specific bodily variables in check. In some instances, we may not fully understand the homeostatic role of HDs – for example, the exact reason why humans want and need so much sleep remains unclear. Yet the disastrous consequences of sleep deprivation on mental function confirm that we do need it, and our capacity to feel sleepy safeguards us against this undesirable outcome. Likewise, we might not know why nausea is provoked by motion, but nausea clearly plays an indispensable role in protecting us from toxins in food – and perhaps from other dangers.

Pains and aches, too, subserve homeostasis (Melzack 1999). Granted, some pains, such as headaches, seem to have little utility – though this assessment may be questioned, too. One might argue, for example, that headaches are future-oriented and teach us to avoid doing the things that bring them on, such as drinking alcohol or foregoing sleep. At any rate, whether or not *all* instances of pain are homeostatically useful, the ability to feel pain plays a crucial part in preserving *bodily integrity*. Consider how pain protects our bodies on a daily basis by ensuring that we withdraw swiftly from hot or sharp objects and take particular care of the body parts that have been bruised, sprained, scratched, or broken. Consider also *congenital insensitivity to pain*. Affected individuals feel no pain and as a result sustain numerous injuries (especially in infancy), develop unusual ailments (e.g. early joint degeneration), and often die prematurely (Melzack and Wall 1996: 3–7). Thus, pain not only enables us to prevent bodily damage and minimise its impact, but its contributions cannot be fully compensated for when it is absent.

When it comes to some interoceptive experiences, however, their homeostatic role is less obvious and thus calls for elaboration. Take itch. Itch directs us to scratch, but is scratching a homeostatic behaviour? In contemporary western societies, most instances of itch-induced scratching do not seem to be beneficial to us. On the contrary, by scratching a mosquito bite or an allergy-induced rash, we risk causing the skin to become infected.

Yet, the presence – or even the preponderance – of useless or harmful itches, much like the existence of useless pains, does not necessarily indicate that itches lack a homeostatic function. Trait selection hinges on whether a trait increases fitness, so even traits that are only occasionally useful may be selected and retained, provided the organism is better off having this trait than not (Millikan 1984: 32–34). Itches, then, may be like fire alarms: mostly false, but still necessary, because the inconvenience of frequent false alarms does not compare with the danger of dying in a fire. Besides, the proper function of a mental state is often understood as the function for which it was initially *selected* (Millikan 1984: 28). Given that the environment we currently inhabit is not the one for which our brains were adapted, it is possible that many of the itches we experience are useless because the environmental shift has almost eliminated the factors triggering useful itches. Indeed, some scholars have theorised that itch ensures avoidance of certain plants, promotes removal of dangerous insects from the skin, or even serves as a last-resort parasite defence, motivating us to self-inflict an injury on an area exposed to pathogens and thereby boost the local immune response (Andrew and Craig 2001; Radjavi 2019). Thus, while we may suffer needlessly from itches in our everyday life, under certain conditions they may turn out to be life-saving.

Next, take tickles. At least some of these phenomena appear to serve a homeostatic role. Tickles in one's throat, for example, motivate one to cough. Coughing, in turn, serves to protect one's airways from obstruction. Now, it may be objected here that coughing is a reflex; but things are not quite as simple as that, since the urge to cough may be voluntarily suppressed by the agent – as, for instance, at a concert. While coughing is often involuntary, this is not always the case, and one may also cough intentionally, to get rid of the discomfort in one's throat.[25]

Finally, take erotic sensations. These interoceptive experiences have affective and motivational features, and while they do not contribute to bodily survival, an argument can be made that they help maintain a homeostatic variable. Indeed, the genitals tend to become aroused, and arousal is resolved by an orgasm, which usually requires genital stimulation; after an orgasm, genitals revert to their non-aroused state. We

[25] See Jackson and others (2011) for an analysis of the differences between reflexes and urges.

may say, then, that genital arousal constitutes a change in the physiological body state which requires behavioural resolution. Accordingly, the state of one's genitalia may be viewed as a homeostatic variable, and erotic sensations as a type of HD whose function is to prompt the agent to restore this variable to its baseline (non-aroused) state. And if this is unconvincing, consider that the mechanisms that evolved to support the stability of homeostatic variables may have been co-opted for erotic sensations.[26]

Consequently, it seems that all (or nearly all) interoceptive experiences that have affective and motivational phenomenology make specific contributions to homeostasis. For example, thirst ensures water intake, itch promotes the removal of irritants from the skin, and nausea helps us expel toxins. This is unlikely to be a coincidence. Rather, given that evolutionary pressures favour mechanisms that maintain the stability of the physiological state of the body, it is plausible that the adaptive value of performing certain behaviours under particular bodily conditions has led to the clustering of certain phenomenal properties. In a nutshell, interoceptive experiences with affective and motivational charge get the homeostatic jobs done – which is why they were selected to have these properties.

Recognising this evolutionary rationale strengthens the case for distinguishing these interoceptive experiences as members of a distinct subclass. I propose that these interoceptive experiences warrant a unified theory – and a unified name. I call them '*homeostatic drives*'. Klein (2015) refers to a similar set of phenomena as '*homeostatic sensations*', while Craig (2015) calls them '*homeostatic emotions'.* Neither term, however, is entirely satisfactory in my view. Klein's term may be misleading: any bodily sensations relating to the physiological condition of the body are homeostatic sensations, but we would not want to include heartbeat sensations in the HDs category. Craig's term, in turn, is controversial. On the one hand, both emotions and HDs may be reasonably viewed as *'evolutionary shortcuts'* that get us to engage in specific behaviours necessitated by worldly or bodily events. On the other hand, trying to force them into a single overarching category on this basis alone is premature, given the important differences between emotions and HDs. For one thing, emotions have cognitive content, something which HDs lack. Thus, I

[26] What about orgasms? These experiences resemble purposeless pains in that they have affective phenomenology but no clear motivational phenomenology. Perhaps, they may be treated as in-between mental states. Regardless, given that it may be controversial to include erotic feelings into the category of HD, I will refrain from discussing them in this thesis.

prefer the term *'homeostatic drives'* to denote the class of interoceptive states that I am interested in exploring.

In the next section – the final section of this chapter – I spell out what we want from a theory of HD and explain how I propose to narrow down the options available to us.

6. The next steps

In this chapter, I have argued that there is a set of interoceptive experiences – I call them HDs – which exhibit remarkable similarities in their phenomenology and function. All interoceptive experiences feel like sensations in the body, but only HDs feel like pleasant or unpleasant sensations in the body that urge us to engage in body-directed behaviours. All interoceptive experiences relate to identifiable homeostatic parameters, but only HDs have the function of promoting behaviours that restore those parameters. Based on all this, I propose that HDs constitute a promising *investigative kind* (Griffiths 2004), which may be refined as the research into it progresses. Examples of interoceptive experiences discussed in this thesis include pain, itch, thirst, hunger and satiety, nausea, breathlessness, fatigue, temperature sensations, and toilet urges. This list provides ample material for analysis, though it is not intended to be comprehensive. The category of HDs is a work in progress, and I am prepared to consider any experience with the right function and phenomenology as a potential member of the category.[27]

My aim is to further our understanding of HD. Given the scope of this project, however, it is necessary to choose a particular focus at the outset. As our constant companions, as mental states with rich phenomenology, and as products of a modality that seems to be unlike any other, HDs pose many philosophical puzzles – too many, in fact, to cover each of them in this thesis. Accordingly, while I consider it essential to discuss HDs as a group, I explore them through a lens of particular approach and omit certain aspects of HDs from the explanandum to avoid stretching myself too thin.

[27] The sensation of eye dryness, for example, is unpleasant and motivates a protective action: blinking. Although blinking, like coughing, is often involuntary, it can be voluntarily suppressed, at least temporarily. Thus, this sensation qualifies as a type of HD.

First, in line with current trends in the philosophy of mind, I focus on the *representational content* of HDs. More specifically, I take the best-known representational theories of pain and examine how they apply to a broader class of pain-like mental states. My objective is to determine which, if any, of these theories is superior to the others – or, if none are, in which direction the inquiry should continue.

Second, I restrict my analysis to the sensory and motivational aspects of HDs. Valence, many scholars now agree, is a *general quality of affective states* – i.e. interoceptive states, emotions, moods, and sensory pleasures – and a product of a domain-general cognitive system used as *common currency* in decision-making (Barlassina and Hayward 2019; Berridge and Kringelbach 2015; Carruthers 2017, 2023; Levy and Glimcher 2012; Tye 2018). This means that valence warrants a separate inquiry, and, in any case, the debate on the intentional content of valence has grown too large in recent years to fit the scope of this thesis. In setting it aside, I am left with sensory and motivational phenomenology and with the function of HDs to explain – more than enough to test any theory.[28]

In the next chapter, I outline what I take to be the most prominent representational theory of pain – pure indicativism – and examine whether it can account for the homeostatic function as well as the sensory and motivational phenomenology of HDs. I then use the same approach with other theories of HDs – namely, pure imperativism and the pushmi-pullyu theory – and compare their explanatory power.

[28] Some philosophers hold that the motivational power of HDs is wholly explained by their affective features: pain and hunger get us to act in virtue of being unpleasant (Bain 2013, 2014; Grahek 2007). Others disagree (Corns 2014b; Klein 2015). In light of this disagreement, I address the motivational aspect of HDs in this thesis, even though I set aside their valence. For what it is worth, I think that Corns (2014b) is correct and the affective and motivational aspects of HDs can dissociate (see my discussion of depersonalisation and the effects of certain brain lesions or surgeries in Section 4).

Chapter 2:

Pure Indicativism

Abstract:

In the previous chapter, I introduced the category of interoceptive experiences and proposed that some of them – *homeostatic drives, or HDs* – are sufficiently similar to each other, and sufficiently distinct from non-HDs, to warrant a unified and systematic investigation of their representational content. Given that in philosophy and the neurosciences, HDs such as pain are often understood as simple representations of changes in the physiological state of the body, I begin this chapter by asking whether this popular conception of HDs might be the right one. After outlining this *pure indicativism* view and the considerations in its favour, I evaluate the arguments that have been advanced against it. I show that these arguments proceed along similar lines and are not especially compelling, though pure indicativism still fails because it cannot account for the motivational aspect of HDs. I conclude that while there is no reason to reject indicativism entirely, pure indicativism must be set aside.

1. Setting the scene

In the previous chapter, I introduced the category of interoceptive experiences – bodily experiences relating to physiological, or homeostatic, parameters (Berntson and Khalsa 2021; Chen and others 2021; Craig 2002, 2009, 2015; Damasio and Carvalho 2021). These experiences, without doubt, constitute a large and diverse group of mental states. Thirst, hunger, satiety, nausea, fatigue, heartbeat and breathing sensations, sensations of distension or movement in the internal organs, pains and aches, itches, temperature sensations, and tickles – these are all examples of interoceptive experiences.

Despite a substantial and rapidly growing body of empirical research, it is not entirely clear how we should classify the interoceptive phenomena (Ritchie and Carruthers 2015). In a bid to make progress in this matter, I have argued that some members of this category (*homeostatic drives, HD*) are sufficiently similar between themselves, and sufficiently distinct from other interoceptive states (*non-HD*), to be treated as an *investigative kind* (Griffiths 2004). First, I have pointed out *the phenomenological distinction* between HD and non-HD. While all interoceptive experiences have sensory phenomenology, i.e. are felt as bodily sensations, only HD (itch, hunger, thirst, nausea) also exhibit affective and motivational phenomenology, i.e. are also felt as (un)pleasant and as urges to engage in specific body-directed actions (scratching, food or water ingestion, vomiting). Second, I have emphasised *the functional distinction* between HD and non-HD. The behaviours motivated by HD play a key role in satisfying the homeostatic needs of the body, which strongly suggests that HD have the function of *promoting homeostatic behaviours* (Carvalho and Damasio 2021). In contrast, non-HD do not motivate any specific behaviours; whatever their function is, it is not to promote homeostatic behaviours.

So far, as I have noted, pain has received much more attention from philosophers than any other HD. In this thesis, I seek to redress this imbalance by recasting the existing theories of pain as candidate theories for a broader class of pain-like mental states. My goal is to determine whether we already have a good working theory of HD — or, failing that, to identify a direction in which the debate should continue. In either case, our understanding of HD will be significantly advanced.

Now, as I have also observed in the previous chapter, the theories that dominate the philosophical landscape with respect to pain are *representational theories*. These theories hold that the phenomenal character of pain can be explained by its representational (intentional) content. There is, however, no agreement among the advocates of this view on what exactly the representational content of pain might be, and in recent years, the debate has stalled as many contributors (Aydede and Fulkerson 2014; Bain 2013, 2014; Barlassina and Hayward 2019; Martínez 2022) have shifted their focus to the representational content of *valence*. On the one hand, philosophers like Tye (1995a, 1995b, 2006), Cutter and Tye (2011),

Bain (2003, 2007), Dretske (1995), Lycan (1996) and Ritchie and Carruthers (2015) hold that pain and similar mental states are just representations of changes in the physiological state of the body. Call it *the pure indicativist view*. On this view, HDs have only indicative content, merely telling us how our bodies are faring. On the other hand, philosophers like Hall (2008), Klein (2015), and Martínez and Klein (2016) insist that pain and similar mental states are instead commands to engage in specific homeostatic behaviours. Call it *the pure imperativist view.* On this view, HDs have only imperative content, telling us what needs to be done rather than providing us with any insight into the physiological condition of the body.

Accordingly, there are two major ways of understanding HDs. On the pure indicativist view, HDs represent changes in the physiological condition of the body. When these informationally rich representations (hunger, thirst, itch, and so on) are processed by the decision-making apparatus, we become motivated to carry out appropriate homeostatic actions (eating, drinking, scratching, and so on). On the pure imperativist view, things are different: HDs command us to perform specific homeostatic actions, and we tend to comply with these commands because we recognise *the authority of the body*. Yet, it is essential to recognise that a third option is also available: HDs may possess both indicative and imperative content, may *both inform and command* us. If pure models do not work, this is the direction in which to look.

Here is another way to say the same thing. It is possible to distinguish between indicativism from *pure* indicativism, and imperativism from *pure* imperativism. Indicativism simply says that HDs have indicative content, but pure indicativism says that HDs have *only* indicative content. Imperativism simply says that HDs have imperative content, but pure imperativism says that HDs have *only* imperative content. Accordingly, one cannot be both a pure indicativist and a pure imperativist about HDs, but one can be both an indicativist and an imperativist about them.

In this chapter, I examine pure indicativism. I begin by outlining this view and detailing its theoretical and empirical support. I then ask whether, as some philosophers have contended, any version of indicativism is ultimately indefensible. I argue that this is not the case. However, pure indicativism cannot be accepted,

since it fails to account for the distinctive motivational profile of HDs. All in all, HDs may have some indicative content, but not indicative content alone.

2. Pure indicativism: Foundations and specifics

The idea that interoceptive experiences – and, by extension, HDs – are *informative* about the physiological state of the body is deeply entrenched in folk psychology. As one study confirmed, lay people and health professionals share this view with respect to hunger, being equally inclined to regard it as an objective sign of energy depletion (Assanand, Pinel and Lehman 1998). Attitudes towards thirst appear to be of the same ilk: on the NHS website, thirst is defined as *'just the brain's way of warning that you're dehydrated'*.[29] Similar assumptions are made about other interoceptive experiences, with those who suffer from medically unexplained pain and fatigue (or other persistent symptoms) often dismissed as 'misinterpreting' what they feel or 'catastrophising' normal levels of pain and fatigue (Barsky and Borus 1999). This unfortunate bias clearly illustrates how confident we are that our interoceptive experiences are fundamentally trustworthy – and how against the grain it is to challenge this conviction.

The idea that interoceptive experiences convey accurate information about the physiological state of the body is also widely endorsed within the scientific community (Barrett and Simmons 2015; Craig 2002, 2008, 2015; Carvalho and Damasio 2021; Seth 2013). As Craig puts it, *'the most basic feelings from the body represent aspects of its physical condition'* (2015: 2). Revealingly, a large body of research has now accumulated in support of the hypothesis that medically unexplained symptoms are the result of *impaired interoception* (Murphy and others 2017; Petzschner and others 2021; Quadt, Critchley and Garfinkel 2018). This means that modern researchers are open to the idea that interoceptive experiences can be inaccurate – but they take such inaccuracy to be necessarily *pathological*, thus reinforcing the folk intuitions detailed above.

[29] https://www.nhsinform.scot/illnesses-and-conditions/nutritional/thirst

From my perspective, this *argument from usage* is the major argument for indicativism. When we feel thirst, we think that we are dehydrated. When we feel the urge to urinate, we assume that the bladder is full. When we feel breathless, we take it for granted that we need oxygen. When we feel our heartbeat, it does not occur to us that we may be misperceiving it. When we feel pain, we become concerned that something is wrong with the part of the body that is hurting – and, if the pain is persistent, we assume that whatever is wrong must be serious. And so on. All in all, under normal conditions, we treat our interoceptive experiences as reliable indicators of the physiological state of the body, using them to ground our decision-making and reasoning. And what would explain this role that interoceptive experiences play in our mental economy? Their indicative content.

A further *argument from simplicity* can be made for *pure indicativism*. Sensory systems are typically seen as representing some aspect of the environment (Macpherson 2011: 23). The interoceptive system is described as responsible for processing homeostatic data (Berntson and Khalsa 2021; Craig 2002, 2008, 2015). It is highly plausible, then, that interoceptive experiences simply represent something like the current state of homeostatic parameters. Given that this view already aligns with both folk and scientific conceptions of the mind, there is no apparent reason to posit anything more elaborate at this stage.

Pure indicativism, then, rests on a combination of considerations concerning our attitude towards interoceptive experiences and our general understanding of the function of sensory systems. If pure indicativism is correct, interoception is on a par with the other senses in that it serves to inform the agent about a specific aspect of (bodily) reality as best it can. The only point of difference is that interoception appears to have evolved to capture *event-like properties* – namely, fluctuations in homeostatic parameters such as fluid levels, oxygen levels, bladder fullness, and bodily integrity. Here is how Ritchie and Carruthers articulate it (2015: 357):

> *The interoceptive senses are primarily sensitive to disturbances, becoming active when things in the body go wrong […] Under normal conditions we are barely aware of any feelings in the*

> *viscera, because most visceral afferents are typically 'silent', and respond only under very specific conditions […] interoception has not evolved to detect osmotic pressure as such, but to detect changes in osmotility that are significant deviations from the homeostatic norm. Thus lower osmotic pressure signals dehydration, causing thirst. Such event-like properties involving deviation from a norm are perfectly legitimate contents of mental representations.*

In philosophy, the idea that at least some interoceptive experiences represent homeostatic events is not new. Its early advocates included David Armstrong (1962, 1968) and George Pitcher (1970). Armstrong, as discussed at length in the previous chapter, argued that an entire class of interoceptive experiences (such as pain, itches, temperature sensations) are best understood as perceptions of '*certain idiosyncratic bodily happenings'* (1968: xxii). Pitcher, for his part, focused on defending the perceptual nature of pain, while acknowledging that other interoceptive states could, in principle, be subjected to the same treatment (1970: 393). It was not, however, until the publication of Michael Tye's *'Representational Theory of Pain'* (1995a) that debate on pain really took off. In this and later writings, Tye famously and convincingly argued that pain represents *'certain sorts of disturbances in the body, paradigmatically, bodily damage'* (2006: 113). Since then, indicativism has gained traction (see Bain 2003, 2007; Cutter and Tye 2011, 2014; Grahek 2007), with Brendan Ritchie and Peter Carruthers going as far as to propose that *all* interoceptive experiences represent fluctuations in homeostatic variables (2015: 357).

Let us now examine indicativism in greater detail. For each type of interoceptive experience, we can identify a homeostatic variable that the experience plausibly *represents*. Thirst may represent fluctuations in fluid levels. Hunger may represent fluctuations in energy levels. Breathlessness may represent fluctuations in oxygen levels. Pain is often viewed as representing bodily damage or disorder (Grahek 2007; Lycan 1996; Tye 1995a), although it may be preferable to frame it as representing departures from bodily integrity. Turning to heartbeat sensations, one would be hard-pressed to identify any candidate for their representational content other than changes in heart rate. Similarly, breathing sensations seem to straightforwardly represent the stages of respiration. Finally, some interoceptive experiences are easiest to

describe as representing the physiological state of a specific organ. For example, the urge to urinate represents bladder fullness, nausea represents indigestion in the stomach, itch represents skin irritation, and muscle fatigue represents muscular exertion.

Crucially, the proposal that the interoceptive system represents homeostatic parameters entails certain assumptions about the functioning of this system. These assumptions might seem obvious, but since they are relevant for later discussion, I will take the time to spell them out. First, the proposal entails that the interoceptive system must have a *parameter-based structure*: each representation it outputs represents the state of some specific homeostatic variable, and only that variable. Furthermore, the interoceptive system must have *homeostatic setpoints* in place. Indeed, if the interoceptive system is to detect fluctuations from homeostatic 'norms', it needs to know what the norms are in order to carry out this task. To give an analogy, I would not be able to tell you that a particular car was speeding even if I knew at what speed it was traveling – not unless I also knew the speed limit of the area and could thus distinguish, on the basis of that knowledge, between permissible and impermissible speeds. Accordingly, if the interoceptive system is to inform the agent that, say, her bladder is getting full, it needs to be built into this system that empty (or nearly empty) bladder is the homeostatic norm, or the *goal value* of the bladder fullness variable. Only a cognitive system which knows the goal values of the variables it tracks can determine when the values of these variables become *erroneous*.

A further implication of pure indicativism is that the interoceptive system does not simply detect errors in homeostatic parameters for its own purposes – it passes this information to the intended recipients. Who are the intended recipients? Well, if the interoceptive system represents homeostatic events, then it plausibly shares this information with the system(s) responsible for the autonomic regulation of homeostatic variables. This form of homeostatic regulation occurs outside our awareness or control, but it is difficult to imagine that it could be successfully carried out without accurate feedback about the current state of the regulated variables. What is important for our purposes, though, is that on the pure indicativist proposal, the interoceptive system also shares the up-to-date information about homeostatic events with

the agent by producing interoceptive experiences that represent homeostatic events. Small-scale fluctuations in homeostatic variables do not give rise to conscious interoceptive experiences: much like spam emails, they would not tell the agent anything important about her physiological body state anyway; so, there is no point in them entering awareness (Garfinkel, Critchley and Pollatos 2017). However, when the interoceptive system detects relevant (i.e. large enough) fluctuations in homeostatic variables, it produces conscious interoceptive experiences. In doing so, it ensures that the agent knows how her body is faring and can use this knowledge to guide her decision-making.

The conception of the interoceptive system as a '*homeostatic error detection machine*' outlined above echoes some popular theories of homeostasis (Berridge 2004: 180). Indeed, it is frequently assumed that the brain keeps homeostatic variables in check by implementing the following two functions: (F_1) tracking any deviations from the narrow range of goal values and then (F_2) taking actions to return the homeostatic variables to their goal values. If pure indicativism is correct, then (F_1) is performed by the interoceptive system and (F_2) is divided between the centres of autonomic homeostatic regulation and the agent. After all, deviations from homeostatic setpoints often require overt behaviour (e.g. foraging, seeking shelter) to be resolved. Since the agent is in control of her motor system(s), the interoceptive system must find a way to 'liaise' with the agent in order for the homeostatic behaviours to be carried out. On the pure indicativism proposal, the interoceptive system has developed the following solution to the *liaison problem*: to represent all significant homeostatic events to the agent.

All this is perfectly plausible. The interoceptive system may well function in this manner. Does it, though? And, more importantly, does pure indicativism provide us with an adequate explanation of HDs? These mental states belong to the class of interoceptive experiences, but they are phenomenally and functionally distinct from non-HDs. Can their distinctive profile be fully explained by their possessing purely indicative content? I address these questions in Section 4. Before turning to that, however, I must deal with a more pressing issue. In recent years, a number of philosophers have been openly critical of indicativism. In Section 3, I examine their objections and show that they are much less compelling than

they may initially appear. This clears the ground for my discussion of pure indicativism, which, I argue, is untenable – though not for the reasons usually cited.

3. Two challenges for indicativism – and why they do not succeed

As the popularity of indicativism grew, so did the number of philosophers who found fault with it (Casser 2021; Corns 2014; Klein 2015). Broadly speaking, there are two kinds of anti-indicativist arguments currently in circulation. Admittedly, these arguments often merge into one another. For the sake of clarity, however, I will to examine them in turn.

Both arguments arrive at the same conclusion – that the cognitive systems producing HDs cannot be in the business of tracking and representing homeostatic events – but they diverge in how they go about reaching it. The first, *the misrepresentation argument*, states that HDs misrepresent homeostatic events too often to be viewed as having the function of representing these properties. It relies on providing examples of misrepresentations which cannot be easily dismissed as rare system glitches, since such cases are common rather than exceptional. The second argument, *the computational* argument, goes further and claims that the very principles underlying the production of HDs rule out the possibility that HDs are meant to represent homeostatic events. It relies on demonstrating that HDs are generated at least partially on the basis of computations which have little or nothing to do with tracking the current state of homeostatic variables.

My aim is not to provide a comprehensive survey of the challenges to indicativism, but to determine the extent to which this position is threatened by them *en masse*. Consequently, I will only have the space to credit some of the authors who espouse anti-indicativist views. Additionally, many of the examples used in this section will concern pain, because this is what philosophers most frequently discuss, but the same points apply to other HDs.

3.1. The misrepresentation challenge

The first argument against indicativism is simple. Its starting point is the assumption that representational systems are, by and large, *good* – that is, generally accurate in representing the worldly or bodily properties that they are supposed to represent. Take, for example, this quote by Akins (1996: 342):

> *If the senses are the brain's window on the world, then any system worth its salt (and functioning correctly) ought to provide an accurate account of just how things are: the brain must be able to tell, from the signals it receives, how things stand in the world.*

On this assumption, if a putative representational system turns out to be *bad* at representing the properties it is supposed to represent, we would have a reason to doubt that the system has this function. Accordingly, if HDs can be shown to be bad at representing homeostatic events – that is, if they regularly misrepresent them – the proposal that HDs represent homeostatic events would be discredited. In line with this, many philosophical attacks on indicativism have focused on showing that the mismatch between HDs and homeostatic events is *sufficiently common* to pose a challenge to this view.

To clarify how the *misrepresentation challenge* works, I will use the following analogy. Imagine that you see me repeatedly throwing a ball in the direction of a basket, and you wish to establish, without asking me, whether my aim is to get the ball into the basket. You know little about me, but since I am wearing the uniform of a well-known basketball team, you infer that I must be a skilled ball thrower. However, after observing me carefully for some time, you realise that I only succeed at getting the ball into the basket about 75 percent of the time. Such a success rate is not what one would ordinarily expect of a top player, so you begin to doubt that my real aim is to get the ball into the basket. Perhaps, you reason, I am simply trying to perfect my throws or merely exercising my arms – for it seems inconceivable that a genuine professional, someone who throws balls for a living, would miss the basket as often as I do.

A clear example of the misrepresentation challenge can be found in a paper by Jennifer Corns (2014a). Corns argues that the correlation between *pain* and *bodily damage* is not sufficiently strong to support the hypothesis that pain represents bodily damage. More specifically, Corns observes that many pains, including ordinary (everyday) pains, arise in the absence of any obvious bodily damage, or else are disproportionate to the degree of damage sustained. Headaches, for example, are not usually related to any damage to the head. Similarly, lower back pain, one of the most common pain complaints, often manifests in the absence of any underlying lower back pathology, and even when pathology is present, the pain often fails to match its severity. The same is true for a number of other chronic pain conditions, such as chronic regional pain syndrome, fibromyalgia, irritable bowel syndrome, temporomandibular disorder (Harte, Harris and Clauw 2018). Corns also cites several other cases of dissociation between pain and bodily damage (such as Couvade syndrome), but since these cases are more rare than the ones detailed above, I will not list them here. After all, the argument is clear. If pains truly represented bodily damage, then *too many* pains would have to be treated as misrepresentations – as illusions when the degree of pain is disproportionate to the degree of damage, or as hallucinations when no bodily damage is present at all. This is a substantial theoretical cost. Accordingly, Corns concludes, pains do not represent bodily damage.

Now, there is no doubt that all sensory systems err from time to time. Well-known visual illusions (e.g. the Müller-Lyer illusion concerning the apparent mismatch in the length of equally long lines) are a case in point. Nevertheless, frequent errors in sensory systems are typically seen as pathological: hearing voices, for example, is a symptom of schizophrenia. Thus, if we agree that pain represents bodily damage, then we must regard pain arising in the absence of bodily damage as a dysfunction of the interoceptive system. The problem is that headaches and lower back pain are observed in the general population. Are we, then, to conclude that we are all routinely undergoing interoceptive experiences that misrepresent homeostatic events? Is the interoceptive system so poorly set up that it regularly fails to perform its proper function? Such a conclusion, Corns (2014a) maintains, is unacceptable. We must avoid it at all costs, and thus we have no choice but to abandon indicativism:

> *The mismatch between the perception of tissue damage at a bodily location and pain appears too prevalent for the perceptual view to remain credible (Corns 2014a: 371).*

I disagree with this verdict. In my view, the problem can be resolved without abandoning indicativism; here is my reasoning. To begin with, one plausible line of response to the argument presented above is to concede that misrepresentation is indeed more prevalent in the interoceptive modality compared with other sensory modalities – but *for good reasons*. First, as noted in the previous chapter, the data available to the interoceptive system is generally thought to be more 'compressed' and less reliable than the data available to exteroceptive systems, which increases the risk of error and leads the interoceptive system to be more reliant on its own predictions (Barrett and Simmons 2015; Feldman, Bliss-Moreau and Lindquist 2024; Seth 2013). Second, a proportion of misrepresentations may be attributable to features of modern lifestyle, in particular, to chronic stress. Recall that autonomic homeostatic control systems are capable of directly modulating homeostatic variables. When the stress response is mounted, large-scale hormonal changes ensue, and when stress becomes chronic, these changes can severely disrupt physiological functioning and make accurate data interpretation tricky (Barrett and Simmons 2015). Together, these points suggest that frequent misrepresentation in the interoceptive system may be reconciled with indicativism. After all, if our interoceptive systems are already prone to error, then we do not need to reconsider their function to explain why they exhibit even greater error rates when processing the data from the bodies of largely inactive and chronically stressed modern humans.

Another reason to resist the conclusion that our interoceptive systems misrepresent homeostatic events too often for this to be their proper function is that our interoceptive systems do not all misrepresent to the same extent. That is to say, medically unexplained pains may be common, but they do not affect us *equally*. Some people – for example, fibromyalgia sufferers – experience far more unexplained pains than others, and these pains are exceptionally severe and debilitating (Clauw 2015). Moreover, different kinds of unexplained pains tend to co-occur with each other, as well as with other unexplained symtoms (e.g.

fatigue and nausea), which has led researchers to posit the existence of common interoceptive disfunction mechanisms for these and related conditions (Barrett and Simmons 2015; Murphy and others 2017; Paulus, Feinstein and Khalsa 2019; Quadt, Critchley and Garfinkel 2018). In general, while it may be counterintuitive to classify unexplained HDs as perceptual errors, the fact that some individuals are particularly susceptible to them makes this interpretation more plausible.

Another reason to push back against the misrepresentation argument comes from the type of misrepresentations typical of the interoceptive system. Take pain. To be sure, it would be a problem if a system thought to be tracking tissue damage consistently failed to represent it when it occurs. However, most instances of pain misrepresentation are those of pain generated in the apparent absence of bodily damages, and the cost of false positives is much lower there than that of false negatives. A hypervigilant pain system would still be selected for, since it successfully performs the critical function – capturing bodily damage when it occurs – on which selection depends. By analogy, overzealous employees may do more work than they are contracted to do, but they are rarely fired from their posts; only lazy ones are. Consider also the social dimension of pain. In virtually all human societies, individuals who exhibit pain behaviours elicit sympathy, care, and support (Steinkopf 2016). Accordingly, far from being maladaptive, a hypervigilant pain system may confer an evolutionary advantage by ensuring that even minor injuries are attended to and thereby reducing the likelihood of more severe injury in the long term.[30]

Finally, the misrepresentation argument itself has an undesirable implication which should prevent us from accepting it. Although some philosophers have objected to the indicativist treatment of HDs, it is hard to imagine any of them denying that non-HDs have (purely) indicative content. What else could heartbeat sensations be about if not heart rate? Yet heartbeat perception accuracy is often strikingly poor, especially when measured at rest (Brener and Ring 2016; Khalsa and Lapidus 2016). If this is not enough to persuade us that non-HDs are devoid of indicative content, then we should apply the same considerations to HDs. Besides, medically unexplained HDs have been found to correlate with poor

[30] See Finlay and Syal (2014) for an evolutionary explanation of labour pain.

performance on tests of non-HDs, such as heartbeat and respiration detection (Di Lernia, Serino and Riva 2016; Horsburgh and others 2024). The most straightforward way to explain these results is by attributing them to a general impairment of the system responsible for generating both non-HDs and HDs. If, however, non-HDs have indicative content and HDs do not, then the two must be the products of different cognitive systems – in which case the straightforward explanation is no longer available.

I conclude that the misrepresentation argument does not pose a serious threat to indicativism. All sensory systems misrepresent from time to time, and it is not obvious how much error is enough to question their presumed functions. As argued above, even relatively frequent misrepresentation can be excused, and the broader implications of endorsing the misrepresentation argument further undermine its force. Nevertheless, the critics of indicativism would not let it off the hook so readily. HD-producing systems, they maintain, are not unintentionally misleading about the current state of the body. Rather, as evidenced by the types of computations they routinely perform, these systems are not even attempting to represent homeostatic events. This *computational argument* is analysed below.

3.2. The computational challenge

So far, I have argued that although the misrepresentation pill may be hard to swallow, there are ways of reconciling indicativism with occasional or even systematic misrepresentation of homeostatic events by HDs. Some philosophers, however, have opposed indicativism on slightly different grounds. Specifically, they have stated that if we look at how the HD-producing systems actually work – that is, if we consider the types of computations they routinely perform and the principles underlying these computations – we will have no choice but to abandon the idea that these systems are in the business of tracking and representing homeostatic events. On this computational argument, the main problem with indicativism is not that it falsely implies that the systems producing HDs are too dysfunctional to have ever been selected for. Rather, the problem with indicativism is that it fundamentally misconstrues the function of these systems.

In order to simplify this argument, I will once again turn to the *'ball and basket'* metaphor introduced earlier. Imagine that you see me throwing a ball in the direction of a basket, and I only manage to get the ball into the basket about 75 percent of the time. You know that I am an experienced ball thrower, but you also know that even experienced ball throwers can err. Perhaps, there is something wrong with my arm, the room is too dark, or else the basket has been placed too far away from me to allow for greater accuracy. So, you decide that on the basis of my performance only, it is not possible to tell whether I am actually trying to get the ball into the basket or not. After observing me closely, however, you realise that I have had my eyes closed the entire time I have been throwing the ball. On the basis of this, you conclude that I cannot possibly be trying to get the ball into the basket. After all, if I were, I would keep my eyes open. I must, then, be exercising my arms or pursuing some other, less obvious goal.

In a similar way, the computational argument against indicativism holds that we can infer *what a cognitive system is for* by analysing the specific ways in which it operates. The computations underlying the production of HDs, the argument goes, are not the sorts of computations we would expect to find in a system whose proper function is to produce accurate representations of homeostatic events. Therefore, HDs are *not* best understood as representations of homeostatic events.

The computational argument can be further clarified through the following two examples from the philosophical literature:

> *Example #1: The non-veridicality of temperature experiences.*
>
> Kathleen Akins (1996) points out that thermal receptors (1) have an uneven distribution pattern on the skin and (2) respond to stimulation in a *non-linear fashion* – that is, their response depends on the initial temperature of the skin. This is why, for example, tepid water feels warmer if one's hands are cold, and why some body areas (such as the groin or the head) feel colder than others upon immersion into cold water. On the basis of this *built-in*

subjectivity of the temperature system, Akins (1996) concludes that it cannot be regarded as representing the objective temperature of the body.

Example #2: Pain modulation.

It is well-established that pain experiences are shaped by attentional focus, affective states, expectations, and personal-level beliefs (Ossipov and others 2010; Thomaidou and others 2021; Tracey and Mantyh 2007). This phenomenon, known as *pain* modulation, refers to the top-down influences that can either facilitate or inhibit pain processing. Some examples of pain modulation – such as the placebo and nocebo effects or the impact of mood on pain intensity – are familiar and commonplace. Others – such as reports of severely injured soldiers experiencing little or no pain during battle – are more striking (Beecher 1946). In a recent paper, Laurenz Casser offers the existence of pain modulation as evidence that pain is the product of a system whose function is *not* to gather information about tissue damage. As he puts it, '*modulating mechanisms in the nociceptive system systematically prevent pain from serving a primarily informative role*' (2021: 364).

How can indicativism respond to the computational argument? I suggest that if HDs are framed as representing *fluctuations from homeostatic setpoints*, the apparent non-objectivity of HD-producing systems can be explained away. Take the first example provided above. If certain body parts (e.g. groin or head) feel colder than others when we enter cold water, it is tempting to conclude that the temperature system does not respond to temperature changes in a consistent, law-like manner and thus does not care about the objective body temperature. Such a conclusion, however, presupposes that optimal temperature values are the same across all body parts – and this assumption may be called into question. Given that the protection of reproductive organs is paramount, and that the head is a major site of heat loss, it seems entirely plausible that different body parts have distinct temperature requirements. And if there are different homeostatic setpoints in place for different body parts, then the same *external event* (entering cold water) can produce different *homeostatic events* in different body parts.

An indicativist can also appeal to the *malleability* of homeostatic setpoints to show that HD-producing systems are only seemingly non-veridical. The needs of the body are not fixed – and neither are homeostatic setpoints, which may undergo temporary or permanent modifications in response to the shifting demands on them (McEwen and Wingfield 2003). Immune activation, for example, is associated with a temporary change in the goal body temperature (Balli, Shumway, and Sharan 2023). Similarly, the researchers have now come to appreciate that obesity and the development type two diabetes may involve a maladaptive permanent increase in the defended level of blood glucose (Alonge, D'Alessio and Schwartz 2021). HD-producing systems, then, may be argued to aim at accurately representing homeostatic events after all – it is just that the properties they track are dynamic and highly complex.

A similar line of defence may be employed with respect to the second example provided above. At first glance, the phenomenon of pain modulation does not sit well with indicativism. If pain-producing systems track disturbances of bodily integrity, why should they be sensitive to beliefs, expectations, and affective states? To this, an indicativist might respond that these factors feature in pain computations because they carry relevant information about bodily integrity. Take the placebo effect. Suppose I suffer from gastritis and am given a pill purported to treat this condition. The pill contains no active ingredients; yet my belief in its efficacy alleviates my stomach ache. One plausible explanation for this is that my pain-producing systems interpret this belief as evidence that my bodily integrity is about to be restored and thus decrease the intensity of pain accordingly. After all, why else would pain-producing systems be sensitive to our beliefs about the efficacy of treatment if not because such beliefs are a source of information about bodily integrity?

Given that HD modulation through *beliefs* may be interpreted as an extension of the indicative function of HDs, it does not constitute a serious threat to indicativism; but what about HD modulation through *affective states*? While this phenomenon has been mainly studied in relation to pain, it also extends to breathlessness, nausea, and fatigue (Constantinou and others 2013; Köteles 2021: 319; Sharma and

others 2016).[31] Specifically, negative affect tends to intensify negatively valenced HDs, although depressed individuals may instead exhibit reduced hunger and elevated pain thresholds (Ben-Tovim and Schwartz 1981; Dickens, McGowan and Dale 2003).

Can we reconcile the affective modulation of HDs with indicativism? I suggest that we can, since affective states alter homeostatic needs. Negative emotions like anger, fear, and anxiety, for example, promote behaviours with high energy costs, thereby amplifying the current energy demands. Chronic stress in particular tends to lead to increased appetite, presumably either through the sensitivity of the hunger-producing systems to cortisol levels, or because it automatically raises the levels of ghrelin, the so-called hunger hormone (Bouillon-Minois and others 2021). In depression, by contrast, diminished hunger and pain may reflect an increased need for *energy conservation* under the conditions of low motivation to act. In general, it would seem that affective states can shape HDs indirectly, by changing the bodily reality that HDs represent.

More generally, the phenomenon of HD modulation strikes me as a useful adaptation for a system representing homeostatic events. The reason for this is straightforward: HDs have the capacity to seriously disrupt our behaviour. For instance, if I am in pain, I may be prevented from foraging, hunting, or taking care of my children. The same point applies to other HDs such as nausea, itch, thirst, and fatigue. To be sure, the experiences generated by other modalities can also be disruptive. For example, if I am smelling a disgusting odour, I may be unable to eat. Or, if I am hearing a loud sound, I may be unable to sleep. The crucial difference, however, is that when we are faced with perceptual experiences of the world that both provide information and simultaneously interfere with our ability to go about our normal life, there are things that we can do *to stop having these perceptual experiences*. If I sense a vomit-inducing smell or hear an aversive sound, I can simply move away from their source. Unfortunately, such a strategy is not available in the case of HDs: we cannot escape our own bodies.

[31] Non-HDs, such as heartbeat sensations, are also subject to affective modulation (Constantinou and others 2013).

Based on this, I suggest that a system representing homeostatic events should be expected to have acquired a mechanism for regulating its outputs. Such a mechanism would not be merely advantageous but virtually indispensable: while it may be good to have highly informative sensory systems, it is maladaptive to have this information provided at the expense of the agent's ability to act on it. Indeed, if I could hear everything within a two-mile radius, I would have a great deal of auditory information. Yet being constantly exposed to it – including numerous loud sounds – would quickly become intolerable and would severely compromise my capacity for normal behaviour.

To further illustrate this proposal, I will make use of the following analogy. HDs, I suggest, are like *bills*. Both are unpleasant, both compel us to act in specific ways, and both can arrive unexpectedly.

Crucially, HDs exert these effects only when they are consciously experienced. Indeed, if I sustain bodily damage but feel no pain, then I experience neither distress nor compulsion to act. Now, in the long run, the absence of pain is harmful: without it, I would eventually inflict too much damage on my body and die. Yet, having too many pains is also detrimental to my survival prospects: if I am constantly attending to the body parts which pain me, I have no time to monitor other homeostatic parameters. The same point applies to thirst, hunger, and other HDs.

Bills, like HDs, are only unpleasant and action-forcing if they are actually received. Indeed, if they are lost in the post or eaten by my dog, they cannot upset me and I do not feel the need to pay them. Once again, never paying bills is unsustainable: eventually, my utilities will be cut off. Yet, paying every bill I receive may not be the optimal solution, especially if I am strained for money. In that case, I might keep the lights on, but I risk starving to death – not to mention that some bills come from scammers and should be ignored. So, what do I do?

I build a robot. The robot has access to all the information about my finances and reads all my letters. At the same time, the robot does not pay the bills: this decision is made by me. The robot, then, opens all my

letters and, when it finds bills, checks if I can afford to pay them. If I can – and there are no other pressing expenditures – the robot passes them on to me. If I cannot – say, I need the money to pay off a thug threatening to kill me – it throws them in the bin. The question is: in this case, do bills still have the function of informing me about the money that I owe? I say that they do: the robot merely interferes (in an adaptive way) with their function.

The computational challenge, then, does not provide a strong reason to reject indicativism either. Of course, a critic may feel that indicativism has been sufficiently compromised by the objections raised against it. Indeed, to save indicativism, I have had to posit that the systems producing HDs function in a rather unusual way. They are prone to errors, and they do not have stable setpoints against which incoming sensory data can be compared. Moreover, a defender of indicativism may have to accept that, rather than representing current homeostatic events, HDs actually represent future ones. As recent research demonstrates, thirst is triggered predictively before dehydration occurs following food consumption, and terminated predictively before fluid balance has actually been restored (Zimmerman and others 2016, 2019). What underlies this predictive capacity is that thirst centres in the brain receive sensory input from the mouth and throat, which means that they 'know' when we drink and eat. As a result, thirst computations – and the experiences they generate – are based not only on the current, but on the *predicted* levels of hydration. None of this, however, is particularly concerning. HD-producing systems could well function in this manner, and as long as the problems of indicativism do not outweigh its advantages, it remains viable as a theory of HD.

I conclude, therefore, that indicativism can be defended from the standard objections. Nevertheless, in the next section, I argue that *pure indicativism* fails to explain every aspect of HDs. The possibility that these mental states have indicative content, I hold, is still on the table – but the claim that they have *only* indicative content should be rejected.

4. The problems of pure indicativism

On the pure indicativist proposal, *all* interoceptive states – from heartbeat sensations to hunger and nausea – represent homeostatic events. What, then, accounts for the phenomenological and functional distinction between HDs and non-HDs? Specifically, what enables HDs, but not non-HDs, to exert motivational pull and successfully promote homeostatic behaviours?

This question can be answered by specifying the pure indicativist proposal as follows. The interoceptive system processes homeostatic data and generates representations of homeostatic events. These representations are motivationally inert, but they contain information that is highly relevant to our survival and well-being. For this reason, they are further processed by a central system whose role is to generate motivations – and through this they acquire motivational character.

In this scenario, the motivational profile of HDs is the product of the central *motivational system*. Presumably, this system accesses interoceptive representations, selects those that warrant action, and outputs motivations to match them: eating, drinking, resting, urinating, vomiting, and so on. As a result, the interoceptive representations marked as actionable by the motivational system – HDs – are experienced by the agent as motivations to perform specific homeostatic behaviours, while interoceptive representations that are ignored – non-HDs – are not. The distinction between HDs and non-HDs thus boils down to whether the homeostatic parameters they represent require behavioural regulation. Heart rate does not, so the motivational system is not interested in its representations. By contrast, fluid and energy levels cannot be maintained without drinking and eating, respectively, so the motivational system has evolved to prioritise their representations, efficiently and reliably outputting appropriate motivations.

Having thus refined the pure indicativist proposal, we can now evaluate it. To begin with, pure indicativism has the advantage of providing a comprehensive account of the sensory phenomenology of interoceptive states, including HDs. First, interoceptive experiences differ in kind because they represent fluctuations in *different homeostatic variables*. For example, what unites pains and sets them apart from itches is that

pains represent bodily damage and itches represent skin irritation. Second, interoceptive experiences of the same kind are distinguishable from one another because they represent *different types of fluctuations* within the same variable. For example, the pain of a pulled muscle is not the same as the pain of a burn, because the former represents muscle strain, while the latter represents thermal damage. Third, interoceptive experiences vary in *intensity* because they represent the extent of fluctuation. For example, a mild urge to urinate represents the bladder as partially filled, while an intense urge to urinate represents the bladder as nearly full. Last, interoceptive experiences have *felt locations* because they represent homeostatic fluctuations as occurring in particular body regions. For example, nausea is felt in the stomach because this is where it represents indigestion as taking place. And if some interoceptive experiences (such as fatigue) are generalised, this is presumably because they represent homeostatic events involving the body as a whole.

At the same time, I contend that pure indicativism cannot account for the motivational profile of HDs. Say that the sensory character of HDs and the motivational character are the products of the interoceptive and the motivational systems, respectively. This two-system proposal does explain how the same type of HD may be experienced as motivating different kinds of actions: for example, how pain can variously motivate withdrawal, inaction, taking a painkiller, rubbing the damaged body part, and crying for help. Yet it also turns HDs into *compound mental states* with *impure content*, which brings pure indicativism to an end.

The proposal itself is also flawed. For one thing, if the motivational component of HDs is the product of the same *domain-general system* that generates all our motivations, then HDs must exhibit the same motivational phenomenology as our other motivations – including, for example, the desire to visit Paris. Yet this is not the case. First, as argued previously, the motivational phenomenology of HDs is that of an *urge* rather than a *desire* to act. While HDs may align with desires, they can exert their motivational force without them and against our will, as exemplified by the intrusive, compulsion-like drive to scratch an itch. This suggests that urges and desires are generated by distinct cognitive systems. Second, while the sensory and motivational phenomenology of HDs are typically *fused*, the same cannot be said of other

perceptual states that give rise to motivations. To illustrate, when I look at a flower and feel drawn to smell it, the visual experience of the flower and the motivation to smell it are distinct parts of my experience. By contrast, the motivational phenomenology of itch or bladder fullness is virtually impossible to separate from its sensory phenomenology.

A closely related issue is that non-HDs do not exhibit the same motivational phenomenology as HDs even when they *do* give rise to motivations. Imagine someone who suffers from heart disease and knows it. Noticing that their heart rate is accelerating would motivate this individual to stop what they are doing and rest, but this motivation would neither feel like an urge nor be fused with the heartbeat sensations.

Furthermore, if the motivational system is *domain-general*, then the computations that determine the motivational force of HDs should involve data beyond what the interoceptive system supplies. This, however, does not appear to be the case, as illustrated by the following examples.

First, although one's desire to visit Paris is responsive to various considerations (such as hotel prices or negative reviews online), one's hunger is immune to one's knowledge of the need to lose weight. Why? Perhaps our motivational systems are simply configured to prioritise biological needs over adherence to beauty standards. This response, however, does not explain why the motivational systems of overweight individuals do not factor in the health risks of obesity when generating hunger.

Second, consider the more dramatic case of pica, a disorder characterised by a strong compulsion to consume inedible items such as chalk, clay, or soap (Parry-Jones and Parry-Jones 1992). Individuals with this disorder do not want to eat chalk, do not consider it nutritious, and know that their behaviour is unusual and potentially dangerous. Yet none of these factors has any effect on the powerful drive to seek and consume chalk. Food cravings are similar: we are all aware that crisps and chocolate are detrimental to our health, yet we cannot stop wanting them. How can the motivational system be so badly set up that it ignores highly relevant information when generating the motivational component of HDs?

Finally, consider, once again, pain asymbolia, a rare consequence of brain injury characterised by indifference to pain and the lack of avoidance or protective behaviour in response to painful stimuli. As noted earlier, philosophers disagree about how to interpret pain asymbolia: some argue that it involves the loss of the affective character of pain (Bain 2014), while others insist that asymbolic patients are simply no longer motivated by their pain, even if it still feels unpleasant (Klein 2015). What is important to us here, however, is fairly uncontroversial: asymbolic patients recognise pain as such, describe its intensity and other sensory features – yet make no attempt to protect themselves. This goes against the proposal under discussion: if asymbolic patients retain the concept of pain, then their decision-making system – which appears to be intact, given that other HDs motivate as normal – should generate appropriate protective motivations. What conclusion can we draw from this, if not that the motivational aspect of HDs is not the product of the domain-general motivational system, but of some domain-specific system which has access only to interoceptive information and is *encapsulated* from non-interoceptive input?

An indicativist could avoid the objections outlined above by proposing a different cognitive architecture of HDs – one in which the motivational component of HDs is generated by a domain-specific *homeostatic urge system* that processes interoceptive representations and outputs urges to match them. Such a cognitive architecture would need to be carefully specified, but it seems possible.[32] A pure indicativist, however, has no recourse. If HDs and non-HDs are phenomenologically and functionally *distinct*, they cannot have *the same representational content*. Further representational content can be posited to explain HDs, but this would entail *rejecting pure indicativism*.

5. Conclusion

While indicativism has been extensively criticised in recent years (Casser 2021; Corns 2014a; Klein 2012,

[32] Note that if such a system exists, we would expect it to be selectively impaired in some individuals. Yet I cannot think of a suitable example. Pica is associated with abnormal eating urges, while pain asymbolia involves the absence of normal protective urges; but these cases do not coincide, which suggests the existence of *multiple* urge systems. Such a cognitive architecture of HDs, however, is so cumbersome that it becomes unattractive.

2015), I have argued that it can withstand the standard objections made against it. It also has considerable explanatory value: not only does it explain why we think of HDs as reliable indicators of the physiological state of the body, but it also clarifies why we use them as such in decision-making and planning. Examples of the latter include relying on fatigue to determine whether one is able to do another set of weight exercises and choosing when to come to the surface for air when free diving on the basis of the experience of air hunger. Another notable example comes from the studies that found that even imagining the movement of the hurting limb is inhibited in chronic pain – and that the degree of this inhibition corresponds to the severity of pain (Schwoebel and others 2002). Together, these examples illustrate that we and the cognitive systems that make up our minds tend to treat HDs as if they have indicative content – and this, in turn, strongly suggests that they do.

Indicativism, then, is tenable – but *pure* indicativism is not. As I have argued, it explains the sensory phenomenology of HDs, but it fails to accommodate their motivational profile. We may, therefore, accept some version of indicativism, but not pure indicativism. The problem is that saying that HDs probably have indicative content does not amount to a theory – and, moreover, there is a clear alternative to this view that deserves our attention. According to Klein (2007, 2012, 2015), indicativism is a mistake: HDs have only imperative content. I consider Klein's *pure imperativism* in the next chapter.

Chapter 3:

Pure Imperativism

Abstract:

In the previous chapter, I discussed the possibility that HDs represent homeostatic events. I concluded that while *pure indicativism* does not fully capture HDs, it is likely that HDs have *some* indicative content. In this chapter, I continue to explore the representational content of HDs by introducing, analysing, and ultimately rejecting the main alternative to indicativism: *pure imperativism*. On this account, HDs do not inform us about specific homeostatic events, but instead command us to engage in specific homeostatic behaviours. I argue that pure imperativism does not provide a quick and easy solution to the problems facing indicativism; on the contrary, it suffers from serious flaws of its own and therefore cannot be accepted as a theory of HDs. Our next step, I suggest, should be to consider *impure theories* of HDs.

1. Pure imperativism as a challenge to indicativism

In the previous chapter, I rejected pure indicativism as a theory of HDs on the grounds that it falls short of capturing every aspect of these mental states. In this chapter, I ask whether we should wholly abandon the idea that HDs have any indicative content in favour of an alternative: *pure imperativism*. In recent years, a number of philosophers – including Hall (2008), Klein (2007, 2012, 2015), and Martínez (2011) – have advocated an imperativist approach to HDs such as pain and itch on the basis of their strong link with action. Most prominently, Colin Klein (2007, 2012, 2015) has argued at length that we need not appeal to indicative content at all to explain the key features of HDs. Klein's position, in other words, is that HDs have *only* imperative content: they are *pure commands,* that is, mental states that direct the agent to perform specific homeostatic actions rather than inform the agent about specific homeostatic events.

Can we really do without indicative content? Although pure imperativism has been around for a few years, it continues to be controversial. This is because it asserts that HDs – which are typically understood to be *regular* perceptual states with indicative content (Carvalho and Damasio 2021) – are something else entirely: mental states with purely imperative content, and thus more akin to motor intentions (such as an intention to lift the arm) that direct our actions than to visual or olfactory experiences that tell us how things stand in the world. This claim may be difficult to accept, and for good reasons. The products of other sensory modalities are generally thought to represent worldly or bodily properties (Macpherson 2011); why would interoception depart from this pattern by issuing commands? And yet, if pure imperativism can fully explain the philosophically interesting aspects of HDs, then it does not matter how revisionary it is. If the theory is sound, it deserves to be taken seriously.

Accordingly, in this chapter, I set out to evaluate whether pure imperativism compares favourably with indicativism in its ability to account for the key features of HDs. I conclude that it does not. My argument proceeds as follows. In Section 2, I introduce pure imperativism and outline its main commitments. In Section 3, I examine how this theory purports to explain the distinctive phenomenology and function of HDs. In Section 4, I show that the explanatory power of pure imperativism is limited, especially with respect to the sensory phenomenology of HDs. In Section 5, I analyse Klein's argument in support of pure imperativism. In Sections 6–8, I show that this argument fails, though HDs do have imperative content of a certain kind. At the end of the chapter, I propose that given that both *pure theories* of HDs have failed, our next step should be to explore the possibility that HDs have both indicative and imperative content.

2. Pure imperativism: An overview

In a book and a series of papers, Colin Klein (2007, 2012, 2015) makes a case that a large number of interoceptive experiences classified in this thesis as HDs have imperative content, and no other type of content. While his arguments mainly focus on pain, Klein maintains that it is akin to other '*homeostatic sensations*' – such as hunger, thirst, itch, breathlessness, fatigue, nausea, temperature sensations, toilet

urges – in that all these mental states are best understood as *commands to carry out specific types of homeostatic behaviours* (eating, drinking, scratching, urinating, seeking shelter, resting, and so on) (2015: 14–21, 31).[33] More precisely, Klein suggests that HDs are commands directed at the agent, whom they motivate to act in a particular way. My thirst, therefore, is not just an instruction to drink; it is an instruction to drink which has a recipient, and the recipient is myself.

> *Hunger, thirst, and itches… are sensations that play a crucial role in keeping us alive and intact, and they do so precisely by commanding us to act in appropriate ways* (Klein 2015: 8).

In order to get a better sense of Klein's proposal before proceeding to evaluate it, let me clarify what it means for HDs to be *commands*. Sentences in natural languages can have indicative (or descriptive) content: they can describe how things are in the world and thus can be true or false (Martínez 2011: 75). For example, the sentence *'This kettle is blue'* has indicative (or descriptive) content. It describes the colour of the kettle; it is true if the kettle is actually blue, and false if the kettle is not blue. Sentences in natural languages, however, can also have imperative (or directive) content: they can be commands to carry out specific actions. For example, a sentence like *'Boil this blue kettle!'* has imperative content. Sentences with imperative content do not have truth values, only satisfaction criteria. That is to say, they can be satisfied (if the blue kettle is boiled) or not satisfied (if the blue kettle is not boiled), but they cannot be true or false.

The content of *mental representations* can likewise be indicative or imperative (Martínez 2011; Millikan 1984; Shea 2018).[34] Perceptual states, for example, are generally thought by philosophers to have indicative content, representing various aspects of the world (Shea 2018: 100). Indeed, when we have a visual experience of a tree, an auditory experience of a child crying, an olfactory experience of rotten

---

[33] To be clear, Klein thinks that the category of homeostatic sensations – which, on his view, all have imperative content – is broader than this list and should remain open to new members. Indeed, he remarks that some homeostatic sensations do not have a name and suggests that they are not necessarily innate, citing cigarette craving as an example (2015: 14).

[34] Klein and Martínez (2016) and Shea (2018) provide more in-depth discussions of what it means for a mental state to have imperative content.

food, or a tactile experience of something slimy, these experiences can all be assessed for truth or falsity. At the same time, some non-perceptual mental states have been conceived by philosophers as having imperative content. Take, for example, motor intentions, such as the intention to lift your right arm (Pacherie 2000). While this mental state concerns the position of your body, it does not seem to be describing it. Instead, it seems to be directing the motor systems to carry out a specific action (lift the right arm). When this action is completed, the intention is implemented, or 'satisfied'; but there do not seem to be any conditions under which intentions can be evaluated as true or false. And this is why these mental states are proposed to have imperative rather than indicative content.[35]

According to pure imperativism, HDs are more like motor intentions than perceptual states. This may seem like an odd notion, and in defending it, Klein certainly goes against the grain. Indicativists (Bain 2003, 2007; Cutter and Tye 2011; Grahek 2007; Pitcher 1970; Tye 1995a) have long emphasised that HDs have *typical bodily causes*. For example, thirst is usually triggered by dehydration, the urge to urinate by bladder fullness, and itch by skin irritation. Based on this, indicativists have assumed that HDs represent the homeostatic events causing them. Yet, HDs also have typical behavioural *effects*. For example, thirst normally results in drinking, the urge to urinate results in urinating, and itch results in scratching. This means that HDs may also be framed as commands to carry out homeostatic actions. On this proposal, all one needs to do to identify the content of a mental state belonging to the category of HDs is to identify the type of behaviour that this mental state typically engenders. Thirst, then, is a command to drink, hunger is a command to eat, itch is a command to scratch, breathlessness is a command to breathe in, temperature sensations are commands to warm up or cool down, and pain is a command to protect some part of the body (Klein 2015). And so on.

But what about the fact that HDs seem to be informative about the physiological state of the body? More to the point, what about the fact that we act as if they do? Klein neither ignores nor dismisses this aspect of HDs. What he denies, however, is that HDs have *the function* to inform (Klein 2012: 919). As Klein sees

[35] Millikan (1995) points out that motor intentions also describe the outcome of an action. On this basis, she argues that these mental states have the kind of content which is at once indicative and imperative – namely, *pushmi-pullyu* content. Pushmi-pullyu representations are discussed in detail in Chapter 4 of this thesis.

it, we merely *infer* what is going on in the body from the homeostatic commands we receive. In his own words, *'pain, like any imperative, may indirectly inform us about the occurrence of the situation that caused it to be issued'* (Klein 2007: 527).

To clarify this point, consider the following analogy. Imagine that we are in the same room and I ask you to open the window. While you can infer from my request that the window is currently closed, the purpose of my request is not to notify you of this state of affairs. The purpose is simply to get you to perform a particular action. By the same token, you can infer from feeling thirsty that you are dehydrated. Most of the time, this inference would be quite correct, too – unless you suffer from a medical condition that causes excessive thirst. Still, according to pure imperativism, thirst does not actually tell you that you are dehydrated. Thirst simply shouts 'Drink!' at you whenever the thirst-issuing cognitive system deems drinking necessary.

Summarising this, the indicativists take it for granted that HDs have evolved to serve as objective markers of homeostatic variables. They believe that HDs are meant to provide us with insight into the internal bodily environment in the same way that visual, auditory, and olfactory experiences are meant to reveal certain aspects of the external world (Klein 2015: 3). Pure imperativism offers a strikingly different perspective on HDs. Specifically, it rejects the popular idea that these mental states are *informative by design* and frames them instead as experiences that are not transparent about their bodily causes. Klein's flagship example here is pain: as he puts it, when his ankle aches, the aching keeps him from walking on it without telling him what exactly is wrong with the ankle (2015: 4). Another example Klein offers is air hunger: this sensation, he stresses, is triggered by the accumulation of carbon dioxide rather than by the lack of oxygen – yet we can only learn this from a textbook, and not via introspection (2015: 18). All in all, for Klein, it is the defining feature of HDs that they *'motivate rather than inform'* (2015: 8) and *'provoke action rather than deliberation'* (2015: 6). In the next section, I examine Klein's pure imperativism in greater detail, focusing on how it purports to account for the function and phenomenology of HDs.

3. Pure imperativism: Under the microscope

Pure imperativism is austere: it affirms that HDs have *only* imperative content. Simple theories are attractive, but can this one explain all the distinctive characteristics of HDs by appealing to a single type of content? In what follows, I answer this question by considering, in turn, how pure imperativism handles (1) the distinctive function and (2) the distinctive phenomenology of HDs. In doing so, I conclude my presentation of this philosophical position and lay the groundwork for its subsequent evaluation.

3.1. Pure imperativism and the function of homeostatic drives

At a glance, pure imperativism offers a straightforward explanation for the function of HDs. As commands, HDs are supposed to get us to act in accordance with our current homeostatic needs by simply telling us to eat, drink, scratch, warm up or cool down, vomit, and so on. Note that in this scenario, we do not need to decipher the meaning of HDs. Since they are *felt imperatives*, we immediately grasp what these experiences are instructing us to do (Klein 2015: 15). The only thing we have to work out is how best to carry out the instructions we receive (Klein 2015: 17). This task is taken care of by our decision-making system(s), which ultimately determine what we eat, when we drink, and which clothes we wear when we feel cold.

What is not immediately obvious in this story, though, is why we should be compliant with the demands of HDs. After all, it is not uncommon for these experiences to be inconvenient and for their fulfilment to go against our goals and desires. A sudden urge to urinate, for example, may interfere with one's enjoyment of a film. Similarly, hunger may undermine one's commitment to losing weight. Given this, why do we listen to what HDs instruct us to do? We do not, after all, obey all the commands we receive. Often, we comprehend them, yet choose to do nothing – or do the exact opposite. In other words, being instructed to do something is not sufficient to spur us into action. Something else is necessary, something that turns a possible course of action into a *motive*. For example, I would not stop eating bread if a stranger (or a

friend, for that matter) urged me to give up carbs in order to slim down. However, I would certainly cut out this product from my diet if my doctor informed me that I had developed a gluten intolerance.

Merely positing that HDs are commands, then, is not enough to explain how these mental states motivate us to act. As Bain (2011) puts it, *'it is precisely because commands don't constitute urges that you can receive commands you have absolutely no urge to obey, e.g. a retaliatory command from your five-year-old that you go to bed without your dinner'* (p. 180). Klein is well aware of this issue, though – and he offers a solution. Its starting point is the idea that homeostatic commands are accepted by the agent as being *issued by the body* (Klein 2015: 74). Given that commands typically have a sender as well as a recipient, Klein suggests that we naturally assume, for want of any other candidate, that the source of homeostatic commands is the body itself. Accordingly, on Klein's account, HDs are not just instructions to act in a certain way – they are *the body's instructions* on how to take care of it.

Building on this idea, Klein further theorises that we obey homeostatic commands because we accept the body as *a practical authority* on its own needs (2015: 80). More specifically, Klein conjectures that HDs are always felt against the backdrop of an *'innate'* and *'difficult to overcome' attitude of acceptance* towards them (2015: 81, 83). As a result of having this attitude in place, whenever one feels thirst, one does not merely register a command to drink – one feels instructed to drink by a trustworthy and authoritative source. And what turns the body into such a source, according to Klein, is our *attitude of care* towards our own bodies. To put it bluntly, we want to survive, and so we are disposed to pay attention to what our bodies are telling us to do. As Klein puts it, *'we accept the body as a minimal practical authority because we care about our bodies […] that care is stable and constant, at least in ordinary cases'* (2015: 143).

Pure imperativism, then, is forced to appeal to something other than the imperative content of HDs to explain how these mental states are able to promote homeostatic behaviours. As mere commands, Klein acknowledges, HDs would fail to consistently move us: there is nothing about imperatives that automatically guarantees compliance. It is the standing attitudes of *bodily care* and *bodily trust* that we

exhibit towards them, Klein suggests, that turn homeostatic commands into fully fledged motives. In his own words, *'against the background of acceptance, the presence of an imperative is sufficient to motivate'* (2015: 83). Later on, I outline my own thoughts about this proposal. For now, however, I reserve my judgment and proceed directly to examining how pure imperativism deals with the distinctive phenomenology of HDs.

3.2. Pure imperativism and the phenomenology of homeostatic drives

HDs, as argued previously, exhibit three types of phenomenology: bodily (sensory), affective, and motivational. In Klein's view, valence, as a shared quality of affective states, is not an intrinsic quality of HDs (2015: ch. 4). This leaves pure imperativism with only the sensory and motivational phenomenology of HDs to explain.

Let us consider how Klein handles the motivational phenomenology of HDs first. As clarified earlier, Klein does not claim that HDs motivate simply in virtue of being commands. Instead, he proposes that HDs motivate because we are disposed to take them seriously. On this view, HDs have motivational phenomenology because they are felt *'against the background of acceptance'* (Klein 2015: 83). If it were not for the attitude of bodily trust – which is, in turn, grounded in the attitude of bodily care – we would understand the demands of HDs, but we would not experience them as urging us to act.

Next, let us look at how Klein addresses the sensory phenomenology of HDs: their felt location, intensity, and quality. Here, Klein focuses almost entirely on pain. Its felt quality, for example, is proposed by Klein to be determined by the type of protection command issued by the relevant cognitive system (2015: 98):

> *The difference between the sharp acute pain when you touch a stove and the later particular dull ache as your burn heals is a difference in what you do to protect the area. When you touch the stove, you protect your hand by removing it as quickly as possible. When your burn is*

healing, you protect the same patch by treating it gingerly. Different demands lead to different feels.

Let us now unpack this idea. As noted earlier, there are many kinds of pains: stinging, dull, throbbing, burning, tearing, and so on. In his earlier work, Klein (2007) characterised pains as *'negative imperatives'* proscribing actions potentially harmful to the body (such as stepping on a blistered heel or keeping one's hand in contact with a very hot object). Later, however, Klein (2015) significantly revised his position, arguing that pains are better understood as commands *to protect* some part of the body. This is why they all feel painful and are classified by us as the same kind of mental state. Yet, for Klein, each pain also specifies *how* the body needs to be protected. This is why pains differ in their felt quality: they instruct the agent to perform different protective actions. Other HDs can be explained in this way, too. For instance, Klein thinks that we can distinguish between hunger for salt and hunger for protein, presumably based on what each instructs us to consume (2015: 14).

When it comes to the felt intensity of HDs, Klein suggests that this phenomenal feature is determined by the ranking assigned to the particular command by the cognitive system that issues it (2015, ch. 8).[36] On this proposal, just as the commands we address to each other can be more or less urgent (e.g. *'Buy milk!'* vs *'Go buy milk right now!'*), so too can the commands sent to us by the body. The more important the relevant cognitive system judges it to be that a particular command is carried out, the more intense the experience is. Mild pain or itch, for example, are low-priority commands, while severe nausea and thirst are high-priority ones.

Finally, Klein seeks to explain the felt location of HDs by highlighting that homeostatic actions have bodily *targets*. The idea is that HDs are located because they command actions aimed at specific parts of the body (Klein 2015: 88–95). Pain in the stomach, for example, is felt there because it commands the agent to protect that region. Itch in the ear, in turn, is felt there because this is what it commands the agent to

[36] This hypothesis is the product of collaboration between Klein and Martínez (2016).

scratch. And so on. As Klein puts it, *'We naturally locate commands in the regions with which they are concerned. While your body says "protect there!", it feels as if a particular region says, "Protect me!"'* (Klein 2015: 89).

All in all, pure imperativism appears to be plausible and reasonably detailed. Despite this, I hold that it cannot be accepted as a theory of HDs. In what follows, I expose the explanatory weaknesses of pure imperativism. Next, I show that Klein's argument for it is based on an incorrect assumption about HDs. Ultimately, I conclude that while HDs have imperative content, they (1) do not have *purely imperative content* and (2) do not have *the kind of imperative content* proposed by Klein.

4. What pure imperativism gets wrong

Pure imperativism, as we have seen, is a well-developed philosophical position. Nevertheless, I hold that it does not provide a satisfactory explanation of HDs. In this section, I support my evaluation by presenting four problems for pure imperativism. Each problem highlights a distinct explanatory gap, and together they reveal the limitations of pure imperativism as a theory of HDs.

4.1. The felt location of homeostatic drives

Pure imperativism states that HDs have *a felt bodily location* in virtue of being commands directed at some part of the body. On this view, if the agent feels pain in the chest, it is because she is commanded to protect this region. Or, if the agent feels an itch in her ear, it is because she is commanded to scratch that part of the body.

The issue I have with this explanation for the felt location of HDs is that while it works for *acute* pains and itches, it does not for many other HDs. Take visceral pains: kidney, uterine, gut pains, and so on. When it comes to these pains, as Maura Tumulty (2009: 165) points out, there is often nothing the agent could do

to protect the relevant body parts. The physiological processes which are experienced as painful – such as uterine contractions, or the passing of a kidney stone – a continue no matter what. Moreover, the agent does not seem to feel any inclination to protect her body either. The only motivation present is the desire for the pain to cease, perhaps accompanied by a generalised motivation to rest or remain still. Given this, it is doubtful that visceral pains are protection commands at all – and if they are, then pure imperativism falsely entails that they should not be located in a specific body part. Yet they clearly are, and this puts pure imperativism into question.

Related issues arise in the cases of thirst and hunger. Since the former is typically experienced as dryness in the mouth and throat, it should be viewed within pure imperativism as a command to *'ingest water with the mouth and make sure it makes its way down the throat'*. By the same token, hunger should be seen as a command to '*put edible objects in the belly*'. These explanations, however, will not do. In one study, rats injected with a drug causing dehydration without an accompanying decrease in salivation still exhibited drinking behaviour characteristic of thirst (Inenaga and Ono 2010). This implies that the sensation of oral dryness can dissociate from thirst, and if this is the case, thirst cannot be a command of the type specified above. The same goes for hunger, which does not have a fixed bodily signature: it may be felt in the belly, but it can also be associated with sensations of weakness and lightheadedness.

Klein (2015) is aware that thirst and hunger pose a difficulty for his view, and he offers a radical solution reminiscent of Armstrong's (1962) classification of these mental states as pure desires (see ch. 1 of this thesis). Namely, Klein states that hunger and thirst are not experienced as located anywhere in the body, although they are *'sometimes confused as such because of accompanying sensations such as dry mouth or stomach cramps'* (Klein 2015: 89). Effectively, Klein sidesteps the problem raised in the paragraph above by removing the felt location of thirst and hunger from his theory's explanandum.

I do not believe, however, that this move is warranted. As argued earlier, the sensory aspect of thirst and hunger is not just non-negligible, but salient. Indeed, much of the normal drinking and eating behaviour is

triggered by located thirst and hunger sensations (Fulkerson 2023). In one study, the mere placing of cotton-wool rolls in the mouth led to a significant increase in the duration of drinking and the amount of water drank by exacerbating the sensation of mouth dryness (Brunstrom, Tribbeck and MacRae 2000). Moreover, it is with these sensations that the negative valence of thirst and hunger is associated. Removing them from the explanandum, then, appears to be unjustified. Besides, Klein's solution does not fully eliminate the problem posed by thirst and hunger. If these experiences are *'undirected commands',* why should we experience them as directed at the body and confuse them with bodily sensations? With this in mind, I conclude that pure imperativism does not capture the felt location of HDs.

4.2. The function of homeostatic drives

Given that pure imperativism portrays HDs as commands – which, for Klein (2015: 17), 'represent a halfway house between mere reflexes and full agential desires' – it may seem particularly well-equipped to explain the motivational force of HD. Yet, I argue that this is not the case. Recall that commands remove the need to deliberate on the part of the receiver but do not themselves provide a *reason to act*. Thus, in order to explain how HDs acquire their motivational force, Klein suggests that we obey the body's commands because we are strongly disposed to treat the body as an authority on its own needs.

The success of the pure imperativist account of the function of HDs is therefore contingent on the existence of a *standing attitude of bodily trust*. Yet there are reasons to doubt that such an attitude exists. First, consider that over the course of our lives, we receive ample evidence that the authority of the body can and should be doubted. Pains, itches, and nausea can appear and disappear without apparent cause. Similarly, hunger, thirst, and the urge to urinate can persist despite having just been satisfied. In light of this, it is difficult to see how our trust that the commands issued by the body are justified coulda remain unshaken. Attitudes, after all, are supposed to be open to revision. Consequently, if Klein is right, then we should expect to see at least some individuals who are poorly motivated by their HDs because they have learnt to mistrust them. Those suffering from unexplained pain and fatigue are obvious candidates for the

role. Yet, although the motivational force of HDs such as pain can be temporarily lessened by drugs like morphine, it does not seem right to say that some of us – especially chronic pain sufferers – are less motivated by pain than others.

Second, consider that our actions are at odds with the existence of the attitude of bodily trust. In fact, we frequently mistrust the body. For example, it is not uncommon for us to say that we feel hungry but do not *really* need to eat. Likewise, it is not uncommon for us to assume that nausea stems from needing to eat rather than indigestion, to resist an itch, or take a painkiller to get rid of pain. If we truly believed that the body 'knows better', we would not exhibit such behaviours. Instead, we would dutifully scratch the itches caused by mosquito bites and chronic eczema despite knowing that it would not resolve any bodily problem. We would not drink coffee in the hope of suppressing fatigue and hunger. And we would not dream of silencing pain because it is inconvenient. In short, not only a rational agent *should not* have absolute trust in the wisdom of the body, but most agents *do not* exhibit such an attitude of bodily trust. This casts serious doubt on Klein's proposal that HDs motivate us in virtue of being authoritative commands and thus weakens the pure imperativist explanation of the function of HD.[37]

4.3. The cognitive architecture of interoception

In addition to its explanatory limitations, pure imperativism entails an unlikely cognitive architecture of interoception. Interoceptive, or homeostatic data, I have argued previously, give rise to two types of representations: HDs and non-HDs. While Klein maintains that HDs have purely imperative content, I take it that we would not dispute that non-HDs such as heartbeat sensations have purely indicative content. Pure imperativism, then, entails that the interoceptive system consists of at least two large subsystems. Both take homeostatic data as input, but System #1 generates representations of homeostatic events (non-HDs), while System #2 generates commands to perform homeostatic behaviours (HDs).

[37] Note that this proposal is at odds with Klein's conception of commands as *'a halfway house'* between reflexes and desires (2015: 17). Neither reflexes nor desires, after all, require certain attitudes to be present to exert their motivational force; so why should this be the case for commands? This asymmetry weakens Klein's case for commands as powerful motivators.

Could the interoceptive system be organised in such a way? It could, but on closer inspection, it appears unrealistic. Recall that the brain maintains the optimal physiological state of the body by promoting homeostatic behaviours and by exerting direct influences on the internal milieu. Both regulatory strategies require accurate sensory feedback about the current state of homeostatic parameters. We may assume, therefore, that however the interoceptive system is set up, it provides the autonomic homeostatic control system with representations of homeostatic events. Indicativism, which treats both HDs and non-HDs as representations of homeostatic events, satisfies this requirement. Pure imperativism, by contrast, does not. On this view, HDs are not informative about their causes, which means that the autonomic homeostatic control system cannot learn from them about the homeostatic events which caused them. Considering the importance of homeostatic control for our survival and well-being, it is unimaginable that our brains would not have evolved to generate representations of homeostatic events causing HDs. And if they have, then HDs do not have purely imperative content.

4.4. 'Pathological' homeostatic drives

In the previous chapter, I introduced the *misrepresentation challenge* for pure indicativism. According to it, if HDs represent homeostatic events, then too many instances of HDs count as misrepresentations of the body state (Corns 2014). Pure imperativism does not face with this particular challenge. Yet this may be a curse rather than a blessing, because some HDs, I suggest, *should* be treated as misrepresentations.

Take the striking examples of severe pains not associated with tissue damage. Fibromyalgia is characterised by widespread musculoskeletal pain in the absence of any apparent pathology (Clauw 2015). Complex regional pain syndrome involves persistent pain in a limb that arises following an injury but does not subside once the injury is healed (de Vignemont 2015). Crucially, if we accept pure imperativism, we would also have to accept that the pains experienced by patients with these conditions are entirely normal, on par with pains associated with actual tissue damage. Since imperatives cannot misrepresent, pains without tissue damage cannot be viewed as misrepresenting the bodily state. At the

same time, it seems to me that we should distinguish ordinary pains from those that are irregular and require specialised treatment. A theory that entails that pains misrepresent often may be seen as implausible, but a theory that entails that pains cannot misrepresent at all is no less objectionable.

I conclude that pure imperativism has limited explanatory power. It struggles to explain the felt location of HDs, offers a convoluted account of their motivational function, paints an unlikely picture of interoception, and does not distinguish between normal and pathological HDs. Despite this, Klein could still maintain that pure imperativism is preferable to indicativism. In the next section, I present Klein's argument in support of this claim, after which, I proceed to show why it does not work.

5. Klein's argument for pure imperativism

Klein is not prepared to accept any version of indicativism and maintains that we are better off treating HDs as pure commands. The indicativist project, according to Klein, was doomed from the start: HDs simply do not correspond to specific types of homeostatic events *sufficiently well* to qualify as their representations (2015: 18–19).

This *mismatch challenge* occupies a central role in Klein's case against indicativism. Let us unpack it. It is generally accepted that mental states with indicative content must '*match the way things are in the world*' (Shea 2018: 177). Thus, reliably matching some worldly state can be seen as a requirement that mental states need to meet in order to fit the indicative mould. Consequently, if some mental state *MS* can be shown to have poor correlation with some worldly state *WS*, then the hypothesis that MS is in the business of representing WS would lose most of its allure. Unless this hypothesis is the only one available, it would not survive for long.

Accordingly, Klein (2015) sets out to undermine indicativism by demonstrating that HDs have limited correspondence with the homeostatic events they are supposed to represent. Klein's argument focuses

primarily on pains. On the one hand, they often occur in the absence of bodily damage (2015: 27–29, 35). Pains of potential damage – such as the pain that one feels when one inadvertently places one's hand on a hot stove – arise *before* any damage has had time to occur. Meanwhile, pains of recuperation persist long *after* the injured body part has healed. On top of this, there are multiple instances of pains either unconnected to bodily damage (e.g. headaches), or not matching the degree of bodily damage in their intensity (e.g. lower back pain).[38] Although some such pains are dismissed as pathological – for example, severe menstrual pain, known as dysmenorrhea, is usually treated as a disorder (but see Serrahima and Martínez 2023) – it would be a tall order to explain them all away in this manner, given how remarkably commonplace they are.

On the other hand, Klein states, bodily damage is not always accompanied by pain (2015: 28). As confirmed by empirical studies, injuries necessitating a trip to the emergency room of a hospital (Melzack, Wall and Ty 1982), as well as those sustained in battle (Beecher 1946) – that is, *serious injuries* – are not necessarily painful, at least not *in the beginning*. This presents a clear problem for indicativism: if the role of pains is to inform us about bodily damage, it is not clear why there should be any serious delay in their onset, especially when the extent of damage is considerable. On this basis, Klein concludes that pains cannot be plausibly viewed as representing bodily damage.

Note that Klein's (2015) argument is partially constituted by a version of the misrepresentation argument addressed in the previous chapter. According to the latter, pains cannot be said to represent bodily damage because they are commonly seen in cases where no bodily damage has taken place. What Klein adds to this, however, is that it is equally problematic for indicativism that pains often fail to register (at least, in a timely manner) certain instances of actual and severe bodily damage. Overall, while Klein does not deny that pains and bodily damage have a connection, he maintains that this connection is not *deep enough* to support the idea that the former represent the latter.

[38] See Corns (2014) for a comprehensive list of examples.

One way in which indicativism could resolve the *mismatch problem* is by reframing pain as representing something other than bodily damage. Yet Klein maintains that there are no good candidates for this role (2015: 35–43). Indeed, take the possibility that pains represent something like *'bodily disturbances'*. For one thing, many homeostatic events, including those that typically underlie itches and tickles, can be plausibly described as bodily disturbances, which means that this account of pain is overly permissive (Klein 2015: 37–38). Moreover, it is unlikely that we would have ever evolved to represent such a vague, uninformative property. To use an analogy, imagine that you receive a text message from your partner which merely says *'There has been a disturbance in the house'*. You would have no idea what your partner means by that. A 'disturbance', after all, could refer to a broken window, a burst pipe, or an altercation with a neighbour – different events that require different actions. You would probably wonder why your partner chose to bother you with a message that tells you nothing of value and cannot guide your decision-making.

The idea that pains represent bodily disturbances is thus unlikely to improve the prospects of indicativism. Of course, a proponent of this view could refuse to budge, arguing that pains represent some non-trivial pain-causing bodily property. The burden of proof, however, would be firmly on them. In the absence of plausible candidates for this elusive property, holding onto indicativism is difficult to justify. In fact, it becomes a mere exercise in motivated reasoning that only hinders philosophical progress.

Where does this leave us? Pain clearly arises as a result of changes in the physiological state of the body. Nevertheless, Klein insists, pain bears only a *degenerate relationship* to its bodily cause (2015: 18). In other words, while it corresponds to a set of homeostatic events, these events have nothing in common that pain could be plausibly viewed as representing. Klein therefore recommends that we look elsewhere for the content of pain – namely, to its typical *effects*. According to Klein, while pain is not always associated with bodily damage, it tends to provoke the same type of behavioural response – a protective response intended to prevent damage or support healing (2015: 32). As Martínez and Klein put it, *'the causes of pain are heterogenous […] while the acts that pains provoke are comparatively unified'* (2016:

284). Furthermore, other HDs are supposed to be resemble pain in being more closely tied to *behavioural outcomes* than to *bodily causes*. For example, itch, Klein argues, may be caused by various skin conditions and parasites, but always promotes scratching as a response (2015: 18). This is why it is natural for us to define HDs as the *needs* to perform specific types of actions. As Klein writes: *'What causes hunger? Needing to eat. What causes thirst? Needing to drink'* (2015: 18).

In corresponding to specific action types, Klein (2015) stresses, HDs are totally unlike perceptual states. Seeing blue, hearing the sound of rain, touching a smooth surface, or smelling a rose – none of these experiences determines what the agent will do about them. The actions they provoke, if any ensue at all, depend on the agent's beliefs and desires and thus may differ from one agent to another. To illustrate, the sound of rain is not associated with any fixed pattern of behaviour. By contrast, itch appears to be inextricably linked to scratching – so much so, in fact, that one would be hard-pressed to describe the former without making reference to the latter (Hall 2008). While we can choose *how* to act on HDs – for example, scratch an itch or what to have for lunch – our choices appear to be significantly *constrained* (Klein 2015: 17). Accordingly, while there is no reason to think that perceptual states have imperative content, it is plain to see that HDs admit of imperativist treatment. In line with this, Klein proposes that instead of forcing a connection between HDs and homeostatic events, we should exploit their clear link with behaviours by treating them as commands.

Let us take stock. Klein insists that the central assumption of indicativism – namely, that there is a sufficiently strong *one-to-one correspondence* between specific types of HDs and specific types of homeostatic events – is misplaced. For example, although pain is typically connected to bodily damage, its bodily causes, Klein argues, are so varied as to *'defy generalisation'* (2015: 38). At the same time, for each type of HD, we can easily identify a type of homeostatic behaviour it plausibly commands the agent to carry out. Klein's strategy, then, is to usher in pure imperativism by presenting it as a desirable alternative to the seriously flawed indicativist position. In what follows, I argue that this strategy fails: pure imperativism is *at least* as problematic as the theories it was meant to replace.

6. My response to Klein's argument

Here is Klein's (2015) argument again:

(A) Mental states have indicative content if they match specific types of states of the world and imperative content if they match specific types of actions.

(B) HDs do not match specific types of homeostatic events, but they do match specific types of homeostatic actions.

(C) Therefore, HDs have imperative, but not indicative content (more precisely, HDs are imperatives to perform specific types of homeostatic actions).

I accept premise (A) but reject premise (B) on the grounds that the mismatch between HDs and specific types of homeostatic events is not as significant as Klein maintains. Since Klein's argument centres on pain, let us start with it. Historically, pain has often been equated with *nociception*, i.e. the processing of noxious stimuli (Baliki and Apkarian 2015; Tracey 2017). In recent years, however, researchers have come to appreciate that the relationship between the two is more nuanced (Melzack 1999). Specifically, while nociceptive processing is constantly ongoing and shaping action planning and execution, the conscious experience of pain arises only when non-conscious nociceptive modulation of behaviour fails to protect the body from injury (Apkarian 2019; Baliki and Apkarian 2015). To illustrate, while nociceptive data about potential joint damage is routinely used for the purposes of postural adjustment, such adjustments are not typically accompanied by pain (Baliki and Apkarian 2015: 475). As a result, we are able to lead largely pain-free lives while simultaneously keeping our bodies safe from harm.

Consequently, Klein's (2015) assertion that the bodily causes of pain are so *heterogenous* as to render indicativism untenable strikes me as ill-founded. For obvious reasons, the brain tracks and acts on potential as well as actual bodily damage. Yet it seems to reserve pain for the more *extreme* cases where damage has either already occurred or is imminent. In short, if actual and potential bodily damage were

located on a continuum, pain would be skewed firmly towards the *actual* end of this continuum.

Contrary to Klein's view, then, the bodily causes of pain appear to be sufficiently unified to support an indicativist theory of its content. The reason pain was selected to capture both actual and imminent bodily damage may be that it is difficult for the brain to differentiate between these types of homeostatic events. After all, where does imminent damage end and actual damage begin? Perhaps, the pain-producing cognitive system simply errs on the side of caution. Alternatively, since bodily damage already sustained predisposes the agent to further bodily damage – cuts and burns are prone to infection, while wounds can bleed out – it could be argued that all pains represent imminent bodily damage. To be sure, some pains do not seem to be associated with *any* noxious events. Such pains certainly require an explanation; but as I argued previously, they do not threaten indicativism. After all, representational systems are allowed to err from time to time, and it is highly revealing that pains which arise and persist for no clear reason are usually regarded as *pathological*.[39] [40]

Now, the mismatch challenge to indicativism about pain has another aspect that needs to be discussed. Namely, not only is pain not always underpinned by bodily damage, but serious bodily damage does not always result in pain. Klein himself offers only two studies in support of this latter claim. The first study was conducted by Beecher (1946) in a WW2 field hospital; it reported that out of more than 200 soldiers wounded in battle, 32% experienced no pain. The second study was conducted by Melzack, Wall and Ty (1982) in a general hospital; it found that out of 138 emergency room patients, 37% did not feel any pain at the time of the incident that precipitated the hospital visit. Do these studies lend appropriate credence to Klein's claim? I am not convinced that they do. For one thing, while the experiments mentioned above are often cited by philosophers of pain – see, for example, Bain (2011), Casser (2021) – their results have not been widely replicated to my knowledge. I do not, however, have to content myself with mere

[39] It is noteworthy that even such mystifying phenomena as fibromyalgia, a chronic condition characterised by constant widespread musculoskeletal pain, often develops as a result of prolonged acute pain caused by a serious injury (Clauw 2015; Harte, Harris and Clauw 2018). If even the most bizarre pains are tied to bodily damage, the two seem to share a deep and meaningful connection.

[40] It has recently been found that nociceptors can be directly stimulated by microorganisms, including those residing in the human gut (Lagomarsino, Kostic and Chiu 2021). If this is the case, then chronic pain conditions like IBS (irritable bowel syndrome) can be reconceptualised as being caused by peripheral factors such as gut dysbiosis rather than used as fuel for the idea that pains in general have a weak relationship to bodily damage.

nitpicking: there are more interesting criticisms to be made.

Let us address Beecher's (1946) study first. From personal experience, I know that injuries do not tend to hurt *constantly*. Instead, they subside if you lie still and flare up if you try to move or touch the injured body part. Crucially, the precise question Beecher asked his patients was: *'As you lie there, are you having any pain?'* It is possible, then, that while the patients felt some pain associated with their injuries, they did not feel any when their doctor inquired about it. A further possibility is that soldiers, being no strangers to hardship, simply did not rate the pain associated with their wounds as serious enough to warrant communicating it. Finally, it is worth noting that a field hospital must be a hectic environment where a doctor may be too busy and tired to conduct a detailed interview with every patient under their watch. Indeed, Beecher himself explicitly states that if the patient denied being in pain, *'that part of the questioning was dropped'* (1946: 97). This, in conjunction with the considerations outlined above, gives us grounds to question the validity of Beecher's findings and thus undermine Klein's attack on indicativism.

Next, let us address the experiment conducted in a general hospital's emergency room (Melzack, Wall and Ty 1982). The first thing to point out is that the participants of this study did feel pain following their injury, albeit after a *delay*. In light of this evidence, it is perfectly plausible that pain does represent bodily damage, even if pain-producing systems are *slow* at times. This may seem odd, but from an evolutionary viewpoint, it makes sense for the brain to suppress pain production when higher-priority tasks – such as running away from a predator or getting to a hospital – need to be carried out. And the mere existence of the pain modulation system, as argued in the previous chapter, is not incompatible with indicativism.

The second noteworthy finding of the Melzack, Wall and Ty's study was that *surface injuries* such as cuts and burns showed a greater association with delayed pain compared to *deep injuries* such as fractures and stabs (1982: 37–38). Given that skin and internal organs are not innervated in the same way (Garfinkel, Critchley and Pollatos 2017), delayed pain onset may have a physiological basis. Combining these two observations, it is also possible that once the brain has tracked the injury as having already taken place,

acute pain may be suppressed as maladaptive, while the pain of recuperation would not yet be produced because the relevant systems are still gathering information about the degree of inflammation and related processes.

Finally, consider that delayed pain is also problematic for pure imperativism. Whatever we might think about the content of the pain experience, we may all agree that injuries so severe that they require urgent medical attention really should be accompanied by pain. To explain why they are not, a proponent of pure imperativism would be forced to appeal to evolutionary or processing factors. But if we have to appeal to these factors *in any case*, then indicativism might do just as well as pure imperativism.

I contend that pain does not, in fact, demonstrate a significant mismatch with the type of homeostatic event it is generally taken to represent. Nor do other HDs. Take thirst – the experience that plausibly represents dehydration. Recent empirical work has demonstrated that thirst tends to subside long before the osmotic environment of the body has had time to change following water intake and before the brain can track these changes (Augustine, Lee and Oka 2019: 27–28; Zimmermann and others 2016; Zimmermann and others 2019). This finding may be taken to imply that thirst does not neatly map onto dehydration. On the basis of this poor correspondence, a pure imperativist could argue that thirst does not in fact represent this bodily state.

To this, I would respond that the thirst-producing cognitive system simply has *multiple channels of information* about hydration levels. Indeed, in addition to tracking the osmolarity of the body, the thirst system is known to receive direct sensory feedback from receptors located in the mouth and throat (Zimmermann and others 2019). As a result, the thirst system is always informed about water ingestion and can register that we are rehydrated before sensing any changes in osmolarity. All this is perfectly consistent with the indicativist view of thirst. In this scenario, thirst tracks and represents dehydration, even if it pools together information from different sources.

What applies to thirst applies also to hunger, which has the same *predictive satiation mechanisms* as thirst: it can be suppressed long before the nutrients have been absorbed (Augustine, Lee and Oka 2019: 29). Other HDs, such as temperature sensations, nausea, and bladder fullness, seem to be immune to the mismatch challenge insofar as there are no obvious examples of mismatch in their case. In general, each type of HD has a strong link to some homeostatic parameter it plausibly represents. While Klein may be right in describing HDs as not *particularly* revealing about their bodily causes, this is not a requirement for the ascription of indicative content. After all, indicative content does not have to be *fine-grained* as long as it it useful to the organism. To illustrate, hunger may not tell us *why* we are low on energy or which nutrients we require, but it still tells us something non-trivial, namely, that we are low on energy. Similarly, coming back to an earlier example, we may not realise that breathlessness-producing systems are tracking the buildup of carbon dioxide, but given that high carbon dioxide reliably signals low oxygen under normal conditions, we can argue that breathlessness still represents the relevant property of low oxygen even if it does not tell us how this property is tracked.

I conclude that the mismatch challenge can be answered by indicativism.

7. A potential rejoinder and my response to it

Klein may have underestimated indicativism; but all is not lost. Indeed, Klein could still maintain that while HDs are not *so bad* at matching specific types of homeostatic events, they are *much better* at matching specific types of homeostatic actions. In this section, I consider this potential rejoinder and argue that it rests on a mistake. Contrary to what Klein takes to be the case, there is no one-to-one correspondence between HDs and specific types of homeostatic motivations. This means that pure imperativism runs into the same problem that Klein has described as fatal for indicativism – and it cannot resolve it.

Now, there is no doubt that HDs are involved in promoting homeostatic behaviours. Klein, however, goes further and asserts that there is a special relationship between HDs and action types. Specifically, he

asserts that HDs are associated with *specific behavioural outcomes* – and thus make excellent candidates for mental states with imperative content (Klein 2015: 23). For pure imperativism to work, therefore, a one-to-one correspondence between types of HDs and specific action types is a non-negotiable requirement.

To better understand Klein's proposal, let us look more closely at how it works in relation to pain:

> *Bodily parameters can deviate from the optimum for a variety of reasons; they are brought back into balance with the very same kind of action. Homeostatic sensations, then, can be thought of as those that connect a heterogenous disjunction of problematic states with a single type of response that would be appropriate for any member of the disjunction […] As a located homeostatic sensation, pain is most analogous to itch: both require taking action with respect to some particular part of the body to eliminate threats (2015: 2, 32).*

As illustrated above, Klein assumes that typical pain behaviours – such as removing one's hand from scalding water or refraining from walking on a bad ankle – can all be interpreted as *protective*. In other words, despite the apparent diversity, Klein insists that they have something crucial in common: namely, they constitute attempts to satisfy the protective motivation. And it is on this alleged link between feeling pain and protecting the body that Klein's entire theory hinges. The idea, in a nutshell, is that if pain is strongly predictive of actions aimed at protecting the body, then this mental state can be plausibly viewed as a command to attain that goal.

I maintain that there is a problem with this proposal. More specifically, I argue that pain is not the only type of HD that promotes behaviours with the same protective objective. Nausea, for example, motivates vomiting, which protects the agent from the toxins ingested. The urge to urinate motivates urination, which protects the agent's bladder from rupturing. Breathlessness motivates seeking oxygen, which protects the agent from suffocation. The tickle in the throat motivates coughing, which protects the agent from the same thing. In fact, *all* HDs can be said to motivate protective behaviours insofar as they all

promote behaviours aimed at restoring homeostatic parameters to their acceptable values.

Klein could try to avoid this objection by tweaking his theory without revising its core tenets. He could concede, for example, that while all HDs command protective behaviours, each type of HD commands a distinct type of protective behaviour. Thirst, for instance, might command behaviours that protect us from dehydration, the urge to urinate might command behaviours that protect the bladder from rupturing, and pain might command something like protective behaviours aimed at restoring the integrity of the body.

Such a modification of Klein's proposal, however, would not be much of an improvement, given that there are plenty of pains which do not appear to motivate anything as *specific* as that. For instance, neither headaches nor menstrual pains promote any behaviours that make essential contributions to the integrity of the body. Usually, we simply *wait* for these pains to pass on their own – or we attempt to silence them by seeking a distraction or taking a painkiller. Klein, therefore, ends up caught between the rock and the hard place. If he insists that pain is a protective command, he faces the objection that many, if not all, types of HDs motivate protective behaviours. If he attempts to circumvent the issue by rebranding pain as a command to engage in a specific type of protective behaviour, he is forced to contend with the inability of his modified proposal to accommodate some perfectly ordinary pains. Either way, the correlation between pain and behaviour is not as tight as Klein needs it to be.

Furthermore, many common pain responses do not obviously serve a protective goal. Examples include stroking or rubbing the hurting body part, crying out, grunting, swearing, grimacing, or lashing out at the object over which one has tripped. In a bid to explain them away, Klein suggests that non-protective pain behaviours of this type are motivated not by the imperative content of pain, but rather by its unpleasantness (2015: 46). In this scenario, one avoids stepping on a blister because the pain associated with it commands one to protect one's foot, but one hisses and swears when stepping on the blister because the pain associated with it happens to be unpleasant.

I find this explanation of non-protective pain behaviours unsatisfactory, however. For one thing, it begs the question: why should we bother to attribute *any* pain behaviours to the imperative content of pain, if unpleasantness of pain alone is sufficient to account for its motivational force (see Bain 2011)? Moreover, although valence is a shared characteristic of many types of mental states, the vocalisations and facial expressions associated with them differ widely *depending on the mental state*. For example, we grunt and hiss in response to pain, but not in response to fear or hunger. What follows from this is that it is the type of mental state, not its unpleasantness, that ultimately determines the nature of the agent's response to it. The conception of pain as protective command, thus, is unsustainable.

Second, take nausea. This mental state promotes vomiting, but it also promotes not eating. This means that nausea motivates two distinct types of behaviour, which is directly at odds with pure imperativism. Now, if vomiting and not eating could be construed as being united by *a common goal*, then pure imperativism would dodge this particular bullet. Say, then, that nausea commands something like emptying the stomach. This can be achieved both by (1) evacuating the contents of one's stomach via the mouth and (2) not adding to the contents of the stomach and thus allowing the current contents of the stomach to be digested and to pass further down the digestive track. Such a view of nausea, however, will not do. Indeed, light nausea does not always evoke a vomiting motivation. It can manifest merely as discomfort in the stomach and a lack of appetite. At the same time, severe nausea is characterised by an almost uncontrollable urge to vomit. And if nausea motivates different behaviours depending on its severity, then it cannot be treated as an imperative.

Finally, take fatigue and sleepiness. These mental states are distinct: being exhausted is not the same as being sleepy, as insomnia sufferers would gladly attest. Despite this, however, fatigue and sleepiness seem to lead to the same type of behaviour: resting. Indeed, sleepiness cannot be treated as motivating sleep, because it is not up to us to fall asleep. It just *happens* that we do, provided we do not actively interfere with this natural process (for example, by listening to loud music or ruminating on unpleasant topics). What sleepiness promotes, then, is hardly different from fatigue, except that sleepiness also

promotes closing one's eyes. Crucially, however you look at it, fatigue and sleepiness present an issue for pure imperativism. If they motivate and therefore command the same behaviour, then they should feel the same way, which is not the case. If fatigue motivates resting, whereas sleepiness motivates resting and closing one's eyes, then sleepiness ends up commanding two different behaviours, which is incompatible with pure imperativism. Finally, if sleepiness commands something like the set of behaviours optimally preparing the agent for the onset of sleep, then we have an overly permissive account of sleepiness predicting that it commands us to get into bed, avoid blue light, and drink chamomile tea. Out of these three options, each is as bad as the others.

Let us recap. Mental states with imperative content, unlike those with indicative content, are not meant to be informative about their causes. Instead, they are meant to simplify decision-making by instructing the agent to act in a certain way (Martínez and Klein 2016). HDs, I have argued, fail to conform to this description and thus cannot be treated as commands. In the end, pure imperativism falls prey to the same *mismatch challenge* which has been part of the rationale for its development.

Can pure imperativism recover from this challenge? I do not think so, but I doubt that Klein would abandon his project so easily. In fact, I can anticipate the sort of rebuttal that Klein might offer. In what follows, I formulate this potential rebuttal and explain why it would not be able to save pure imperativism.

8. Another potential rejoinder and my response to it

Klein (2015) argues that HDs exhibit a stronger one-to-one correspondence with homeostatic actions than with homeostatic events. I have sought to defuse this argument by turning it on its head and showing that many types of HDs — pain, nausea, fatigue, and sleepiness — do not, in fact, conform to it. However, an obvious line of defence is available to Klein. Namely, even if he were to accept my criticisms, he could still insist that in being *consistently motivational* (i.e. usually resulting in behaviour) as well as *rigidly motivational* (i.e. usually resulting in more or less the same set of behaviours), HDs differ so markedly from

regular perceptual states that they cannot possibly have the same indicative type of content.[41]

Consider, for example, the experience of seeing the colour green. By itself, this experience is not motivationally charged and is not associated with any particular course of action. Indeed, if happen to like this colour, I may be disposed to look at green objects for long periods of time; but you might not be, and even in my case it would depend on my mood and a myriad other factors. A purely imperativist account of colour would therefore be untenable. At the same time, a purely imperativist account of HDs is very attractive. Why? Well, HDs seem to be unlike regular perceptual states in that they motivate *consistently* and rigidly, while the latter motivate *inconsistently* and *flexibly*. Klein, therefore, could still defend his account of HDs on the grounds that these mental states exhibit too meaningful a link with behaviour to be denied imperative content. In other words, Klein could argue that despite being vulnerable to the mismatch challenge, pure imperativism is nevertheless preferable to indicativist views.

I agree that HDs exhibit a strong link with behaviour, but I deny that this link provides sufficient support for the idea that HDs have imperative content. The reasons for this conclusion are listed below.

First, the claim that HDs motivate consistently cannot be used against pure imperativism, since it is perfectly compatible with indicativism. Indeed, consider that unlike regular perceptual states, HDs are always *relevant* to us. One sees a wide range of green objects over the course of one's life, most of which have nothing to do with us or our aims (or rather, have nothing to do with us or our aims in virtue of being green). The same is true of the majority of objects in the world. By contrast, most events happening in our bodies matter to us because they are caused by homeostatic fluctuations and are therefore likely to have far-reaching consequences for our survival and well-being. It is hardly surprising, then, that while many visual or auditory experiences fail to consistently move us, this is not the case for HDs.

Second, the apparent distinction between HDs that motivate *rigidly* and regular perceptual states that

[41] When I say '*regular*' perceptual states, I refer to all the perceptual states whose indicative content is not typically disputed.

motivate *flexibly* is not as clear-cut as it appears to be. Perceptual states whose indicative content has never been disputed may, in fact, exhibit a strong correlation with a specific type of behaviour. Music perception is a case in point, as it reliably promotes rhythmic movement, starting from a very early age (Kim and Schachner 2023). Similarly, seeing familiar faces usually leads us to raise our eyebrows and smile, while touching something slimy causes us to jerk away. Accordingly, it is not entirely clear that HDs motivate *significantly* more rigidly than regular perceptual states.

Third, even if HDs do motivate particularly rigidly, this still does not constitute decisive evidence in favour of pure imperativism. Take pain. Suppose I concede to Klein that pain behaviours have something in common insofar as they all contribute to restoring bodily integrity, whether directly or indirectly. Indeed, there are many ways to achieve or at least to get nearer this particular goal. Withdrawing one's hand from scalding water, for instance, prevents or minimises burns. Pain vocalisations provide the agent with assistance from other people. Rubbing a cramping muscle helps it relax. And so on. On its own, however, this fails to support the conclusion that pain has imperative content, since an alternative explanation is available. That is to say, a good number of typical pain responses are likely to belong to the brain's *innate repertoire of motor programs* activated at some point during pain processing, meaning that a good number of typical pain responses do not need an in-virtue-of-content explanation in the first place. When one burns oneself on a hot object, one reflexively withdraws one's hand from it. By the same token, pain vocalisations are carried out habitually and sometimes initiated *before* the agent feels any pain. Accordingly, it seems best to treat them as *hard-wired, automatic* responses. And if this is the case, even if pain behaviours have something in common, it does not follow from this that they *stem from the same single command*.

Let me make this point in a slightly different way. Klein argues that pain should be treated as a command because this would explain the similarities between the actions it promotes. I object on the grounds that it is doubtful whether any such command needs to be posited. If, say, some pain behaviours increase the agent's fitness – like pain vocalisations do via soliciting help from conspecifics (Finlay and Syal 2014) – we

might simply posit that they have been *selected* as automatic responses to pain. The point generalises to other types of HD, many of which are associated with hard-wired motor programs that can and will be carried out even if our decision-making centres were destroyed. Vomiting, breathing, coughing, blinking, and urinating – these are just some examples of such programs.

None of this implies, of course, that appealing to states with imperative content is completely out of the question. Indeed, it could still be suggested that, in addition to sending a number of direct instructions to the motor centres, the interoceptive system also sends commands to the agent. Such a cognitive architecture of HDs, however, strikes me as awkward and unlikely. Take tickles. Depending on where a tickle is felt, it can promote different actions. If one is tickled in the ribs, one tends to laugh or make uncontrollable movements designed to shake off the attacker. If one feels a tickle in the throat, one has an urge to cough. If one feels a tickle in the nose, one has an urge to sneeze. Now, in order to explain why all tickles feel similar to the agent, pure imperativism must posit that tickles command the agent to perform some specific action. What action could this be, though? And why should we bother trying to come up with something that tickles could possibly command when the behaviours they lead to are already explained by hard-wired motor programs? Imperative content seems to be *explanatorily superfluous*.

9. Conclusion

Pure imperativism offers a fresh perspective on HDs by treating them as commands to perform homeostatic behaviours rather than as representations of homeostatic events. Ultimately, however, pure imperativism does not live up to its promises and does not constitute an improvement on indicativist views. Indeed, it fails to capture all the aspects of HDs, struggling, among other things, to explain their sensory phenomenology – in particular, their felt location. Since the foundation of pure imperativism – the idea that HDs correspond more strongly to specific types of homeostatic behaviours than to specific types of homeostatic events – is also flawed, I conclude that pure imperativism is not viable. Having thus established the inadequacy of pure accounts of HDs, I now turn to exploring impure accounts.

Chapter 4:

The Pushmi-Pullyu Theory

Abstract:

As argued in the previous chapters, the pure approach to HDs has not delivered the desired results. While pure indicativism and pure imperativism do not fail for the same reason, they fail nevertheless, and their problems cannot be easily resolved. With this in mind, I suggest we explore the hypothesis that HDs have *impure* – that is, both indicative and imperative – content. In this final chapter, I outline the simplest version of an impure theory of HDs – *the pushmi-pullyu (PP) model* – and show that it does not constitute an improvement on its predecessors. In tracing the fault lines of the PP model, I determine that many of its problems stem from *the same source* as the problems of pure models. In order to get anywhere, I conclude, we have to abandon the assumption that HDs may be explained by positing a single representation.

1. The motivation for the impure approach

This thesis began with the idea that a set of experiences relating to the *physiological condition of the body* are sufficiently alike in terms of their function and phenomenological profile to warrant a unified explanation. Upon making the argument that these sporadically addressed experiences, which I call *homeostatic drives* (HDs), should be investigated as a *distinct type of mental state*, I proceeded to try and pin down their intentional content by considering two obvious candidates: *pure indicativism* and *pure imperativism*. What I have come to determine, however, is that neither of these accounts handles HDs with much success. Indeed, while pure indicativism and pure imperativism certainly have *something* going for them, both also suffer from serious flaws. Ultimately, neither view is preferable to the other, which means that we are effectively at an impasse. So where do we go from there?

When a philosophical debate shows signs of having run aground, there are several ways of getting it back on track. The most radical, though often the most realistic, option is to assume that unsatisfactory progress is a symptom of a general problem with the *research question* and thus to have the latter revised. Accordingly, faced with the lacklustre performance of the pure models, we could simply abandon the project of drawing out the essence of HDs, dismissing our earlier efforts to this end as having been *destined* for failure. Perhaps the bodily experiences taken here to be plausibly constituting a natural kind do not, in fact, belong together. Perhaps, however we may arrive at the putative intentional content of hunger, it is not by reflecting on thirst, nausea, or pain.

I reject this option. The case for a joint theory of HDs was set out in the first chapter of this thesis. The analysis of the two main contenders exposed their respective weaknesses, but it neither unearthed any threats to the case itself nor gave us any cause to be tempted by the alternative. After all, it is not as if pure indicativism or pure imperativism worked for some types of HDs but not for others. If this were so, then tackling these phenomena *separately* would make sense. Who knows: temperature feelings could be bodily perceptions, itches could be bodily commands, and thirst could be something else – say, a desire.

Except, as things stand, both pure indicativism and pure imperativism have as much trouble dealing with, say, hunger and nausea as they have with itch or pain. Indeed, the *real letdown* of the pure accounts is not that they fail to accommodate *all* HDs at once – it is that they struggle to explain *each* mental state from this category *for the same sort of reason*. Pure indicativism, for example, does not have a firm grip on the motivational aspect of HDs. Meanwhile, pure imperativism, though it too falters in this respect, is stumped primarily by the sensory features of HDs – namely, by their felt locations. Moreover, the misrepresentation challenge, however much weight one places on it, applies to purely indicativist treatments of *all* types of HDs, just as my argument that the various behaviours engendered by HDs are not attributable to a single command applies to *all* purely imperativist narratives. Based on this, I conclude that a joint theory of HDs is still worth pursuing – especially since the reasonable lines of inquiry have not yet been exhausted.

This brings us to a more promising solution to the frustrating stand-off between pure indicativism and pure imperativism. In this chapter, I do not deviate from my initial project but *broaden* my discussion of HDs by introducing a new class of potential models. Note that in choosing to move forward in this direction, I am not suggesting that the prospects of the pure accounts could never be improved. Such an outcome is not inconceivable, though it is unlikely. During my detailed examination of the pure accounts, I found them to be deeply inadequate. No matter how they are tweaked, something always goes wrong, and some part of the explanandum is always left out. A continued allegiance to the pure research programme, therefore, strikes me as unwise. The time has come, I think, to change gears and *go impure*.

Here is how this chapter unfolds. The attempts to elucidate the nature of HDs by appealing *only* to indicative or imperative content have not been productive. What about appealing to *both* these types of content, though? In the next section, I put this plan into practice by testing out the least complicated version of the impure theory. More precisely, I assess the possibility that HDs are *pushmi-pullyu representations (PPRs)*, i.e. mental states that describe and direct at the same time (Millikan 1995; 2004; 2023). Such a model has not been fully articulated in contemporary philosophy, but it has been acknowledged as tenable and even endorsed in passing by several authors (Bayne 2011: 232; Hall 2008: 534; Nanay 2017: 495; Millikan 2004: 93, 159). I flesh out the *pushmi-pullyu (PP)* model by setting down its principal tenets and commitments – and then argue that it fares no better than its pure counterparts. Afterwards, I ask *why* HDs have proven so elusive. My response is that both the pure theories and the impure pushmi-pullyu theory err in treating HDs as simple mental states with one layer of content: indicative, imperative, or pushmi-pullyu. In order to get to a good theory of HDs, I conclude, we must not only *go impure* but also *go complicated* by positing more than one layer of intentional content.

2. Pushmi-pullyu representations: An analysis

The pure theories that have been the focus of my attention up to this point fail to explain HDs in their entirety. Yet they do not miss the mark *completely*. First, despite their limitations, each of the theories is

capable of explaining *a major aspect* of HDs. Pure indicativism, for example, is well-equipped to handle their sensory characteristics: location, intensity, and temporal profile. Pure imperativism, in turn, is thoroughly suited to dealing with the motivational side of things, i.e. the robust link with body-directed behaviours. Second, although the two theories are grounded in vastly different conceptions of the role played by HDs in our mental economy, these conceptions are equally plausible. Pure indicativism has legitimate reasons to view HDs as reports about the physiological state of the body, but pure imperativism is no less convincing in framing the same experiences as instructions to act in a certain way. Both theories, then, manage to capture something important about HDs, though individually they fall short of providing us with a full picture of this type of mental state. This invites the following question: if two theories struggle to meet all our requirements, but still seem to have a kernel of truth in them, could they not be *combined* to yield a superior explanation of the phenomenon that they both target?

The answer is: yes, they could. Pure indicativism and pure imperativism may be rivals, but their core ideas are not incompatible. Indeed, it is entirely plausible that HDs have *impure* – i.e. both indicative and imperative – content. The issue lies in figuring out how best to explore a proposal that may be specified in a number of ways. As I see it, the optimal navigation strategy is to identify the *most basic version* of the proposal and lead with that. If it takes care of all the salient aspects of HDs, then we are in luck and do not need to do anything else. If, however, the skeleton model has any defects, we can attempt to rectify this by *building the model up*, i.e. by introducing further layers of representational content into the mix.

The most straightforward way to marry pure indicativism and pure imperativism is doubtless to posit that HDs are *pushmi-pullyu representations* (henceforth PPRs). The existence of this type of representation was famously defended by Ruth Millikan (1995; 2004; 2023) in the course of her work on a general theory of intentionality.[42] While they came under increased scrutiny in recent years (Artiga 2014; Bauer 2020; Rescorla 2013), Millikan's insights on PPRs have, on the whole, fared remarkably well, becoming part and parcel of the mainstream philosophy of cognitive science (Bayne 2011; Papineau 2003; Shea 2018;

[42] Millikan's use of the concept of PPR dates from her eponymous paper published in 1995, but similar ideas appear in earlier writings, including 'Biosemantics' (1989: 296) and *Language, Thought, and Other Biological Categories* (1984: 118).

Sterelny 2003). Defined as representations whose content is *simultaneously descriptive and directive*, PPRs are supposed to have both a mind-to-world and a world-to-mind direction of fit, thus facing in opposite directions at the same time – hence their curious name (Millikan 1995: 186; 2004: 17).

The most striking examples of PPRs are derived from animal communication. Take the rabbit thump (Millikan 2004: 89). When a rabbit detects the presence of a predator, it thumps hard on the ground with its powerful hind legs. This, in turn, induces other rabbits in the vicinity to take cover or freeze. Now, since the thump always has this (and only this) useful effect on rabbits, it could easily be seen as an imperative signal that has evolved to protect the members of the species from predation.[43] Millikan concurs; but she thinks that it would be a mistake to stop there. The thump, Millikan reasons, only fulfils its *proper function* of aiding predator escape if there are actual predators lurking nearby and posing a threat to the rabbits (1995: 190). As such, it has not *one*, but *two* satisfaction conditions and cannot be interpreted as a signal devoid of indicative content. As far as Millikan is concerned, the thump *both* reports the danger of predation taking place right here, right now and commands the behaviours appropriate to this kind of situation – quickly, reliably, without any additional inferences on the part of the receiver.[44]

PPRs, then, are envisaged not as *conjunctions* of purely indicative and purely imperative representations, but as something wholly different, more *primitive* than either (Millikan 1995: 192; 2004: 80). Let me expand on this point. As detailed previously, indicative representations have the function of notifying their consumer about a specific state of affairs, but they do not prescribe a specific course of action. Put another way, they are *coupled* with the objects or events causing them (i.e. their inputs), but *decoupled*

[43] Chances are, of course, that rabbits are often sent scattering by 'false-alarm' thumps. Overall, however, the species is better off with the thump as a warning signal than without it, which explains why the signal was selected and retained. Note also that the rabbits performing the thumps incur a greater personal risk of predation in doing so: if they simply fled without alerting conspecifics, they would be more likely to survive on that particular occasion. Still, it seems that the benefits of receiving thump alerts from other rabbits outweigh the costs of a hard-wired thumping response to predators.

[44] That is to say, Millikan thinks that rabbit thumps have descriptive as well as directive content because it is a *condition on the realisation of their proper function* that predators be present in the vicinity of the rabbits (1995: 190). After all, if this state of affairs does not obtain and the rabbits are quite safe, then the thump cannot result in rabbits *escaping predation* – it can only result in them *running* or *taking cover.* Artiga challenges this claim, protesting that Millikan's criteria for the ascription of descriptive content to representations that would otherwise be treated as directive are too permissive (2014: 553–554). Millikan (2023) disagrees. For my part, I suspect that Artiga's criticism stems from overlooking the distinction between the *function* and the *goal* of a directive representation. I will not, however, elaborate on this here. A full breakdown of Millikan's account of PPRs is a huge task. Pursuing it would derail my evaluation of the PP model, and since I ultimately reject this model, the outcome of the debate on the status of PPRs makes no difference to the conclusions of this thesis.

from the behaviours they generate (i.e. their outputs) (Shea 2018: 190). Conversely, imperative representations do not aim to inform their consumer about anything in particular, but they do specify exactly what is to be done. Thus, imperative representations exhibit the opposite pattern to indicative representations, being decoupled from their inputs rather than from their outputs (Shea 2018: 190).

PPRs, however, are in a class of their own in that they are *not decoupled from either their inputs or their outputs*. Instead, when it comes to PPRs, all three – the input, the representation, and the output – are glued together in a unique representational structure (Sterelny 2003: 30). Take the so-called 'dances' of honeybees, social insects that cooperate during foraging. Experts in animal behaviour recognise that dances, i.e. sequences of repetitive movements accompanied by sounds, vibrations, and the release of chemicals, are performed by honeybees to communicate the location of nectar (or water sources, or nest sites) to their conspecifics (Gruter and Farina 2008: 242–43). Honeybee dances, therefore, appear to have content, but what exactly is that content? Millikan's position is that they simultaneously (1) signify the location of nectar to the spectator honeybee and (2) instruct the spectator honeybee to fly to this location (1995: 190). This is because, just like rabbit thumps, honeybee dances exhibit a strong and consistent coupling with both (1) specific environmental states (nectar at the location L) and (2) specific consumer behaviours (spectator honeybee flying to the location L). And, again like rabbit thumps, honeybee dances only fulfil their biological purpose (which is to help other honeybees find nectar) if *both couplings are intact*, i.e. if (1) if nectar *is* present at the location L and (2) the spectator bee *does* fly to the location L.

The presence of one-to-one correspondence on both the input and the output side is what sets PPRs apart from other types of representations, the distinguishing mark of PPRs. It is also, I take it, what leads Millikan to assert that pushmi-pullyu mental representations (or *inner PPRs*) must predate their purely indicative and purely imperative counterparts in evolutionary terms (Millikan 2004: 158). Whereas humans set and pursue their goals in highly versatile ways, simpler animals have a finite and more rigid behavioural repertoire reflective of the processing power of their brains.

Evidently conscious of this, Millikan theorises that the earliest examples of mental representations were, by necessity, *blunt tools,* latching only onto the states of affairs that warranted *a fixed response* from the consumer and thus '*undifferentiated between description and direction*' (2004: x, 157–58). Under such a representational arrangement, relevant variations in the state of the environment would translate directly and reliably into predetermined variations in the behaviour of the animal (Millikan 1995: 190). Light, for example, would always cause an earthworm to move in the opposite direction; a chick's widely open mouth would always trigger chick-feeding in an adult bird; the approach of winter would always set off food hoarding in a squirrel (Millikan 2004: 168–69). At the end of the day, this form of representing was better than nothing – and it kept the computational costs low. Yet it imposed serious constraints on how the consumer could act in any given situation and provided a fairly fragmented, narcissistic view of reality by representing it only in relation to the consumer's capacities for action. As Millikan puts it, a purely pushmi-pullyu animal '*represents only facts that it already knows how to use, and represents them only in the context of their use*' (Millikan 2004: 168).

Eventually, Millikan suggests, the limitations of the representational systems set up in the manner described above must have spurred the development of decoupled mental representations (Millikan 2004: 157). Presumably, the two sides of PPRs were gradually *'pulled apart'* (Artiga 2014: 552–53). Purely indicative representations (i.e. separate '*pushmis*') could inform consumers about facts without telling them to do something about these facts – and they could also store facts for future use (Millikan 2004: 168). Purely imperative representations (i.e. separate *'pullmis')* could direct behaviours not compelled by the current state of the world (Millikan 2004: 168). Together, these new forms of representing afforded the animal far greater versatility and sophistication of behaviour than could have been achieved without them. Indeed, complex decision-making and reasoning in humans are underpinned by mental states with purely indicative content (beliefs) and purely imperative content (intentions, desires).

Nevertheless, and this is an integral part of Millikan's proposal, PPRs are not meant to have been *fully* supplanted by the more advanced 'software' in animal brains (2004: 158). Why not? Well, in Millikan's

(2023) assessment, some of the challenges routinely faced by both simple and complex animals do not require much behavioural flexibility for cost-effective resolution (Millikan 2023: 415–16). Accordingly, the cognitive systems that respond to such challenges *might as well* run on PPRs.

Take the example of the earthworm from earlier in this section. Whether or not there could be any occasions when it would be advantageous for worms to move towards the light, if moving away from the light works out just fine *most of the time*, then it would not be very adaptive for worms to deliberate about the best course of action every time they detected a fluctuation in light levels. Deliberation, after all, is energy-draining and time-consuming. Moreover, if the situations in which the worm really *should* approach the light are fairly *uncommon*, then deliberation would not even be worth the hassle. On the grand scale of things, then, it would make sense for worms to *always* move away from the light – and this is something that could be facilitated with appropriate PPRs in place. Likewise, if social norms are rules that humans generally need to follow, transmit, and enforce, then a pushmi-pullyu cognitive system issuing PPRs such as *'We do not microwave fish in the office kitchen'* would be sufficient to ensure norm adherence in society (Millikan 1995: 193). *'If it's not broken, don't fix it'* – this is the crux of Millikan's idea.

Let us now take stock and reflect on the implications of the material covered in this section for the study of HDs. Millikan introduces PPRs as the enduring predecessors of purely indicative and purely imperative representations, not yet fine-tuned for information-gathering or action-triggering, but offering *both for the price of one* (2004: 157). Plenty of animal species, according to Millikan, rely on PPRs for all their representational needs (2023: 416). When it comes to humans, PPRs share the workload with other types of representations, but still take part in an array of functions, from perception to decision-making. For example, sweet tastes *'tell of nutritive value on one hand and direct continued eating or seeking more on the other'* (Millikan 2004: 159). And here is what Millikan says about intentions (2023: 416):

> *A fully settled intention is a pushmi-pullyu, the same representation token serving as a directive for the inner interpreters involved in planning and forming action and at the same*

time as a descriptive, say, as a belief, representing what a certain part of your future will be like so you can plan around it.

Is Millikan right to claim that PPRs have carved out *a representational niche* for themselves and continue to be extensively used by advanced cognisers? I am not sure. On the one hand, it is not entirely clear why human brains should maintain the archaic pushmi-pullyu 'software' when they have the most recent 'updates' installed and regularly churn out representations decoupled from either input or output. To rephrase this worry: what exactly is preventing PPRs from becoming extinct? This form of representing may provide a cost-effective way of organising behaviour *in simple animals*, but, precisely because of its rigidity, it is not without shortcomings.

What is more, the slashing of processing costs is unlikely to have been a major source of selective pressure on human brains. Our brains, after all, boast an impressive capacity for data handling and are not in the least frugal in terms of representation production. Mind-wandering is a case in point, showing that it is not unusual for our brains to devote cognitive resources to tasks of dubious importance. With purely indicative and purely imperative representations produced constantly, and with their combined representational powers far exceeding that of PPRs, would the preservation of pushmi-pullyu representational systems really yield any tangible benefits? Moreover, it is possible that retaining these systems would only serve to *increase* the processing burden on the brain. Human behaviour, after all, is subject to a great deal of cognitive control, which means that in our case, the use of PPRs would come with the processing baggage of multiple imperatives to suppress. And if cognitive control is a finite resource, as some researchers suggest (Muraven and Baumeister 2000), then the demands of top-down suppression of impulses triggered by the current environment might constitute a hindrance worth *selecting against*.

On the other hand, Millikan is not left without recourse against the arguments outlined in the previous paragraph. She could point out that, relative to all the representational activity in which our brains engage

on a daily basis, the cost of PPRs is merely a drop in the ocean, an expense too minor to lead to the elimination of PPRs from our cognitive repertoire. In the same way, a person picking up a large restaurant bill would not notice that they have been charged for an extra bottle of water. Alternatively, Millikan could speculate that the speed with which PPRs can both inform and instigate broadly appropriate action is enough to keep PPR-producing systems in operation. From my perspective, the second potential line of defence is stronger than the first, since it frames PPRs as having been retained because they are *adaptive*, rather than merely *insufficiently maladaptive*. At the end of the day, however, it is not essential for us that Millikan supply a convincing evolutionary story in support of her proposal. If some of our mental states display the hallmarks of PPRs, we can identify them as such without a full grasp of the benefits associated with preserving this form of representing. Should the pushmi-pullyu (PP) model of HDs prove definitively superior to its rivals, any uncertainty with respect to the role of PPRs in human cognition would recede into the background, outweighed by their explanatory power.

The question on which we should focus, then, is whether HDs *really* qualify as PPRs. Millikan suggests that they do, though she does not provide an explicit justification (2004: 93, 159, 164). The appeal of this view, however, is plain to see. HDs *seem to fit the mould* of PPRs by exhibiting links to both homeostatic events and homeostatic behaviours, and by giving the impression of being both informative and motivationally charged. It also helps that homeostatic parameters are often restored by the same type of action (e.g. drinking water or urinating) executed again and again: reliance on the slightly 'dated' PPR software is less surprising when the task at hand is straightforward. All in all, I think that the PP model of HDs deserves our attention; I just do not think that it deserves our *acceptance*. In the next section, I expound upon this model and its key commitments. Afterwards, I proceed to point out where it goes wrong, and how.

3. The pushmi-pullyu account of homeostatic drives

Let us assume, for the sake of argument, that HDs are members of the class of PPRs. What would follow from this assumption? In Millikan's own words, the content of PPRs is at once descriptive and directive

because it is interpreted in two different ways by two independent receivers, each with access to their own rules and databases (2023: 415):

> *A pushmi-pullyu representation would need to be a member of each of two different representational systems at once. It would need to have one satisfaction condition as a descriptive and a different satisfaction condition as a directive. The descriptives and the directives would be written in different languages, languages that contained the same set of representation forms but with different meanings… When read as descriptives, the set of forms in a system of pushmi-pullyu representations would map onto the set of states of affairs that these forms would represent as being actual. Read as directives, the set of these forms would map onto a set of states of affairs represented as to-be-brought-about.*

Here is an analogy to clarify the above. Imagine being in a restaurant with an open-plan kitchen. When the chef shouts *'Service!',* the waiting staff take it as a command to collect the dishes and bring them to the right tables. The customers, however, take the same utterance as a sign that they are about to receive their orders. The chef's exclamation, then, has different meanings for different categories of people in the restaurant, and what determines the meaning in each case is the person's purpose in coming to the restaurant. If you are an employee whose job is to take care of customers, *'Service!'* means *'Come serve these dishes now!'* If you are a customer looking forward to a meal, *'Service!'* means *'Your food is about to arrive!'*

The example above highlights that the content of pushmi-pullyu mental states is meant to be consumed by *two separate cognitive systems*. Here is how I understand the setup. One system – call it [DES] – speaks the descriptive language and reads PPRs as reports of states of affairs. Another system – call it [DIR] – understands only directives and reads PPRs as instructions to be carried out. [DES], then, is told how things stand (in relation to the agent) and [DIR] is told what is to be done about it – but all this is accomplished through the production of *a single representation* with *two types of uses*. As a result, while

the receiver systems may not have the tools to communicate with each other, they are kept perfectly 'in sync' because they receive their data from the same source.

Take intentions as an illustration. Millikan classifies them as PPRs, stating that they serve both to guide actions (e.g. turning on the oven) and to give rise to beliefs about what will happen once the actions are performed (e.g. the oven will start heating up) (2023: 416).[45] Since action guidance is best executed by imperatives (think of the language of manuals), while representing the potential outcomes of one's actions requires the use of indicatives, we may assume that the two processes are completely separate. Still, whenever the agent's intentions are altered, both the agent's actions and beliefs about the future automatically undergo a *coordinated modification* (Millikan 2004: 160). For example, should one decide to cook dinner in the microwave rather than in the oven, one would neither approach the oven nor expect it to start heating up. Instead, one would go to the microwave and expect the light to come on inside it.

What follows from this is that if HDs are pushmi-pullyu mental states, they should also feed into two separate cognitive systems of the types stipulated above. [DES], I take it, would interpret HDs as signalling something like homeostatic events and would use this information for the purposes of belief formation, learning, planning, and so on. [DIR], in turn, would interpret HDs as homeostatic commands and would make the arrangements for the required behaviours to be carried out (unless these are suppressed, since not every PPR has to result in overt behaviour). Each type of HD, then, would have a descriptive (pushmi) and a directive (pullyu) aspect, but being a *single representation*, it would be experienced by the agent in a unified way.

Accordingly, within the PP framework, the content of pain would be something like *'Damage; protect the body!'* The content of thirst would be *'Dehydration; drink water!'* The content of nausea would be 'Indigestion; vomit!' The content of fatigue would be *'Low energy; rest!'* The content of temperature feelings would be *'Body temperature too low/high; warm up/cool down!'* And so on. I expand on this in

[45] Millikan classifies intentions as PPRs on the basis that they both direct actions and describe their outcomes, and their descriptive aspect is only rendered true upon the satisfaction of their directive aspect (2004: 200).

the next section, where I argue that the PP model specifies the contents of HDs in a way that removes too many regular instances of these mental states from the explanandum. Following this, I tabulate the more general flaws of the PP account, showing that every version of it is bound to run into serious problems.

4. The problems of the pushmi-pullyu account of homeostatic drives

In this section, I propose three problems for the pushmi-pullyu account of homeostatic drives, each of which highlights a serious limitation of the model. I conclude by arguing that HDs do not qualify as PPRs.

4.1. Content ascription

Keeping in mind that PPRs are at once descriptive and directive, let us focus for a moment on their descriptive aspect. According to Millikan, PPRs have the capacity to describe past, current, or future events (Millikan 2004: 163, 197–98). For example, the food calls made by hens to their young represent that there is food available for consumption right now. Meanwhile, the inner PPRs that guide birds to stock up on food in preparation for winter represent that winter will arrive soon. In Millikan's own words, *'the nutcracker's perception is a representation of winter coming in that it will serve its own or proper function in a normal way only if winter is indeed coming'* (2004: 197):

Earlier, I suggested that if HDs are PPRs, they describe something like homeostatic events. This suggestion is uncontroversial: no one who endorses some form of indicativism has claimed that HDs inform us about anything other than changes in the physiological body state, broadly construed. Yet there is a caveat here, and it has to do with the *temporal orientation* of HDs. As I argue below, both philosophical considerations and neuroscientific evidence strongly imply that it is part of the proper function of HDs to describe *future homeostatic events*. The PP model should be able to accommodate this, but it fails to do so – and this failure, I maintain, calls the entire model into question.

The temporal orientation of HDs is by no means obvious. On the one hand, some HDs could be conceived as representing *past* homeostatic events. For example, pains of recovery (e.g. cuts, sprained ankles), as well as visceral pains, could be seen as informing us about bodily damage that has *already* taken place. Presumably, there would be learning value in representing past bodily damage: pains are unpleasant, and prolonged discomfort in the aftermath of an injury would ensure that the agent is more careful in the future. On the other hand, barring pain, few types of HDs would permit such interpretation. At a stretch, fatigue could be construed as representing past exertion, but I can think of no other example.

The proposal that HDs represent *current* homeostatic events has a wider scope. All the instances of HDs seemingly representing past events may be easily reclassified as instances of HDs representing current events. For example, *why does my knee hurt? Because it was dislocated and I am damaging it right now by climbing up the stairs. Why does the cut on my finger hurt? Because the inflammation in my finger is ongoing.* Similarly, breathlessness may be plausibly interpreted as indicating a current lack of oxygen. By the same token, nausea seems to indicate current indigestion, an urge to cough seems to indicate a current tickle in the throat, an urge to urinate seems to indicate current bladder fullness, and feeling cold seems to indicate a currently low body temperature.

Yet even this proposal does not capture *all* instances of HDs, since there are many cases of HDs apparently representing *future* events. As recent experiments confirm, thirst is consistently triggered *in anticipation* of a drop in hydration levels and satiated *'tens of minutes before'* the fluid balance is restored (Zimmermann and others 2016: 680; Zimmermann and others 2019). Hunger shows the same pattern of *predictive* activation and deactivation: the group of hypothalamic AgRP neurons that promote the drive to eat are 'switched on' by cold exposure and 'switched off' by the presentation of food, before ingestion is even initiated (Betley and others 2015; Chen and others 2015; Deem, Faber, and Morton 2022).[46] And pain often corresponds to *potential* rather than actual bodily damage (Klein 2015: 2).

[46] Since keeping the body warm is energy-consuming, increased hunger following cold exposure serves as a compensatory strategy.

Crucially, these examples cannot be dismissed as outliers. Rather, they seem to reveal *the basic principles* underlying the production of HDs. As Woods and Ramsay (2007) point out in their literature review, the mechanisms of homeostatic maintenance are not limited to the processing of negative feedback about the current state of homeostatic variables. In their own words, *'regulated variables commonly achieve this stability in the absence of any obvious error signal or deviation'* (2007: 6). This suggests that the brain routinely exerts direct influences on regulated variables (e.g. lowering glucose levels before a meal to avert a spike following food ingestion) and generates HDs (e.g. producing hunger at regular meal times) to *prevent* as well as to correct homeostatic fluctuations (Lowell 2019; Skvortsova and others 2021; Woods and Ramsay 2007).

Far from merely *reacting* to deviations from the homeostatic norm, then, HD-producing systems appear to *anticipate* such deviations by tracking internal and external predictive cues, some of which are learnt. Anticipatory nausea, for example, is a common side-effect of chemotherapy: after a few treatments, the unpleasant sensations caused by chemotherapy can be triggered by conditioned stimuli such as entering the clinic or seeing the nurse (Kamen and others 2014). Anticipatory breathlessness is another well-documented phenomenon, reproducible in laboratory settings by pairing affective cues with elevated levels of CO_2 in the air (Stegen and others 1999).[47] Importantly, HD-producing systems are not indiscriminate in their learning: they learn only reliable cues, i.e. cues that *make sense* as signs of upcoming homeostatic fluctuations. In the study by Stegen and others (1999), for instance, neutral images previously paired with inhaling CO_2–enriched air did not produce anticipatory breathlessness, whereas fear-inducing images did. More recently, errors in predictive models of interoception have been proposed to underlie the development of medically unexplained symptoms (Barrett and Simmons 2015; Paulus, Feinstein and Khalsa 2019; Quadt, Critchley and Garfinkel 2018). If this hypothesis is correct, our HD-producing systems may be prone to dysfunction as a result of their disposition to foresee and pre-empt homeostatic perturbations.

[47] As noted earlier, the cognitive systems generating the feeling of breathlessness track CO_2 levels rather than oxygen levels.

In sum, representing future events appears to be part of the *proper function* of HDs.[48] The PP model should be able to accommodate this; but it does not. Recall that in Millikan's nutcracker example, the PPR that correlates with both the stockpiling of food and the imminent arrival of winter is future-oriented insofar as the stockpiling only performs its proper function if winter does arrive (2004: 197). Now, let us see if we can draw an analogy with thirst here. Thirst, I argued above, represents future dehydration at least some of the time; call it predictive thirst. Yet, on the PP model, thirst also directs drinking – which means that predictive thirst directs the agent to do the very thing that would *prevent* the future event it describes from ever taking place. It is difficult to believe that any representation could have such content. After all, if predictive thirst fulfils its function, dehydration does not take place.

As far as I can tell, the only way to solve this problem is to remove the requirement of representing future events from the explanandum. This, however, would remove a substantial number of regular instances of HDs and thus weaken the PP model. And, as we will now see, there are bigger problems facing the model there are bigger problems facing the model once we combine everything we have discussed so far and look at it together.[49]

4.2. The problems with phenomenology

Both pure accounts, I have argued, fail to accommodate some aspect of the phenomenology of HDs. How does the PP account compare to them in this respect? At a glance, it may seem that it has a decided advantage over its competitors. Indeed, it may seem that the PP model is capable of explaining both the sensory and the motivational phenomenology of HDs by simply appealing to the purported pushmi-

[48] An argument could even be made that HDs *always* represent future (catastrophic) homeostatic events. We can survive for days without drinking, but thirst arises in response to very small levels of fluid depletion. Likewise, many feel the urge to urinate when the bladder is nowhere near full. The *sensitivity* of HDs to minute changes in the body state may then lead us to wonder if HDs are not totally *future-oriented*. In this framework, thirst would not be signalling dehydration, but rather signalling upcoming dehydration, with dehydration being the catastrophic end result of the process which should be avoided.

[49] Strictly speaking, it could also be posited that HDs, much like intentions in Millikan's interpretation, describe the homeostatic events that follow the successful performance of the actions they direct, i.e. that HDs describe their own *satisfaction conditions*. For example, on this version of the PP model, thirst would signal the restoration of fluid levels in the near future. The flaws of such a model, however, are impossible to overlook. First, the restoration of homeostatic parameters is not a certainty, but something for which to strive. Second, there does not seem to be any value in the information that a homeostatic parameter is probably going to be restored soon, which makes it extremely improbable that any system should have evolved to extract such information.

pullyu nature of these mental states. For example, *why does breathlessness feel both like a bodily event and an urge to breathe? Because breathlessness is a representation that informs us about the lack of oxygen and instructs us to inhale all at the same time*. More specifically, the sensory features of HDs could be attributed to their descriptive aspect. For example, *why does thirst have a felt intensity? Because thirst indicates a fluctuation in fluid levels that has a certain magnitude*. In turn, the motivational force of HDs could be attributed to their directive aspect. For example, *why does one feel like vomiting when nauseous? Because nausea commands this specific type of behaviour*.

On this occasion, however, appearances are deceiving. Consider *the felt location* of HDs – for example, itch. From the representationalist perspective, itch has a felt location because it represents this location (in some way or another) to the cognitive system by which it is consumed. Now, if itch is a PPR, it is consumed by *two* cognitive systems, [DES] and [DIR]. [DES] interprets itch as describing something like skin irritation; [DIR] interprets itch as an order to scratch. In order to explain why itch has a felt location, we only need *one* of these systems to consume the location data of itch – but it seems that both do. Indeed, [DES] would be interested in the location data because there is no use in knowing *that* skin is irritated without also knowing *where* the irritation is. By the same token, [DIR] would be interested in the location data because a command to scratch is meaningless if it does not specify *where* to scratch. We are thereby obliged to conclude that on the PP model, itch has a felt location in virtue of *both* reporting an event occurring in some bodily location and instructing an action targeting this bodily location.[50]

While this explanation for the felt location of itch is rather convoluted, the real challenge for the PP model is to provide *any* explanation for the felt location of some other types of HDs. Generally speaking, the PP model would be stumped by any member of this group that could be framed as informing us about events at location L_1 and instructing us to act on location L_2. Take nausea.[51] Also referred to as 'queasiness'

[50] Honeybee dances may be subjected to a similar analysis, as they both describe the location of nectar and direct the spectator bee to fly to this location.

[51] Hunger and thirst are two other types of HDs whose felt location may be used to challenge the PP model. I chose to focus on nausea because hunger and thirst already featured in my arguments against pure imperativism, which ran along similar lines as the arguments laid out here. I return to this point later in the chapter.

or 'stomach awareness', nausea is typically felt in the upper digestive tract (Koch 2017; Napadow and others 2013; Quigley, Hasler and Parkman 2001). Can we accept, by analogy with itch, that nausea is felt in the general stomach area because it describes indigestion at this location and directs vomiting aimed at this location? Vomiting does serve to evacuate the contents of one's stomach, but I hesitate to say that this is where it is *directed*. A complex set of movements involving the throat, the glottis, the diaphragm, and the abdominal muscles is involved in this process, and it is these parts of the body, not the stomach, that are at the centre of one's attention during vomiting and retching (Quigley, Hasler and Parkman 2001). This suggests that one does not vomit *with* one's throat, diaphragm, etc in the same way as one scratches an itch *with* one's nails. In the case of scratching, one is clearly aiming for the location of the itch. Yet, in the case of vomiting, one's actions are not guided by one's stomach.

Let me rephrase and expand on this last point. For a location to be part of an imperative, this particular location needs to matter for action execution. To illustrate, if you ask me to throw the ball into the basket, it may be your purpose that the ball lands on the ground beneath the basket. Still, the ground beneath the basket does not feature (and does not need to feature) in your request – only the basket does.

When it comes to vomiting, the desideratum I have sketched here seems to apply to the throat, the glottis, the diaphragm, and the abdominal muscles – but not to the stomach, the emptying of the contents of which *naturally* follows from certain muscular contractions. This poses an issue for the PP model, which cannot accept that nausea is directing us to act on all these bodily locations, because nausea can hardly be framed as *reporting* any events occurring in them. To circumvent the issue, a PP theorist would have to grant that (1) [DES] and [DIR] can extract *different bits of data* from HDs and (2) [DIR] is not *always* interested in their location data. Yet these additions to the theory do not place it on firmer footing. Body-directed commands need to be specified in terms of bodily locations, and if [DIR] is not extracting this data from HDs, then it is not clear where it is getting it. Moreover, a question arises: if the PP model has to concede that [DES] and [DIR] do not access *the same aspects* of HDs, then what are the benefits of holding on to the idea that HDs are *single* representations accessed by two different cognitive systems?

This question steadily gains in prominence as we look at how the PP model handles the motivational phenomenology of HDs. Recall that HDs are typically experienced as *urges to act* in a certain way which differ *in degree of their felt urgency*. For example, hunger is experienced as an urge to eat, but this urge is at times overwhelming, at times moderately powerful, and at times only slightly nagging. What does the PP account have to say about that?

The PP model has options here – but none of them are good. Take hunger. From the representationalist perspective, hunger has a felt urgency because it represents the urgency (i.e. the extent of the importance) of eating to the cognitive system that decodes it. Now, if hunger is a PPR, then it is decoded by *two* cognitive systems, [DES] and [DIR]. While [DES] reads hunger as something like energy depletion, [DIR] reads it as a command to eat. Out of these two systems, [DIR] is the obvious candidate for explaining the felt urgency of hunger, as [DIR] would certainly care about the importance of fulfilling the commands it receives. What is equally obvious, however, is that [DES] would care about the severity, or the magnitude of the events of which it is informed. After all, what would be the use of knowing that one's energy levels are no longer the same as they were without knowing *how much* they have changed?

And here we get to the sticking point. As noted previously, *the felt urgency* (a motivational characteristic) of HDs appears to be distinct from *the felt intensity* (a sensory characteristic) of HDs. If this is the case, it is detrimental to the PP model, since it entails that HD-producing system(s) separately encode, *in the same representation*, the magnitude of a homeostatic event *and* the importance of acting on it. Not only is this a lot of content to cram into a single representation, but it is not clear how the two types of content could be separately extracted by [DES] and [DIR] with sufficient accuracy – or, for that matter, why a system capable of such complex computations would be generating PPRs, the most primitive of representations.

And this is not to mention how implausible it is in the first place that both tasks would be managed by the same system. The magnitude of homeostatic events may be estimated purely on the basis of the sensory

data about them, but determining the urgency of addressing these events requires different types of data. Consider that one may be low on nutrients and water, suffering exhaustion and bodily damage, but have even more pressing concerns – such as saving one's child from hypothermia or vomiting up a poisonous mushroom. What we may call the *'objective' urgency* of homeostatic behaviours thus depends on the magnitude of the homeostatic events prompting them, *as well as* on the relative urgency of engaging in other behaviours. And since the algorithms that include all these factors may be expected to fall within the remit of complex centralised systems, there is no reason to believe that they are managed by the same system that only needs sensory information to determine the magnitude of homeostatic events. The bottom line is that the two tasks are *not* on the same level.

Of course, a PP theorist could accept that the severity of fluctuations in homeostatic variables does not always match the urgency of correcting them, but deny that the HD-producing system(s) are in charge of the urgency computations. Such a move, however, would spawn the fresh problem: identifying the cognitive system that *is* responsible for them. As [DIR] only speaks the language of imperatives, it would not be able to support any computations that involve comparing the outcomes of various actions (and thus involve representing indicative content). The decision-making system(s) would not be a good candidate for the job either. Given that one can choose not to act on a strong urge, urgency is not something the decision-making system(s) determines, but something it factors into its calculations. The only possibility that remains is that the urgency of HDs (and other motivational states) is the output of a dedicated system that takes various commands as input and ranks them in terms of their relevance to the needs of the agent. Yet if such a system exists – and I think it might – its very existence serves to undermine the PP model. Indeed, if the felt urgency of HDs is generated separately and after homeostatic commands (or, in that case, something like *homeostatic recommendations*) are issued by [DIR], then HDs do not have a felt urgency *in virtue* of having pushmi-pullyu content. And if the PP model does not offer an explanation for the motivational phenomenology of HDs, then it is no better than the models it was meant to improve upon.

All in all, it looks like the cognitive architecture of HDs entailed by the PP model simply has no place for urgency computations. In keeping with this, a PP theorist could, as a last resort, deny that our brains engage in these computations. That is to say, a PP theorist could argue that while we are able to report how urgent HDs feel to us, in doing so, we are not reporting a distinct quality of HDs, but rather making a judgement on the basis of their felt intensity.

To this, I would reply that there are phenomenological and non-phenomenological reasons to think that our brains compute *both* the magnitude and the importance of acting on homeostatic events. This position was defended in Chapter 1, and I will not repeat it here. I will only note that, given how advanced our brains are, they should at least *try* to approximate the 'objective' urgency of homeostatic behaviours. This would greatly simplify our decision-making (which already involves choosing between doing things we feel compelled to do, things we would like to do, and things we think we should do) and prevent us from being overwhelmed by multiple urges demanding action. Contemporary neuroscientists seem to agree with this general picture, suggesting that '*an urgency signal would be expected to originate from a region that projects to a wide range of cortical areas to influence both decision-making and action execution'* (Carland, Thura and Cisek 2019: 9).

Let us take stock. The PP model has not provided an improved account of the phenomenology of HDs. While it might have been expected to gain considerable 'wiggle room' in portraying the representational content of HDs as simultaneously descriptive and directive, the PP model has proved just as vulnerable to objections about its treatment of the phenomenology of HDs as its pure predecessors. In fact, the key commitment of the model – that the representational content of HDs is unpacked by two cognitive systems – has turned out to be particularly problematic. In what follows, I offer further confirmation of this by showing that the two-systems requirement seriously complicates matters when it comes to making sense of homeostatic regulation.

4.3. The problem with homeostatic regulation

The explanation for the function of HDs that follows from the PP model is fairly straightforward. On this model, [DIR] reads HDs as commands and arranges for these commands to be carried out. At the same time, [DES] reads HDs as reports of homeostatic events – and in doing so, affords the agent valuable insight into the physiological state of her body. This, in turn, opens up the possibility of carrying out more creative and flexible homeostatic behaviours in addition to the rigid behaviours already prescribed by the directive aspect of HDs. For example, the agent can address back pain not only by means of movement restriction and rest, but also by altering her posture, massaging the sore spots, or seeing a doctor and using the felt characteristics of pain to inform diagnosis and treatment.

Once again, however, it is not all plain sailing for the PP model. Recall that the brain has a number of tools in its arsenal for controlling the physiological body state. In addition to generating experiences that motivate the agent to act in accordance with homeostatic priorities, the brain also sends direct commands to the motor centres, releases hormones that serve as 'chemical messages' to organs and tissues, and directly adjusts homeostatic variables such as the metabolic rate (Berntson and Khalsa 2021; Craig 2008). Even the putative setpoints for homeostatic variables appear to be subject to short-term and long-term modifications (e.g. body temperature during illness, hormonal levels during a sustained stress response).

Crucially, in configuring these processes, the brain relies on its own estimations of the current and predicted physiological state. For example, the reduction in blood glucose levels before regular meals is not entirely 'natural', since it is engineered by the relevant cognitive systems to prevent a spike in blood glucose following food consumption (Ramsay and Woods 2016). Overall, the regulatory influences described here are highly intricate and ever-present, although the brain does not always get things right. For example, the development of type II diabetes may be at least partially attributable to impaired glucose sensing and the resulting dysregulation (Modell and others 2015; Myers and others 2021). Such findings highlight the key role of accurate data and accurate predictions in homeostatic regulation.

Keeping this in mind, let us return to the PP model. It has something to say about *one* of the aspects of homeostatic regulation – but what it has to say is not compatible with *the full picture*. In assuming that HDs are single representations interpreted by [DES] and [DIR] in two different ways, the PP model weds itself to the idea that the system for *understanding* and the system for *correcting* changes in the physiological body state *cannot speak to each other* for want of a shared language.

And herein lies the problem. HDs *may* have content that is at once descriptive and directive, but non-HDs (e.g. heartbeat sensations and sensations that accompany breathing or swallowing) certainly do not. These mental states have never been attributed any content other than descriptive, and for good reason: they are not felt as motivationally charged, and there is no specific response associated with them. This means that, on the PP model, non-HDs would be read by [DES], the system that speaks the descriptive language, but not by [DIR], the system that speaks the directive language. Bizarrely, this would deny non-HDs – and any other purely indicative representations of homeostatic events, which may include representations of predicted homeostatic events – any role in informing homeostatic regulation. The system capable of exerting regulatory influences would not be capable of reading non-HDs, while the system capable of reading non-HDs would not be capable of exerting regulatory influences. Consequently, for example, the heart rate data would not be used by the brain to support heart rate regulation – and this is just not true.[52]

Can the PP model avoid this absurd conclusion? To do so, the PP model would have to posit the existence of additional cognitive system(s) within the domain of homeostatic regulation. Specifically, it would have to posit the existence of system(s) that can take representations of homeostatic events as input and produce appropriate responses to these events as output. These systems would need to represent both current states of affairs, possible future states of affairs, and expected outcomes of homeostatic regulation; as such, they would speak both the descriptive and the directive language. There is nothing strange about positing the existence of such system(s); except that their representational capacity

[52] See Taggart and others (2016) for examples of of higher brain centres involvement in cardiac control.

exceeds that of pushmi-pullyu representational systems and makes it unclear what possible benefit the latter could bring to homeostatic regulation.

4.4. Homeostatic drives do not qualify as pushmi-pullyu representations

Given the problems of the PP model listed in this section, it is doubtful that HDs qualify as PPRs. I argue that they do not. Here are Millikan's conditions for the ascription of pushmi-pullyu content to representations, summarised by me:

(1) The representation reliably corresponds to a certain type of input and a certain type of output.

(2) The fulfilment of its proper function by the representation hinges on this two-sided correspondence.

Take the familiar example of the rabbit thump:

(1) The thump reliably corresponds with the presence of a predator and the running/freezing of rabbits;

(2) Aiding predator escape hinges on the presence of a predator and the running/freezing of rabbits.

Do HDs meet these conditions? Let us start with (2) and use thirst as an example. Say thirst corresponds to low fluid levels and commands drinking. Say the proper function of thirst is to help maintain fluid levels. The issue is that drinking would contribute to the maintenance of fluid levels even if the fluid levels are normal. Likewise, even if your bladder is only a 10 percent full, urinating still performs its proper function of emptying the bladder. Or, if you feel pain when is no bodily damage, pain still performs its proper function of protecting the structural integrity of the body.

Here, I may be accused of being nitpicky. After all, if the bladder is completely empty, then the urge to urinate cannot perform its biological function. To this, I would reply that if the bladder is *completely* empty, urination cannot happen. This is the same as the rabbit not being able to run, rather than the same as rabbits running but not escaping the predator. I say, then, that it is impossible, under normal conditions, to urinate without emptying the bladder, or to consume food without replenishing energy levels. As such, the indicative content of HDs ends up being too broad and uninformative. If all that is required is some

tiny fluctuation from the normal values of a homeostatic variable, then the indicative content of the urge to urinate is *'the state of your bladder is such that you could urinate'*.The indicative content of hunger is *'the state of your body is such that you could eat'*. And so forth.

Now, consider (1). Here is a helpful quote by Millikan (2023: 415):

> *A pushmi-pullyu representational system would survive only if a high enough proportion of the affairs it could represent really were paired with just one thing that should be always be done about it, one thing that would be maximally beneficial some critical proportion of the time to both the producing and the consuming side of the system.*

Condition (1) renders the PP model more problematic than either of the pure theories. Pure indicativism has been criticised on the grounds that there is no *single type of bodily event* corresponding to each type of HD. At the same time, I have argued that pure imperativism ends up in a similar bind by incorrectly positing that there is a *single type of body-directed action* corresponding to each type of HD. Given that the PP model states that each type of HD corresponds to a single type of bodily event and to a single type of behavioural response, it inherits both criticisms – and, as shown in this chapter, it also acquires additional ones. The PP model, then, not only repeats the mistakes of its predecessors but amplifies them, which makes it an even less convincing account of HDs.

5. Conclusion

In this chapter, I began to explore the impure approach to HDs by considering the possibility that HDs are pushmi-pullyu representations. This theory has been mentioned in the literature as a viable option but has not been properly developed or analysed until now. In filling this gap in the research, I found that instead of combining the strengths of pure indicativism and pure imperativism, the PP model inherits their problems and acquires new ones. In the extended conclusion, I offer a way forward for the inquiry into the

nature of HDs. Pure models, I argue, fail not simply because they attempt to explain HDs with one type of content, but because they attempt to do so with a single representation. The PP model repeats the same mistake; in fact, by trying to force two types of content into a single representation, it fares worse than its predecessors. We should learn from this failure. I conclude that in order to get to a good working theory of HDs, we need to abandon the idea that HDs are simple mental states that may be explained by positing a single representation. We need to go *impure*, and we need to go *complex*.

Conclusion

1. Taking stock

What, if anything, do bodily experiences represent? This debate has long centred on pain, and lately, it has split into two: namely, the debate on *valence* and the debate on *what remains* of pain once valence is removed. With respect to the latter, pure indicativism and pure imperativism have emerged as the main contenders. Despite their opposing takes on pain, both accounts assume it to be a relatively *simple* mental state, that is, a mental state with *one type of content*. Of course, valence is meant to *attach* to this representation, but in essence, pain is still viewed as simple. The only question is whether its content is purely indicative or purely imperative.

This debate is now showing signs of stagnation. The misrepresentation challenge to pure indicativism continues to be reiterated (Casser 2021; Corns 2014a), and objections have been raised against pure imperativism (Tumulty 2009; Bain 2011), but there have been no serious attempts in recent years to significantly improve the prospects of either theory or to develop an alternative. In general, it is not clear where we go from here, and this is what has inspired this thesis. Pain, I have argued, is only one member of a broader class of *homeostatic drives (HDs)*: thirst, hunger, nausea, itch, breathlessness, fatigue, temperature sensations, toilet urges, and so forth. Despite sharing the key characteristics of pain, these mental states remain poorly understood and only sporadically addressed. By applying existing theories of pain to them as a group, I have suggested, we can clarify the nature of these mental states and bring either resolution or a new direction to the inquiry.

What my analysis has shown is that neither pure indicativism nor pure imperativism succeeds in explaining HDs, though each has its merits. Accordingly, I have explored the possibility that the content of these mental states is *at once* indicative and imperative. This pushmi-pullyu theory, however, has proven equally inadequate at capturing every aspect of HDs. In inheriting the weaknesses rather than combining the

strengths of its predecessors, and in acquiring new problems on top of that, it has not offered an improvement in terms of its explanatory power.

So what is the lesson to draw from this? Why have these theories failed? We must answer these questions to determine the best way forward. One option is that these theories have failed simply because they are *representational*. Perhaps feelings of pain, thirst, nausea, fatigue, and related mental states are devoid of representational content. Or perhaps they do have a content, but this only gets us so far in understanding them. While this is a possibility, it is not an exciting one, and other options remain. In particular, other versions of impure theories may be developed – more complex versions. My position, in short, is that the theories discussed in this thesis are not wrong in assuming that HDs have representational content; they just do not posit enough layers of content to accommodate all aspects of these mental states. In what follows, I show how this can be done by offering a sketch of an *impure and complex* theory of HDs.

2. Moving on

I will use the example of nausea throughout this discussion for the sake of simplicity.

On the pure indicativist account, nausea is a representation of a specific homeostatic event (namely, indigestion) generated by the interoceptive system and then processed by the *motivation-forming* system, which matches it with an appropriate motivation (such as vomiting). Consider, however, the following points. First, vomiting may occur quickly and suddenly, without being preceded by nausea. Second, nausea may occur without an accompanying urge to vomit. It would therefore seem that in addition to vomiting that is motivated by nausea, the interoceptive system (specifically, its effector arm) can also send direct commands to the motor centres to initiate vomiting, commands not causally attributable to nausea but rather generated in parallel with it. Given the protective value of vomiting, and neuroscientific findings suggesting that urge suppression is *'an active process of response suppression'* (Sundby and others 2019: 1404), the existence of such *direct interoceptive commands* to the motor centres seems likely. But then

what is the point of the interoceptive system sending a separate signal to the motivation system? Such an arrangement strikes me as redundant.

On the pure imperativist account, nausea is a command (namely, to vomit) generated by the interoceptive system and then processed by the decision-making system, which determines how and whether the homeostatic commands are carried out. As argued in Chapter 3, however, pure imperativism struggles to accommodate direct interoceptive commands to the motor centres. Indeed, given the dissociation of nausea and vomiting, it would have to posit that the interoceptive system sends separate, yet *identical*, commands to the motor centres and to the decision-making system. Once again, the cognitive architecture of HDs entailed by pure imperativism leads us to doubt whether this theory is on the right track.

In my view, the direct interoceptive commands to the motor centres, so problematic for pure theories, may serve as a foundational building block for *impure theories* of HDs. Here is the idea: I propose that we, as agents, have direct experiential access to these commands: that is, we feel them *as urges*. As previously argued, urges are distinct from desires, being generated subpersonally, and are characteristic of HDs, though they are also observed in the context of emotions (for example, an angry person might feel an urge to shout, while a sad person might feel an urge to cry).[53] Pathological urges — those that are inappropriate and exceptionally difficult to suppress — are hallmark features of obsessive-compulsive disorder (OCD) and Tourette syndrome (Lerner and others 2009).

As a component of an impure theory of HDs, urges can account for at least some of the motivational profile of these mental states. In simple animals, I surmise, interoceptive commands to the motor centres may constitute the sole mechanism of behavioural homeostatic regulation. In most adult humans, however, all commands to the motor centres are automatically intercepted and suppressed. The potential extent of top-down motor suppression is revealed in catatonia, a neuropsychiatric syndrome

[53] See Hall (2008) for a distinction between *'central desires'* and *'sensory urges'*.

characterised, in addition to motor symptoms such as prolonged involuntary posturing, by excessive urinary retention that may reach dangerous levels and by a lack of normal pain and fatigue behaviours (Beld, Philbrick and Rummans 2004; Northoff 2002).

> *No catatonic patient reported any feeling of pain or tiredness even if he postured and remained in the same position for hours, days, or weeks […] For several days this patient stood in front of her wardrobe remaining in the same quite uncomfortable position with raised arms and standing tip-toe […] denying any feeling of tiredness* (Northoff 2002: 557–8).

The development of the ability to intercept and suppress direct interoceptive commands to the motor centres, I suggest, is what has necessitated the evolutionary emergence of the ability to experience urges. As the example of catatonia illustrates, motor control can be dangerous – and this is why it needs to be *counterbalanced by urges*. In a nutshell: while motor control is highly adaptive, permitting flexible and complex behaviour, it can also result in the excessive suppression of the vital interoceptive commands such as those to breathe, drink water, or urinate. Urges serve to prevent this by motivating us not to delay what needs to be done.[54]

In principle, this proposal – that interoceptive commands to the motor centres are routinely suppressed and represented as urges – could be combined with pure indicativism or pure imperativism to yield a two-component theory of HDs. Yet it does not seem to fit well with either. If HDs are felt commands plus urges representing interoceptive commands to the motor centres, then there is nothing to explain the sensory aspect of HDs (and the value of felt commands in this scenario is unclear). And if HDs are representations of homeostatic events combined with urges representing interoceptive commands to the motor centres, then there is nothing to explain how HDs are individuated. Given that HDs such as hunger do not have a fixed bodily signature, it may be tempting to suggest that HDs are distinguished on the

[54] Of course, our ability to suppress interoceptive commands is not *absolute*: for example, we cannot keep a hand in scalding water or hold our breath indefinitely, and we cannot always suppress vomiting either. Urges, then, may be viewed as having evolved to motivate us to address our homeostatic needs before it is either too late to act or before we lose control over their satisfaction (and thus have our goal-directed activities seriously disrupted).

basis of the urge they contain rather than the event they represent. This suggestion, however, does not work, since, as argued in Chapter 3, some HDs lack a specific urge by which they could be individuated.

Here is what I propose instead. The interoceptive system processes the homeostatic data and outputs commands to the motor centres and representations of homeostatic events. We, as agents, have experiential access to both: interoceptive commands are felt as urges and representations of homeostatic events on their own are felt as non-HDs, affectively and motivationally neutral bodily sensations. HDs, in this scenario, are *representations of homeostatic needs* formed by a separate cognitive system (or a collection of systems), which receives both commands to the motor centres and representations of homeostatic events as input. The idea is that, starting from birth, HD-producing systems are constantly learning to *infer* the needs of the body by matching our actions with their homeostatic consequences. By connecting homeostatic fluctuations to commands which follow and precede them, HD-producing systems can eventually come to represent bodily states which necessitate action — that is, homeostatic needs — before the interoceptive system issues the relevant commands. This is why, for example, we can feel nausea without feeling an urge to vomit.

On this tentative proposal, HDs such as nausea, thirst, and pain are distinct because they represent different homeostatic needs, which are bodily properties inferred from homeostatic fluctuations and interoceptive commands. Nausea, for example, represents something like the need for an empty stomach and is inferred from changes in the body state which have reliably preceded vomiting in the past. Thirst, in turn, represents the need to restore optimal fluid levels, and pain represents the need to restore the integrity of some body part. The reason why HDs have affective phenomenology, I suggest, is that they represent homeostatic needs as being *frustrated*, while the reason HDs have sensory phenomenology is that they integrate representations of homeostatic events from which they are inferred. For example, thirst is experienced as unpleasant throat and mouth dryness because it represents the frustrated need for

water causing throat and mouth dryness.[55] And when it comes to the motivational profile of HDs, these mental states, I suggest, motivate us in a complex way: not only by integrating within them urges to perform specific homeostatic behaviours, but also by being *our* needs and thus providing us with an intrinsic reason to act.

This conjecture finds support in case studies. Take Tourette syndrome, which is a tic disorder. Sufferers typically develop tics first and then gradually acquire the ability to *'sense'* oncoming tics in their bodies (Houghton and others 2014). These sensations are commonly known as *'premonitory urges'* and are described as muscle tension in the location of the tic; for example, the tic to blink is felt as tension in the eye. Crucially, my proposal can explain the transition from pure urges to the sensory phenomena characteristic of Tourette syndrome. Specifically, this can be understood as a case of the homeostatic needs system maladaptively learning to represent the need to tic by associating the recurrent urge with tension in the muscle that the agent is attempting to control.[56] Note that in this scenario, the system is not malfunctioning: rather, the the intensity and frequent occurrence of the urge to tic, combined with tension in the relevant area, mislead the system into encoding something that is not a genuine homeostatic need as if it were one. The fault, then, lies with the mechanism producing the tic, not with the homeostatic needs system.

Something similar, I suggest, takes place in OCD, where compulsions may be preceded not by obsessive thoughts but by sensory phenomena (Bragdon and others 2023; Ferrao and others 2012). Indeed, as

---

[55] To use Armstrong's (1962) terminology, I propose that representations of homeostatic events generated by the interoceptive system are *transitive* sensations of pressure, pulsations, throat dryness, stomach emptiness or fullness, bladder fullness, stomach contractions, and so on. These sensations may acquire the *intransitive* qualities of painfulness or itchiness by being integrated into the *homeostatic needs representations*. This explains the felt bodily location of HDs: while homeostatic needs are not the properties that can themselves be located, they integrate within themselves homeostatic events that are. Pain, for example, represents the need to restore the integrity of the body, but can be felt as pulsating pressure in the head because this is where the homeostatic events evidencing the frustration of this particular need are unfolding. Hunger, in turn, represents the need to restore energy levels, but can be felt as an emptiness in the stomach or as lightheadedness for the same reason.

[56] My proposal also clarifies why HDs such as pain are subject to extensive modulation (as discussed in Chapter 2, §3.2). Take central sensitisation – the amplification of pain signalling within the central nervous system, which manifests as the lowering of pain thresholds (Harte, Harris and Clauw 2018). On my proposal, this phenomenon has not evolved to make us feel more pain under certain circumstances. Rather, it has evolved to facilitate the issuing of relevant interoceptive commands when the body requires extra protection (for example, under conditions of existing injury). However, the homeostatic needs system, being sensitive to interoceptive commands, matches the intensity of HDs to their strength, which is why we also experience increased pain as a result of central sensitisation.

Ferrao and others note (2012: 256), *'many OCD patients report some kind of skin sensation prior to washing (a sensation that their hands are greasy, sweaty, or dusty).* Interestingly, the insula, the brain region commonly identified as a key component of the interoceptive cortex (Craig 2015: ch.6; Livneh and Andermann 2021), has also been found to be active during the suppression of *'natural'* urges (such as the urge to blink) (Lerner and others 2009). The overlap between the urge-suppression and interoceptive circuitry lends further support to the hypothesis that HDs are mental states that integrate urges and interoceptive representations.

If I am on the right track, then HDs are complex, inference-based representations, probably unique to humans and developing during the first years of life. The wide use of predictive cues in generating HDs and the presence of HD-like mental phenomena in conditions such as Tourette syndrome and OCD support the view that the homeostatic needs system retains the capacity to learn throughout life. The adaptive value of this system lies in increasing our behavioural flexibility. In addition to satisfying pressing urges, we can anticipate and address our homeostatic needs while the urge to act is still weak, as well as plan to address them without having our activity disrupted. And by recognising multiple urges and unpleasant bodily sensations as arising from the same need, we can make sense of our experiences and resolve these more easily than we could otherwise.

I suggest that this sketch of a proposal can serve as a useful starting point for the inquiry into impure theories of HDs.

Bibliography

Andrew, David, and Arthur D. Craig. 2001. 'Spinothalamic Lamina I Neurons Selectively Sensitive to Histamine: A Central Neural Pathway for Itch', *Nature Neuroscience*, 4.1: 72–77 <https://doi.org/10.1038/82924>

Akins, Kathleen. 1996. 'Of Sensory Systems and the 'Aboutness' of Mental States', *Journal of Philosophy*, 93.7: 337–72 <https://doi.org/10.2307/2941125>

Alonge, Katherine M., David A. D'Alessio, and Michael W. Schwartz. 2021. 'Brain Control of Blood Glucose Levels: Implications for the Pathogenesis of Type 2 Diabetes', *Diabetologia*, 64.1: 5–14 <https://doi.org/10.1007/s00125-020-05293-3>

Ambrosino, Nicolino, and Maurizia Serradori. 2006. 'Determining the Cause of Dyspnoea: Linguistic and Biological Descriptors', *Chronic Respiratory Disease*, 3.3: 117–22 <https://doi.org/10.1191/1479972306cd110ra>

Apkarian, Vania A. 2019. 'Definitions of Nociception, Pain, and Chronic Pain with Implications Regarding Science and Society', *Neuroscience Letters*, 702: 1–2 <https://doi.org/10.1016/j.neulet.2018.11.039>

Artiga, Marc. 2014. 'Teleosemantics and Pushmi-Pullyu Representations', *Erkenntnis*, 79.S3: 545–66 <https://doi.org/10.1007/s10670-013-9517-5>

Armstrong, David M. 1962. *Bodily Sensations* (London: Routledge)

—— 1968. *A Materialist Theory of the Mind* (New York: Humanities Press)

Assanand, Shazma, John P. J. Pinel, and D. Ross Lehman. 1998. 'Personal Theories of Hunger and Eating', *Journal of Applied Social Psychology*, 28.11: 998–1015 <https://doi.org/10.1111/j.1559-1816.1998.tb01664.x\>

Augustine, Vance, Soohyun Lee, and Yuki Oka. 2020. 'Neural Control and Modulation of Thirst, Sodium Appetite, and Hunger', *Cell*, 180.1: 25–32 <https://doi.org/10.1016/j.cell.2019.11.040>

Aydede, Murat. 2001. 'Naturalism, Introspection, and Direct Realism about Pain', *Consciousness and Emotion*, 2.1: 29–73 <https://doi.org/10.1075/ce.2.1.03ayd>

—— 2009. 'Is Feeling Pain the Perception of Something?', *Journal of Philosophy*, 106.10: 531–67 <https://doi.org/10.5840/jphil20091061033>

—— 2023. 'Is Pain Representational?', *Belgrade Philosophical Annual*, 36.2: 25–39 <https://doi.org/10.5937/BPA2302025A>

Aydede, Murat, and Matthew Fulkerson. 2014. 'Affect: Representationalists' Headache', *Philosophical Studies*, 170.2: 175–98 <https://doi.org/10.1007/s11098-013-0206-7>

Aziz, Qasim, and James K. Ruffle. 2019. 'The Cutaneous Borders of Interoception: Active and Social Inference of Pain and Pleasure on the Skin', in *The Interoceptive Mind: From Homeostasis to Awareness*, ed. by Manos Tsakiris and Helena De Preester (Oxford: Oxford University Press), pp. 138–159.

Bain, David. 2003. 'Intentionalism and Pain', *The Philosophical Quarterly*, 53.213: 502–23 <https://doi.org/10.1111/1467-9213.00328>

—— 2007. 'The Location of Pains', *Philosophical Papers*, 36.2: 171–205 <https://doi.org/10.1080/05568640709485198>

—— 2011. 'The Imperative View of Pain', *Journal of Consciousness Studies*, 18.9–10: 164–85

—— 2013. 'What Makes Pains Unpleasant?', *Philosophical Studies*, 166.1: 69–89< https://doi.org/10.1007/s11098-012-0030-9>

—— 2014. 'Pains that Don't Hurt', *Australasian Journal of Philosophy*, 92.2: 305–20 <https://doi.org/10.1080/00048402.2013.822399>

Baliki, Marwan N., and A. Vania Apkarian. 2015. 'Nociception, Pain, Negative Moods, and Behavior Selection', *Neuron*, 87.3: 474–91 <https://doi.org/10.1016/j.neuron.2015.06.005>

Balli, Selin, Shobhit Sharan, and Kelsey R. Shumway. 2023. 'Physiology, Fever', *StatPearls* (Treasure Island, FL: StatPearls Publishing) <https://www.ncbi.nlm.nih.gov/books/NBK562334/>

Barlassina, Luca, and Max K. Hayward. 2019. 'More of Me! Less of Me!: Reflexive Imperativism about Affective Phenomenal Character', *Mind*, 128.512: 1013–44 <https://doi.org/10.1093/mind/fzz035>

Barrett, Lisa Feldman, and W. Kyle Simmons. 2015. 'Interoceptive Predictions in the Brain', *Nature Reviews Neuroscience*, 16.7: 419–29 <https://doi.org/10.1038/nrn3950>

Barsky, Arthur J., and Jonathan F. Borus. 1999. 'Functional Somatic Syndromes', *Annals of Internal Medicine*, 130.11: 910–21 <https://doi.org/10.7326/0003-4819-130-11-199906010-00016>

Bauer, Michael. 2020. 'The Explanatory Breadth of Pushmi-Pullyu Representations', *Biology and Philosophy*, 35.3: 35 <https://doi.org/10.1007/s10539-020-09751-5>

Bayne, Tim. 2011. 'Agentive Experiences as Pushmi-Pullyu Representations', in *New Waves in Philosophy of Action*, ed. by Jesús H. Aguilar, Andrei A. Buckareff, and Keith Frankish (London: Palgrave Macmillan), pp. 219–36 <https://doi.org/10.1057/9780230304253_11>

Beecher, Henry K. 1946. 'Pain in Men Wounded in Battle', *Annals of Surgery*, 123.1: 96–105 <https://doi.org/10.1097/00000658-194601000-00012>

Beld, John Thomas, Kemuel Philbrick, and Teresa Rummans. 1994. 'Periodic Catatonia', in *Catatonia: From Psychopathology to Neurobiology*, ed. by Stanley N. Caroff, Stephan C. Mann, Andrew Francis, and Gregory L. Fricchione (Washington, DC: American Psychiatric Association Publishing), pp. 93–104

Ben-Tovim, David I., and Morris S. Schwartz. 1981. 'Hypoalgesia in Depressive Illness', *The British Journal of Psychiatry*, 138.1: 37–39 <https://doi.org/10.1192/bjp.138.1.37>

Berntson, Gary G., and Sahib S. Khalsa. 2021. 'Neural Circuits of Interoception', *Trends in Neurosciences*, 44.1: 17–28 <https://doi.org/10.1016/j.tins.2020.09.011>

Berridge, Kent C. 2004. 'Motivation Concepts in Behavioral Neuroscience', *Physiology and Behavior*, 81.2: 179–209 <https://doi.org/10.1016/j.physbeh.2004.02.004>

Berridge, Kent C., and Morten L. Kringelbach. 2015. 'Pleasure Systems in the Brain', *Neuron*, 86.3: 646–64 <https://doi.org/10.1016/j.neuron.2015.02.018>

Betley, J. Nicholas, Shijun Xu, Zhen F. H. Cao, Rui Gong, Christopher J. Magnus, Yuanyue Yu, and Scott M. Sternson. 2015. 'Neurons for Hunger and Thirst Transmit a Negative-Valence Teaching Signal', *Nature*, 521.7551: 180–85 <https://doi.org/10.1038/nature14416>

Billman, George E. 2020. 'Homeostasis: The Underappreciated and Far Too Often Ignored Central Organizing Principle of Physiology', *Frontiers in Physiology*, 11: 200 <https://doi.org/10.3389/fphys.2020.00200>

Block, Ned. 2006. 'Bodily Sensations as an Obstacle for Representationism', in *Pain: New Essays on Its Nature and the Methodology of Its Study*, ed. by Murat Aydede (Cambridge, MA: MIT Press), pp. 137–42

Bouillon-Minois, Jean-Baptiste, Marion Trousselard, David Thivel, Brett A. Gordon, Jonathan Schmidt, Farah Moustafa, Claire Oris, and François Dutheil. 2021. 'Ghrelin as a Biomarker of Stress: A Systematic Review and Meta-Analysis', *Nutrients*, 13.3: 784 <https://doi.org/10.3390/nu13030784>

Bragdon, Lindsey B., Joseph A. Nota, Gloria K. Eng, Nicholas Recchia, Polina Kravets, Kathryn A. Collins, and Emily R. Stern. 2023. 'Failures of Urge Suppression in Obsessive-Compulsive Disorder: Behavioral Modeling Using a Blink Suppression Task', *Journal of Obsessive-Compulsive and Related Disorders*, 38: 100824 <https://doi.org/10.1016/j.jocrd.2023.100824>

Brener, Julian, and Christopher Ring. 2016. 'Towards a Psychophysics of Interoceptive Processes: The Measurement of Heartbeat Detection', *Philosophical Transactions of the Royal Society B: Biological Sciences*, 371.1708: 20160015 <https://doi.org/10.1098/rstb.2016.0015>

Brunstrom, J. M., P. M. Tribbeck, and A. W. MacRae. 2000. 'The Role of Mouth State in the Termination of Drinking Behavior in Humans', *Appetite*, 34, pp. 197–207.

Byrne, Alex. 2001. 'Intentionalism Defended', *Philosophical Review*, 110.2, 199–240 <https://doi.org/10.1215/00318108-110-2-199>

Cameron, Oliver G. 2002. *Visceral Sensory Neuroscience: Interoception* (New York: Oxford University Press)

Carland, Matthew A., David Thura, and Paul Cisek. 2019. 'The Urge to Decide and Act: Implications for Brain Function and Dysfunction', *The Neuroscientist*, 25.5: 491–511 <https://doi.org/10.1177/1073858419841553>

Carruthers, Peter. 2017. 'Valence and Value', *Philosophy and Phenomenological Research*, 97.3: 658–80 <https://doi.org/10.1111/phpr.12395>

—— 2023. 'On Valence: Imperative or Representation of Value?', *The British Journal for the Philosophy of Science*, 74.3: 533–53 <https://doi.org/10.1086/714985>

Carvalho, Gil B., and Antonio Damasio. 2021. 'Interoception and the Origin of Feelings: A New Synthesis', *BioEssays*, 43.6: 2000261 <https://doi.org/10.1002/bies.202000261>

Casser, Laurenz C. 2021. 'The Function of Pain', *Australasian Journal of Philosophy*, 99.2: 364–78 <https://doi.org/10.1080/00048402.2020.1735459>

Ceunen, Ellen, Johan W. S. Vlaeyen, and Ilse Van Diest. 2016. 'On the Origin of Interoception', *Frontiers in Psychology*, 7: article 743 <https://doi.org/10.3389/fpsyg.2016.00743>

Chen, Wendy G., Donald Schloesser, Ann M. Arensdorf, J. M. Simmons, Chunling Cui, Rita Valentino, James W. Gnadt, L. Nielsen, C. St. Hillaire-Clarke, V. Spruance, T. S. Horowitz, Y. F. Vallejo, and Helene M. Langevin. 2021. 'The Emerging Science of Interoception: Sensing, Integrating, Interpreting, and Regulating Signals within the Self', *Trends in Neurosciences*, 44.2: 153–66 <https://doi.org/10.1016/j.tins.2020.10.007>

Clauw, Daniel J. 2015. 'Fibromyalgia and Related Conditions', *Mayo Clinic Proceedings*, 90.5: 680–92 <https://doi.org/10.1016/j.mayocp.2015.03.014>

Constantinou, Eleni, Katrien Bogaerts, Ilse Van Diest, and Omer Van den Bergh. 2013. 'Inducing Symptoms in High-Symptom Reporters via Emotional Pictures: The Interactive Effects of Valence and Arousal', *Journal of Psychosomatic Research*, 74.3: 191–96 <https://doi.org/10.1016/j.jpsychores.2012.12.015>

Cornman, J. W. 1977. 'Might a Tooth Ache but There Be No Toothache?', *Australasian Journal of Philosophy*, 55.1: 27–40 <https://doi.org/10.1080/00048407712341021>

Corns, Jessica. 2014a. 'The Inadequacy of Unitary Characterizations of Pain', *Philosophical Studies*, 169.3: 355–78 <https://doi.org/10.1007/s11098-013-0186-7>

—— 2014b. 'Unpleasantness, Motivational Oomph, and Painfulness', *Mind & Language*, 29.2: 238–54 <https://doi.org/10.1111/mila.12048>

—— 2018. 'Recent Work on Pain', *Analysis*, 78.4: 737–53 <https://doi.org/10.1093/analys/any055>

Craig, Arthur D. 1996. 'An Ascending General Homeostatic Afferent Pathway Originating in Lamina I', in *Progress in Brain Research*, Vol. 107, ed. by G. Holstege, R. Bandler, and C. B. Saper (Amsterdam: Elsevier), pp. 225–42 <https://doi.org/10.1016/S0079-6123(08)61867-1>

—— 2002. 'How Do You Feel? Interoception: The Sense of the Physiological Condition of the Body', *Nature Reviews Neuroscience*, 3.8: 655–66 <https://doi.org/10.1038/nrn894>

—— 2003. 'A New View of Pain as a Homeostatic Emotion', *Trends in Neurosciences*, 26.6: 303–07 <https://doi.org/10.1016/S0166-2236(03)00123-1>

—— 2008. 'Interoception and Emotion: A Neuroanatomical Perspective', in *Handbook of Emotions*, 3rd edn, ed. by Michael Lewis, Jeannette M. Haviland-Jones, and Lisa Feldman Barrett (New York: Guilford Press), pp. 272–92

—— 2015. *How Do You Feel? An Interoceptive Moment with Your Neurobiological Self* (Princeton NJ: Princeton University Press)

Critchley, Hugo D., and N. A. Harrison. 2013. 'Visceral Influences on Brain and Behavior', *Neuron*, 77.4: 624–38 <https://doi.org/10.1016/j.neuron.2013.02.008>

Cutter, Brian. 2017. 'Pain and Representation', in *The Routledge Handbook of Philosophy of Pain*, ed. by Jessica Corns (Abingdon: Routledge), pp. 290–309

—— and Michael Tye. 2011. 'Tracking Representationalism and the Painfulness of Pain', *Philosophical Issues*, 21.1: 90–109 <https://doi.org/10.1111/j.1533-6077.2011.00199.x>

—— and Michael Tye. 2014. 'Pains and Reasons: Why Is It Rational to Kill the Messenger?', *The Philosophical Quarterly*, 64.256: 423–33 <https://doi.org/10.1093/pq/pqu025>

Davenport, Paul W. 2008. 'Urge-to-Cough: What Can It Teach Us about Cough?', *Lung*, 186.Suppl. 1: S107–11 <https://doi.org/10.1007/s00408-007-9045-7>

Deem, Jeffrey D., Christine L. Faber, and Gregory J. Morton. 2022. 'AgRP Neurons: Regulators of Feeding, Energy Expenditure, and Behavior', *The FEBS Journal*, 289.8: 2362–81 <https://doi.org/10.1111/febs.16176>

de Vignemont, Frédérique. 2015. 'Pain and Bodily Care: Whose Body Matters?', *Australasian Journal of Philosophy*, 93.3: 542–60.

de Vignemont, Frédérique. 2018. *Mind the Body: An Exploration of Bodily Self-Awareness* (Oxford: Oxford University Press)

Di Lernia, Daniele, Silvia Serino, and Giuseppe Riva. 2016. 'Pain in the Body: Altered Interoception in Chronic Pain Conditions: A Systematic Review', *Neuroscience & Biobehavioral Reviews*, 71: 328–41 <https://doi.org/10.1016/j.neubiorev.2016.09.015>

Dickens, Chris, Linda McGowan, and Susan Dale. 2003. 'Impact of Depression on Experimental Pain Perception: A Systematic Review of the Literature with Meta-Analysis', *Psychosomatic Medicine*, 65.3: 369–75 <https://doi.org/10.1097/01.psy.0000041622.69462.06>

Dretske, Fred. 1995. *Naturalizing the Mind* (Cambridge, MA: MIT Press)

Dretske, Fred. 2003. 'Experience as Representation', *Philosophical Issues,* 13.1, 67–82 <https://doi.org/10.1111/1533-6077.00005>

Dubin, Adrienne E., and Ardem Patapoutian. 2010. 'Nociceptors: The Sensors of the Pain Pathway', *Journal of Clinical Investigation*, 120, 3760–72 <https://doi.org/10.1172/JCI42843>

Engelen, Tahnée, Marco Solcà, and Catherine Tallon-Baudry. 2023. 'Interoceptive Rhythms in the Brain', *Nature Neuroscience*, 26, 1670–84 <https://doi.org/10.1038/s41593-023-01425-1>

Feldman, Mallory J., Eliza Bliss-Moreau, and Kristen A. Lindquist. 2024. 'The Neurobiology of Interoception and Affect', *Trends in Cognitive Sciences*, 28, 643–61 <https://doi.org/10.1016/j.tics.2024.01.009>

Ferrão, Ygor Arzeno, and others. 2012. 'Sensory Phenomena Associated with Repetitive Behaviors in Obsessive-Compulsive Disorder: An Exploratory Study of 1,001 Patients', *Psychiatry Research*, 197.3: 253–58 <https://doi.org/10.1016/j.psychres.2011.09.017>

Finlay, Barbara L., and Supriya Syal. 2014. 'The Pain of Altruism', *Trends in Cognitive Sciences*, 18, 615–17 <https://doi.org/10.1016/j.tics.2014.08.002>

Fulkerson, Matthew. 2023. 'How Thirst Compels: An Aggregation Model of Sensory Motivation', *Mind & Language*, 38, 141–55 <https://doi.org/10.1111/mila.12369. Wiley Online Library>

—— 2014. 'Rethinking the Senses and Their Interactions: The Case for Sensory Pluralism', *Frontiers in Psychology*, 5, 1426 <https://doi.org/10.3389/fpsyg.2014.01426.>

Garfinkel, Sarah N., Hugo D. Critchley, and Olga Pollatos. 2017. 'The Interoceptive System: Implications for Cognition, Emotion, and Health', in *Handbook of Psychophysiology*, 4th edn, ed. by John T. Cacioppo, Louis G. Tassinary, and Gary G. Berntson, pp. 427–43 (Cambridge: Cambridge University Press)

Graham, George, and G. L. Stephens. 1985. 'Are Qualia a Pain in the Neck for Functionalists?', *American Philosophical Quarterly*, 22.1: 73–80

Grahek, Nikola. 2007. *Feeling Pain and Being in Pain*, 2nd edn (Cambridge, MA: MIT Press)

Griffiths, Paul E. 2004. 'Emotions as Natural and Normative Kinds', *Philosophy of Science*, 71.5: 901–11 <https://doi.org/10.1086/425944>

Grüter, Christian, and William M. Farina. 2009. 'The Honeybee Waggle Dance: Can We Follow the Steps?', *Trends in Ecology & Evolution*, 24.5: 242–47< https://doi.org/10.1016/j.tree.2008.12.007>

Hall, Robert J. 2008. 'If It Itches, Scratch!', *Australasian Journal of Philosophy*, 86.4: 525–35 <https://doi.org/10.1080/00048400802346813>

Harshaw, Christopher. 2008. 'Alimentary Epigenetics: A Developmental Psychobiological Systems View of the Perception of Hunger, Thirst, and Satiety', *Developmental Review*, 28.4: 541–69 <https://doi.org/10.1016/j.dr.2008.08.001>

Harte, S. E., R. E. Harris, and D. J. Clauw. 2018. 'The Neurobiology of Central Sensitization', *Journal of Applied Biobehavioral Research*, 23: e12137 <https://doi.org/10.1111/jabr.12137\>

Houghton, Danielle C., Margaret R. Capriotti, Cara A. Conelea, and Douglas W. Woods. 2014. 'Sensory Phenomena in Tourette Syndrome: Their Role in Symptom Formation and Treatment', *Current Developmental Disorders Reports*, 1.4: 245–51 <https://doi.org/10.1007/s40474-014-0026-2>

Inenaga, K., and K. Ono. 2010. 'Oral Dryness and Thirst', *Journal of Oral Biosciences*, 52.4: 344–51 <https://doi.org/10.1016/S1349-0079(10)80015-5>

Jackson, Stephen R., A. Parkinson, So-Young Kim, Martin Schüermann, and Simon B. Eickhoff. 2011. 'On the Functional Anatomy of the Urge-for-Action', *Cognitive Neuroscience*, 2.3–4: 227–43 <https://doi.org/10.1080/17588928.2011.604717>

Kamen, C., M. A. Tejani, K. Chandwani, M. Janelsins, A. R. Peoples, J. A. Roscoe, and G. R. Morrow. 2014. 'Anticipatory Nausea and Vomiting due to Chemotherapy', *European Journal of Pharmacology*, 722: 172–79< https://doi.org/10.1016/j.ejphar.2013.09.071>

Khalsa, S. S., and R. C. Lapidus. 2016. 'Can Interoception Improve the Pragmatic Search for Biomarkers in Psychiatry?', *Frontiers in Psychiatry*, 7: 121 <https://doi.org/10.3389/fpsyt.2016.00121>

Khalsa, S. S., R. Adolphs, O. G. Cameron, H. D. Critchley, P. W. Davenport, J. S. Feinstein, J. D. Feusner, S. N. Garfinkel, R. D. Lane, W. E. Mehling, A. E. Meuret, C. B. Nemeroff, S. Oppenheimer, F. H. Petzschner, O. Pollatos, J. L. Rhudy, L. P. Schramm, W. K. Simmons, M. B. Stein, … N. Zucker. 2018. 'Interoception and Mental Health: A Roadmap', *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, 3.6: 501–13 <https://doi.org/10.1016/j.bpsc.2017.12.004>

Khalsa, S. S., D. Rudrauf, J. S. Feinstein, and D. Tranel. 2009. 'The Pathways of Interoceptive Awareness', *Nature Neuroscience*, 12.12: 1494–96 <https://doi.org/10.1038/nn.2411>

Kim, M., and A. Schachner. 2023. 'The Origins of Dance: Characterizing the Development of Infants' Earliest Dance Behavior', *Developmental Psychology*, 59.4: 691–706 <https://doi.org/10.1037/dev0001436>

Klein, Colin. 2007. 'An Imperative Theory of Pain', *Journal of Philosophy*, 104.10: 517–32 <https://doi.org/10.5840/jphil2007104104>

—— 2010. 'Response to Tumulty on Pain and Imperatives', *Journal of Philosophy*, 107.10: 554–57 <https://doi.org/10.5840/jphil20101071038>

—— 2012. 'Imperatives, Phantom Pains, and Hallucination by Presupposition', *Philosophical Psychology*, 25.6: 917–28 <https://doi.org/10.1080/09515089.2011.625121>

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

Koch, K. L. 2017. 'Physiology of Nausea', in *Nausea and Vomiting: Diagnosis and Treatment*, ed. by K. L. Koch and W. L. Hasler, pp. 1–13 (Cham: Springer International Publishing)

Köteles, F. 2021. *Body Sensations: The Conscious Aspects of Interoception* (Cham: Springer International Publishing)

Lagomarsino, Valentina N., Aleksandar D. Kostic, and Isaac M. Chiu. 2020. 'Mechanisms of Microbial-Neuronal Interactions in Pain and Nociception', *Neurobiology of Pain*, 9: 100056 <https://doi.org/10.1016/j.ynpai.2020.100056>

Landrieu, P., G. Said, and C. Allaire. 1990. 'Dominantly Transmitted Congenital Indifference to Pain', *Annals of Neurology*, 27.5, 574–78 <https://doi.org/10.1002/ana.410270520>

Legrain, Valéry, G. D. Iannetti, Luc Plaghki, and André Mouraux. 2011. 'The Pain Matrix Reloaded', *Progress in Neurobiology*, 93.1: 111–24 <https://doi.org/10.1016/j.pneurobio.2010.10.005>

Levy, Daniel J., and Paul W. Glimcher. 2012. 'The Root of All Value: A Neural Common Currency for Choice', *Current Opinion in Neurobiology*, 22.6: 1027–38 <https://doi.org/10.1016/j.conb.2012.06.001>

Livneh, Yoav, and Mark L. Andermann. 2021. 'Cellular Activity in Insular Cortex across Seconds to Hours: Sensations and Predictions of Bodily States', *Neuron*, 109.22: 3576–93 <https://doi.org/10.1016/j.neuron.2021.08.036>

Lowell, Bradford B. 2019. 'New Neuroscience of Homeostasis and Drives for Food, Water, and Salt', *The New England Journal of Medicine*, 380.5: 459–71 <https://doi.org/10.1056/NEJMra1812053>

Lumpkin, Ellen A., and Michael J. Caterina. 2007. 'Mechanisms of Sensory Transduction in the Skin', *Nature*, 445.7130: 858–65 <https://doi.org/10.1038/nature05662>

Lycan, William G., *Consciousness and Experience* (Cambridge, MA: MIT Press, 1996)

Macpherson, Fiona. 2011. 'Introduction: Individuating the Senses', in *The Senses: Classic and Contemporary Philosophical Perspectives*, ed. by Fiona Macpherson, pp. 3–43 (Oxford: Oxford University Press)

Martínez, Manolo. 2011. 'Imperative Content and the Painfulness of Pain', *Phenomenology and the Cognitive Sciences*, 10.1: 67–90 <https://doi.org/10.1007/s11097-010-9172-0>

—— 2022. 'Imperative Transparency', *Mind*, 131.522: 585–601< https://doi.org/10.1093/mind/fzab009>

Martínez, Manolo, and Colin Klein. 2016. 'Pain Signals Are Predominantly Imperative', *Biology & Philosophy*, 31.2: 283–98< https://doi.org/10.1007/s10539-015-9514-y>

Matthen, Mohan. 2024. 'Hunger, Homeostasis, and Desire', *Mind & Language*, 39.3: 397–414 <https://doi.org/10.1111/mila.12496>

Maund, Barry. 2006. 'Michael Tye on Pain and Representational Content', in *Pain: New Essays on Its Nature and the Methodology of Its Study*, ed. by Murat Aydede, pp. 295–312 (Cambridge, MA: MIT Press)

McGinn, Colin. 1996. *The Character of Mind,* 2nd edn. (Oxford: Oxford University Press)

Melzack, Ronald. 1999. 'From the Gate to the Neuromatrix', *Pain*, 82.Suppl. 1: S121–26 <https://doi.org/10.1016/S0304-3959(99)00145-1>

Melzack, Ronald, and Patrick D. Wall. 1996. *The Challenge of Pain: A Modern Medical Classic*, 2nd edn (London: Penguin Books)

Melzack, Ronald, Patrick D. Wall, and T. C. Ty. 1982. 'Acute Pain in an Emergency Clinic: Latency of Onset and Descriptor Patterns Related to Different Injuries', *Pain*, 14.1: 33–43

Millikan, Ruth Garrett. 1984. *Language, Thought, and Other Biological Categories: New Foundations for Realism*(Cambridge, MA: MIT Press)

—— 1989. 'Biosemantics', *Journal of Philosophy*, 86.6: 281–97 <https://doi.org/10.2307/2027123>

—— 1995. 'Pushmi-Pullyu Representations', *Philosophical Perspectives*, 9: 185–200 <https://doi.org/10.2307/2214217>

—— 2004. *Varieties of Meaning: The 2002 Jean Nicod Lectures* (Cambridge, MA: MIT Press)

—— 2023. 'Comment on Artiga's 'Teleosemantics and Pushmi-Pullyu Representations'', *Erkenntnis*, 88.1: 409–17< https://doi.org/10.1007/s10670-020-00354-w>

Modell, Harold, William Cliff, Joel Michael, James McFarland, Mary Pat Wenderoth, and Adam Wright. 2015. 'A Physiologist's View of Homeostasis', *Advances in Physiology Education*, 39.4: 259–66 <https://doi.org/10.1152/advan.00107.2015>

Muraven, Mark, and Roy F. Baumeister. 2000. 'Self-Regulation and Depletion of Limited Resources: Does Self-Control Resemble a Muscle?', *Psychological Bulletin*, 126.2: 247–59 <https://doi.org/10.1037/0033-2909.126.2.247>

Murphy, Jennifer, Rebecca Brewer, Caroline Catmur, and Geoffrey Bird. 2017. 'Interoception and Psychopathology: A Developmental Neuroscience Perspective', *Developmental Cognitive Neuroscience*, 23: 45–56 <https://doi.org/10.1016/j.dcn.2016.12.006>

Myers, Martin G., Jr., Ann H. Affinati, Nicholas Richardson, and Michael W. Schwartz. 2021. 'Central Nervous System Regulation of Organismal Energy and Glucose Homeostasis', *Nature Metabolism*, 3.6: 737–50 <https://doi.org/10.1038/s42255-021-00408-5>

Nanay, Bence. 2017. 'Pain and Mental Imagery', *Monist*, 100.4: 485–500 <https://doi.org/10.1093/monist/onx024>

Napadow, Vitaly, Jason D. Sheehan, Jieun Kim, Lauren T. Lacount, Kyungmo Park, Ted J. Kaptchuk, Bruce R. Rosen, and Braden Kuo. 2013. 'The Brain Circuitry Underlying the Temporal Evolution of Nausea in Humans', *Cerebral Cortex*, 23.4: 806–13 <https://doi.org/10.1093/cercor/bhs073>

Northoff, Georg. 2002. 'What Catatonia Can Tell Us about 'Top-Down Modulation': A Neuropsychiatric Hypothesis', *Behavioral and Brain Sciences*, 25.5: 555–604 <https://doi.org/10.1017/S0140525X02000109>

Oppenheimer, Stephen M., and David F. Cechetto. 2016. 'The Insular Cortex and the Regulation of Cardiac Function', *Comprehensive Physiology*, 6.2: 1081–1105 <https://doi.org/10.1002/cphy.c140076>

Ossipov, Michael H., Gregory O. Dussor, and Frank Porreca. 2010. 'Central Modulation of Pain', *Journal of Clinical Investigation*, 120.11: 3779–85 https://doi.org/10.1172/JCI43766>

Pacherie, Elisabeth. 2000. 'The Content of Intentions', *Mind & Language*, 15.4: 400–32 <https://doi.org/10.1111/1468-0017.00142>

Papineau, David. 2003. *The Roots of Reason* (Oxford: Oxford University Press)

Paulus, Martin P., Justin S. Feinstein, and Sahib S. Khalsa. 2019. 'An Active Inference Approach to Interoceptive Psychopathology', *Annual Review of Clinical Psychology*, 15: 97–122 <https://doi.org/10.1146/annurev-clinpsy-050718-095617>

Petzschner, Frederike H., Sarah N. Garfinkel, Martin P. Paulus, Christof Koch, and Sahib S. Khalsa. 2021. 'Computational Models of Interoception and Body Regulation', *Trends in Neurosciences*, 44.1: 63–76 <https://doi.org/10.1016/j.tins.2020.09.012>

Pezzulo, Giovanni, Francesco Rigoli, and Karl J. Friston. 2015. 'Active Inference, Homeostatic Regulation and Adaptive Behavioural Control', *Progress in Neurobiology*, 134: 17–35 <https://doi.org/10.1016/j.pneurobio.2015.09.001>

Pitcher, George. 1970. 'The Awfulness of Pain', *Journal of Philosophy*, 67.14: 481–92 <https://doi.org/10.2307/2024109>

Quadt, Lisa, Hugo D. Critchley, and Sarah N. Garfinkel. 2018. 'The Neurobiology of Interoception in Health and Disease', *Annals of the New York Academy of Sciences*, 1428.1: 112–28 <https://doi.org/10.1111/nyas.13915>

Quigley, Eamonn M., William L. Hasler, and Henry P. Parkman. 2001. 'AGA Technical Review on Nausea and Vomiting', *Gastroenterology*, 120.1: 263–86 <https://doi.org/10.1053/gast.2001.20516>

Quigley, Kathleen S., Scott Kanoski, Warren M. Grill, Lisa Feldman Barrett, and Manos Tsakiris. 2021. 'Functions of Interoception: From Energy Regulation to Experience of the Self', *Trends in Neurosciences*, 44.1: 29–38 <https://doi.org/10.1016/j.tins.2020.09.008>

Radjavi, Azar. 2019. 'The Itch-Scratch Reflex Generates Protective Immunity', *bioRxiv*: 808477 <https://doi.org/10.1101/808477>

Ramsay, Douglas S., and Stephen C. Woods. 2016. 'Physiological Regulation: How It Really Works', *Cell Metabolism*, 24.3: 361–64 <https://doi.org/10.1016/j.cmet.2016.08.004>

Reich, Adam, Sonja Ständer, and Jacek C. Szepietowski. 2009. 'Drug-Induced Pruritus: A Review', *Acta Dermato-Venereologica*, 89: 236–44

Rescorla, Michael. 2012. 'Millikan on Honeybee Navigation and Communication', in *Millikan and Her Critics*, 1st edn, ed. by David Ryder, James Kingsbury, and Kenneth Williford, 87–106 (Chichester: Wiley) <https://doi.org/10.1002/9781118328118.ch4>

Ritchie, J. Brendan, and Peter Carruthers. 2015. 'The Bodily Senses', in *The Oxford Handbook of Philosophy of Perception*, ed. by Mohan Matthen, pp. 353–70 (Oxford: Oxford University Press)

Röhricht, Frank, and Stefan Priebe. 2002. 'Do Cenesthesias and Body Image Aberration Characterize a Subgroup in Schizophrenia?', *Acta Psychiatrica Scandinavica*, 105.4: 276–82 <https://doi.org/10.1034/j.1600-0447.2002.1107.x>

Schilder, Paul F., and Erwin Stengel. 1931. 'Asymbolia for Pain', *Archives of Neurology & Psychiatry*, 25: 598–600

Schwoebel, John, H. Branch Coslett, Judith Bradt, Rhonda Friedman, and Cheryl Dileo. 2002. 'Pain and the Body Schema: Effects of Pain Severity on Mental Representations of Movement', *Neurology*, 59.5: 775–77 <https://doi.org/10.1212/wnl.59.5.775>

Serrahima, Clara, and Manolo Martínez. 2023. 'The Experience of Dysmenorrhea', *Synthese*, 201.5: 173 <https://doi.org/10.1007/s11229-023-04148-9>

Seth, Anil K. 2013. 'Interoceptive Inference, Emotion, and the Embodied Self', *Trends in Cognitive Sciences*, 17.11: 565–73 <https://doi.org/10.1016/j.tics.2013.09.007>

Sharma, Prabha, Norman R. Morris, and Laurence Adams. 2016. 'Effect of Experimental Modulation of Mood on Perception of Exertional Dyspnea in Healthy Subjects', *Journal of Applied Physiology*, 120.2: 114–20 <https://doi.org/10.1152/japplphysiol.00122.2015>

Shea, Nicholas. 2018. *Representation in Cognitive Science* (Oxford: Oxford University Press)

Shevlin, Hugh, and Patrick Friesen. 2021. 'Pain, Placebo, and Cognitive Penetration', *Mind & Language*, 36.5: 771–91 <https://doi.org/10.1111/mila.12292>

Sierra, Mauricio. 2009. *Depersonalization: A New Look at a Neglected Syndrome* (Cambridge: Cambridge University Press)

Sierra, Mauricio, and German E. Berrios. 1997. 'Depersonalization: A Conceptual History', *History of Psychiatry*, 8.30: 213–29 <https://doi.org/10.1177/0957154X9700803002>

Simeon, Daphne, and Jeffrey Abugel. 2009. *Feeling Unreal: Depersonalization Disorder and the Loss of the Self* (Oxford: Oxford University Press)

Skvortsova, Anastasia, Daisy S. Veldhuijzen, I. E. M. Kloosterman, Gabriel Pacheco-López, and Andrea W. M. Evers. 2021. 'Food Anticipatory Hormonal Responses: A Systematic Review of Animal and Human Studies', *Neuroscience & Biobehavioral Reviews*, 126: 447–64 <https://doi.org/10.1016/j.neubiorev.2021.03.030>

Stampe, Dennis W. 1988. 'Need', *Australasian Journal of Philosophy*, 66.2: 129–60 <https://doi.org/10.1080/00048408812343251>

Stegen, Kristien, Katleen De Bruyne, Wim Rasschaert, Karel P. Van de Woestijne, and Omer Van den Bergh. 1999. 'Fear-Relevant Images as Conditioned Stimuli for Somatic Complaints, Respiratory Behavior, and Reduced End-Tidal pCO_2', *Journal of Abnormal Psychology*, 108.1: 143–52 <https://doi.org/10.1037/0021-843X.108.1.143>

Steinkopf, Lutz. 2016. 'An Evolutionary Perspective on Pain Communication', *Evolutionary Psychology*, 14.2: 1474704916653964 <https://doi.org/10.1177/1474704916653964>

Sterelny, Kim. 2003. *Thought in a Hostile World: The Evolution of Human Cognition* (Oxford: Blackwell)

Stevenson, Richard J. 2024. 'The Psychological Basis of Hunger and Its Dysfunctions', *Nutrition Reviews*, 82.10: 1444–54 <https://doi.org/10.1093/nutrit/nuad092>

Strigo, Irina A., M. Catherine Bushnell, Mélanie Boivin, and Gilles H. Duncan. 2002. 'Psychophysical Analysis of Visceral and Cutaneous Pain in Human Subjects', *Pain*, 97.3: 235–46 <https://doi.org/10.1016/S0304-3959(02)00023-4>

Strigo, Irina A., and A. D. (Bud) Craig. 2016. 'Interoception, Homeostatic Emotions and Sympathovagal Balance', *Philosophical Transactions of the Royal Society B: Biological Sciences*, 371(1708): 20160010.

Sundby, Kristine K., Julia Wagner, and Adam R. Aron. 2019. 'The Functional Role of Response Suppression during an Urge to Relieve Pain', *Journal of Cognitive Neuroscience*, 31.9: 1404–21 <https://doi.org/10.1162/jocn_a_01423>

Taggart, Peter, Hugo Critchley, Suzanne van Duijvendoden, and Pier D. Lambiase. 2016. 'Significance of Neuro-Cardiac Control Mechanisms Governed by Higher Regions of the Brain', *Autonomic Neuroscience*, 199: 54–65 <https://doi.org/10.1016/j.autneu.2016.08.013>

Talbot, J. D., J. G. Villemure, M. Catherine Bushnell, and Gilles H. Duncan. 1995. 'Evaluation of Pain Perception after Anterior Capsulotomy: A Case Report', *Somatosensory & Motor Research*, 12.2: 115–26 <https://doi.org/10.3109/08990229509101503>

Thirst. [n.d.]. NHS Inform. Retrieved 5 February 2025, from <https://www.nhsinform.scot/illnesses-and-conditions/nutritional/thirst/>

Thomaidou, M. A., K. J. Peerdeman, M. I. Koppeschaar, Andrea W. M. Evers, and Daisy S. Veldhuijzen. 2021. 'How Negative Experience Influences the Brain: A Comprehensive Review of the Neurobiological Underpinnings of Nocebo Hyperalgesia', *Frontiers in Neuroscience*, 15: 652552 <https://doi.org/10.3389/fnins.2021.652552>

Toussaint, Bastien, Jan Heinzle, and Klaas E. Stephan. 2024. 'A Computationally Informed Distinction of Interoception and Exteroception', *Neuroscience & Biobehavioral Reviews*, 159: 105608 <https://doi.org/10.1016/j.neubiorev.2024.105608>

Tracey, Irene. 2022. 'Why Pain Hurts', *Trends in Cognitive Sciences*, 26.12: 1070–72 <https://doi.org/10.1016/j.tics.2022.09.020>

Tracey, Irene, and Patrick W. Mantyh. 2007. 'The Cerebral Signature for Pain Perception and Its Modulation', *Neuron*, 55.3: 377–91 <https://doi.org/10.1016/j.neuron.2007.07.012>

Tsakiris, Manos, and Critchley, Hugo D. 2016. 'Interoception beyond Homeostasis: Affect, Cognition and Mental Health', *Philosophical Transactions of the Royal Society B: Biological Sciences*, 371.1708: 20160002 <https://doi.org/10.1098/rstb.2016.0002>

Tumulty, Moira. 2009. 'Pains, Imperatives, and Intentionalism', *Journal of Philosophy*, 106.3: 161–66 <https://doi.org/10.5840/jphil200910635>

Tye, Kay M. 2018. 'Neural Circuit Motifs in Valence Processing', *Neuron*, 100.2: 436–52 <https://doi.org/10.1016/j.neuron.2018.10.001>

Tye, Michael. 1995a. 'A Representational Theory of Pains and Their Phenomenal Character', *Philosophical Perspectives*, 9: 223–39 <https://doi.org/10.2307/2214219>

—— 1995a. *Ten Problems of Consciousness: A Representational Theory of the Phenomenal Mind* (Cambridge, MA: MIT Press)

—— 2006. 'Another Look at Representationalism about Pain', in *Pain: New Essays on Its Nature and the Methodology of Its Study*, ed. by Murat Aydede, pp. 99–120 (Cambridge, MA: MIT Press)

Vance, E. Bruce, and Norman N. Wagner. 1976. 'Written Descriptions of Orgasm: A Study of Sex Differences', *Archives of Sexual Behavior*, 5.1: 87–98 <https://doi.org/10.1007/BF01542242>

Wang, Rui-Li, and Ryan B. Chang. 2024. 'The Coding Logic of Interoception', *Annual Review of Physiology*, 86.1: 301–27 <https://doi.org/10.1146/annurev-physiol-042222-023455>

Weinstein, Edwin A., Robert L. Kahn, and William M. Slote. 1955. 'Withdrawal, Inattention, and Pain Asymbolia', *Journal of Nervous and Mental Disease*, 121.3: 242–57

Wong, Hong Yu. 2017. 'In and Out of Balance', in *The Subject's Matter: Self-Consciousness and the Body*, ed. by Frédérique de Vignemont and Adrian J. T. Alsmith, pp. 1–20 (Cambridge, MA: MIT Press)

Woods, Stephen C., and Douglas S. Ramsay. 2007. 'Homeostasis: Beyond Curt Richter', *Appetite*, 49.2: 388–98 <https://doi.org/10.1016/j.appet.2006.09.015>

Zimmerman, Christopher A., Elwood L. Huey, Joseph S. Ahn, Lindsey R. Beutler, Chen Liang Tan, S. Kosar, Lin Bai, Yi Chen, Thea V. Corpuz, Loren Madisen, Hongkui Zeng, and Zachary A. Knight. 2019. 'A Gut-to-Brain Signal of Fluid Osmolarity Controls Thirst Satiation', *Nature*, 568.7750: 98–102 <https://doi.org/10.1038/s41586-019-1066-x>

Zimmerman, Christopher A., Yi-Chieh Lin, David E. Leib, Lin Guo, Elwood L. Huey, Gregory E. Daly, Yi Chen, and Zachary A. Knight. 2016. 'Thirst Neurons Anticipate the Homeostatic Consequences of Eating and Drinking', *Nature*, 537.7622: 680–84 <https://doi.org/10.1038/nature18950>