Radboud University

Thesis for obtaining a “Master of arts” degree in philosophy

Article

On a Confusion about Access Consciousness:

How scientists and the brain access information about consciousness

and

Research proposal

Mind, Meaning, Information:

Integrating Enactivism and Haig’s Information

Student name: Jochem Giel Koopmans

Student number: S4529901

Email: Jochem.koopmans@ru.nl

Supervisor : prof. dr. Marc V. P. Slors

Date of submission: 20 June 2022

Radboud University Nijmegen

Radboud University

I hereby declare and assure that I, Jochem Giel Koopmans, have drafted this thesis independently, that no other sources and/or means other than those mentioned have been used and that the passages of which the text content or meaning originates in other works – including electronic media – have been identified and the sources clearly stated.

Place: Nijmegen date: 20 June 2022

Radboud University

Table of contents

Article

On a Confusion about Access Consciousness: How Scientists and the Brain Access Information about Consciousness

Abstract

The distinction between phenomenal consciousness (P) and access consciousness (A) is strongly contested. Ambiguity in the concept of A could partially explain these discussions, so a refined formulation of A is proposed. It respects the conceptual, epistemological and empirical criticism on the distinction between P and A and presents a new perspective to the neural realisation of A. Common views about neural representations and their availability in the brain imply that information for neuronal populations and for researchers is not equivalent. The introduction of two concepts of information provides tools to directly relate access consciousness to brain functioning. These developments provide a new perspective on the distinction between P and A for both proponents and opponents of the distinction. *(119 words)*

Keywords: Access, consciousness, information, neuronal population, phenomenal, representation

Introduction

The phenomenon of consciousness is one that can be investigated in a variety of disciplines, from artificial intelligence to neuroscience to psychology to philosophy. A problem, however, is that these different approaches seem to conceptualise consciousness in their own ways. This is troublesome because it makes findings from different fields hard to compare and integrate. In an attempt to structure consciousness research, Ned Block (1995) identified two separate ways

in which people used the term and proposed to acknowledge that by creating two different concepts: phenomenal consciousness (P) to refer to the phenomenon of subjective experience and access consciousness (A) to refer to ways in which such experience can be measured or demonstrated.

The distinction between P and A has allowed people in the field to specify which aspect of consciousness they investigated, no doubt avoiding certain debates where the two uses are conflated. Another benefit of the distinction, namely separating the neural correlates of access consciousness from phenomenal consciousness proper (e.g. Block, 2007), has not resulted in the fundamental advances in understanding that Block hoped for. The main reason for this is that the distinction has been called into question as arguably introducing an unnecessary obstacle to understanding consciousness (e.g. Cohen & Dennett, 2011; Kouider et al., 2012; Naccache, 2018).

While most of this critique has focused on the (supposed) relation between P and A, less attention was devoted to structurally developing the notion of access. Discussants generally tend to agree on a relatively broad notion of access consciousness. It will be argued that this broadness hides the possibilities for a more precise notion of access consciousness to provide a new impulse to discussions about the use of the P/A distinction.

To arrive at this conclusion, the paper is structured as follows. In section 1, the distinction between phenomenal and access consciousness will be introduced, as well as the critique against it. Section 2 will elaborate on access consciousness, discussing how it has been used and operationalised so far. The discussion will show how a more precise definition of access captures this variety of uses. As the goal of this paper is to show how this improvement might advance theories about the way consciousness relates to the brain, section 3 will provide a short overview of a dominant view on brain functioning. This overview will introduce two related perspectives on the information present in brain activity, which each support their own idea of access in section 4. The fifth section will ask the question of how the two kinds of access consciousness relate to Block's distinction between P and A.

Section 1: The P/A distinction and its critique

Section 1.1: Phenomenal consciousness

When Block introduced the concept of phenomenal consciousness, he tried to avoid his description from generating controversy. For this reason and because a non-circular definition has been hard (if not impossible) to give, Block used pointers to circumscribe one use of the concept of 'consciousness'. These pointers include (rough) synonyms, like 'experience' or Nagel's (1974) 'what it is like'. Nagel argued that an organism can only be conscious 'if and only if there is something that it is like to *be* that organism—something it is like *for* the organism' (p. 436, original emphasis). Block continues by giving (human) examples of phenomenal consciousness like 'sensations, feelings, and perceptions, … thoughts, desires, emotions.' (Block, 1995, p. 230).

These synonyms and examples share properties that Block thinks help characterise phenomenal consciousness. P is often considered to be representational (being about something else, also called intentionality), and a difference in representational content is often accompanied by a difference in phenomenal consciousness. For example, an experience as of a pen is different from an experience as of a sheet of paper.[1]

In addition, and this is what makes Block's distinction controversial, Block 'take[s] P-conscious properties to be distinct from any cognitive, intentional or functional property.' (Block, 1995, p. 230). It also introduces ambiguity: P could be fundamentally distinct from these properties, or it could be taken to suppose that conscious states may have a cognitive, intentional, or functional property but that the state's being P is independent of those properties. The former would not allow a conscious state to have a certain function *and* be phenomenally conscious, the latter would.

[1] This formulation ('as of') allows for the represented object to be different from the way it is experienced. What I see/experience as a pen might not actually be a pen.

Block himself seems to support the latter interpretation, since he claimed that differences in intentional content can make a difference in P. Others seem to support the former interpretation, perhaps most evidently when they relate phenomenal consciousness to qualia (e.g. Chalmers, 1995). 'Quale' (plural: qualia) is the term to refer to an experience itself, independent of what it may be about. The concept of qualia drives the 'inverted colour spectrum' thought experiment (cf. Byrne, 2020). In this experiment, two people might look at the same object, say a ripe tomato, and both call it red. It is still possible for their experience of the tomato's redness (their qualia) to differ but that they learnt to refer to that experience with the word 'red'.

Qualia have a number of properties (described in Dennett, 1988). They are ineffable, as the inverted colour spectrum thought experiment shows. They are intrinsic, which means that an experience's character or 'feel' itself is not determined by what it is an experience *of*. Their intrinsicness and ineffability makes that they are private to the person experiencing them. Finally and perhaps trivially, qualia are directly apprehensible in consciousness. These properties explain more directly why P is considered independent of functional, intentional or cognitive properties. The experience associated with a certain intentional relationship does not take its nature from that relationship, so it is not determined by its intentional object nor any functional of cognitive function of that object.

In sum, phenomenal consciousness as introduced by Block refers to the phenomenon of a person's having (conscious) experiences. The nature of these experiences themselves is established independently of any cognitive, intentional or functional properties that they may be associated with. This independency could indicate P to be a property of an experience that may also have such other properties, or that P as an experience is independent on but can sometimes be associated with such properties.

Section 1.2: Access consciousness and the reasons for the distinction

Access consciousness refers to states that are '(1) … poised for use as a premise in reasoning, (2) poised for rational control of action, and [or] (3) poised for rational control of speech' (Block, 1995, p. 231). The explicit requirement of such states to be under rational control (of action or speech) serves to exclude unconsciously-driven behaviours from also indicating conscious awareness. Immediately, Block mentions that it is not meant as a precise definition: 'I see A-consciousness as a cluster concept' (*ibid*). At the outset, then, access consciousness was conceived of as a term encompassing various possible functions. It also makes this term one that is open to different interpretations, discussed in section 2. In other works, Block (2007) also refers to 'cognitive accessibility' or 'reportability' instead of access consciousness.

Block (1995) discussed two main reasons to distinguish between two uses of 'consciousness'. First, he showed how researchers use the term to refer to the phenomenal or experiential aspect of consciousness in describing what they set out to explain, but use a functional notion related to access consciousness in their eventual answer. More generally, some philosophers discussing the concept of consciousness could discuss the relation between qualia and experience whereas others could discuss how human subjects produce rationally controlled language. They would use the same word 'consciousness', but as Block showed it is far from evident that they also refer to the same concept.

Moreover, Block's reference to A as 'cognitive accessibility', dropping the notion of consciousness altogether, suggests that he sees them as referring to two different categories. Philosophers taking the strict interpretation of phenomenal consciousness as qualia also make it fundamentally distinct from access consciousness. Failing to distinguish between these two types of consciousness could lead to unnecessary debates that hinder progress on our understanding of consciousness in general.

A second reason to distinguish between P and A was the various studies that were argued to empirically demonstrate the two to be different. Block (1995) discusses the finding that in comparatively complex visual displays, subjects seem to have a

richer visual experience of such displays than what they can report about them (Landman et al., 2003; Sligte et al., 2008, 2010; Sperling, 1960). Block argues that people's visual (phenomenal) experience overflows their capacity to access (report) that experience (Block, 2011).

Another empirical demonstration of the separability of phenomenal from access consciousness that Block names is that of change blindness, a paradigm in which people are shown two nearly identical images in alternation. A blank screen in between the alternations makes it hard for people to spot the difference between the two images. Block (2007) argues that people do experience the item that changes between alternations, but that they are unable to report or access that item for a report.

Section 1.3: Critique on the P/A distinction

In the more than 25 years since Block's proposal to distinguish between phenomenal and access consciousness, it has been criticised on diverse grounds. The criticism discussed below, broadly grouped as taking a conceptual, epistemological, and empirical approach, is not necessarily exhaustive. Rather, it serves to demonstrate the point that progress on discovering the neural realisation of phenomenal (and access) consciousness is difficult under the current conceptualisation of P and A.

The strongest conceptual critique on the P/A distinction (in terms of its implications) claims that phenomenal consciousness is not a real phenomenon in its own right. Dennett (1988), for example, argues that qualia (which under some interpretations are identified with phenomenal consciousness) do not exist in the way that they are often described. Phenomenal consciousness might be ineffable and private, but only comparatively (rather than fundamentally) so. When two people experience a complex sound, Dennett argues they pick up the same, shared, property (e.g. the combination of individual tones) that may be harder for the musical novice to pick out or describe than for an experienced musician. The

'privacy' of the experience merely describes that each person has their own 'property detectors', which detected the sound for our musicians.

Extending this proposal further, Frankish (2016) proposes that experience does not have the properties associated with qualia at all. Without such properties, the idea of phenomenal consciousness having such properties is an illusion that should be investigated as such. Regarding the proposed distinction between P and A, P being an illusion reduces it to functions regarding people's convictions that there are properties described by the concept of qualia. Such functions would likely fall in the category of access consciousness, collapsing the distinction entirely. This critique is itself hotly contested (e.g. Balog, 2016; Niikawa, 2021), explaining why the suggested collapse of P and A has not pervaded literature on consciousness (yet).

A second argument against the conceptual distinction made by Block is that access consciousness is not a form of consciousness (Schlicht, 2012). According to Schlicht, the decision by Block to define access consciousness independent of the subjectivity or 'what-it's-likeness' that is characteristic of consciousness, makes that it should not be considered a form of consciousness. This critique is not as influential because Block's recognition of access consciousness was merely descriptive. Block is therefore not committed to calling it a form or type of consciousness, as demonstrated by his using the term 'cognitive accessibility' (e.g. Block, 2007).

Epistemological arguments against the distinction between P and A focus on the idea of experiences that are not accessed or even inaccessible. The conceptual separation of P from A makes this a theoretical possibility and the empirical findings described above arguably show that it is a real phenomenon. A clear objection is that without access consciousness of putative experiences, there is no way to be certain about the phenomenal consciousness of them (Cohen & Dennett, 2011, 2012; Kouider et al., 2012; Naccache, 2018). Information that is not accessed could equally be unconscious (Kouider et al., 2010, 2012) and without an access

conscious report to arbitrate, there is no way to distinguish between such information being phenomenally conscious or unconscious (Schlicht, 2018).

Conversely, when determining the neural substrates of experiences that are both phenomenally and access conscious, the conjunction of P and A makes it impossible to fully separate the contribution of A from the contribution of P to that substrate (Block, 2007). A solution that has been proposed is to claim that access consciousness is independently constitutive of a conscious experience (Dehaene et al., 2006, 2011; Cohen & Dennett, 2011, 2012; Mashour et al., 2020), which amounts to collapsing the distinction between P and A and proceeding with the definition of access consciousness. However, changing the definition of consciousness can obviously go two ways: one could also exclude access consciousness as a constitutive part of consciousness (e.g. Lamme, 2006; Schlicht, 2012). This solution to finding the neural correlate of consciousness depends on consensus about a definition of consciousness, a seemingly impossible endeavour given the diametrically opposed opinions in this respect.

Block (1995, 2007) foresaw these epistemologically oriented arguments against distinguishing between P and A and suggested to decide on the phenomenal consciousness status of inaccessible contents based on inference to the best explanation: if the functions related to access can be investigated, then the remainder of neural activation related to accessed and phenomenally conscious experiences can be attributed to P. Isolated occurrences of that remainder, without activation related to functions of access, should then be inferred to be conscious as well.

However, such an account requires the separability of A from such A-and-P experiences. Given the nature of A as indicating P, there cannot be A in isolation (i.e., without P; Cohen & Dennett, 2011, 2012). The theoretical idea of separating the neural activity related to A from activity related to P is therefore concluded to be misguided. Even granted the conceptual distinction between P and A, it is an empirically inseparable distinction.

Empirically oriented critique problematises the studies purportedly demonstrating P without A. In these studies, participants are presented with a relatively rich visual stimulus, for example a scene or grid filled with letters. Participants indicate to have an experience of the entire stimulus, but when asked to report it they either fail to notice substantial changes in otherwise identical scenes (e.g. Simons & Ambinder, 2005) or their report of presented letters ends after four or five out of twelve letters (e.g. Sperling, 1960). The crux of these studies is that subjects' access to or report of the stimulus is more limited compared to what one would expect given their experience of the whole stimulus: phenomenal consciousness of the stimulus is rich, but the mechanisms of access are too slow to convey all that information before the experience fades away (Block, 2007).

Others argue that the evidence for rich phenomenal consciousness is mistaken because a subject's claim of having experienced the whole stimulus is itself also an accessed experience. Crucially, this might be access to a different experience than what subjects are tasked to report (Naccache, 2018; Richards, 2015). For example, it might reveal access to a summary statistic about the stimulus having consisted of 12 letter-like stimuli. Subjects need not have had a phenomenally conscious experience of all individual letters for them to report to have seen a grid filled with letters (Kouider et al., 2010). This means that their phenomenal consciousness might correspond to what they can access about that experience, avoiding an overflow account.

The three approaches to criticising the distinction between phenomenal and access consciousness all impede the desired progress on understanding the neural realisations of phenomenal and access consciousness. By collapsing the distinction to contain either only A or only P, as proposed by several opponents of the distinction, such progress could be amenable to scientific investigation. It is a heavily criticised suggestion, however, with people defending both alternatives, so such 'progress' would not be widely supported. Even when accepting a distinction between the concepts of P and A, there seems to be an epistemological problem regarding the attributability of neural activity to either P or A. Finally, studies

purportedly demonstrating the separation of P (rich) from A (sparse) have been reinterpreted in a way that P need not 'overflow' A. Neural activity elicited by such studies could therefore only be related to A-and-P.

Section 2: Towards a precise formulation of access consciousness

Access consciousness was a cluster concept from its first main presentation in 1995 onward. Correspondingly, it has been defined in various ways during the debate about its relation to phenomenal consciousness. The concept is referred to as access consciousness (e.g. Block, 1995), but also as cognitive access (Block, 2011; Schlicht, 2012) or simply reportability (e.g. Block, 2007; Cohen & Dennett, 2011). In addition, the concept has been interpreted as a property of 'states' (Block, 1995), 'contents' (Kouider et al., 2010), 'representations' (Kouider et al., 2010), and 'information' (Cohen & Dennett, 2011), even when it is not clear that they mean the same in all cases. The property constituting access has been operationalised as broadly as the tripartite premise for reasoning, rational control of action or speech (Block, 1995), but also as a requirement for such reasoning and control to occur (Naccache, 2018), or as simply reportability (Cohen & Dennett, 2011, 2012).

This list shows how the concept of access consciousness is used in a variety of different ways, probably causing confusion in some of the debates surrounding it (Kouider et al., 2010). One author even argued that one might settle the debate on whether the distinction should be collapsed or not, depending on the specific definition of access consciousness one uses (Overgaard, 2018). When looking for the neural correlate of access consciousness, this underdetermination only adds ambiguity to the already difficult task of investigating whether the neural substrates of A can be distinguished from P. These are all incentives to develop a more precise formulation of A, which is what this section aims to achieve.

The recurrent use of reportability as either a measure of or constituent of access consciousness is no coincidence, especially in empirical studies purportedly showing that P is richer than A. Reports are important because they are taken to convey information about what a person was phenomenally conscious of: when

someone says 'I see a red tomato', we generally take that as an indication that this person is having a phenomenally consciousness experiences as of a red tomato.

When viewed from this perspective, not just someone's verbal reports but also their (rational) behaviour is access conscious in virtue of its conveying information about what that person is phenomenally conscious of. A person might be asked to press a button when they have a phenomenally conscious experience as of a red tomato, for example. It also explains why a state's being use as a premise in reasoning was included in Block's original formulation of access consciousness: assuming that one cannot reason unconsciously (which might even be considered a *contradictio in terminis*), someone's use of 'red tomato' as a premise in reasoning, about making a salad for example, would indicate that person to have had a phenomenally conscious experience as of a red tomato.

Furthermore, this perspective would explain why access consciousness is often related to higher-level cognitive functions like attention (e.g. Dehaene & Changeux, 2011; Schlicht, 2012) or rational as opposed to automatic or heuristic-driven behaviour (Block, 1995). Higher-level processes might be considered to be less controversially conveying information about one's phenomenally conscious experiences than lower-level sensory processing. Indeed, this might underlie some researchers' suggestion that lower-level sensory processing in the absence of access consciousness is unconscious rather than phenomenally conscious.

Still, formulating access consciousness as what a person reports about their own phenomenally conscious experiences is arguably too limited (Block, 2007). 'Reasoning' could be another way in which a person might be inferred to have phenomenally conscious experiences, and there are more. A vegetative-state patient was able to selectively modulate her brain activity (as measured by MRI) in response to questions from researchers (Owen et al., 2006). To the extent that one considers this to be evidence for the presence of some phenomenally consciousness experience (e.g. Block, 2007), one might also take MRI-data to convey information about a person's phenomenally conscious experiences. In line with this idea, Cohen and Dennett (2011) even see these MRI-data as a report of sorts.

To avoid ambiguity about what constitutes a report, it would be better to formulate access consciousness without referring to reportability. A precise formulation might therefore be something like the following: Something is access conscious when it is taken (by an observer) to convey information about that person's phenomenally conscious experiences. This information could be about the mere occurrence of a phenomenally conscious experience or to point to its intentional object (e.g. the redness of a tomato), but to the extent that P is private, intrinsic and ineffable it cannot be about the qualities of P itself.

This formulation of access consciousness explicitly decouples *de facto* phenomenal consciousness from communication about them. Simultaneously, the formulation also captures the idea that researchers can have reasons to believe a person has certain phenomenally conscious experiences, corresponding to the idea that access consciousness does in certain cases have some relation to phenomenal consciousness.

In line with Schlicht (2012), I would agree that the term access *consciousness* is distracting. Instead, one might call it 'access *of/to* presumed (phenomenal) consciousness'. To the extent that such access is cognitive, the terms cognitive access or cognitive accessibility (e.g. Block, 2007, 2011) would also apply. Formulating access consciousness this way does of course not end all debates on such a complicated topic. Evidently, it will not solve debates about when something should be taken to convey information about a phenomenally consciousness experience. For example, the no-report paradigm investigates brain activity in the absence of behavioural report (Tsuchiya et al., 2015), but whether such activity reflects phenomenal consciousness is hotly debated (e.g. Block, 2019; Overgaard & Fazekas, 2016; Pitts et al., 2018).

Additionally, it is worth mentioning that the current formulation is compatible with an illusionist's view about phenomenal consciousness (cf. Dennett, 2003; Frankish, 2016). The information that is conveyed would then pertain to the illusion of P instead of the actual occurrence of P. The main benefit of it will become apparent

when relating access consciousness to brain functioning, which will be discussed below.

Section 3: A framework of brain functioning

Section 3.1: Predictive processing

Before access consciousness can be related to brain functioning, it will be necessary to explain how the brain is generally thought to function in the first place. I will do so using a currently dominant framework called predictive processing (or predictive coding; Clark, 2015; Friston, 2010; Hohwy, 2020), but the important aspects in this framework are not particular to predictive processing.

The central idea of the predictive processing framework is that the brain processes information (stimuli, input) in a hierarchy of processing steps. At every processing step, top-down predictions about the input are integrated with the actual input. Because stimuli are inherently noisily transduced into neural signals, the integration of that noisy input signal with top-down predictions can make estimations about the 'true' input more accurate (e.g. Summerfield & de Lange, 2014). Such integration can be modelled effectively using the mathematical models called Bayesian inference (de Lange et al., 2018).

As a model of cognition, predictive processing proposes a hierarchy of such processing steps (e.g. Huang & Rao, 2011). Visual object processing, for example, is modelled as representing visual stimuli in terms of orientation and spatial frequency in primary visual cortex (V1), which combine to form edges in V2, which in turn form shapes in inferior temporal (IT) cortex. There, these shapes can be recognised as objects (Perry & Fallah, 2014; they describe another hierarchical route for visual motion processing).

When a visual stimulus is processed in V2, the representational content of V2-activity will be determined by the integration of top-down 'predictions' from IT with bottom-up input from V1. For example, if I am marvelling at a tall skyscraper, the representation of rectilinear shapes and a generally rectangular building in IT will predict, via its neural connections to V2, mostly right (90-degree) angles to be

the source of signal entering V2. If unexpectedly a friend jumps in my view, bottom-up evidence will instead drive a representation of the curvilinear edges of their face in V2. The integration of that bottom-up signal with the predicted rectilinear edges results in a large mismatch, causing areas upwards in the visual hierarchy to update their subsequent 'predictions' to seeing more of the face.

Clearly, the term 'prediction' is not used in a fully cognitive sense but rather as a neuronal population modulating other populations that feed into it. The exact mechanism of how predictions modulate these other populations is debated (de Lange et al., 2018), but the view one favours will not affect the way access consciousness relates to brain functioning.

The neural implementation of predictive processing is assumed to be the actual connections between neuronal populations (e.g. de Lange et al., 2018), which greatly restricts the scope of availability of representations generated at a certain processing stage. The representations in one population are only available to neural populations that it sends its output to. If V1 only sent axons to V2, information about the orientation and spatial frequency of visual stimuli in V1 cells' receptive fields (the area in the visual field to which they respond) would only be directly available to neurons in V2. This nuances the idea of information *flow* in the brain, as each processing step in the processing hierarchy involves a qualitatively different type of representation. Nonetheless, there is a sense in which lower-level representations can be inferred from representations at a higher-level neuronal population, that will be discussed in the next subsection.

As mentioned, this view of representation in neural populations does not depend on the particulars of predictive processing. Human cognition has been modelled as hierarchical processing since the 1970s (e.g. Craik & Lockheart, 1972), and predictive processing theories assign similar roles to neural activity as do traditional theories of brain processing (Cao, 2020). That is, more traditional models might interpret neural output as 'informing' the next processing step rather than presenting the integration between predictions and bottom-up input to that population. Still, both interpretations suggest that neural populations are representational vehicles

and distribute representations using axonal connections to other neuronal populations, a view that has a concrete neural explanation in the current predictive processing framework.

Section 3.2: Information and representation

Using the predictive processing framework, it is possible to determine relatively precisely which kinds of representation exist in the brain's different neural populations. The terms 'information' and 'representation' are used ubiquitously by neuroscientists, often interchangeably. Without further qualification, they both refer to a property that a certain cell or neural population is selectively responsive to (Cao, 2020). To relate access consciousness to brain functioning, however, it will be useful to separate both terms and assign them their own meaning.

The reason why this is useful, is because there are differences between a researcher's and a neuronal population's accessing information. Neuronal populations are 'blind' to the properties of the external world that they respond to. That is, the neuronal activity that is their input does not specify which property in the environment caused that input. Researchers can, but the neurons in V1 cannot, tell whether their input comes from light waves or from the skin's pressure receptors. This is a recognised phenomenon, also called the 'labelled lines hypothesis' (Pennartz, 2009), experimenter- versus cortex-as-receiver (de-Wit et al., 2016), or 'looking-down' versus 'looking-up' in mechanistic explanations (Bechtel, 2009).

Researchers can do this because often there is a statistical relationship between a neuronal population's activity pattern and a certain real-world category or property. For example, the fusiform face area (FFA) and parahippocampal place area (PPA) have been found to respond to the categories of faces and scenes, respectively (Peelen & Downing, 2005). As a result, the FFA can be said to represent or contain information about faces. However, a possible correlation between the category of faces and, say, curvilinearity of lines, could be exploited by statistical analyses that decode neuroimaging data (de-Wit et al., 2016; Ritchie et al., 2019). They could

show that FFA activity patterns can also be used to dissociate rectilinear from curvilinear shapes, for example. Consequentially, researchers would also be able to claim that line curvilinearity information is also present in FFA activity patterns.

In other words, there is a difference in information between neuronal populations and researchers. A neuronal population may be responsive to a single real-world property, but correlations between that and other properties allow researchers to extract additional information from that population. It presents a complication that is relevant to access consciousness because that is about a researcher's interpretation of information (as being about phenomenally conscious experiences).

The distinction between information and representation announced earlier will be applied precisely to make this difference in information explicit. Haig's idea about information as uncertainty reduction (Haig & Dennett, 2017) might help in clarifying this meaning of the information. Haig suggests to define information as uncertainty reduction in a sensor. A sensor could take a number of values, depending on the state of the property it responds to. This range of values is its uncertainty, which will be reduced by the sensor taking one specific value within that range (e.g. a temperature sensor reading 22 degrees). A sensor could be anything that is selectively responsive to some property, so neuronal populations can also be considered a sensor in this sense.

Interpreted this way, a sensor's information is highly unconceptualized. A light detector does not in and of itself contain information as of 'light waves'. That will only be the case when interpreted as such by an interpreter. When the information in a neuronal population is described as 'responding to faces' (e.g. in the FFA) or 'skyscraper' (e.g. in IT cortex), it refers to the name that we as interpreters give to the property (face) or property value (skyscraper) that it selectively responds to. This uncertainty reduction is how I will propose to use the term 'information'. As

such information pertains only to one (complex) real-world property[2], it is a relatively thin notion.

The broader interpretation of information as what researchers might interpret from conceptualising response properties will be referred to as 'representation'. When a neuronal population in IT cortex is maximally responsive to skyscrapers as opposed to other buildings, it will contain information about skyscrapers but also cardinal line orientations, straight (90-degree) angles and probably grey colours. From here on, it will be said to represent those types of information, whereas its information will be something like 'taking value or state x (skyscraper) among value- or state space Y (a certain set of shapes that it is responsive to)'.

Section 4: Relating access consciousness to brain functioning

When relating the concept of access consciousness to neural activity, it will be necessary to switch from a researcher's interpretation of information (representation) to the brain's thinner notion of information. This section will discuss some considerations pertaining to such a switch.

First, a notion of informational access (IA) will be developed. IA does not in itself convey information about consciousness, but it will provide a background against which to interpret a neural account of access consciousness. The notion of information as a neuronal population's response to its response property, combined with the idea that such information is transferred via neuronal connections, implies that information is not necessarily widely available in the brain. Rather, the information contained in one population's response pattern is only available to those populations that it sends its output to. Those populations can be said to access that information and use it to determine their own response pattern.

One might argue that for a given neuronal population, not just its (bottom-up) inputs but also top-down 'predictions' count as access. A bi-directional relationship of

[2] The real-world property that a neuronal population selectively responds to need not neatly correspond to a property or category that researchers acknowledge. A population might respond to a certain combination of established properties, which could be described as a complex property.

feed-forward and feedback connections between two neuronal populations would therefore count as both populations accessing the information contained in the other population's response pattern (but see Kok et al., 2016, for evidence that a neuronal population might distinguish its feedforward and feedback connections). While this may complicate the view on brain processing and informational availability somewhat, the same ideas should apply.

From the notion of IA and the vast interconnectedness of the brain's neuronal populations, it follows that practically all the information contained in neuronal populations' response patterns is accessible to some other neuronal population. All that is required, is that a population's output is input to another population. Hypothetically, one could prevent access by surgically severing all a population's output connections (Cohen & Dennett, 2011).

Now, back to the formulation of access consciousness: something is access conscious when it is taken to inform (a researcher) of a person's phenomenally conscious experiences. IA describes the way neuronal populations inform other such populations, so the neural realisation of access consciousness is a subtype of IA. Access consciousness is that (neuronal) information that a researcher takes to be informative about a person's phenomenally conscious experiences.

However, the sorts of things that have traditionally been interpreted as conveying information about such experiences, (verbal) reports and button presses, would count as representations; the concept of A has traditionally been representational access consciousness (RA). Confusion between the notion of access consciousness as pertaining to representations or information could explain Block's intuition that (representational) access consciousness is an impure measure of consciousness but also the intuition by Cohen and Dennett (2011, 2012) that no information about P can be communicated without it having been accessed somehow.

In addition, it captures the debate about how to interpret supposed demonstrations of a rich phenomenally conscious experience with sparse access consciousness. By distinguishing between information and representation it becomes clear that the two

intuitions are not incompatible. In terms of representational access consciousness, one might infer a subject to have a rich phenomenal experience (e.g. being aware of a grid of letters), but in terms of informational access consciousness (i.e. which individual letters), the subject would be more limited.

Identifying the relevant information in a representation requires a closer analysis of what it is that is informative about a person's phenomenally conscious experiences than has hitherto been done. In this sense, the critique by Kouider et al. (2010), Naccache (2018), and Richards (2015) that one should not interpret a participant's claim about rich experience to in itself indicate phenomenally conscious experience of all a scene's details is a first step in the right direction. Formalising a procedure to reduce access consciousness in representational terms to its information is beyond the scope of this article, but some pointers will indicate how one could go from representation to information.

Consider the example of subjects indicating a rich experience of a grid of letters. Following Kouider et al. (2010), the specific utterance matters: 'I experienced a grid of letters' conveys a different representation and information compared to 'I experienced twelve individual letters composing a grid'. Moving that aside as an issue that more detailed questioning will disambiguate, further considerations must include which kinds of information are present in the utterance. There will be information about loudness, speech timing, stress, syllable- and word pronunciation, word selection, and the semantics of the utterance. Probably, the information that is informative about that person's phenomenally conscious experience should be on the level of semantics.

The next question to answer is which response property the semantics is information of, in other words which response property's uncertainty was reduced to/by this utterance's meaning. In this example, it could be the various arrangements of symbols, the types of symbols or (perhaps more likely) a combination of both: the utterance *could* have conveyed any kind of arrangement of various symbol types, but *did* convey specifically a grid-arrangement of letter-symbols. Finally, such information is in principle relatable to information in neural populations'

response patterns, say some neuronal population in Broca's area (e.g. Heim et al., 2009). This area would be the neural correlate of access consciousness pertaining to the perception of the grid.

Section 5: How does IA relate to the P/A distinction?

The refined formulation of access consciousness developed in the previous sections has implications for the way scientists would look for the neural correlates of access consciousness. It will therefore be interesting to find out how the conceptualisation of access consciousness and informational access presented here relates to the P/A distinction from the various perspectives on the P/A distinction. In section 1, critique on the P/A distinction was summarised into conceptual, epistemological, and empirical approaches, but of course there are also proponents of the P/A distinction who might have their own view on how to relate informational access to the P/A distinction. This section will explore how the ideas presented in this paper might apply to their views.

Conceptual critique on the P/A distinction mainly targeted the concept of phenomenal consciousness. Illusionists claim that it is not the phenomenon of P itself that needs explaining, but rather the fact that people think it is a phenomenon. The developed concept of access consciousness presented here is compatible with illusionism as it emphasises a researcher's interpretation of information concerning a person's (illusory) phenomenally conscious experiences. The relation of this notion to information in the brain, using the concept of IA, also does not take the notion of phenomenal consciousness to be a real phenomenon. Hence, there could be reasons for illusionists to agree with the concepts of access consciousness and IA developed here.

Epistemological critique came twofold: the possibility of P without A is epistemically opaque and the definition of A as the only source of evidence for P makes the case of P without A impossible. These criticisms were argued to make a distinction between P and A superfluous in terms of their neural realisations.

The introduction of informational access presents a new perspective on the idea of inaccessible phenomenally conscious experiences. The pervasiveness of IA in the brain suggests that inaccessible P would not occur in practice, at least to the extent that P has a relation to brain functioning. Phenomenal consciousness that is completely unrelated to brain functioning could be informationally inaccessible, those would be experiences that one could in no way convey information about, not even when they occur. Further research should investigate what such cases would mean and which lessons to take away from them.

The development of access consciousness and informational access did not directly affect the fundamental connection between P and A. The introduction of IA and its relation to RA does however provide a way to investigate to which extent information in the brain is access conscious as they can be established independently. In that sense, it might provide a new impulse to debates about whether reports of (representational) access consciousness indicate rich or sparse information processing (e.g. Block, 2011).

Empirical critique on the P/A distinction contested ostensive demonstrations of P being richer than A. The development of RA and IA has not aimed to support demonstrations of P being independent of A. It does, however, provide an indication on how to disambiguate such putative demonstrations of P being richer than A, by moving from RA as a notion of access consciousness to an IA notion of access consciousness.

The concepts of IA and RA might also be interesting for proponents of the P/A distinction to relate to P and A. The notion of information access is conducive to the idea that, using the variety of research methods available to consciousness scientists, more information in the brain can be taken to be access conscious than hitherto thought using a representational notion of access consciousness. A case for rich phenomenal consciousness, topic of so much debate, might be made on the basis of rich access. In addition, the notion of IA could capture the idea that the neural activity traditionally related to access consciousness might include processes that are unrelated to the information that is accessed (Block, 2007).

These considerations might provoke proponents to explore possible additional parallels between IA and P. One might consider whether properties like intrinsicness, ineffability, and privateness could be applicable to IA, for example.

Summary and conclusion

The distinction between phenomenal and access consciousness was introduced as an attempt to structure and disambiguate discussions about consciousness. Although it may have been partially successful in doing so, it also created additional debates about the distinction itself. These debates slowed or even stalled progress on identifying the neural correlates of consciousness, an area of research that the distinction has direct implications for.

In this paper, part of the debates about the justifiability and relevance of the P/A distinction were considered to be caused by ambiguity concerning the notion of access consciousness. This has allowed a variety of different formulations and operationalisations to be used under the same concept, providing a potential reason for controversies concerning the neural realisation of access consciousness and, by extension, also phenomenal consciousness. In an attempt to capture the main idea of access consciousness, a more precise formulation was proposed that nonetheless respects concerns about the reality or investigability of phenomenal consciousness: something is access conscious when it is taken to convey information about a person's phenomenally conscious experience.

This formulation of access consciousness was subsequently related to brain functioning. Common ways to think about brain functioning revealed that 'information' may have a subtly different meaning for researchers and neuronal populations, which were named 'representation' and 'information'. The difference between representation and information was suggested to have played a role in current debates about empirical demonstrations of the separability of P and A. The reformulation of access consciousness and the introduction of (neuronal) informational access provides the tools to possibly reconcile competing views. In addition, it suggests new avenues for empirical research to investigate the extent to

which neural processes are access conscious and hence associable with occurrences of phenomenal consciousness.

The development of the access consciousness concept presented here provides a new perspective on the role of access consciousness in consciousness research. Consequentially, it might cause both critics and proponents of the phenomenal/access consciousness distinction to refine the way they see the relation between the concepts of access consciousness, informational access, and phenomenal consciousness. Such refinements might allow philosophers and scientists with rivalling views to find new points of agreement, from which further understanding of consciousness can be built.

7220 words

References

Balog, K. (2016). Illusionism's Discontent. *Journal of Consciousness Studies, 23*(11-12), 40-51. https://www.ingentaconnect.com/content/imp/jcs/2016/00000023/f002001 1/art00003

Bechtel, W. (2009). Looking down, around, and up: Mechanistic explanation in psychology. *Philosophical Psychology*, *22*(5), 543-564. https://doi.org/10.1080/09515080903238948

Block, N. (1995). On a confusion about a function of consciousness. *Behavioral and Brain Sciences*, *18*(2), 227-247. https://doi.org/10.1017/S0140525X00038188

Block, N. (2007). Consciousness, accessibility, and the mesh between psychology and neuroscience. *Behavioral and Brain Sciences*, *30*(5-6), 481-499. https://doi.org/10.1017/S0140525X07002786

Block, N. (2011). Perceptual consciousness overflows cognitive access. *Trends in Cognitive Sciences*, *15*(12), 567-575. https://doi.org/10.1016/j.tics.2011.11.001

Block, N. (2019). What is wrong with the no-report paradigm and how to fix it. *Trends in Cognitive Sciences*, *23*(12), 1003-1013. https://doi.org/10.1016/j.tics.2019.10.001

Byrne, A. (2020, August 26). *Inverted Qualia*. Stanford Encyclopedia of Philosophy Archive. https://plato.stanford.edu/archives/fall2020/entries/qualia-inverted/

Cao, R. (2020). New labels for old ideas: Predictive processing and the interpretation of neural signals. *Review of Philosophy and Psychology*, *11*(3), 517-546. https://doi.org/10.1007/s13164-020-00481-x

Chalmers, D. J. (1995). Facing up to the problem of consciousness. *Journal of Consciousness Studies, 2*(3), 200-219.

https://www.ingentaconnect.com/content/imp/jcs/1995/00000002/00000003/653

Clark, A. (2015). Radical predictive processing. *The Southern Journal of Philosophy*, *53*(S1), 3-27. https://doi.org/10.1111/sjp.12120

Cohen, M. A., & Dennett, D. C. (2011). Consciousness cannot be separated from function. *Trends in Cognitive Sciences*, *15*(8), 358-364. https://doi.org/10.1016/j.tics.2011.06.008

Cohen, M. A., & Dennett, D. C. (2012). Response to Fahrenfort and Lamme: Defining reportability, accessibility and sufficiency in conscious awareness. *Trends in Cognitive Sciences*, *16*(3), 139-140. https://doi.org/10.1016/j.tics.2012.01.002

Craik, F. I., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, *11*(6), 671-684. https://doi.org/10.1016/S0022-5371(72)80001-X

de Lange, F. P., Heilbron, M., & Kok, P. (2018). How do expectations shape perception? *Trends in Cognitive Sciences*, *22*(9), 764-779. https://doi.org/10.1016/j.tics.2018.06.002

Dehaene, S., & Changeux, J. P. (2011). Experimental and theoretical approaches to conscious processing. *Neuron*, *70*(2), 200-227. https://doi.org/10.1016/j.neuron.2011.03.018

Dehaene, S., Changeux, J. P., & Naccache, L. (2011). The global neuronal workspace model of conscious access: From neuronal architectures to clinical applications. In S. Dehaene, & Y. Christen (Eds.), *Characterizing consciousness: From cognition to the clinic?* (pp. 55-84). Springer. https://doi.org/10.1007/978-3-642-18015-6_4

Dehaene, S., Changeux, J. P., Naccache, L., Sackur, J., & Sergent, C. (2006). Conscious, preconscious, and subliminal processing: A testable

taxonomy. *Trends in Cognitive Sciences*, *10*(5), 204-211. https://doi.org/10.1016/j.tics.2006.03.007

Dennett, D. C. (1988). Quining qualia. In A. J. Marcel & E. Bisiach, (Eds.), *Consciousness in modern science*. Oxford University Press.

Dennett, D. C. (2003). Who's on first? Heterophenomenology explained. *Journal of Consciousness Studies*, *10*(9-10), 19-30. https://www.ingentaconnect.com/content/imp/jcs/2003/00000010/F0020009/art00003

de-Wit, L., Alexander, D., Ekroll, V., & Wagemans, J. (2016). Is neuroimaging measuring information in the brain? *Psychonomic Bulletin & Review*, *23*(5), 1415-1428. https://doi.org/10.3758/s13423-016-1002-0

Frankish, K. (2016). Illusionism as a theory of consciousness. *Journal of Consciousness Studies*, *23*(11-12), 11-39. https://www.ingentaconnect.com/content/imp/jcs/2016/00000023/f0020011/art00002

Friston, K. (2010). The free-energy principle: A unified brain theory? *Nature Reviews Neuroscience*, *11*(2), 127-138. https://doi.org/10.1038/nrn2787

Haig, D., & Dennett, D. C. (2017). Haig's 'strange inversion of reasoning' (Dennett) and Making sense: Information interpreted as meaning (Haig). http://philsci-archive.pitt.edu/id/eprint/13287

Heim, S., Eickhoff, S. B., Friederici, A. D., & Amunts, K. (2009). Left cytoarchitectonic area 44 supports selection in the mental lexicon during language production. *Brain Structure and Function*, *213*(4), 441-456. https://doi.org/10.1007/s00429-009-0213-9

Hohwy, J. (2020). New directions in predictive processing. *Mind & Language*, *35*(2), 209-223. https://doi.org/10.1111/mila.12281

Huang, Y., & Rao, R. P. (2011). Predictive coding. *Wiley Interdisciplinary Reviews: Cognitive Science*, *2*(5), 580-593. https://doi.org/10.1002/wcs.142

Kok, P., Bains, L. J., van Mourik, T., Norris, D. G., & de Lange, F. P. (2016). Selective activation of the deep layers of the human primary visual cortex by top-down feedback. *Current Biology*, *26*(3), 371-376. https://doi.org/10.1016/j.cub.2015.12.038

Kouider, S., de Gardelle, V., Sackur, J., & Dupoux, E. (2010). How rich is consciousness? The partial awareness hypothesis. *Trends in Cognitive Sciences*, *14*(7), 301-307. https://doi.org/10.1016/j.tics.2010.04.006

Kouider, S., Sackur, J., & de Gardelle, V. (2012). Do we still need phenomenal consciousness? Comment on Block. *Trends in Cognitive Sciences*, *16*(3), 140. https://doi.org/10.1016/j.tics.2012.01.003

Lamme, V. A. (2006). Towards a true neural stance on consciousness. *Trends in Cognitive Sciences*, *10*(11), 494-501. https://doi.org/10.1016/j.tics.2006.09.001

Landman, R., Spekreijse, H., & Lamme, V. A. (2003). Large capacity storage of integrated objects before change blindness. *Vision Research*, *43*(2), 149-164. https://doi.org/10.1016/S0042-6989(02)00402-9

Mashour, G. A., Roelfsema, P., Changeux, J. P., & Dehaene, S. (2020). Conscious processing and the global neuronal workspace hypothesis. *Neuron*, *105*(5), 776-798. https://doi.org/10.1016/j.neuron.2020.01.026

Naccache, L. (2018). Why and how access consciousness can account for phenomenal consciousness. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *373*(1755), Article 20170357. https://doi.org/10.1098/rstb.2017.0357

Nagel, T. (1974). What is it like to be a bat? *The Philosophical Review*, *83*(4), 435–450. https://doi.org/10.2307/2183914

Niikawa, T. (2021). Illusionism and definitions of phenomenal consciousness. *Philosophical Studies, 178,* 1-21. https://doi.org/10.1007/s11098-020-01418-x

Overgaard, M. (2018). Phenomenal consciousness and cognitive access. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *373*(1755), Article 20170353. https://doi.org/10.1098/rstb.2017.0353

Overgaard, M., & Fazekas, P. (2016). Can no-report paradigms extract true correlates of consciousness. *Trends in Cognitive Sciences*, *20*(4), 241-242. https://doi.org/10.1016/j.tics.2016.01.004

Owen, A. M., Coleman, M. R., Boly, M., Davis, M. H., Laureys, S., & Pickard, J. D. (2006). Detecting awareness in the vegetative state. *Science*, *313*(5792), 1402. https://doi.org/10.1126/science.1130197

Peelen, M. V., & Downing, P. E. (2005). Within-subject reproducibility of category-specific visual activation with functional MRI. *Human Brain Mapping*, *25*(4), 402-408. https://doi.org/10.1002/hbm.20116

Pennartz, C. M. (2009). Identification and integration of sensory modalities: Neural basis and relation to consciousness. *Consciousness and Cognition*, *18*(3), 718-739. https://doi.org/10.1016/j.concog.2009.03.003

Perry, C. J., & Fallah, M. (2014). Feature integration and object representations along the dorsal stream visual hierarchy. *Frontiers in Computational Neuroscience*, *8*, Article 84. https://doi.org/10.3389/fncom.2014.00084

Pitts, M. A., Lutsyshyna, L. A., & Hillyard, S. A. (2018). The relationship between attention and consciousness: An expanded taxonomy and implications for 'no-report' paradigms. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *373*(1755), Article 20170348. https://doi.org/10.1098/rstb.2017.0348

Richards, B. (2015). Advancing the overflow debate. *Journal of Consciousness Studies*, *22*(7-8), 124-144. https://www.ingentaconnect.com/content/imp/jcs/2015/00000022/f0020007/art00008

Ritchie, J. B., Kaplan, D. M., & Klein, C. (2019). Decoding the brain: Neural representation and the limits of multivariate pattern analysis in cognitive neuroscience. *The British Journal for the Philosophy of Science*, *70*(2), 581-607. https://doi.org/10.1093/bjps/axx023

Schlicht, T. (2012). Phenomenal consciousness, attention and accessibility. *Phenomenology and the Cognitive Sciences*, *11*(3), 309-334. https://doi.org/10.1007/s11097-012-9256-0

Schlicht, T. (2018). A methodological dilemma for investigating consciousness empirically. *Consciousness and Cognition*, *66*, 91-100. https://doi.org/10.1016/j.concog.2018.11.002

Simons, D. J., & Ambinder, M. S. (2005). Change blindness: Theory and consequences. *Current Directions in Psychological Science, 14*(1), 44-48. https://doi.org/10.1111/j.0963-7214.2005.00332.x

Sligte, I. G., Scholte, H. S., & Lamme, V. A. (2008). Are there multiple visual short-term memory stores?. *PLOS One*, *3*(2), Article e1699. https://doi.org/10.1371/journal.pone.0001699

Sligte, I. G., Vandenbroucke, A. R., Scholte, H. S., & Lamme, V. A. (2010). Detailed sensory memory, sloppy working memory. *Frontiers in Psychology*, *1*, Article 175. https://doi.org/10.3389/fpsyg.2010.00175

Sperling, G. (1960). The information available in brief visual presentations. *Psychological Monographs: General and Applied*, *74*(11), 1-29. https://doi.org/10.1037/h0093759

Summerfield, C., & de Lange, F. P. (2014). Expectation in perceptual decision making: Neural and computational mechanisms. *Nature Reviews Neuroscience*, *15*(11), 745-756. https://doi.org/10.1038/nrn3838

Tsuchiya, N., Wilke, M., Frässle, S., & Lamme, V. A. (2015). No-report paradigms: Extracting the true neural correlates of consciousness. *Trends in Cognitive Sciences*, *19*(12), 757-770. https://doi.org/10.1016/j.tics.2015.10.002

Radboud University

Research proposal

Mind, Meaning, Information: Integrating Enactivism and Haig's Information

Summary of the theme and aim of the project

Enactivism understands the mind as an agent's making sense of the world through its interactions with it. Autopoietic (AE) and sensorimotor enactivism (SME) present competing approaches to sense-making, each with their own domains of application. Radical enactivism (RE) criticises theories of sense-making for staying within a former, problematic framework of thinking about the mind but does not propose an alternative theory. These developments impede progress in the enactivist framework towards a general understanding of the mind. This project proposes to investigate whether a novel concept of information could be used to develop a unifying account of enactivism. A unified account is necessary to determine to what extent enactivism can provide a satisfactory explanation of the mind in general. To arrive at such an account, existing theories of sense-making will be investigated for commonalities in their concepts of sense-making. In addition, an overview of RE's critique will make clear which concepts are problematic and why. A novel concept of information will be compared against these commonalities in sense-making and analysed for potentially using such problematic concepts. If an informational notion of sense-making is adequate, it will bring the enactivist framework one step closer to realising its potential to reform the mind. *(200 words)*

Keywords: autopoietic, sensorimotor, radical, enactivism, information, sense-making

Introduction: the problem

Enactivism is relatively recent framework to explain the mind which is rapidly gaining in popularity and impact. For enactivists, the mind is the way an organism makes sense of the world that it perceives. The notion of sense-making, central to enactivism, was first formulated with respect to living organisms. Their sense-making is theorised to happen in terms of its relation to an organism's continued survival (autopoietic enactivism).[3] Later, a second formulation of sense-making posited that it is inherent in a contingency between perception and action (sensorimotor enactivism).[4] Finally, radical enactivism aims to rid all enactivist theorising about mind from references to content or meaning, thereby undermining the use of sense-making to explain the mind in the first place.[5]

What started out as a novel approach to thinking about the mind is rapidly formulated and interpreted in different ways, distracting away from the general development of enactivism as a framework that can reconceptualise the mind. The present project aims to establish the shared idea behind the two formulations of sense-making developed so far and investigate whether a recently proposed notion of information can figure in a general theory of sense-making. At the same time, this proposal will be evaluated against the points raised by radical enactivism. This way, the project will contribute to realising enactivism's philosophical, scientific and societal impact (see also section 4).

Philosophical and scientific background

According to enactivism, the mind should be understood as the way an organism makes sense of its world. When an external stimulus impinges on the organism's receptors, the organism's stimulus processing (e.g. neural activity) is not inherently informative of what caused it; for humans, temperature, light, and sounds are all converted into neural activity patterns. Enactivism maintains that organisms are nevertheless able to (re)assign meaning to stimuli by behaving differentially with

[3] Varela et al. (1991)
[4] Degenaar & O'Regan (2017); O'Regan & Noë (2001a)
[5] Hutto & Myin (2012)

respect to them. To illustrate, for something to mean 'edible' we should tend to eat and digest it.

As a novel way to understand the mind, enactivism has potentially wide-ranging implications. Understanding the mind has not just been a long-standing problem in philosophy, a fundamentally different view on the mind could affect psychology[6] and the (cognitive) neurosciences by changing our thinking about cognitions as their objects of scientific investigations[7]. Furthermore, the large role of action in the enactivist's conception of the mind will likely change the way we look at physical and mental disorders, as well as how we should relate to them.[8] As the framework is still developing, it is unclear how enactivism will influence these areas.

The main obstacle preventing the framework from developing its central idea of sense-making into a way to understand the mind is that enactivism has been interpreted in various different ways. The different approaches that all fall under the umbrella of enactivism compete for acknowledgment and priority, distracting away from the important development of sense-making and its role in explaining the mind.

One point of contention is the operationalisation of sense-making. Autopoietic enactivism (AE) applies enactivism to living organisms. Organisms need to stay alive, which gives them a reason to respond differentially to what they perceive and therefore make sense of their perceptions. As a simple example, they will approach things that are beneficial to their health and move away from things that impede survivability.[9] This context explains an important AE claim that 'life is mind', or in other words that life gives on organism a reason to engage in sense-making.

Sensorimotor enactivism (SME) applies enactivism to (intelligent) systems.[10] These could be designed to discriminate between stimuli in terms of the system's

[6] Krueger (2021)
[7] Segundo-Ortin & Hutto (2021)
[8] de Haan (2020b); Krueger (2021)
[9] Varela (1997)
[10] Degenaar & O'Regan (2017); O'Regan & Noë (2001a)

behaviour, meaning they would be making sense of their 'perceptions'. This emphasis allows them to investigate when certain systems engage in sense-making, moving towards questions about artificial intelligence and artificial minds. Such a focus required proponents of SME to adopt a different operationalisation of sense-making, namely the existence of sensorimotor contingency loops: perception-guided behaviour that influences the same perception (consider a tracking system that changes its orientation to keep the target in its centre of 'view').

There are even indications that further developments in enactivism will result in another suggestion on how to explain sense-making. The application of enactive ideas to cases where there is no overt behaviour have prompted several authors to suggest sensori-attentional contingencies as a replacement.[11] Although this variety has not been developed into a full account of sense-making (yet), it would be another example of enactivism being applied to explain some limited set of cases.

The various enactivist theories each have their specific backgrounds from which they develop their view on sense-making. These are subsequently presented as a general theory of sense-making, which naturally leads to debates about which one should be considered the better alternative. Such debates emphasize differences between them and no doubt develops and improves their theories[12], but also proceeds to delineate each theory's boundaries of application. Without efforts to unify the different proposals, the development of a generally applicable theory of sense-making will only be of secondary concern. The much-anticipated potential of enactivism to provide an alternative understanding of the mind is therefore slowed in its realisation.

A third main strand of enactivism, radical enactivism (RE), has taken on the task of developing the enactivist approach. Enactivism is a fundamentally different way of thinking about the mind, so RE attempts to make sure enactivist theories do not accidentally revert to previous frameworks' (unsatisfactory) conceptions of the

[11] Nusbaum et al. (2001); O'Regan & Noë (2001b); Virsu & Vanni (2001)
[12] E.g. Di Paolo (2005); Degenaar & O'Regan (2017)

mind.[13] As such, it does not present a theory but rather a set of enactive-oriented proposals and arguments. These proposals try to reconceptualise cognitions and the mind in enactivist terms, rejecting terms that it considers to be problematic left-overs from previous attempts at explaining the mind.

The theories of sense-making discussed earlier use concepts like meaning, content, and information in their operationalisations of sense-making. However, RE considers these terms to be problematic[14], hence rejecting such theories. Where RE does laudably fulfil a watchdog function, protecting the enactivist framework from slipping into using problematic concepts and arguments, it does not provide an alternative theory.[15] Instead of evaluating strengths of both theories and developing an improved synthesis, this critique ultimately leads to debates about whether existing theories are enactivist enough (i.e. sufficiently radical).[16] Again, such debates distract away from developing a broadly applicable notion of sense-making to work out how enactivism would affect philosophy of mind, the sciences of mind, and societal issues.

Developments in enactivism, both in terms of specific theories that work out a concept of sense-making as well as radical enactivism's critique on such theories, seem to have lost focus on the general development of the enactivist idea to understand the mind. Such focus can be restored by investigating alternatives that unify these various strands of enactivism. Alternatives must be acceptable to radical enactivism or otherwise comply with the conditions it places on explanations of the mind, as well as capture a generally applicable operationalisation of sense-making.

Research project

Stage 1: Overview of enactivism

The various approaches to enactivism each come with their own assumptions and definitions. An overview of the enactivism framework will reveal on which aspects

[13] Hutto & Myin (2012); Hutto (2017)
[14] Hutto (2005, 2011, 2017)
[15] Hutto (2017)
[16] E.g. Kee (2021)

the enactivist theories of sense-making differ and, more importantly, agree. Their points of agreement will specify the requirements that an alternative concept should adhere to if it is to capture a shared idea behind competing formulations of sense-making.

To produce such an overview, I will mainly analyse the concepts used in current approaches to enactivism. This means that first the concepts used in autopoietic, sensorimotor, and possibly sensori-attentional enactivism (insofar as its limited development allows for such an analysis) will be mapped and compared. Concepts that recur between theories will provide important indications about which new formulations could capture shared intuitions behind AE and SME. As a concrete example, a possible result could be that AE and SME share the idea than an agent must engage in certain specific interactions with its environment. The analysis will additionally indicate what the two views require of the concepts 'agent' and 'interaction' to make their theories work.

Radical enactivism does not present a single theory, so it cannot be part of such a conceptual analysis. Nevertheless, a proposal that hopefully unifies enactivist approaches must take RE into account. The second part of the overview developed in stage 1 will investigate why RE criticises the use of certain concepts in AE and SME. This investigation is necessary to understand which concepts and lines of reasoning were considered unsatisfactory and for which reasons, in order to avoid a unificatory account of enactivism from not being enactivist enough.

Stage 2: Investigating a possible role for information in sense-making

After having elucidated the requirements for a unificatory account of enactivism, it will be possible to investigate new ways to conceive of sense-making. In this stage, a recently developed notion of information[17] will be considered as possibly figuring in a generally acceptable notion of sense-making.

[17] Haig & Dennett (2017)

Information is one of the terms that RE tries to avoid, so at first sight it might seem odd to propose to unify sense-making through such a notion. Yet, there are reasons to believe such an investigation will be worthwhile. Haig proposes a radically new view on information in a time where its traditional limitations are well-known and Haig's notion of information seems to share important characteristics with current conceptions of sense-making.

In order to investigate whether Haig's proposal could unify enactivist theories, I will first perform a conceptual analysis of Haig's proposal. Similar to that analysis of enactivist theories, it should indicate which concepts play important roles and how they are defined. Before applying them to sense-making, however, I will first investigate how they relate to RE's critique on those concepts. If Haig's revision of information and meaning does not fully escape such critique, I will try to find out if Haig's proposal can be adapted by proposing slight reformulations.

When the concepts used by Haig are successfully protected against such critique, they will be compared against the requirements that an operationalisation of sense-making should adhere to. This might require additional reformulation and comparisons: Haig's notion of 'interpreter' does not necessarily overlap with that of 'agent' in AE, even though they seemingly occupy a similar role in their respective theories.

Stage 3: Developing 'informational enactivism'

Assuming that some (adapted) version of Haig's notion of information can figure in a reformulation of sense-making, this third stage will work that notion of sense-making out in more detail. In the case that it cannot, it will be useful to identify which aspects were ultimately bottlenecks. They can then inform further attempts to understand what makes them problematic.

The judgment that Haig's information is compatible with existing theories of sense-making is merely a first step in unifying enactivist approaches to the mind. A second step is to create and evaluate an informational account of sense-making. Its general

applicability should be confirmed and there might be new issues that did not apply to other theories of sense-making.

'Informational sense-making' can then be developed into a more extensive proposal of how to understand the mind. This will require addressing problems that have been foreseen for the enactivist framework but that AE and SME proposals have not yet addressed. These include the cognitive gap[18], which suggests that enactivism may have trouble explaining 'higher' cognitions for which the role for behaviour is less evident, and possibly an explanatory gap in understanding why the mind's supposed physical realisation creates the mind.[19]

Depending on the available time at this stage of the project, further developments of what might at this point be called 'informational enactivism' will address implications for fundamental philosophical, psychological and (neuro)scientific issues. Currently, these implications are highly speculative, but a successful unification of enactivist views of the mind likely has consequences for the way we understand the minds of other people, organisms and maybe systems, for how we divide the mind into cognitions, and for how the (cognitive) neurosciences operationalise and investigate mental states.

Philosophical, scientific, and societal relevance

The current project proposes to develop the enactivist framework in such a way that fully explores its potential. The unification of the various strands of enactivism under a single shared concept of sense-making will be a large step towards investigating to what extent enactivism can explain the various aspects of the mind. In addition, it will lay the foundation to working out the concrete ways in which enactivism affects our thinking about the minds of other organisms and systems.

A reconception of mind will change the way we look at and investigate cognitions. Until recently, cognitions have often been thought of as manipulations on mental

[18] de Bruin & de Haan (2012); de Bruin & Kästner (2012)
[19] See Levine (1983); Kirchhoff & Hutto (2016)

representations of the outside world.[20] With the mind conceived of in enactivistic terms, it is likely that the idea of mental representations will be revised. More concretely, the large role for behaviour in the constitution of mind will have consequences for (cognitive) neuroscientific research methods. Most require subjects to move as little as possible so as to pick up on the tiny fluctuations in magnetic fields or electric currents caused by neural activation, but with action being integral part of the enactivistic mind, it might require such methods to be re-evaluated.

The reconception of cognitions in enactivism has already made its influence on psychiatry.[21] De Haan's enactive view on mental disorder changes the focus from physical malfunctioning to disturbances in the way a person makes sense of their physical and social environment. Along with it, there are implications patients' understanding of themselves[22] and the types of interventions[23] that contribute to patients' well-being. The current project will unify the enactivistic understanding of minds, facilitating and disambiguating the application of enactivism to psychiatric disorders and their treatments.

2485 words

[20] See e.g. van Gelder (1995)
[21] de Haan (2020b, 2021)
[22] de Haan (2020c)
[23] de Haan (2020a); de Haan et al. (2013)

Bibliography

de Bruin, L. C., & Kästner, L. (2012). Dynamic embodied cognition. *Phenomenology and the Cognitive Sciences*, *11*(4), 541-563. https://doi.org/10.1007/s11097-011-9223-1

de Bruin, L. C., & de Haan, S. (2012). Enactivism & social cognition: In search of the whole story. *Cognitive Semiotics*, *4*(1), 225-250. https://doi.org/10.1515/cogsem.2012.4.1.225

de Haan, S. (2020a). Enactive causality: Interventions, cakes, and clockworks: A reply to Gallagher and Donovan and Murphy. *Philosophy, Psychiatry, & Psychology*, *27*(1), 31-33. https://doi.org/10.1353/ppp.2020.0004

de Haan, S. (2020b). *Enactive psychiatry*. Cambridge University Press.

de Haan, S. (2020c). The need for relational authenticity strategies in psychiatry. *Philosophy, Psychiatry, & Psychology*, *27*(4), 349-351. https://doi.org/10.1353/ppp.2020.0044

de Haan, S. (2021). Bio-psycho-social interaction: An enactive perspective. *International Review of Psychiatry*, *33*(5), 471-477. https://doi.org/10.1080/09540261.2020.1830753

de Haan, S., Rietveld, E., Stokhof, M., & Denys, D. (2013). The phenomenology of deep brain stimulation-induced changes in OCD: An enactive affordance-based model. *Frontiers in Human Neuroscience*, *7*, Article 653. https://doi.org/10.3389/fnhum.2013.00653

Degenaar, J., & O'Regan, J. K. (2017). Sensorimotor theory and enactivism. *Topoi*, *36*(3), 393-407. https://doi.org/10.1007/s11245-015-9338-z

Di Paolo, E. A. (2005). Autopoiesis, adaptivity, teleology, agency. *Phenomenology and the Cognitive Sciences*, *4*(4), 429-452. https://doi.org/10.1007/s11097-005-9002-y

Haig, D., & Dennett, D. C. (2017). Haig's 'strange inversion of reasoning' (Dennett) and Making sense: Information interpreted as meaning (Haig). http://philsci-archive.pitt.edu/id/eprint/13287

Hutto, D. D. (2005). Knowing what? Radical versus conservative enactivism. *Phenomenology and the Cognitive Sciences*, *4*(4), 389-405. https://doi.org/10.1007/s11097-005-9001-z

Hutto, D. D. (2011). Enactivism: Why be radical? In H. Bredekamp & J. M. Krois (Eds.), *Sehen und handeln* (pp. 21-44). Akademie Verlag. https://doi.org/10.1524/9783050062389.21

Hutto, D. D. (2017). REC: Revolution effected by clarification. *Topoi*, *36*(3), 377-391. https://doi.org/10.1007/s11245-015-9358-8

Hutto, D. D., & Myin, E. (2012). *Radicalizing enactivism: Basic minds without content*. MIT press.

Kee, H. (2021). Phenomenology and naturalism in autopoietic and radical enactivism: Exploring sense-making and continuity from the top down. *Synthese*, *198*(9), 2323-2343. https://doi.org/10.1007/s11229-018-1851-3

Kirchhoff, M. D., & Hutto, D. D. (2016). Never mind the gap: Neurophenomenology, radical enactivism, and the hard problem of consciosuness. *Constructivist Foundations*, *11*(2), 346-353. http://constructivist.info/11/2/346

Krueger, J. (2021). Enactivism, other minds, and mental disorders. *Synthese*, *198*(1), 365-389. https://doi.org/10.1007/s11229-019-02133-9

Levine, J. (1983). Materialism and qualia: The explanatory gap. *Pacific Philosophical Quarterly*, *64*(4), 354-361. https://doi.org/10.1111/j.1468-0114.1983.tb00207.x

Nusbaum, H. C., Skipper, J. I., & Small, S. L. (2001). A sensory-attentional account of speech perception. *Behavioral and Brain Sciences*, *24*(5), 995-996. https://doi.org/10.1017/S0140525X01000115

O'Regan, J. K., & Noë, A. (2001a). A sensorimotor account of vision and visual consciousness. *Behavioral and Brain Sciences*, *24*(5), 939-973. https://doi.org/10.1017/S0140525X01000115

O'Regan, J. K., & Noë, A. (2001b). Authors' response: Acting out our sensory experience. *Behavioral and Brain Sciences*, *24*(5), 1011-1031. https://doi.org/10.1017/S0140525X01000115

Segundo-Ortin, M., & Hutto, D. D. (2021). Similarity-based cognition: Radical enactivism meets cognitive neuroscience. *Synthese, 198*(1), 5-23. https://doi.org/10.1007/s11229-019-02505-1

van Gelder, T. (1995). What might cognition be, if not computation? *The Journal of Philosophy, 92*(7), 345-381. https://doi.org/10.2307/2941061

Varela, F. J. (1997). Patterns of life: Intertwining identity and cognition. *Brain and Cognition*, *34*(1), 72-87. https://doi.org/10.1006/brcg.1997.0907

Varela, F. J., Thompson, E., & Rosch, E. (1991). *The embodied mind: Cognitive science and human experience*. MIT Press.

Virsu, V., & Vanni, S. (2001). Perceptions as hypotheses of the outside world. *Behavioral and Brain Sciences, 24*(5), 1009-1010. https://doi.org/10.1017/S0140525X01000115

Timetable

Stage	*Year*	*Description*	*Expected product*
1	1	Conceptual analysis of theories of sense-making	Overview article
1		Overview of critique on representationalism	(Review article)
2	2	Analysing Haig's notion of information in terms of critique on representationalism	Paper
2		Comparing Haig's notion of information to requirements for sense-making	(Paper)
3	3	Developing 'informational sense-making'	Paper
3		Developing 'informational enactivism (IE)'	Paper
3	4	Working out implications of IE	Paper
4		Synthesis, integration of above projects	PhD thesis

Summary for non-specialists

A popular way to think about our minds is as the presentation of a multisensory 'movie' to the self. Our senses detect which objects we encounter in our environment and re-present those in our mental movie. Enactivism challenges this view and conceives of the mind as actively constructing the objects and categories that we ascribe to the environment. The mind is seen as continuously making sense of the sensations impinging our bodies in terms of objects or things that caused those sensations by behaving differentially to them. Shifting our thinking about the mind will have important implications for the way we think about human and other minds.

There are various views in enactivism that explain sense-making. Autopoietic enactivism applies these ideas to living organisms, who use their health to categorise their sensations. External influences on an organism can be beneficial or detrimental to its health, so at a basic level the organism can make sense of those influences by respectively approaching or withdrawing from them. Sensorimotor

enactivism applies enactivism mostly to non-human systems, which can be designed in such a way that they make sense of their inputs. Radical enactivism aims to maintain enactivism as a fundamentally new conception of the mind. This means that it should avoid using concepts and ways of thinking that were found problematic in previous conceptions, which autopoietic and sensorimotor enactivism do not comply with.

These different developments in enactivism each have their own goals with the enactivist framework, slowing the framework's development in becoming a serious alternative to representational thinking. A unification of these developments should accelerate this development, which is the goal of this project. In the first stage, autopoietic and sensorimotor enactivism will be analysed to reveal commonalities in their concepts of sense-making that a unified account should be based on. In addition, an overview of radical enactivism's critique on current theories of sense-making will reveal which kinds of concepts and argumentation are flawed and, more importantly, why. These steps will provide the necessary background to consider a new concept to unify theories of sense-making in stage 2.

Recently, a new proposal connects a revised concept of information to the concept of meaning. Although these terms are considered problematic in radical enactivism, their revised versions of them might avoid the issues with previous formulations. At the same time, the connection between information and meaning might make it possible to construct a theory of sense-making that unifies autopoietic and sensorimotor proposals. These suggestions need to be investigated for adequacy, which the analyses in stage 1 will be used for.

If this new notion of information is considered appropriate to support a new theory of sense-making, stage 3 will develop this theory of sense-making and confirm that it indeed unifies autopoietic and sensorimotor theories. With a unified notion of sense-making, a full enactivist explanation of the mind can be developed. That will bring the enactivist framework a step closer to realising its potential in revising the way we think about the mind.

498 words

Curriculum Vitae

1. *Personal data*

Full name:	Jochem Giel Koopmans
Date of birth:	2 September 1997
Address:	Graafseweg 8, 6512CB Nijmegen (the Netherlands)
Phone number:	+316 48495562
Email address:	Jochem.koopmans@ru.nl
Nationality:	Dutch
Gender:	Male
Academic interests:	Enactivism, consciousness, representationalism, information, meaning, difference, brain

2. *Education*

Year	Description	GPA
2020-2022 (expected)	M.A. Research Master Philosophy (Philosophy of Mind) *Radboud University Nijmegen*	7.8 (20 June 22)
2019-2022 (expected)	M.Sc. Research Master Cognitive Neuroscience *Donders Institute for Brain, Cognition and Behaviour;* *Radboud University Nijmegen*	8.0 (20 June 22)
2019-2020	Premaster Philosophy *Radboud University Nijmegen*	7.8
2016-2018	Radboud Honours Academy for Bachelor students *Radboud University Nijmegen*	8.5
2015-2018	B.Sc. Psychology *Radboud University Nijmegen*	Cum laude
2015-2016	Radboud Honours Academy for first-year students	N/A

	Radboud University Nijmegen	
2015-2016	Minor Sociology *Radboud University Nijmegen*	7.5

3. Relevant academic experience

Internships

2020- 2022	Research internship (master's thesis) *Visual Cognitive Neuroscience lab, Donders Institute for Brain, Cognition and Behaviour*
2017- 2018	Research internship (bachelor's thesis) *Brain Stimulation and Motivational Control lab, Donders Institute for Brain, Cognition and Behaviour; Philosophy of Cognition and Language, Radboud University Nijmegen*
2016- 2017	Research internship (Radboud Honours Academy) *Sensorimotor Research Lab, Donders Institute for Brain, Cognition and Behaviour*

Conferences

2022 (July)	Attendee of the Association for the Scientific Study of Consciousness 25th conference *(https://theassc.org/)*
2018	Attendee of the Association for the Scientific Study of Consciousness 22nd conference *(https://theassc.org/)*

Teaching

2022	Teaching assistant *Philosophy of Mind, Brain, and Behaviour (Psychology B2)*

2017	Teacher for Athena Studies (https://athenastudies.nl/) *Statistics 1 (Psychology B1)*

Other relevant positions

2021-2022	General member of the Donders Institute's Diversity and Sustainable Science Committee *(https://www.ru.nl/donders/about-us/diversity-sustainable-science/)*
2019-2021	Board member of Dondrite, Study association Cognitive Neuroscience *(https://dondrite.ruhosting.nl/)*

4. Relevant academic output

2022 (expected)	On a confusion about access consciousness: How scientists and the brain access information about consciousness *Research master's thesis Philosophy* *Supervisor: prof. dr. Marc V. P. Slors*
2022 (expected)	How our predictions cause unexpected letters to (not) be illusorily perceived *Research master's thesis Cognitive Neuroscience* *Supervisors: prof. dr. Marius V. Peelen, dr. Genevieve Quek, Sushrut Thorat M.Sc.*
2022	Exploring the role of action in sense-making theories of mind *Essay in partial completion of course 'Neurophilosophy, Subjectivity, and Phenomenology' at the VU Amsterdam* *Professor: prof. dr. Gerrit Glas*
2018	A review of arguments and theories that distinguish between consciousness and unconsciousness *Review article in partial completion of the Bachelor's thesis Psychology*

	Supervisor: dr. ing. Léon C. de Bruin
2018	A comparison between introspection and the P3 event-related potential in measuring consciousness *Bachelor's thesis Psychology* *Supervisor: dr. Dennis J. L. G. Schutter*
2017	The contribution of visual and motor vectors in action selection *Research article* *Supervisors: prof. dr. W. P. (Pieter) Medendorp, Lonneke Theunissen M.Sc.*

5. Languages

Dutch	Fluent (native)
English	Fluent
Spanish	Good
French	Conversational
German	Basic