

-

## BALKAN JOURNAL OF PHILOSOPHY

*Vol. 17, Issue 1, 2025*

The issue is sponsored by the Bulgarian National Science Fund, competition “Bulgarian Scientific Periodicals – 2025”, project No: КП-06-НП6/61 from 05.12.2024.

## TABLE OF CONTENTS

Special issue devoted to the topic of “VARIETIES OF REPRESENTATIONS”

Varieties of Representations / 3

*Madelaine Elchinova – Angelova, Guest Editor*

### **Articles on “Varieties of Representations”**

Invited Paper

What Do Emotions Represent / 6

*Jesse Prinz*

Invited Paper

Multilevel-Grounded Semantics: Representation, Intension and Referentiality in Music / 22

*Mihailo Antović*

Where representations begin: from basic to decoupled representations / 38

*Anastasia Garbayo de la Torre*

Presentation and Content / 54

*Ivan V. Ivanov*

Rosenthal’s Representationalism: Is Consciousness a Higher Order Thought? / 67

*Ksenija Puškarić*

Revisiting Searle's Refutation of Sense Datum Theory of Perception / 80

*S. Sreenish*

Understanding 'Representation-Eschewing' Approaches: An 'Ancient' View / 92

*Hari Narayanan V and Yogeshwar Joshi*

## VARIETIES OF REPRESENTATIONS

Madelaine Angelova - Elchinova\*, Guest editor

The notion of representation has been in use in a wide variety of scientific discourses as well as in our everyday language. We can talk about maps as providing symbolic representations of real or fictional places. We can discuss the role of randomized clinical trials as representing the efficacy of a given medical treatment or highlighting how certain groups were underrepresented in a given trial. We can deliberate on the idea of representational art as depicting objects or events that we can identify. We can insist that artefacts like thermometers or speedometers represent the outside temperature or the speed of our car. We can make claims to how our language represents reality in a very broad sense. Representations are also part of the vocabulary of cognitive scientists, psychologists, neuroscientists and philosophers. In all these fields representational talk has been accompanied by descriptions of a given representation as being *about* an object or as carrying and providing *information* to the system.

Because of the wide variety of applications and usages that the term ‘representation’ has, it is important to clarify that this journal issue discusses representations in a narrower sense. We were looking for and we received submissions that were discussing representational theories particularly in philosophy. In the field of philosophy, representations have been mostly associated with the Representational Theory of Mind (RTM). It can be debated for how long exactly RTM has been on the market. Some may find its origins in Aristotle’s philosophy, while others would settle for moving the timeline up to John Locke’s imagistic ideas.

Regardless of its roots, RTM has been motivated by a couple of simple assumptions that can be regarded as its cornerstones. First, mental states like beliefs, desires, imaginings or emotions have *aboutness* or they *refer* to something. Consequently, if they successfully fulfil their representational function, such mental states can be appraised as true, accurate or successful. If something is off with a given representational “device”, we can describe what is going on as a misrepresentation that fails to fulfil the required role. To give an example, if beliefs are described as propositional representational states, a false belief would amount to a misrepresentation of the world. Different representational theories may largely vary about the details, but they tend to agree on the point that representations are *intentional* in the aforementioned sense. Second, representations can be about or be representations of real or fictional objects. If one represents Mordor as a coastal region in Middle Earth, we can still argue that one is misrepresenting Mordor even if it is a fictional land. Further, even when representations are representations of real objects, unlike presentations that would necessarily require the presence of the object, representations can be just as accurate if the represented object is not currently present. Take one’s memory of what kind of cake one had for their last birthday. If it was a chocolate cake with sprinkles on top of it, one can have an accurate memory representation of the cake as a chocolate cake with sprinkles on top of it many years later. Finally, it is now almost universally accepted amongst philosophers that representations may be conceptual or non-conceptual. Take the sensory representations of a given perceptual system – say one is walking around the centre of the city while entertaining

\*Department Philosophy, Sofia University “St. Kliment Ohridski”; email: [madelaine.angelova@gmail.com](mailto:madelaine.angelova@gmail.com);

ORCID ID: <https://orcid.org/0000-0002-5961-0327>

a thought about universal properties. At the same time, one is just as capable of navigating around the crowd as to not collide with other pedestrians or street lamps. While conceptually representing a propositional content about universal properties, one is non-conceptually representing their environment.

Regardless of the common grounds, there are yet many points of disagreement amongst holders of representational views. In the current volume one can find some inside in regards to the following ongoing debates:

In the first paper, Jesse Prinz raises the question of what the content of emotional experiences represents and how it represents it. Existing cognitive as well as noncognitive accounts of emotional representation are examined and scrutinised and an imperatival account of emotional content is put forward. On the account defended by Prinz, emotions do not represent external things like dangers, but rather internal demands in the sense of to-be-doneness, e.g. fearing a bear may represent the object as demanding escape.

The second paper by Mihailo Antović offers a glimpse into the way the concept of representation is used in cognitive linguistics. It proposes an examination of semantic representations and more precisely – the ones related to musical meaning. Antović offers a defence of multi-level grounded semantics on which music representation is examined as amodal functionalist construct relating music perception to an extramusical experience.

In the third paper, Anastasia Garbayo de la Torre takes us on a journey into the evolutionary history of representations by examining the question, “Where do representations begin?” The paper proposes an account of the minimal conditions that an organism must fulfil in order to qualify as representational, with respect to both basic and decoupled representations. The case of box jellyfish is examined as a target organism.

The next paper by Ivan V. Ivanov engages with a classical debate in the philosophy of perception between the two predominant views – representationalism and disjunctivism (or naïve realism). A novel defence of disjunctivism is proposed on the grounds that it fares better than intentionalist theories in accounting for the person-level representational contents of experiences.

Ksenija Puškarić provides a reconstruction of Rosenthal’s representational theory of consciousness. It provides an insight into the so-called HOT theories of consciousness and addresses the status of qualia on such a theory. Puškarić also attempts to provide an answer to the question: if transitive consciousness is factive, and thoughts are not, how can all HOTs constitute transitive awareness?

The sixth paper returns to debates in the field of philosophy of perception as it offers a critical review of Ayer’s sense datum theory by appealing to Searle’s direct realism. While S. Sreenish’s exposition does not directly address contemporary intentionalist accounts, it aims to provide reasons to consider the sense datum theory as related to representationalism in terms of it being a form of indirect realism about perception.

The final paper by Hari Narayanan V and Yogeshwar Joshi brings us back to the ancient roots of ‘representation-eschewing’ approaches (e.g. enactivism) that attempt to provide an account of cognition without directly appealing to representations. The authors claim that such approaches offer objective participant models of understanding the self and contrast

them to the subjective individualist models which are presented as featuring in representational theories. The argument appeals to cultural differences in ancient modes of living and proposes that ‘representation-eschewing’ approaches were intuitively more plausible in ancient times.

*Invited paper*

## WHAT DO EMOTIONS REPRESENT

Jesse Prinz\*

**Abstract:** It is widely agreed that emotions have intentional objects. In fact, they are said to represent both formal objects (e.g., fear represents danger) and particular objects (e.g., one might be afraid of a spider or an exam). But there is little agreement about how emotions represent. The challenge of explaining how emotions represent is deepened on theories that associate emotions with somatic states of some kind (e.g., William James's view that emotions are feelings of bodily change). Emotions represent things outside the body, but somatic states, such as bodily perceptions, are usually thought to represent bodily things. This puzzle (called here the Somantodox) has been addressed in a variety of ways. A number of proposals are reviewed here: embodied appraisal theory, affordances, Sartre's existential theory, and enactivism. Building on the strengths and weaknesses of these, a new proposal is offered called the embodied imperative theory, according to which the body imparts a sense of to-be-doneness. Rather than saying fear represents danger, it would be more perspicuous, on this proposal, to say fear represents an object or situation as something to-be-escaped. This is termed a core-dynamic theme.

**Keywords:** emotions, formal and particular objects, enactivism, existentialism, affordances, embodiment, imperatives

### 1. Introduction

Suppose, to take a time-honored example, you encounter a bear while hiking in the mountains. You feel afraid. You must decide what to do: Stand still? Run? Serve yourself up as a tasty meal? The fear prods us to reflect, to make a decision, and act. To do so, it seems plausible to think of the fear as a mental representation. It presents the bear to us as being a certain way – being fearsome, one might say. What does fearsome mean? What is the content of this experience such that it can factor into our decision making? What exactly do emotions represent? And how do they represent?

There are many possible answers to these questions. My aim here is to review some of the answers that can be culled from the literature and to suggest an alternative, building on what has come before. I will propose a theory of what emotions represent, or a theory of emotional content (since, as we will see, the term “represent” may carry baggage that we would do well to offload). I will begin with cognitive approaches to emotional representation, and reject them on general grounds – the implausibility of cognitivism. Then I will survey several noncognitive alternatives, arriving at a final recommendation, which posits imperativational content.

\*CUNY Graduate Center; jesseprinz@gmail.com; ORCID ID: <https://orcid.org/0000-0001-6855-0023>

Before embarking, it will help to draw a distinction familiar in the philosophy of emotions that will help us clarify what we are after. It is often noted that emotional episodes have two semantic objects, not just one (Kenny, 1963). Your fear of the bear represents the bear. It is about the bear. That is the “particular object” to which it refers. In addition, there is a formal object – the kind of thing that fear in general is in the business of designating. A standard view, which I will ultimately challenge, is that the formal object of fear is danger. On this analysis, your fear of the bear represents the bear (its particular object) and danger (its formal object). More accurately, it represents the bear as dangerous.

Much of the emphasis in emotion research has focused on formal objects. One implicit reason for that is one might think of emotional episodes as composite mental states that adjoin emotions to representations of particular objects. Perhaps you have a visual representation of the bear, and you combine it with your state of fear. On this breakdown, the visual state represents the bear, and the fear represents the danger you are in. If so, emotions really only represent formal objects. Particular objects are represented by other kinds of mental representations. Though this is broadly correct, I think it is a mistake to neglect particular objects entirely. For somehow, the content of the emotion must be attached to those objects. We need a story of how this works. If we ignore particular objects, we might arrive at a theory of formal objects that cannot be easily attached to account for the content of an emotion episode considered as a whole. In addition, there may be ways in which the sharp distinction between formal and particular objects breaks down psychologically. The two might be interdependent in ways that deserve theoretical attention. Therefore, in seeking an answer to the question about what emotions represent, it is important to keep both formal and particular objects in mind.

## 2. Cognitive theories

### 2.1. Construals and Appraisals

If your fear of a bear can be glossed as a representation of the bear as dangerous, it is tempting to think of emotion on the model of predication. The fear serves as a property attribution to the bear: that bear is dangerous. From this perspective, emotional episodes look a lot like ordinary judgments. “That bear is dangerous” is akin to “that bear is carnivorous”. Both can be analyzed as concatenations of concepts, where each component concept might be further analyzed as structured mental representation (e.g., definitions, prototypes, clusters, or something along those lines).

Taking this as a starting place makes it tempting to adopt the view that emotions are cognitive. For present purposes, a cognitive theory of emotions can be defined as one that says emotions necessarily have conceptual structure. By that I mean that emotions are constituted in part by concepts. Cognitive theories have been popular in philosophy and psychology. The range from simple – e.g., emotions are “construals” (Armon-Jones, 1991) – to complex (emotions comprise multiple judgments (e.g., Smith & Ellsworth, 1985)). On a simple theory, fearing a bear can be analyzed as appending the concept of danger to bear representation as a predicate (or perhaps as an adjectival modifier).

A more complex view might deploy the machinery of an appraisal theory of emotions (Arnold, 1960). In psychology, appraisal theories equate emotions with constellations of

judgments. Each judgment represents an organism-environment relationship that bears on well-being. For example, Scherer (2001) identifies 14 different appraisal dimensions, which he characterizes as “stimulus evaluation checks”. When you see the bear, you unconsciously assess whether it is novel, pleasant, goal relevant, in urgent need of response, under your control, certain in outcome, and so on. At no point in this process do you form the explicit judgment that the bear is dangerous, though that might arise later. The emotion consists in these finer-grained judgments. As Smith and Lazarus (1993) put it, appraisal consists of these very specific “molecular” judgments, and the concept of danger is a “molar” appraisal that theoreticians use to summarize these more complex mental states. Lazarus (1991) refers to the content of these moral appraisals as “core relational themes.” Danger is the theme of fear, loss is the theme of sadness, offense is the theme of anger, and so on. Many appraisal theories grant that emotions are also attended by physiological changes, but these are consequences of the appraisals in question. The representational content of an emotion derives from the appraisals, not from the somatic sequelae.

In sum, cognitive theories regard emotions as conceptual states, where the concepts involved may be simple or complex. To settle what emotions represent, we would then need to settle what these cognitive states represent, and we might ultimately adopt a theory of how cognitive states represent, drawing on the psychosemantics of concepts.

Simple construal theories and multi-dimensional appraisal theories will differ in the answer to the questions of what emotions represent, but they will agree that they represent in ways that parallel the nature of representation. At least they implicitly agree on this, since their proponents rarely provide any details about how concepts represent. Within the literature on concepts, there is no consensus view. One family of views, however, has been especially popular. According to these views, concepts get their content through an indication relation (Dretske, 1988). This can be cashed out in many different ways, but the basic idea is that concepts are like detectors. They are set up to be set off by whatever it is they represent, in much the way that the sound emitted by a smoke detector is set up so that it will sound off when fires are detected.

Cognitivist theorists could adopt a story like this. For construal theories, the idea would be that there is a mental representation that can function as a danger detector. When the bear is encountered, that concept is set off, but the concept does not represent the bear as such. That concept is not a bear detector. Rather it represents the property that sets it off in general, which is not bears but danger, since many different dangerous things will set it off.

It is interesting to observe that appraisal theories could also adopt a story like this, and they could do so in one of two compatible ways. First, they might think of each appraisal as a detector. For example, the appraisal that I have little control over the situation could work by detecting brain systems that register a loss of control. They could also say that each of these detectors, when taken collectively, detects certain properties in the world. The sum of all the fear appraisals might turn out to qualify as a danger detector. This would allow authors like Lazarus, to say that emotions represent both the specific features implicated by molecular appraisals, as well as the molar properties that they collectively detect: namely, core relational themes. When you fear a bear then, you would be said to combine a representation of that bear (itself a bear detector) with detectors for lost control,

goal incongruence, unpleasantness, and so on; and these might qualify as danger detectors when they all go off together. Expressed in language, the content of your fear would include both levels of details: that bear is incongruent with my goals, etc., and it is dangerous.

Properly elaborated, such an approach would provide a rich and attractive story about what emotions represent. It would help explain why emotions function like judgments in deliberation: they are judgments! Your fear represents the bear as dangerous and that becomes a premise for practical reasoning.

## 2.2. Problems with Cognitive Theories

The difficulty with the approach is that much evidence weighs against cognitive theories. Even if such theories could offer an answer to what emotions represent, that answer cannot be right because the theories can be ruled out on other grounds. Building a case against cognitive theories is beyond the scope of this discussion. In what follows, I will simply assume they are wrong, but let me briefly mention some familiar concerns (for more, see Prinz, 2004; Prinz, forthcoming).

Perhaps the most obvious concern is that cognitive theories are overly demanding. This is especially clear when we think about emotions in human infants or non-human animals. Crediting simpler creatures with concepts such as danger seems risky, and the demands only increase with dimensional appraisal theories, which imply that emotions encompass the concept of such as goal and control. Some appraisal theories, such as Scherer, argue that there are nonconceptual analogues of appraisals, which would shift such theories toward the non-cognitive side, but, even so, appraisal theories imply an enormous amount of mental processing every time we have an emotion. We must constantly be evaluating everything we perceive among the many dimensions they posit.

It is not just that cognitive theories are demanding; these demands also outstrip what seems necessary in many cases. Often, we can explain the elicitation of an emotion by sensory states. Consider injury, body horror, falling, sudden sounds, being tickled, hearing someone sob or scream, sexual stimulation, rotten smells, soothing colors, and uplifting music. In each case, it seems parsimonious to posit an intervening stage of judgment or construal. If you stub your toe and yell an expletive, it is unlikely that you first formed a judgment that the mishap was goal-incongruent, outside your control, and so on. Much less that you construe it an offense. The flash of ire seems to follow immediately upon impact. This is not just speculation. Research suggests the emotions can be elicited subcortically and via sensory pathways, bypassing brain centers associated with judgment (LeDoux, 1996).

This is not to deny that emotions can be cognitively elicited. That happens frequently, as when we meditate on the daily news, or stress over an ever-growing to-do list. But it seems very likely that emotions can be elicited without concepts or judgments, as in the aforementioned cases. The thrill of a sunset or fright from a crashing sound seem likely to arise without the aid of conceptualization. Such considerations have led many emotion theorists away from cognitive theories. And, if concepts are not necessary for having emotions, then concepts cannot be necessary for explaining how emotions represent.

### 3. Four Non-Cognitive Theories

#### 3.1. Embodiment and the Somantodox

According to non-cognitive theories, emotions can occur without any cognitive state (here understood as the deployment of concepts). One can also define cognitive theories as those that say (at least some) emotions have no cognitive components. Emotions themselves are not cognitive on these views. What, then, are they?

The most popular answer (or class of answers) is that emotions are in some sense embodied. These terms cover three possibilities, along with their combinations. It can refer to the idea that emotions are constituted by bodily perceptions (e.g., sensing a racing heart, arrested respiration, muscle tension, and so on), constituted by bodily commands (the control structures that regulate bodily organs and/or behavior), or constituted by states of the body itself.

Again, this is not the place to fully review evidence for bodily theories. But there is a large body of empirical work linking emotions and the body. For example, the brain areas associated with emotion overlap strikingly with brain areas involved in our body senses (Oosterwijk et al., 2012), and pattern classifiers show that individual emotions correlate with specific brain patterns associated in such brain areas (Kragel & LeBar, 2015) and specific patterns in our autonomic nervous systems (Kreibig, 2010). There are also expressive behaviors associated with a great many emotions (Keltner et al., 2019). When we look for somatic symptoms associated with any given emotion, we tend to find them. As actors can confirm, even highly specific emotions can be displayed through the body.

All this has led many to equate emotions with bodily states. But this conclusion is hard to square with the question at the heart of the present discussion. We began with the observation that emotions seem to represent things. And what they seem to represent are features of external objects, like bears, or more accurately, external relations, such as danger and loss. These are relations we bear to external objects. The difficulty arises when we reflect on what bodily states represent. Bodily perceptions and commands are generally assumed to represent states of the body. And states of the body are not presumed to represent anything at all.

This leaves us with a puzzle that can be formulated as an inconsistency. We have seen reasons for accepting each of the following claims:

- Affective Somaticism:* Emotions are bodily states
- Affective Externalism:* Emotions represent things outside our bodies
- Somatic Internalism:* Bodily states can only represent states of our bodies

Each of these seems true, but they cannot all be true. If emotions are somatic, and somatic states represent bodily states (if they represent at all), then emotions do not represent things outside our bodies. Let's call this the Somantodox, a portmanteau for somatic/semantic paradox.

As we will see, the usual way to address the Somantodox is to challenge Somatic Internalism. The scholars who have noticed this puzzle tend to reply by trying to show that

bodily states can refer to things outside the body, such as core relational themes. I will turn to such theories now and argue that they face serious objections.

### 3.2. Embodied Appraisal Theory

My own earlier solution to the (then-unnamed) Somantodox was syncretic. I argued that we could combine a bodily theory of what emotions are with a detector theory of how emotions represent. Recall that many theories of how concepts refer adopt the Dretskean (1988) adage that they refer to what they are set up to be set off by. Dretske himself often compares concepts to simple artifacts such as thermostats. What's striking about these examples is that artifacts do not contain concepts. So, implicit in his defense of the theory, is a recognition of the fact that the relation underlying reference does not hinge on the conceptual status of state that serves as the representational vehicle. In the case of the smoke detector, that vehicle is a high-pitched sound. In principle, then, anything can stand in a detection relation.

With that as my starting place, I proposed that bodily states can detect, and thereby represent, core relational themes (Prinz, 2004). A racing heart, with tense muscles, constrained breath, piloerection of the hair follicles, a spike in cortisol levels, a wrinkled brow, and a gaping jaw, collectively comprise a pattern that occurs in the presence of danger. When this pattern occurs, it thereby serves as a danger detector. (More specifically, I suggested that the inner perception of this pattern is a danger detector, since I was concerned to identify emotions with internal states that could also explain how emotions give rise to feelings and inform our decisions.) This struck me as a way to have my cake (Affective Somaticism) and eat it too (by accommodating Affective Externalism). I concluded that cognitive theories are right about emotional content, but wrong about the vehicles. I called this the Embodied Appraisal Theory.

I now think I was too hasty (see Shargel & Prinz, 2017; Prinz, forthcoming). For one thing, I've lost confidence in the entire program of detector semantics. One concern is that, as versions of the theory proliferated, it became increasingly clear to me that there was no way to empirically adjudicate between competing variants. Should we say that the tone of a smoke alarm represents smoke, or fire, or the possibility of fire? It detects all three, and it's not clear whether there would be any way to settle which proposal is right. These theories were advertised as "naturalistic" – reducing semantic properties to physical relations – but closer examination reveals that they are not testable claims about the physical world.

A further worry is that the relations in question are external. What confers meaning, on the Dretskean view, is outside the head. But things that are outside the head cannot directly guide the behavior of an organism. So that's a kind of meaning that is psychologically inconsequential. At the very least, detector theories are incomplete. They need to be augmented by theories of how organisms come to know the meaning of their own concepts. That requires that meaning be in some readable off features of the mental representations. But, if we had such a theory, it's not clear how much work the external relations would be doing. I don't mean to suggest that the external world is semantically irrelevant. It's rather that the external world can be relevant only to the extent that it can be re-inscribed internally, and the detector story offers little guidance about how that occurs.

A related worry is that detection makes the specific form and format of our mental representations irrelevant. At first, I thought that was precisely what makes such theories powerful (they could be combined with concepts or somatic states equally well), but I came to see this as a liability. The tone emitted by a smoke alarm is arbitrary. It could have been a gong, a melody, or a flashing light. But the bodily states associated with emotions are not arbitrary. The flight pattern would not be a good embodiment of affection, and cuddling would not be a good way to express fear. A number of authors recognized this limitation of my approach, and they were right (Slaby, 2008; Deonna & Teroni, 2012; Colombetti, 2014; Hufendiek, 2016).

Another concern is that the detector approach assumes there is some objective feature out there in the world to be detected. Committed to this, I claimed that emotions represent objective properties. Fear, I claimed, represents danger, and danger is a mind-independent property, albeit a relational one (Prinz, 2004). I now think this is mistaken. Against objectivism, I would point out that there are many dangerous circumstances in which fear would not be appropriate. A firefighter ought not be afraid (and often would not be afraid) when facing a dangerous blaze, even knowing that it could cause serious harm. A tightrope walker, would likewise, do well not to fear teetering precariously over a great height. Are these individuals wrong when they don't experience fear? Are they making a mistake? It's far from obvious that they are. From a functional perspective, fearlessness is apt. Conversely, was Darwin making a mistake when he was frightened by a snake in a terrarium? Would it have been a mistake for him to fear a snake he encountered in the wild, knowing that most snakes are harmless or unlikely to attack. Objectivism assumes that the answers to such questions are straightforward. But they are not. It seems reasonable to say that fear is apt even in response to harmless snakes (perhaps because it's good to err on the side of caution). But even someone who has no rational reason to be afraid, such as a person who experiences social anxiety, might not be making an error when they experience such fear. It is often enough to say, I don't like big social gatherings because they are outside of my comfort zone. An objectivist theory demands criticism in this case, but such criticism may not be justified. Concerns about objectivism magnify when we consider other emotions: Is it a mistake if someone doesn't feel jealous about infidelity? Is it a mistake when Buddhists withhold anger at offenses? Is it wrong to feel sad when watching a fictional tragedy? Often emotions do not arise in lockstep with the objective properties that they are alleged to represent. Hufendiek (2016) argues that the aptness of fear depends on the abilities of the organism. Aptness, I would argue, also depends on responsibilities (as in the firefighter case), values (as with a Buddhist who eschews negative emotions), and personality traits (neither timidity nor courage are erroneous). If we allow such subjectivity, then the attraction of detector semantics is greatly attenuated.

For all these reasons, I think the Embodied Appraisal theory, as originally formulated, is ripe for reconsideration. I will now consider three other attempts to assign representational content to embodied emotions, each of which has something helpful to offer.

### 3.3. Affordances

I cited Hufendiek (2016) as a critic of the detector approach. In developing an alternative account, she draws on J. J. Gibson's psychology of visual perception. According to

Gibson (1979), when we see things in the external world, we do not only register their form or identity; we also recognize their potential for action. He calls these potentialities “affordances”. Scissors afford snipping, chairs affording sitting, and bowling pins afford tipping. These are things we can see. With this idea in hand (so to speak), Hufendiek proposes that emotions represent affordances as well. Fear of a bear represents a bear as affording flight, and this is grounding in the body’s preparation for action. Hufendiek also retains the thesis that emotions represent core relational themes. But instead of saying that fear represents danger simpliciter, she says it represents a danger-to-be-fled (2016: 168).

This is an improvement over Embodied Appraisal theory in some key respects. It makes the bodily response non-arbitrary. Fright is embodied, at least sometimes, by a motor response that prepares the body for taking flight. That response can be recognized as such; it makes itself known internally to the organism. And, if we postulate that flight preparation represents the affordance in question, it follows that embodiment is semantically relevant. If fear represents the affordance of flight, it can be embodied by a cuddle reflex.

I think this suggestion is a leap forward in the right direction, but I think there are reasons for resisting the affordance view (cf. Shargel and Prinz, 2017). The concerns stem from two commitments that Gibson makes in advancing his theory of affordances. First, Gibson insists that affordances are not in us, as perceivers; they are out there in the world (see Scarantino, 2004). He is an objectivist about affordances. I have already suggested that we might want to resist objectivism in our theories of what emotions represent. Hufendiek preserves Gibson’s objectivism, though she grants that affordances are relative to organisms and skills, and she also says they exist only for organisms that can recognize them. This concession to mind-dependence falls short of subjectivism, however. For Hufendiek affordances are objective relations out there to be recognized. I think the content of emotion is more subjective, because it can depend on traits and values. For this reason, I think we might also need to abandon, or greatly qualify, the thesis that emotions represent core relational themes. Hufendiek retains that thesis, saying fear represents danger, but, in modern life, we fear many things that pose no direct physical threat: an exam, a new government, dinner with in-laws, singing karaoke, etc. These threaten academic goals, values, mood, and ego, but it seems inflated to call them dangers, and they hinge on our attitudes. Thus, if we insist on saying fear represents danger, we must add an asterisk specifying that the danger in question is a subjective impression that something you care about is under threat.

The second reason for resisting affordances is that, according to Gibson, affordances are best characterized as possibilities for action. Whenever I see chairs, I see that they afford sitting. If I walk past a street lined without door cafes, I will see this property instantiated repeatedly. But I may have no interest in sitting at that time. Perhaps I am late for work. Perhaps I just left a café up the road. Perhaps I haven’t a penny to spare. So the sittability that is manifest to me need not guide my actions. Like chairhood itself, it is just a property I observe, one toward which I could be indifferent or even inimical.

Emotional content does not seem to be like that at all. When you see a bear, you don’t simply notice that you could take flight. There is a difference, one might say, between potential and potentiation. Flight is not simply an available response; it is one that takes on a kind of urgency. Therefore, while the affordance approach has much to recommend it, the

construct comes with commitments that prevent direct application to emotions. We should incorporate what it gets right, while abandoning the Gibsonian theory that lies at its core.

### 3.4. Existential Magic

Another approach that moves in a promising direction can be found in a small book by Jean-Paul Sartre (1939/1993). Let's begin with one of his examples. In discussing a case of fear that is highly germane our hackneyed bear case, he writes:

I see a wild animal coming toward me. My legs give way, my heart beats more feebly, I turn pale, I fall and faint. Nothing seems less adapted than this...yet it is a behavior of escape. Here the fainting is a refuge...[L]acking power to avoid the danger by the normal methods...I denied it...The urgency of the danger served as motive for an annihilating intention which demanded magical behavior. And, by virtue of this fact, I did annihilate it... (p. 62)

There is a lot to unpack here. First, though sometimes described as cognitivist, Sartre clearly regards emotions as inherently bodily. He details bodily symptoms. Moreover, these are crucial to what comes next. The body becomes weak, causing the individual to faint. Then Sartre says something puzzling: he refers to this behavior as magical. Why? It is like a disappearing trick. Confronted by a threat that cannot be escaped through flight, the individual manages to escape in another way: by somatic magic, the animal is annihilated. The animal isn't killed in reality, but rather removed from phenomenal experience when the person faints. Sartre refers to an annihilating intention, but we need not interpret that in cognitive terms. Seeing the wild animal causes a bodily response that, in a subjective way, eliminates the threat.

Sartre's magical theory can be described as a somatic theory of what emotions represent. Through the bodily response, the person who encounters a wild animal first transforms it into something requiring elimination and then succeeds in eliminating it. That can be described as an act of magic. As Sartre put it, "the true meaning of fear...is a consciousness which, through magical behavior, aims at denying an object of the external world" (p. 64). Notice that, for Sartre, the threat is subjective. Had this wanderer been a hunter, or a thrill seeker, or a wild-life photographer, or a grizzly man, the animal would not have been represented requiring elimination.

In a grand finale, the animal is magically erased. One might say the disappearance was only an illusion, but if we recall that the threat itself was subjective, we can recognize a sense in which disappearance is a successful transformation of the (subjective) world. We might tend to dismiss Sartre's invocation of magic as a bit of poetic excess, but I think we should take it seriously. The term captures the idea that emotions do not reveal the mind-independent world. Rather, they bring new things in and out of existence (Shargel & Prinz, 2017). This insight helps free us further from the temptation to see emotions as detectors. They respond to external circumstances, but not like the retina does, reacting mechanically to rays of light. Rather, they imbue things in the external world with subjective meaning. Magic is a term that can help free us from the objectivism of Gibsonian affordances as well. Sartre is not saying that the animal affords escape. He is saying that the animal is

experienced as an insufferable presence in consciousness that we feel urgently compelled to eliminate.

The contrast between Sartre and Gibson can also be gleaned from a further remark. Strikingly, he says we can offer essentially the same analysis of the content of fear in cases where a person manages to take flight instead of fainting:

Flight is a fainting which is enacted; it is a magical behavior which consists of denying the dangerous object with our whole body by subverting the vectorial structure of the space we live in by abruptly creating a potential direction on the other side (p. 63).

The spatial and geometric language might bring Gibson to mind: there is a potential for escape to the other side. But the verb choice is crucial. This potential is “created”. It would not have been created by someone with a different mindset. Sartre is also breaking with Darwin here. Emotional behaviors are not adaptive responses that ensure survival, in his view. Rather they are ways of coping with desirable and undesirable aspects of experience.

This last bit of the magic story might take things a step too far. As a phenomenologist, Sartre often seems overly concerned with experience. The meaning of emotions is not just subjective but is somehow about subjectivity. Here we are at danger of losing a feature that the affordance account handles well. If emotions primarily serve to alter conscious experience, their grounding in the body becomes harder to explain. If the aim of fear is the removal of unpleasant experiences, then why does it make us run instead of close our eyes, or daydream, or project a positive veneer on unsightly beasts? As a phenomenologist, Sartre is very observant about the somatic nature of emotions, and he arrives at clever ways of putting embodiment in the business of experiential amelioration, but at great cost. The link between emotion and action is severed. Here the affordance view has an advantage.

### 3.5. Enactivism

The final embodied approach that I want to consider in this review is enactivism (Colombetti, 2007; Colombetti & Thompson, 2008; Colombetti, 2014). Enactivism integrates a non-objectivist and phenomenological orientation, akin to Sartre, with an emphasis on action, akin to the affordance view. It developed as a general framework for studying the mind that challenged the prevailing approaches in cognitive science based on the classical computer analogy (Varela et al, 1991). As a starting place for understanding the approach, it is helpful to think about simple organisms, like a bacterium in a chemical gradient (Thompson, 2007). Bacteria are “autopoietic” systems: they engage in ongoing activity to perpetuate their own maintenance and survival. To do so, they must have ways of distinguishing themselves from the environments they inhabit (enactivists call this “autonomy”), and also for identifying things in those environments that are noxious or notorious (enactivists call this “meaning-making”). For example, a bacterium will swim toward sucrose, which provides it with energy. In so doing, it assigns meaning and value to the sucrose, and it does so through its interactions with the world. Applied to emotions, this approach suggests that bodily responses are modes of embodied meaning-making. Just as a bacterium assigns positive value to sucrose rich regions in a chemical gradient, human emotions assign different kinds of value to things in our world, and they do so via the whole body and our experience thereof (the lived body).

Some enactivists reject the idea that emotions represent anything (Hutto, 2012). Usually this amounts to a rejection of the approach to representation found on computer-inspired approaches in cognitive science: the idea that the mind contains inner symbols that store information about the world, or recreate the world in rich detail. Other enactivists recognize that the term “representation” can be used in ways that are compatible with the approach (Colombetti, 2014: 214 n. 5). They will insist, however, that meaning-making involves the dynamic activity of a whole organism, and can involve both body and world. It is a kind of skill or know-how. Di Paolo, et al. (2017) reject internal representations but say we can instead use the verb “representing” to capture what enactivists have in mind.

Enactivists also have another reason, often implicit, for resisting talk of representations. They reject the idea that the world comes divided into entities and kinds that we then simply detect. Recall the bacterium. In treating sucrose as a food, the bacterium is imposing significance that sucrose does not have independent of being treated this way. One might say the bacterium presents sucrose in the way rather than re-presenting it. The world we inhabit is not an objective reality but an environment of our own making (Thompson, 2007). In this respect, we might compare the enactivist approach to Sartre’s idea of magic.

Enactivists are not always specific about the meaning or content of emotions. Colombetti (2014), who develops the most detailed enactivist theory of emotions, takes up the language of appraisals. She challenges traditional cognitive appraisal theories in two ways. First, appraisal systems in the brain, she argues, cannot be separated from sensorimotor systems – brain systems are highly distributed and dynamically coupled, and emotions engage a wide range of brain systems in an integrative way. Second, emotions are not just inside the brain, but also supervene on our bodies as we interact with the world. Still, by preserving the term “appraisal” she implies that these distributed, dynamic, and embodied responses may have meanings of the kind that cognitive appraisal theories emphasized, such as core relational themes.

As discussed above, I am no longer convinced that core relational themes are the best way to think about the meaning of emotions. I think we would do well to think about alternative accounts of emotion content, and I offer one in section 4. I also think there is a slight tension between Colombetti’s two critiques of cognitivism. On the one hand, she emphasizes the integration of cognitive and sensorimotor brain systems. This gives the impression that bodily states become sources of meaning because they are yoked to cognitive systems. On the other hand, Colombetti emphasizes the role of the body and body-world interactions in the genesis of meaning. If the latter is true, it’s not clear why emotions need cognitive systems to provide a basis of meaning. Colombetti does not commit to the necessity of cognitive systems in meaning-making, but her invocation of cognitive systems distracts from the task of specifying how bodies create meaning in the case of emotions. The details on how that works remain under-specified.

As a final quibble, I think some of the core theoretical commitments of enactivism can be misleading when applied to emotions. The concept of autopoiesis, which derives from biology, places too much emphasis, for my taste, on adaptivity. While organisms may be self-maintaining overall, human emotions can be self-destructive (as with some forms of anger and hate). They can also be removed from matters of survival (as will some of our aesthetic emotions). The concept of autonomy too, though quite different from its dominant uses in

philosophy, may distract us from the many ways emotions are both collective and externally manipulated. Enactivists have done much to explore these aspects of emotionality (e.g., von Maur, 2021; Colombetti & Krueger, 2015), and that suggests they could introduce better vocabulary to articulate their core commitment. That said, enactivists do much to improve on the limitations of Sartre and Gibson: they provide a framework for understanding the importance of action in meaning-making, while avoiding commitments to objectivism.

## 4. The Embodied Imperative View

### 4.1. Formal and Particular Objects

Taking stock, section 1 introduced that idea that emotions represent formal and particular objects, 2 surveyed some reasons for preferring non-cognitive approaches to emotion over cognitive approaches, and 3 began with a puzzle that emerges if we take both of these lessons on board. Non-cognitive theories emphasize the body (Affective Somaticism), and it's not clear how bodily states can represent the kinds of external things that emotions are generally taken to represent. I called that puzzle the Somantodox. The standard solution, I said, has been to reject Somatic Internalism, i.e., show that bodily states can indeed represent things out there in the world. Then we looked at some theories that try to show how this is possible. The Embodied Appraisal theory and affordance approach both fall into that category, but face objections. Colombetti's enactivism may also be classified this way, insofar as she implies that emotions have the content of appraisals, but this may not be the best way to read her or other enactivists. Enactivism and Sartre's magic theory implicitly invite a different approach to the Somantodox. They assume that emotional content is subjective in some sense. Rather than insisting that emotions represent something external (Affective Externalism), they could argue that emotions represent (or present) something internal. Here I propose a theory of that kind.

Enactivists generally eschew internal/external distinctions, and, in a sense, they are right to do so. Phenomenologically, we experience emotional significance as something out there in the world: that sculpture is beautiful, the leaking faucet is irritating, and the bear is scary. But such phenomenological externality does not entail that these properties are inherent in external objects. Instead, they might be regarded as projections from our lived bodies onto the world. If so, we can say that the meaning of emotion is actually internal. This would offer a new way of addressing the Somantodox. To succeed, we need to specify what these internal contents are such that they can be (1) grounded in the body and (2) provide a satisfying analysis of what emotions (re)present.

The approach I'd like to recommend draws attention to an observation that I made passingly in discussing affordances: emotions do not simply tell us what actions are possible in a given situation; they have an urgency that compels us to act. Elsewhere, building on this, I have advanced the view that emotional content is imperatival (Shargel & Prinz, 2017; Prinz, forthcoming). On this view, emotions impart a sense of to-be-doneness. They present things in the world as demanding responses from. Fear, for example, presents things as demanding escape. Anger presents things as demanding an aggressive response. Sadness represents things as demanding withdrawal, inertia, and perhaps expressions of mourning. In a phrase, emotions are Embodied Imperatives.

Notice that this way describing emotional content blurs the distinction between formal and particular objects. We perceive things in the world (particular objects) as demanding something of us (formal objects). The formal object of an emotion, on this view, is an action-type, and that action-type experienced as issuing from some entity or situation in the world.

The link between formal and particular objects unfolds in real time, in a dynamically coupled way. When you see the bear, your body seizes up, your muscles become tense, your breathing stops, you grow pale. When this happens, you experience the bear as an object of terror. That is not an inference or appraisal, but rather a projection of your freezing response onto the bear. At a level of neural implementation, we might expect to see strong reciprocal activations in both perceptual streams and brain circuits that regulate the body. Despite your effortful stillness, the bear sees you; its head perks up, it glares at you, a glint of salivation is visible as it begins to expose its teeth. Now, in a panic, your heart begins to race, your eyes dart about looking for a path of escape, and, in a moment, you find yourself in flight through the dense woods, panting and glancing back at every turn. If you see the bear, you accelerate, and, if it ends the pursuit, you slow down. The way fear is embodied changes, moment to moment, depending on how the particular object of that fear reacts to those embodiments. Fear is not a single bodily state, but a repertoire of escape strategies, improvised anew in each episode and precisely tuned to its elicitors.

#### **4.2. Compared to the Embodied Appraisal View and Other Theories**

The Embodied Imperative account differs from the Embodied Appraisal view in a number of ways. First, the Embodied Appraisal view identifies emotions with body perceptions. On the Imperative view, it makes more sense to say emotions can be embodied in all of the three senses mentioned above: bodily perceptions, body control systems, and the body itself. In most cases, all three work in concert, and there is little to be gained from circling one of them and saying the emotion resides here. Second, the Embodied Imperative view has introduced a very different perspective on the content of emotions. On the Embodied Appraisal view, emotions represent in a way that is externalist, objective, and indicative. For the Imperative view, emotional content is internal, subjective, and imperative. For these reasons, it is also less clear whether the term representation even applies. In addition, the Imperative view also overcomes objections to the Embodied Appraisal view. The bodily states that constitute an emotion are not arbitrary: they contribute directly to emotional meaning. Here the story is straightforward and highly resistant to the objections against detector theories. Emotions (re-)present to-be-doneness, and they do so because the bodily states they encompass are states of action preparation. The bodily is literally compelling us to act.

One consequence of this approach is that the meaning of an emotion can be described at different levels of specificity. I said that fear presents things as to-be-escaped. But I also said that the specific embodiment of fear is highly variable, dynamic, and context sensitive. At one moment, fear presents its object as something that demands freezing, and, at the next, it represents its object as demanding flight. The form that flight takes depends on circumstances (running through the woods or, say, turning a corner when someone you dislike is spotted down the street). When talking about the content of emotions, we can mention

any or all of these levels. Each action description accurately captures what the emotion demands of us: e.g., escaping the bear or running away from the bear or dashing through the woods in a zig-zag pattern and such-and-such speed. At a high level of abstraction, we can say fear represents the imperative to escape, but we can also say these more specific things. The term “fear” is an abstraction used for a wide range of embodiments that impel escape or evasion.

If this view is right, emotions do not represent core relational themes, unless we take the pains to explain that, by “danger” we mean demanding-escape. We might improve on the old vocabulary by saying emotions (re-)present “core dynamic themes”. Dynamic themes are relational, but the term dynamic emphasizes their active and interactive nature. We can translate lists of core relational themes into core dynamic themes, preserving the spirit of that approach, while breaking from cognitivist and objectivist associations.

In emphasizing meaning-making through dynamic interaction, the Embodied Imperative View shares much with enactivism, and, elsewhere I call it an enactive theory (Shargel & Prinz, 2017). But it does not share all of enactivism’s commitments (e.g., autopoiesis), and offers a specific account of what emotions represent. That account emphasizes actions, much like the affordance view, but these are not objective possibilities; they are subjective demands. They are potentiated, not mere potentialities. The creation of subjective properties puts the Embodied Imperative view into alignment with Sartre’s existential magic. The main difference is that Sartre relates emotional meaning to consciousness (fear presents conscious experiences as the things we need to escape), and the Imperative view locates meaning in action.

## 5. Implications and Conclusions

We began with the question: what do emotions represent? The standard answer is that they present both formal and particular objects. The formal objects have been analyzed in different ways, but the most popular answer is that fear represents danger, sadness represents loss, anger represents offense, and so on. These have been called core relational themes. This approach is most associated with cognitive theories of emotion, but such theories have been challenged by defenders of embodied views. Assuming those challenges succeed, this saddles embodied views with a challenge: if emotions represent external things like dangers it’s not clear how they can be bodily states (the Somatodox). To address this challenge, some embodied theories try to establish that bodily states can, in fact, represent external things like anger. Here, I proposed an alternative. Instead of insisting that bodily states represent external things like dangers, we can say they represent (or present) something internal. In particular, I proposed that emotions present their particular objects as demanding things of us. Emotional content in imperative.

Much more would need to be said to develop and depend on the Embodied Imperative view. I offer it here as a direction for research rather than a final theory. It builds on lessons from other embodied views but also breaks from them, and future work will have to assess whether the shift to imperatives can be sustained. It might have implications for domains in which emotions contribute to judgments and decision making. For example, if moral judgments have an affective dimension, and emotions have imperative content, does it follow that moral judgments are commands? Is that tenable?

Other potential implications stem from the subjectivism of the Embodied Imperative view. If emotional meaning is just a projection onto the world, can our emotions ever be criticized as false, mistaken, or unfitting? What should we say about emotional disorders? Can the person with anxiety or phobias be faulted for fearing too much? Can depression be described as misrepresenting the world? Have psychopaths erred when they do not feel guilt? Here, I think the Embodied Imperative view might make progress by shifting from norms of truth to norms of action. In particular, rather than looking for truth or fittingness, we might assess emotions for practical rationality. Future work should assess the prospects of this move.

The approach will also face apparent counterexamples. Some emotions have no obvious link to action: some are passive like boredom, regret, and awe, or future-directed like hope, or cognitive like suspicion, doubt, and interest. Such examples threaten any embodied theory. The Embodied Imperative approach invites new lines of response, asking us what each of these emotions demands. The concept of action might have to include symbolic actions (we knock on wood in hope) and cognitive actions (scrutiny follows from suspicion). Can the theory withstand such extensions? There is always a danger that one's paradigm case (fear of a bear in this discussion) will lead to a theory that is poorly suited to other examples. On the other hand, seeing these other emotions on the model of escape imperatives could lead to fresh insights. The main thesis here requires considerable rethinking. Emotions do not tell us how things are, I've suggested, but rather what to do. I hope the hypothesis is intriguing enough to compel further investigation.

## References

- Armon-Jones, C. (1991). *Varieties of affect*. Toronto: University of Toronto Press.
- Arnold, M. B. (1960). *Emotion and personality*. New York, NY, US: Columbia University Press.
- Colombetti, G. (2007). Enactive appraisal. *Phenomenology and the Cognitive Sciences*, 6, 527-546.
- Colombetti, G. (2014). *The feeling body: Affective science meets the enactive mind*. Cambridge, Massachusetts: The MIT Press.
- Colombetti, G., & Thompson, E. (2008). The feeling body: Towards an enactive approach to emotion. In W. F. Overton, U. Müller, and J. L. Newman (eds.), *Developmental Perspectives on Embodiment and Consciousness* (pp. 45-68). Mahwah, NJ: Erlbaum.
- Colombetti, G., and Krueger, J. (2015). Scaffoldings of the affective mind. *Philosophical Psychology*, 28, 1157-1176.
- Deonna, J. A., & Teroni, F. (2012). *The emotions: A philosophical introduction*. New York: Routledge.
- Di Paolo E., Buhrmann T., Barandiaran X. E. (2017). *Sensorimotor life. An Enactive proposal*. Oxford: Oxford University Press.
- Dretske, F. (1988). *Explaining behavior*. Cambridge, MA: MIT Press.
- Gibson, J.J. (1979). *The ecological approach to visual perception*. Boston: Houghton Mifflin.
- Hufendiek, R. (2016). *Embodied emotions: A naturalist approach to a normative phenomenon*. New York: Routledge.
- Hutto, D. D. (2012). Truly enactive emotion. *Emotion Review*, 4, 176-181.
- Keltner, D., Sauter, D., Tracy, J., & Cowen, A. (2019). Emotional expression: Advances in basic emotion theory. *Journal of nonverbal behavior*, 43, 133-160.
- Kenny, A. (1963). *Action, emotion, and will*. London: Routledge.
- Kragel, P. A., & LaBar, K. S. (2015). Multivariate neural biomarkers of emotional states are categorically distinct. *Social Cognitive and Affective Neuroscience*, 10, 1437-1448.
- Kreibig, S. D. (2010). Autonomic nervous system activity in emotion: A review. *Biological Psychology*, 84, 394-421.

- Lazarus, R. S. (1991). *Emotion and adaptation*. New York: Oxford University Press.
- LeDoux, J. E. (1996). *The emotional brain: The mysterious underpinnings of emotional life* (p. 384). New York: Simon & Schuster.
- Oosterwijk, S., Lindquist, K. A., Anderson, E. C., Dautoff, R., Moriguchi, Y., & Barrett, L. F. (2012). States of mind: Emotions, body feelings, and thoughts share distributed neural networks. *NeuroImage*, 62, 2110–2128.
- Prinz, J. J. (2004). *Gut reactions: A perceptual theory of emotion*. Oxford: Oxford University Press.
- Prinz, J. J. (forthcoming). *Cambridge elements: Emotion*. Cambridge: Cambridge University Press.
- Sartre, J.-P. (1939/1993). *The emotions: Outline of a theory*. New York: Carol Publishing Group.
- Scarantino, A. (2004). Affordances explained. *Philosophy of Science*, 71, 949-961.
- Scherer, K. R. (2001). Appraisal considered as a process of multilevel sequential checking. In K. R. Scherer, A. Schorr, & T. Johnstone (eds.), *Appraisal processes in emotion* (pp. 92-120). New York: Oxford University Press.
- Shargel, D., & Prinz, J. J. (2017). An enactivist theory of emotional content. In F. Teroni & H. Naar (Eds.), *The ontology of emotions* (pp. 110–129). Cambridge: Cambridge University Press.
- Slaby, J. (2008). Affective intentionality and the feeling body. *Phenomenology and the Cognitive Sciences*, 7, 429–444.
- Smith, C. A., & Ellsworth, P. C. (1985). Patterns of cognitive appraisal in emotion. *Journal of Personality and Social Psychology*, 48, 813–838.
- Smith, C. A., & Lazarus, R. S. (1993). Appraisal components, core relational themes, and the emotions. *Cognition and Emotion*, 7, 233–269.
- Thompson E. (2007). *Mind in life: Biology, phenomenology, and the sciences of mind*. Cambridge, MA: Harvard University Press.
- Varela, F.J., Thompson, E., & Rosch, E. (1991). *The embodied mind: Cognitive science and human experience*. Cambridge, MA: MIT Press.
- von Maur I. (2021). Taking situatedness seriously: Embedding affective intentionality in forms of living. *Frontiers in Psychology*, 12, 599939.

*Invited paper*

## MULTILEVEL-GROUNDED SEMANTICS: REPRESENTATION, INTENSION AND REFERENTIALITY IN MUSIC

Mihailo Antović\*

**Abstract:** The article introduces the author's model of contextual constraints on the generation of musical meaning in an effort to suggest how the approach may be of interest to philosophers pursuing the problem of representation. Since the model is based on insights of linguistic semantics, representation is defined as an amodal functionalist construct relating music perception to an extramusical experience, which assumes varying degrees of imagistic and/or referential specificity. The theory proposes six such degrees, or "grounding boxes": perceptual, cross-modal, affective, conceptual, culturally rich, and personal. In doing so, it turns classical dichotomies in musical semantics into continua, which lead the cognizer from more general to more concrete, more intra- to more extramusical, more formal to more referential, more perceptual to more conceptual, and finally more ineffable to more representational sets of connotations. The system is illustrated with the analysis of the Ming the Merciless leit-motif from the 1980 science-fiction film *Flash Gordon*, composed by Freddie Mercury and performed by Queen.

**Keywords:** music, language, semantics, conceptual blending, grounding, representation

This paper aims to present the theory of multilevel-grounded semantics (Antović, 2022), which is intended as a contribution to the topic of musical meaning. Developed from a cognitive linguist's perspective, the approach may relate well to the problem of mental representation as considered by philosophers, since precisely the question of (non-)existence of representations and their exact nature has caused serious divides among linguistic semanticists in the past decades. My proposal is that the retention of the linguistic perspective, yet with a focus on the non-linguistic phenomenon of music, can not only provide a new way to look at musical semiosis, but also shed new light on the debates on representation in cognitive science in general. To specify what I have in mind, in this paper I provide a brief theoretical presentation of the model, illustration of how it relates to some previous contributions to linguistic and musical semantics (in particular, the conceptual blending theory), and a full analysis of a short piece of film music by means of the model's instrumentarium.

### 1. Introduction

If one takes mental representation to be a hypothesized formal integration in the mind/-brain by which the organ "translates" between an external event – for example, sensory stimulus – and an introspective experience of this event – for example, my hunch that

\*Department of English and Center for Cognitive Sciences, University of Niš;  
mihailo.antovic@filfak.ni.ac.rs; ORCID ID: <https://orcid.org/0000-0002-5308-6747>

what I hear must be a Chopin nocturne, this postulated construct will likely function as a domain- and modality-neutral intermediary. Thus, whatever relates “nocturne” the word and “nocturne” the instrumental excerpt has a nature of its own, unrelated to either musical pitches or linguistic phonemes. At the current state of knowledge, we need to represent such a mental connection functionalistically, as a series of abstract relations which are in some sense descriptive of whatever is happening neuronally in the brain during the process. Such an “abstract” / “amodal” position on meaning is typically defended by formal approaches to linguistic semantics, e.g. generative (Chomsky, 2000; Pustejovsky, 2016) and conceptual schools (Jackendoff, 2022). Conversely, cognitive linguists, who, in spite of their name, are typically opposed to Chomskyan methodologies in the language science, largely affiliate with an “embodied” and “enactive” epistemology in studying cognition. Hence they either claim that there is no need for mental representations at all, since our experience is embedded and direct, or endorse a weak sense of representations, claiming that they are not computational, but rather “geared towards the action an organism performs” (following up on the work of philosophers such as Chemero, 2013). Such “proto”-representations directly link to, and derive from, the modality in which they manifest, e.g. bodily motion (Lakoff and Johnson, 1999; Gibbs, 2005).

Music provides an interesting domain for a “reconciliatory” approach to the problem. On the one hand, the meaning it generates is obviously much more iconic than in language (Bierwisch, 2009). Therefore, semantic interpretation in music strongly depends on the structure of the musical stimulus and the underlying activity producing it, for instance, a composer’s use of loud notes with low frequencies and a performer’s movement while playing them. These may make a musical section intuitively appear as “strong”, “forceful”, “deep”, “masculine”, or “torpid” to experienced audiences. While my approach is based on linguistic semantics, and therefore usually looks at *descriptions* of music, by either participants in experiments or authors in written pieces of criticism, one must recall that such “structural” musical impressions need not be *verbalized* at all. Thus, the strength and relative inactivity of the musical stimulus may be experienced directly by the body, and converted into either a *disposition* for a motor reaction (in the line of still influential behaviorist approaches to meaning, e.g. Osgood, Suci and Tannenbaum, 1957), or to the actual motor response, such as the tapping of feet to the musical meter. Arguably, one does not need representations – at least not of the linguistico-semantic kind – to interpret, and thus “semiotinize”, music in this direct fashion. Such an approach, of course, relates strongly to schools in cognitive science arguing that “embodiment” is an important, if not the sole, means of grasping meaning.<sup>1</sup>

On the other hand, music perception has an equally strong potential to “break away” from the sheer mimesis of corporeal action and become much more referential, e.g. in familiar tunes used in internet memes to enhance social satire. In this case, human cognizers often

---

<sup>1</sup>In fact, the proposal that an organism just *responds* to the music with the body, even if directly, i.e. without representations, comprises a rather weak sense of embodiment, which I happen to endorse. Its other variant that I find attractive is the direct invocation of bodily terminology in free-form descriptions of music, which shows, time and again, that music is typically experienced by humans as being *about* their bodies in motion (cf. Antović; Küssner, Kepmf, Hashim, Omigie, and Schiavio, 2023). Stronger epistemological affiliations with the movement of embodiment may go much further than mere corporeal reaction, so far as to claim that *all cognition*, from perception to conceptualization, from reasoning to action planning, is embodied (e.g. Varela, 1991).

cannot help but *enrich* the relatively simplistic, reactive interpretation, such as the one from the previous paragraph. In such a process, the “torpid” music may well become much more specified, and relate to actual extramusical images: perhaps an old-fashioned, somewhat verbally “slow” politician with whose policies we do not agree, or the movement of an albatross’ wings from a cartoon we are familiar with. This aspect of musical meaning is much more contextual and interpretive so that, it appears, it can hardly proceed without resorting to a classic form of representation, which follows the functionalist definition I offered above: that there is a relatively stable, yet amodal, mental entity that binds together the experience (here: the musical structure) and the interpretation of this experience (here: the referentially rich imagery and/or linguistic description).

Multilevel-grounded semantics turns this apparent duality into a continuum, by instituting a six-level system, where the generation of musical meaning occurs piecemeal, through several hierarchical, if not fully recursive, stages of constraint, or “grounding boxes”. The last concept comes from cognitive linguistics as well. Namely, one of the main issues in interpreting novel meanings (as in metaphors or counterfactual statements) is the way in which we explain the problem of “ground” – the set of assumptions which stand in background cognition, and which we implicitly share with the interlocutor when we negotiate a “proper” interpretation (Searle, 1978; Langacker, 1999). Following up on a line of thinking that dates back at least to the classical structural linguist J. R. Firth (1957), cognitive linguists Coulson and Oakley (2005) proposed that, at a minimum, such a “grounding box” should contain the knowledge of participants, their roles, and the forum/circumstances in which they operate. For instance, in order to interpret the counterfactual statement “In France, the Lewinsky affair would not have hurt Clinton”, one needs to go much further than the formal process of “conceptual blending” (Fauconnier & Turner, 2002). In this well-known epistemological framework of cognitive linguistics, conceptual elements associated with an American president (e.g. the fact that Clinton reigns the country in a parliamentary democracy) are mapped onto his French counterpart (at that time, Mitterrand, who equally reigns in France, and in an equally strong “presidential” system), to create an emergent concept of a (non-existent, imaginary, but also imaginative!) American president reigning in France. This complex mapping alone represents a powerful tool of semantic generation. Yet to fully grasp the subtle joke one needs to know a lot about who Clinton and Lewinsky at the time were, what Mitterrand was like with regard to alleged relationships with women, what moral expectations one should have from politicians in the two countries, indeed even what “politician”, “France”, and “America” denoted. The more one knows, in fact, the better the interpretation, and the stronger the joke (that, ultimately, the French “lack moral clarity” as compared to politicians in the United States; Coulson & Oakley, 2005: 1518). All these prior assumptions, relevant to the interpretation, yet preceding the formal process of blending, are allocated to the grounding box.

Multilevel grounded semantics elaborates on this construct in an effort to provide a more detailed explanation of contextual factors contributing to the construction of musical meaning. By doing so, it also provides a, hopefully novel, perspective on (musical) representations. Its contribution is found in the proposal that the construction of such representations also works along a continuum: at the outset, the process is likely purely perceptive and not semantic, yet as it develops, clearly referential, representational elements begin to kick in.

## 2. Mapping the Musical onto the Extramusical: Conceptual Blending in Film Scoring

Multilevel-grounded semantics (Antović, 2022) proposes that conceptual blending, enriched by the notion of the grounding box, is a useful tool to explain the basics of the meaning construction process in music, yet that the *single* grounding box is not sufficient to capture many of the intricacies of the process. The new suggestion the theory offers is that the grounding box should be iterative, i.e. transformed into a series of interrelated, hierarchical, if not fully recursive boxes to which individual conceptual blending sequences “attach”. Along the way, the cognizer is taken from simpler to more complex aspects of musical meaning construction: perceptual to conceptual; biologically-determined to strongly socially-grounded; preverbal to verbal; obligatory to optional; and ultimately – non-representational to representational.

Let us illustrate the process by focusing on an apparently simple case of conceptual blending in programmatic music. In classical Hollywood film scoring, whose inspiration lies in the late Romantic operatic canon, characters or events are often associated with their musical *themes*, a phenomenon that Wagner once named the *leit-motif* (an excellent recent source on connections between cognitive linguistics and film music is Chattah, 2023). When superstar rock band Queen wrote their music for *Flash Gordon* (1980), a campy and in many ways ridiculous SciFi flick, which is today strangely praised by dedicated fans and critics alike, they came up with a rather disharmonious theme to “stand for” the main villain, Gordon’s arch-adversary Ming the Merciless, masterfully portrayed by Max von Sydow<sup>2</sup>.

The theme revolves around perceptual musical *instability* and extramusical *menace* that it connotes. The mode appears to be minor, but it is actually understated and ultimately ambiguous: due to the avoidance of the 3rd degree in the scale (which determines the mode), rather than proverbial (and oftentimes too simplified!) “seriousness” or “sadness” evoked by the minor, what we experience is almost perpetual uncertainty as to where we tonally “stand” in the tune. The melodic line is full of out-of-key tones, with lots of chromatic aberrations which keep disappointing the listener’s musical expectancies, thus creating the notion of suggestive, if campy horror. The deep register of the piercing, yet potent, and therefore somewhat irritating, synthesized timbre adds to the uneasiness of the experience, which is topped by frequent intervallic connections between the tonic and the augmented fourth – a musical leap historically known as the “tritone”, so devastating to the refined ear that, ever since the Middle Ages it was oftentimes “banned” by many treatises in music theory as the “diabolus in musica”. This historical extramusical connotation, and its relation to the movie’s anti-hero, are obvious to the musically-trained listener but may, to an extent, be intuitively grasped by the untrained spectator as well. All of these musical features – and I listed them here as illustration only, there certainly are more – appear to nicely support the qualities of the antagonist that the music is meant to evoke.<sup>3</sup> So Ming the Merciless is

<sup>2</sup>The theme (without the visuals) can be played at the following link: <https://www.youtube.com/watch?v=VBfuXXGqFL8>. The music analysis that follows revolves around the first instantiation of the main motif, ca. 0:05-0:28. Elements of the visual scene are analyzed based on the full video clip (the visual and the music) available here: [https://www.youtube.com/watch?v=a7KgR\\_H9VsI](https://www.youtube.com/watch?v=a7KgR_H9VsI). The relevant portion is Ming’s entrance, ca. 0:35-1:00.

<sup>3</sup>I safely use the verb “evoke” here as the theme’s title, and its original use in the film – at the moment in which Ming appears before the subjects in his court – make it obvious that the invocation is intended by the

unstable, menacing, horrific to the point of diabolism and, ultimately, *ugly*. The synthesized theme thus provides a series of *analogies* with the relevant traits of this antihero, which are aimed at *enhancing* the attributes of his character, as presented in the conceptual blending diagram below (Figure 1):

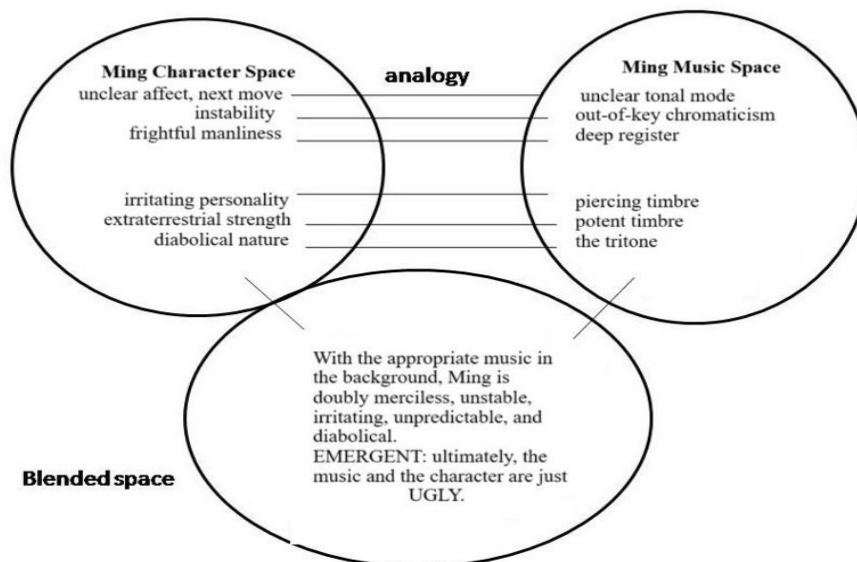


FIGURE 1. The conceptual blending process in which the character and the music integrate: the case of Ming the Merciless

Therefore, blending theory operates with two conceptual information packets, functionalistic “mental spaces”, one of which hosts our knowledge of the character and the other of the music. When elements of one selectively project onto the corresponding elements of the other, a new, third structure is created, known in the theory as the “blended” mental space. Importantly, this is not a mere amalgam of the music and the character, but a venue in which *novel* information emerges from the process of interaction: in the case of good program music, this is at the very least an effect of *augmentation* (or “complementation”, cf. Cook, 2001), in which the music strengthens the connotations implied by the extramusical narrative. Yet, very often, the structure emerging in the blended space does not originate in the inputs at all. In this particular case, one can argue that the notion of “ugliness” arises as a complex inference coming from the perception of various musical and visual structures described above, thus creating a sense of the lack of aesthetic appeal related to the character in question.

Yet the formal process of interaction of conceptual elements (from the input spaces) is just one part of the story. A more detailed interpretation of the extramusical event, and the way in which music interacts with this event, requires a lot of *background* knowledge. Indeed, if we had no idea of context(s) important to the story, Ming the Merciless, the *Flash Gordon* universe, or courts of villains in campy 1980s science fiction movies, we would have a hard

composer. Oftentimes, however, proving such a connection – between the composer’s intentions and music analysis – is not at all easy. In fact, the very existence of an intention is often questioned, by composers and music historians alike.

time “unlocking” the intended connotations of the music in the first place. And knowledge of the genre and the “canon” related to the comic book are just some of the very obvious, surface-level contextual elements that we need. In fact, layers of context can be much deeper, including, for instance, the knowledge of what “synthesized” means, what “villains” are, not to mention what “instability” in music is based on (even if we can sense it only intuitively!) or what some of the historical connotations of using “the tritone” are. Ultimately, to interpret some of the connotations we need to also have an idea of what an exaggerated, frowning facial expression typically denotes in our culture. It is precisely such elements that Coulson and Oakley (2005) proposed to be allocated to “the grounding box”, as they say – not really a mental space on its own, but rather a “set of background assumptions” that the interlocutors need to share in order to proceed with the meaning generation process. At the very least, the two authors claim, these should include “the participants”, “the forum” and “the circumstances” of the communicative act in question. Here, thus, we need to know who Ming the Merciless is, where the scene in question occurs, and what the contents of the take are, i.e. what the filmmaker’s general intent while shooting the scene may have been. Arguably, the composer – in this case Freddie Mercury, the exquisitely creative, though not formally educated, musical prodigy – had all of these elements, and many more, hidden in the layers of his subconscious mind while composing the piece and then proposing the extramusical connections. Thus the grounded blending process involving music and the extramusical in some sense aims to unpack this process in the composer’s mind. A possible blending diagram for the Ming scene, with the grounding box added, is presented in Figure 2.4

Thus the construct of the grounding box is very useful as a supplement to the process of conceptual blending. Yet, as a monolith constituent, it is not sufficient to capture many of the intricacies of musical meaning construction. For instance, why exactly is it the case that this music, with all of its “unstable” elements mentioned above, is “appropriate” to the narration involving Ming’s frightening appearance in front of his courtiers? More particularly, what is it *in the music itself* that motivates such a possible interpretation (among many others, of course, the Ming story is obviously just one viable visualization of the extramusical narrative)? If this issue is not considered, one may erroneously conclude that musical meaning is predominantly about “arbitrary association” (or at least “also about it”, e.g. Juslin, 2013), where the Mercury instrumental is related to the Ming character solely due to the fact the two were used together in the film. While association is always a relevant factor – if anything, repeated exposure to the theme in the same extramusical context strengthens the connection and constrains the range of the listener’s subsequent expectations – it is never the case that just “anything goes”, i.e. that *any* music can suit *any* film character or story. To suggest why this may be so, one needs to elaborate on the “grounding box” construct.

---

<sup>4</sup>The reader is advised to pay attention to the qualifier “possible” here: just like composing, the musical meaning construction process is open-ended, so there is never a “correct” interpretation. One can rather discuss levels of saliency, dependent on numerous intra- and extramusical factors. And this is again a gradient: the more familiar the listener is with the elements of intra- and extramusical signification, the closer his or her interpretation to the “ideal” (perhaps the one “envisaged” by the composer) will be.

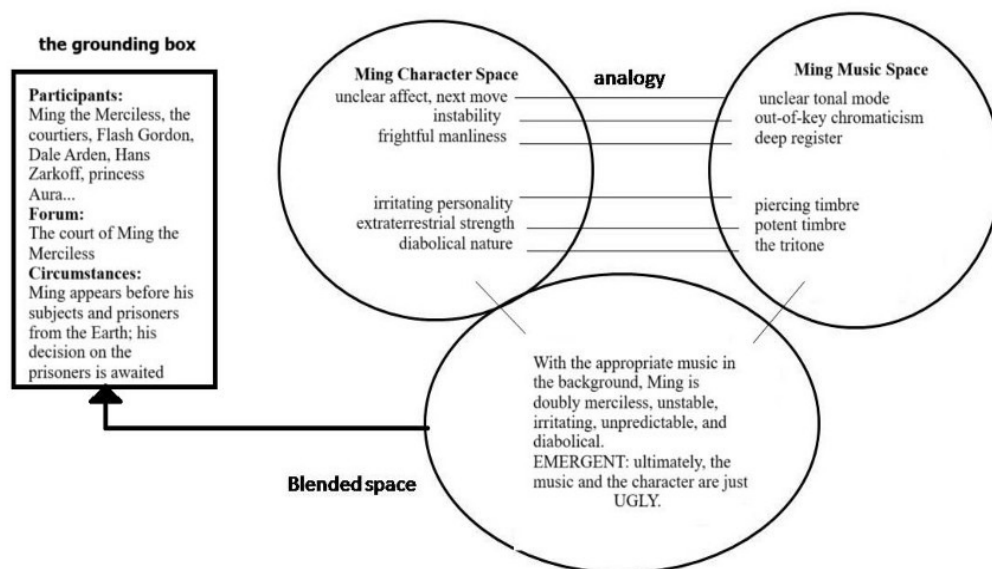


FIGURE 2. The conceptual blending process for music and language with the grounding box added: the case of Ming the Merciless

### 3. Conceptual Blending Becomes Iterative: Multilevel Grounded Semantics of Music

The central proposal of multilevel grounded semantics is that the grounding box should be viewed as *iterative*. More particularly, instead of one, this theory proposes a series of six interrelated grounding boxes, constraining the possibilities of musical interpretation – so that implausible scenarios are ruled out – yet equally allowing for astounding creativity of expression – where connotations that are apparently very different on the surface may prove to be quite similar on an underlying level. This last thesis, in fact, explains well the old “paradox of musical semantics”: that musical meanings are ubiquitous, yet we find it hard to either exactly pinpoint those meanings or explain why many more interpretations are possible than, for instance, in linguistic denotation (cf. Swain, 1997). Importantly, the process is still based on conceptual blending, i.e. the combination of “appropriate” musical and extramusical elements, which remain at the core of the construction of meaning. The grounding boxes do not represent the meaning generation process in themselves, but rather only constrain the options for plausible elements partaking in the mapping between the two input spaces. In other words, they stand as unconscious, experiential “distillers”, implicitly informing the cognizer of the range of scenarios that may be “appropriate” on the particular level of blending. By doing so, however, they carry out an important underlying function, in which they turn the old musico-semantic dichotomies, including the one related to representation, into gradients.

In its current form, the theory proposes six such levels, yet one should keep in mind that they are envisaged as just points on a continuum, thus intended as useful “hinges” for the cognitive scientist to discretize the otherwise anything-but-discrete process of semiosis: (1) The perceptual level parses the musical stimulus into formal gestalten or meaningful

“chunks”, which assume varying degrees of dominance over one another; (2) The cross-modal tier motivates schematic correspondences between these formal structures and the listener’s spatial, haptic, or more generally embodied experience, which includes the conceptualization of movement and force; (3) The affective layer ascribes to this embodied appreciation dynamically changing sensations closer to consciously experienced emotions, e.g. where tense and lax parts of the musical flow combined with the bodily mappings provide stable invocations of satisfying or dissatisfying emotions, perhaps yearning for a resolution; (4) The conceptual level draws analogies between such schematic and affective appreciation and elementary experiential imagery, resulting in very basic outlines of narratives, e.g. of unspecified characters moving in various directions and with various levels of effort; (5) The culturally rich tier checks such a narrative outline against the recipient’s cultural knowledge, considerably specifying the *who*, *what*, and *why* of the constructed referential activity; and (6) The individual level adds idiosyncratic recollections from the participant’s personal experience.

An important governing principle of multilevel grounding is that lower levels partake in the motivation of higher ones. In other words, *results* of the blending process on a lower tier usually become *inputs* in the higher-stage integration process, making the entire system at least partially recursive. For instance, if the cognitive system formally deduces that a musical segment functions as the peak of the stimulus on level one, and then cross-modally interprets the path to this peak as upward (thus VERTICAL) movement on level two, on the third level *both* the generalized notion of the highest tension *and* the specification of this tension as the highest point on the vertical axis will be instrumental in creating the “peak” affect, and thus, potentially, the highest emotional satisfaction, on level three. Equally importantly, the higher levels can piggyback on lower ones, so that, for instance, a personal circumstance from level six (e.g. our knowledge that the theme comes from a favorite childhood movie) backfires on the affective level three in a feedback loop (e.g. increasing our positive affect associated with the score). The process is thus ever-changing, “pulsating”, since it proceeds in both hierarchical and circular fashion. An important consequence is that the interpretation of a musical segment *never ends*, which is a possible reason why we tend to return to the same tunes time and again, sometimes till the end of our lives.

Yet, in the ideal scenario, and for presentation purposes, it is best to consider the hierarchical aspects first. A logical consequence of such an approach is that one needs to go through integration on the lower levels in order to reach the higher ones. The lower tiers are thus more “obligatory” and the higher ones tend to be “optional”; the lower ones are more perceptive, physiological, and thus “natural” than the higher ones, which tend to be more “connotational”, “interpretive” and “social”; the lower ones are more embodied, yet tend to be difficult to verbalize, while the higher ones lend themselves more easily to linguistic description, but require more conscious effort by the cognizer; the lower ones are relatively simplistic (e.g. the music is just “tense”) and the higher ones can be better elaborated, either imagistically or also verbally (e.g. “it sounds like a very dramatic scene from a film in which someone is chasing the main character who is desperately trying to escape”); the lower ones lean on the formal, self-referential, and “intramusical” end (and are thus hardly “about” anything at all), while the higher ones are more extraneous, referential, and extramusical (and therefore quite extensional); ultimately – the lower ones are not at all or are barely representational, while the higher ones do exhibit representational properties taken in the

linguistico-semantic sense, that I defined in the first section: relating a sensory experience to a relatively stable range of reference.

By approaching semiosis in this way, the theory aims to turn classical semantic dichotomies (e.g. sense/reference, intension/extension, perceptual/conceptual, literal/metaphorical) into *continua*: rather than clear-cut boundaries, one may more fruitfully consider the *amounts* of the one or the other in every instance of musical meaning construction. With regard to the phenomenon of representation, the theory exhibits the exact same gradience: from very little need for representation in the beginning to full representations (in the classical linguistic sense) in the end. In the main analysis that follows, I present how the approach works in practice on the film music example referred to in the previous section.

So what happens with the Ming theme on level one? As cognitive musicologists have claimed for decades (at least since Lerdahl & Jackendoff, 1983) music is not a series of auditory stimuli but *a mental construct*. A sequence of pitches in and of itself is just a mish-mash of sound waves, ultimately air molecules moving around and hitting one another. In a mentalist approach, and almost all linguistic semantics is mentalist nowadays, these sounds need to be *perceived* as a sequence first, and then assessed as more or less important than other sequences in the given musical context. Once such sequences proliferate in the score, and get perceived as formal gestalten, i.e. sensical “chunks” rather than meaningless series of pitches, one can proceed to construct intricate intramusical relationships among them – repetitions, variations, added or reduced stability, open- or closed-ended sets, peaks and troughs, the same or different arrangements, and many more. These structures, of course, need to be somehow processed by the human brain – and thus may be representational in a very general sense (in which pretty much *any* mental processing forms a representation). But they do not comprise representations in the stronger linguo-semantic usage from the beginning of this article, that of “translating” one sensory mode (music) into another (imagery, language). This is why level one remains predominantly “perceptual”, formal, intramusical and self-referential.

In the Ming case, the musical elements include the aforementioned deep register, synthesized timbre, a unison melody in a close-to-minor mode, the abundant use of chromatic pitches outside of the main key, and the insistence on the tritone, which in fact opens the motif, holding on to the augmented fourth in the form of an “announcement” for several seconds before the main melodic sequence starts. Twice in the section, one also hears a triple thud of a deep strike of the percussions, likely a floor tom from the drum kit played by Queen drummer Roger Taylor. Finally, the synthesized timbre of the lead melody gets significantly altered (typically toward a more piercing and darker, thicker “hue”) as the tune progresses, only to return to the original waveform towards the beginning of the repeated motif. When these elements start to blend with one another (here: intramusically, e.g. the percussive hits with the meter, forming the rhythm; the rhythm with the melody; the melody with the harmony), what one gets is the ineffable sensation of almost perpetual musical *instability* – except for the first and final tone in the sequence (occurring on the tonic, i.e. the most stable position in the key), pretty much the entire musical motif played here is melodically and harmonically very *tense*.

In turn, for the musico-visual augmentation proposed in Figures 1 and 2 to work, this level-one blended “tenseness” in the music needs to be mapped onto the appropriate tenseness

caused by the targeted character in the visual scenario. Indeed, Ming’s entrance supports this level-one, “formal” or perceptual musical structure in one main respect, that is – the pace of his walking is (loosely) synchronized with the perceived rhythm of the piece. Likewise, the tension in the music is at least partly mapped onto the exaggerated, raised eyebrows forming an important part of Max von Sydow’s makeup, providing a rather “unstable” expression in Ming’s face. The campy outfit, consisting of a dark cape and red-and-golden shirt with a huge collar, only adds to the sensation that Ming is barely able to walk without falling, and this really is a good visual counterpart to instability, so strongly suggested by the music. Further, the floor tom is heard at the moment in which the film director replaces the (relative) close-up of Ming’s face with a broader take, which focuses on the size of the hall through which Ming is making his entrance (and such a reverberating drum sound is naturally expected to be heard in large halls of that kind). Therefore this entrance, pace of movement, and tense elements in Ming’s face and apparel blend with one another to suggest the character’s psychological instability, even without the musical accompaniment. Another important effect here is that of color: the entire cavern is painted bright red, which, of course, causes a strong perceptual effect of tension, adding to the (formal) appreciation of instability. Yet, without the music, the effect would not have been so convincing (which can be tested rather easily – by playing the scene with the volume turned down). It appears, on this level, that the musical tension is in fact *crucial* to the construction of such an (intuitive, ineffable, hardly representational) “instability” of character. Of course, what at this point looks like just a hint of this person’s unstable nature will only get specified in further layers of blending.

On level two, results of the blending process from level one further blend with the spatial, haptic, and more generally “forceful” embodied behavioral dispositions sparked by the music, resulting in slightly more tractable, no longer just “intramusical”, but cross-modal, and therefore a bit more “representational” meaning. The drums that reverberate, followed by the slow-paced, but firm rhythm of the piece, provide the notion of REGULAR SEQUENCING and PATH of the movement; the ascending and descending pitch progressions occasionally specify this path as VERTICALLY oriented (going up or down, sometimes in a JAGGED manner); the auditory timbre of the synthesized waveform gradually changes its central property that can cross-modally best be described as physical SHARPNESS (which is an ultimately haptic quality), and the harmonic instabilities (for instance, in the opening tritone) result in the sensation of MOTILE INSTABILITY: the ascending or descending movement never ends on an appropriate, stable point, but rather results in a “miscalculation”, in which the steps of the moving agent always end up on the wrong, unstable point of support.

The Ming character equally goes through comparable blending operations on level two, where the formal tension in his facial expression and rhythmicity of walking from level one get enriched here by the PATH that he progresses through (the hall or cavern mentioned above), the VERTICAL movement at the beginning of his entrance (he ascends to the hall while standing on a hidden platform; it effectively lifts him up in a simulation of floating). The walk through the cavern suggests another cross-modal effect: that of exaggerated SIZE. First, the hall itself is rather big. Yet, more importantly, the camera stands at one end of the hall, shooting Ming from the front, so that, as he walks toward it, the merciless “ruler of the universe” becomes bigger and bigger. Finally, the entire setting is red: rather than just irritating the perceptual system (as on level one), here on level two the color causes a strong

cross-modal effect of increased EMBODIED TENSION (perhaps again “piercing” through the visual system like a haptically perceived physical object).

Again, the important bit for our analysis is the augmented effect which further comes from blending the results of the visual and the musical operations from the previous two paragraphs. On level two, the drums are blended with the reverberation in the hall of Ming’s entrance. Here they no longer function as just intramusical clues. Rather, their amalgamation with the visuals of the hall suggests the physical SIZE of the cavern that Ming progresses through. Metaphorically, this immediately projects on the size of the court, and of Ming’s power in general. Ming’s walking blends with the pace of the music to further stabilize the cross-modal notion of the PATH. The character’s “floating ascent” from the beginning of the scene is fully supported by the first two upward jumps in the musical melody (the tritone from the tonic to the augmented fourth, and then immediately the perfect fifth from the tonic to the upper dominant). Ming’s size is augmented by the size of some of the musical intervals, but also – by the occasionally “THICKENING” quality of the synthesized wave-form, which appears to pulsate, from a less toward a more dense, or sizeable, entity that it mimics, and may also partly relate to the exaggerated features in Ming’s appearance, such as the huge eyebrows and ornate collars. Finally, the HAPTICALLY piercing red color is augmented by the second timbral quality of the synthesized waveform, which creates the blended cross-modal effect of a (sharp) SHAPE.

In turn, all of these musical and visual elements combine with one another to create the cross-modal effect of slow and elevated movement, physical and metaphorical size (and thus, power), and a stabbing, piercing nature of the character presented. This level is still relatively underspecified, “pre-conceptual”, as cognitive linguists call it (Johnson, 1987; Sinha, 1999), and thus relatively poor in terms of representation. Yet, the first gist of referential tractability starts to appear.

On level three, such resultant cross-modal effects further blend with the ineffable tension inferred from level one, to result in the affective interpretation of the music accompanying the character. As the main trait of this music is formal instability, followed by the cross-modal inconsistency of inferred musical movement through a PATH, the blended affect can be only related to *insecurity*. This is an even more tractable – and thus representational – property: one needs to know quite well semantically what “insecurity” is in order to register such an emotion or report it in experiments. It is not yet fully conceptual, of course – if anything, it denotes a *psychological state* rather than a referential object – but it still slowly approaches a fully utterable semantic meaning. Yet, the relative lack of specificity of emotional responses results in the question often posed by music emotion theorists: *whose* insecurity does this music evoke? Is it the listener’s – so that *we* are supposed to be frightened by what we hear? Or the composer’s – so that he is communicating *his own* personal states? Or really (and most likely, in film scoring at least) “the music’s” – where the inferred musical structure “intends” to evoke in us the emotions of *characters* it will be telling us about on levels four and five? In such a scenario, the tonal and cross-modal instability blend to suggest that, whoever will be participating in the “musical action” on levels four and five, feels quite frightened and insecure during the slow, socially-elevated and “piercing” movement of the character, which level two has motivated. The same logic applies to the second affective state this music appears to connote all along – that of *tension*, which is not quite the same

as instability, but the two certainly motivate an additionally blended *uneasiness*, if not the ultimately aesthetic *ugliness* during the perception of the score.

A similar effect happens in the visual scene, though, of course it is much more clearly referential in nature. One’s cognitive system infers there is a human(oid) character involved right away, so this makes the process much quicker and more intuitive. Still, Ming’s facial expression and slow and careful walking, blended with the appropriate music in the background, augment the sensations of instability, tension, and uneasiness, to ultimately motivate a fearful response. The red color that the film take bathes in no longer simply irritates the perceptual system or motivates a cross-modal, piercing sensation, but becomes more strongly associated with a negative, “fight-or-flight” emotion and, generally, with danger. So the ultimate blend of the visual and musical instability, walking, the character’s physical augmentation and color suggests to us that what approaches is evil and dangerous, and may cause a lot of harm. Therefore, one is supposed to either run away or prepare for a fight.

Level four finally becomes tractable (and representational) enough to motivate a story. It is still rather unspecified and generic, but remains descriptive enough to comprise a meaningful event around which further interpretations can mount on levels 5 and 6. When it comes to the musical structure alone, it is this generic nature of the construct that allows for numerous interpretations – which is why the semantic interpretation of even programmatic music is always open-ended. The question to be asked here is, therefore, who or what would be a good candidate for the perceptual tensions and instabilities, cross-modal forces, paths, disbalances, sizes and sharpnesses, and affective uneasiness, ugliness, ultimately fear, motivated by the musical score? All of these results of blending operations from the first three levels again blend with one another here on level four to create a set of “generic” agents who or which can perform in practice the activities with such musical connotations. A dangerous and ugly villain walking slowly and carefully, augmenting in size, showing off his power, can indeed be a good option “suitable” for this music. But other scenarios are very viable, too. For instance, while listening to this theme, I have no problems visualizing a spherical geometrical object moving slowly along a path and increasing in diameter, with edges becoming sharper when the synth timbre becomes more piercing, and approaching me from the screen in a most dangerous fashion when the tritone plays. And, of course, countless other options are possible: a “budding-and-expanding” alien (perhaps an orc from the *Lord of the Rings*), a dangerous, lurking snake (maybe from *Conan the Barbarian*), or a slowly approaching killer-truck (as in *The Hitcher*, starring the all-too-convincing Rutger Hauer). This level provides the referential stage-set, makes the semantic construct representational enough to motivate manageable scenarios, and ultimately puts the imagination into motion.

The visual segment once again narrows down the options due to its nature, which is much more denotational than that of music. So the “something” that moves effortfully in order to overcome instability, occasionally ascends and augments, changes its shape from a relatively dull to a quite sharp, if not fully piercing, posture, which is ugly and thus causes uneasiness and fear, obviously relates to a human being in strange clothing frowning around in an apparent effort to intimidate whoever is watching. Note that on this level the contextual specification of the character is still not available: without a richer knowledge in background cognition (e.g. of the *Flash Gordon* canon), we still interpret the fellow at

hand in a rather generic fashion. Of course, the grounding levels proposed here are just an approximation, so this “visual” level four cannot be *totally* decontextualized. In fact, even for such a “bare” interpretation of a dangerous agent moving around, we need to know quite a lot from experience – for instance, what makes someone’s walking unstable, or what a frowning facial expression tends to indicate. Still, compared with the “culturally rich” level five, the interpretation here is still rather basic, and may relate to different scenarios: a supernatural, intergalactic villain as much as someone’s angry father ready to give his child a dressing-down for some recent mischief.

And again – the musical and the visual “conceptual” levels four get mutually blended here to augment (through the music) the image that has been constructed (mostly through the visual clues) – of a strongly dangerous, intimidating, and angry character (who makes us “see red”), walking around in a pronounced, if not exaggerated, fashion, frowning at us, and becoming both physically and metaphorically bigger as he approaches. Again, this last blending process makes the musico-visual semantic interpretation a step more representational, setting the stage for the fully contextual, experiential, and “culturally rich” understanding on level five.

The fifth level blends all of the resultant, blended structures from the prior tiers, in particular the affective level three and the conceptual level four, with appropriate elements from the cognizer’s background knowledge, in effect asking the following question: *who or what, from our experience, saliently realizes the formal, embodied, affective, and conceptual specifications that we have reached on the lower levels?* An answer to this question requires quite specific imagery and a lot of clear denotations, so that the meaning construction process becomes fully representational here, also in the “linguistico-semantic” sense that I have adhered to in this article. Now the music and the visuals start to integrate with our, ideally structurally identical, and therefore (formally, cross-modally, etc.) “appropriate” experiential, social, and cultural knowledge to provide fully specified imagery. The tension, piercing and carefully-walking qualities, uneasiness, ugliness, and fear, and the character walking toward us here clearly relate to some culturally-laden specifications: the exaggerated walking in a flamboyant suit implies the character is *regal*; the tritone tells us he is *diabolical*; the red color implies *anger*, but also *blood* with all of its connotations – sorrow, suffering, battlefields; the piercing quality of the music suggests a *stabbing weapon*. And these clues now help us go much further, in concentric sub-levels specifying the “culturally-rich”, level-five, context: if we know enough of the *Flash Gordon* story, we infer that the situation at hand involves Ming the Merciless in his court, appearing for the first time before a bunch of subjects, all too happy to intimidate them and thus reinforce his sadistic reign.

Of course, further sub-levels are also possible, i.e. a situation in which the *Flash Gordon* saga is satirized. In such a deliberately changed context, the Ming theme might create an effect totally opposite from fear. On the contrary, it may be used to *undermine* the scary qualities of a Ming alter-ego, as it often happens with satirical memes (one may think of the much more famous Darth Vader theme from *Star Wars*, and hundreds of online jokes [ab]using it, to get a clearer idea of what I have in mind here).

And finally, all of these frequent uses in the same or similar extramusical contexts, which have entrenched the representational qualities of “meaning” associated with the theme, get ultimately elaborated on the *personal* level six. Importantly, this is not a matter of arbitrary

association, either. While it is completely possible to hear the Ming theme in a totally unrelated context, e.g. while having a cup of coffee in a bar, the connection between that particular bar and the music, even if our cognitive system should make it, is not in the least “semantic” in the sense which I am proposing here. Rather, to become a real instance of personal, yet strongly connotational, referential, and representational musical meaning, the theme needs to undergo interpretation on *all the lower levels* before reaching the personal one. So it must be perceived as unstable, involve some excessively careful movement, cause a bit of uneasiness if not full fright, relate to a male character menacingly moving around in a royal fashion, and ultimately to Ming the Merciless from *Flash Gordon*. Only then can a personal circumstance become relevant as well, for instance, the fact that I became interested in similar villains and Max von Sydow’s other performances upon seeing the film for the first time. As mentioned, this can now piggyback on lower levels, e.g. the affective layer, adding this sentiment for the actor to my appreciation of the music, or the conceptual one, where my subsequent knowledge of Max von Sydow’s other roles – perhaps the knight in *The Seventh Seal* – may alter my adult appreciation of the Ming-with-the-music entrance. In effect, these changed interpretations can now influence level five, in further concentric circles of meaning generation. During the process, of course, the meaning becomes ever more specific, if not idiosyncratic, making one’s personal musical interpretation relatively connotationally restricted, and thus strongly representational in the linguistico-semantic sense. A possible graphic sketch of the first five levels of the multilevel-grounded blending process is given in Figure 3<sup>5</sup>.

#### 4. Conclusions

This article has provided an overview of the author’s approach to musical meaning, which builds upon the cognitive linguistic conceptual blending theory, and proposes to extend it with a “multilevel-grounded” semantics. With regard to the problem of representation, it has attempted to show that the concept in question, in music at least, is not an “either-or” phenomenon, but can rather be observed as a gradient. The more the cognizer progresses from lower to higher tiers of grounded blending, the more his or her appreciation of musical meaning becomes representational in the linguistico-semantic sense that I have adopted in this paper: that of translating between the musical stimulus and the rich, referential, imagistic and/or linguistic interpretation of this stimulus.

In conducting such an analysis, I hope to have also suggested an important path for researchers to move forward on the problem of representation in cognitive science in general. It appears to me that an approach based on such “continua” may help break some unnecessary epistemological boundaries in the field. For instance, much as they are fiercely debated in the applicable field(s), I think that embodiment and amodality do not need to

---

<sup>5</sup>Note that the figure here refers to the *musical* interpretation alone (without the visual segment from the film). Analyzing the visuals would require an additional such grounded blending diagram, and the further amalgamation of the music *with* the visuals would require yet another one, and quite complex, with two entire grounding continua interacting with one another. Due to the limitations of space, such a combined analysis is not possible in the present paper. An interested reader may refer to my book for a corresponding analysis (Antović, 2022: chapter 7, especially pp. 106-108).

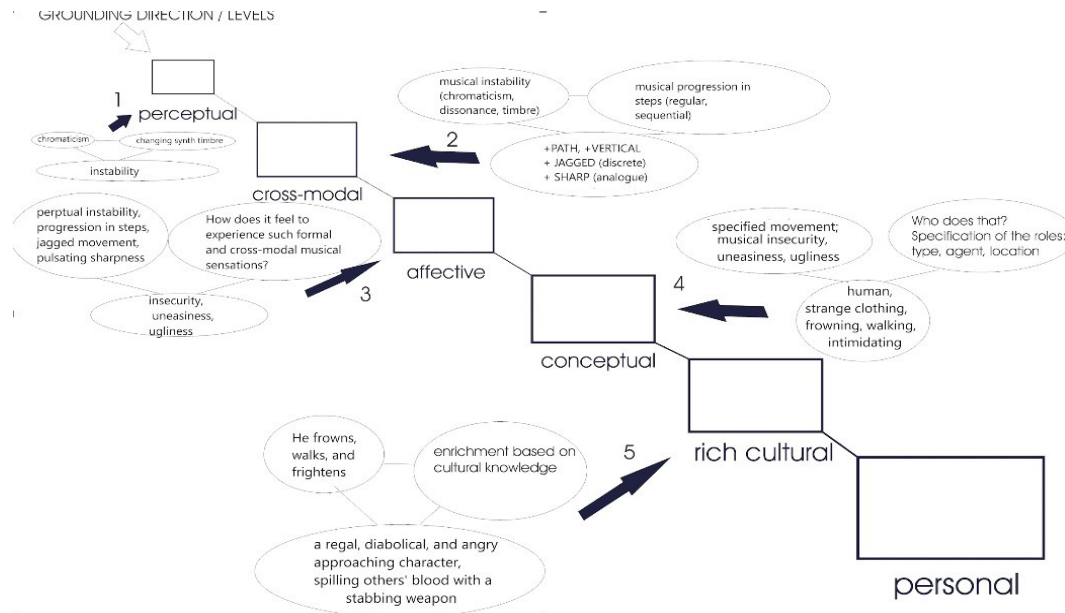


FIGURE 3. A multilevel-grounded interpretation of the Ming the Merciless theme

be taken as contradictory constructs. In the case of (musical) meaning, at first glance, embodied operations appear to prevail, but there may be additional amodal principles beneath this embodiment (e.g. the construction of formal gestalten or cross-sensory correspondences motivating an urge to “move” to some music), which are later specified as instances of more tractable, connotational meaning (e.g. a description of “moving armies”). So it may be that the factors providing *options* for interpretation are amodal, whereas embodied cognition followed by cultural circumstances serves to further “stabilize” an option selected by a culture, a community, or an individual. Such a reconciliatory approach can perhaps take us forward in the attempt to solve the riddle of mental representation.

## Acknowledgements

This research was supported by the following sources: Science Fund of the Republic of Serbia, Grant No. 7715934, Structuring Concept Generation with the Help of Metaphor, Analogy and Schematicity -SCHEMAS; Branch in Niš of the Serbian Academy of Sciences (project no. O36-24, coordinator, academician Aleksandar Kostić; project manager, prof. Mihailo Antović); and the Internal Project of the Faculty of Philosophy, University of Niš, “Scientific Publications in Teaching English Linguistics and Anglo-American Literature and Culture” (No. 300/1-14-1-01).

## References

Antović, M. (2022). *Multilevel grounding: A theory of musical meaning*. Abingdon and New York: Routledge.

- Antović, M., Küssner, M. B., Kempf, A., Omigie, D., Hashim, S., & Schiavio, A. (2023). “A huge man is bursting out of a rock”: Bodies, motion, and creativity in verbal reports of musical connotation. *Journal of New Music Research*, 52(1), 73-86.
- Bierwisch, M. (2009). Sprache – Musik – Bild. Zeichentypen und ihre Konsequenzen. *Zeitschrift für Literaturwissenschaft und Linguistik*, 39, 8-34.
- Chattah, J. (2023). *Film Music: Cognition to Interpretation*. Abingdon and New York: Routledge.
- Chemero, A. (2013). Radical embodied cognitive science. *Review of General Psychology*, 17(2), 145-150.
- Chomsky, N. (2000). *New horizons in the study of language and mind*. Cambridge: Cambridge University Press.
- Cook, N. (2001). *Analyzing musical multimedia*. Oxford: Oxford University Press.
- Coulson, S., & Oakley, T. (2005). Blending and coded meaning: Literal and figurative meaning in cognitive semantics. *Journal of Pragmatics*, 37(10), 1510-1536.
- Fauconnier, G. & Turner, M. (2002). *The way we think*. New York: Basic Books.
- Firth, J. R. (1957). *Papers in linguistics 1934–1951*. Oxford: Oxford University Press.
- Gibbs, R. (2005). *Embodiment and cognitive science*. Cambridge: Cambridge University Press.
- Jackendoff, R. (2002). *Foundations of language*. New York: Oxford University Press.
- Johnson, M. (1987). *The body in the mind*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Juslin, P. N. (2013). What does music express? Basic emotions and beyond. *Frontiers in Psychology*, 4, Article 596. <https://doi.org/10.3389/fpsyg.2013.00596>.
- Lakoff, G. & Johnson, M. (1999). *Philosophy in the flesh*. New York: Basic Books.
- Langacker, R. (1999). Virtual reality. *Studies in the Linguistic Sciences*, 29(2), 77–103.
- Lerdahl, F. & Jackendoff, R. (1983). *A generative theory of tonal music*. Cambridge, MA: MIT Press.
- Osgood, C., Suci, G. & Tannenbaum, P. (1957). *The measurement of meaning*. Champaign: University of Illinois Press.
- Pustejovsky, J. (2016). Lexical semantics. In: Aloni, M. & Dekker, P. (eds). *The Cambridge handbook of formal semantics* (pp. 33-64). Cambridge: Cambridge University Press.
- Searle, J. R. (1978). Literal meaning. *Erkenntnis*, 13, 207-224.
- Sinha, C. (1999). Grounding, mapping and acts of meaning in Janssen, T. & Redeker, G. (eds). *Cognitive linguistics: Foundations, scope and methodology* (pp. 223-256). Berlin: De Gruyter Mouton.
- Swain, J. P. (1997). *Musical languages*. W. W. Norton.
- Varela, F. (1991). *The embodied mind*. Cambridge, MA: MIT Press.

## WHERE REPRESENTATIONS BEGIN: FROM BASIC TO DECOUPLED REPRESENTATIONS

Anastasia Garbayo de la Torre\*

**Abstract:** This paper is concerned with the question of where representations begin, i.e., tries to answer the question of which are the simplest organisms that possess representations. This question is addressed to two types of representations: basic representations, which coordinate sensory input with motor output, and decoupled representations, which collect environmental information without being directly tied to specific behaviors. To answer this question, I will adopt an evolutionary approach by which each kind of representation will be mapped onto different stages of the evolution of the nervous system. More precisely, I will argue that basic representations emerged when a simple neural system appeared, while decoupled representations will be linked to the formation of a central nervous system that enables off-line simulation abilities. Such an evolutionary framework will provide empirical reasons to establish which are the minimal conditions that an organism must fulfill in order to qualify as representational, with respect to both basic and decoupled representations. In other words, I will argue that the biological grounding of both types of representations serves as the demarcation element that helps answer the question of where basic and decoupled representations begin.

**Keywords:** biological grounding, decoupled representations, evolution, learning, nervous system

### 1. Introduction

Representations are used both in philosophy and in the cognitive sciences as a separate level of analysis that must be postulated in order to explain an organism's behavior. That is, a representation is an internal state that represents external features and conditions, and whose function is to guide or trigger behavior. In this paper, I will accept the general notion of representation by which a representation is an internal state that can be true/fulfilled or false/unfulfilled in virtue of its content. However, many different kinds of representations have been distinguished, including decoupled representations.

Decoupled representations have been defined as those whose function is to track features of the environment, so they are not directly tied to a specific behavior. In other words, they serve to identify and gather information from the environment that can be used to guide different behaviors as appropriate, or no behavior at all.

Philosophers have long debated the nature of representations (Dretske, 1981; Millikan, 1984; Papineau, 1984), making it one of the main topics in the philosophy of mind. In the relevant literature, we can find many different accounts of various representational statuses, attempting to specify the minimal conditions that an organism must fulfill in order to qualify as having basic and decoupled representations (Shea, 2018; Artiga, 2021; Butlin, 2021).

\*Universidad de Valencia (UV); garbayonastia@gmail.com;

ORCID ID: <https://orcid.org/0009-0003-8202-4881>

©2025 Anastasia Garbayo de la Torre. This article is distributed under a Creative Commons CC BY-NC 4.0 license.

Such accounts would also help establish a demarcation element between different kinds of representational and non-representational systems.

Nevertheless, there is no consensus on this issue, and it is unclear what might resolve the debate, or whether it is simply a matter of definitions. That is to say, there are no obvious or straightforward factors that will determine a final decision, and the debate among the most important current accounts seems less about empirical evidence and more about how the terms in question are framed.

For this reason, in this paper, I propose a different approach. More precisely, I shall look to the evolution of cognitive abilities to establish how both basic and decoupled representations appeared during the evolutionary process. Such an account will make clear that there are demarcation elements between different representational and non-representational systems, namely, the biological grounding from which each type of representation emerged. Furthermore, it will provide empirical evidence and reasons to distinguish between different types of representations.

This is the main goal of this paper: to establish the biological conditions under which different kinds of representations became possible and how both basic and decoupled representations were derived from a biological framework, in order to answer the question about where representations begin.

## 2. Basic Representations

In this paper, I will adopt the general notion of basic representational states or representations. According to this notion, “a representational state is an internal state that has semantic content, and that possesses truth-, correctness- or fulfillment-conditions in virtue of having that content” (Schulte, paper draft, 2). It is also commonly accepted that representations must satisfy the so-called explanatory constraint, which states that an adequate account of representational status should ascribe representational states to a system if, and only if, these ascriptions have explanatory purchase (Burge, 2010; Ramsey, 2007; Sterelny, 1995). In other words, ascriptions of representational states are legitimate only if they can play a role in genuine representational explanations.

Although this definition of basic representations is broadly accepted, there are many competing attempts to specify the conditions under which something qualifies as a representation or, put differently, the criteria an organism must fulfill to be ascribed representations. I will now present a specific account of representational status, as it has received significant attention and has been widely adopted in recent years. I will then argue why it is preferable to first examine how representations emerged during evolution and how such an evolutionary approach can help answer the demarcation question and determine where representations begin.

An increasingly influential account of representations is the one proposed by Souza Filho (2022). He advocates a dual proposal for the minimal conditions that a state must meet to be considered representational, establishing two criteria to address the demarcation question: the success pattern conditions and the constancy mechanism condition. According to the success pattern condition, a state is representational if a success pattern exists in the system’s behavioral output and the system uses the state as a proxy for the presence of an

external feature to achieve the desired result. The constancy mechanism condition, in turn, states that “a sensory state represents an external feature provided that the sensory system employs a constancy mechanisms in its production, such that the production of the state remains constant when the external feature obtains, despite the variety of proximal stimuli reaching the system’s sensory apparatus” (Souza Filho, 2022, p. 13).

This account attempts to resolve the demarcation question by defining two minimal criteria that an organism must satisfy to qualify as representational. However, both criteria seem somewhat arbitrary. For instance, one might ask why both criteria must be satisfied, why not just one, or why these criteria are prioritized over others. The answer to these questions seems to rely largely on unjustified reasons. Let me break down these statements.

Given the generally accepted definition of representations, for a state to qualify as representational it must possess accuracy conditions and be capable of accounting for the possibility of misrepresentation (Grice, 1957; Fodor, 1987; Millikan, 2004). That is, the representation must be evaluable as correct or incorrect based on whether it corresponds to the real world, and it must not only represent accurately but also allow for and account for error.

Thus, any account that seeks to establish the minimal criteria for a state to qualify as representational must justify that these criteria are sufficient and necessary to explain both the accuracy conditions and the possibility of misrepresentation of the state. Of course, Souza Filho’s account must also demonstrate this explanatory capacity.

His dual proposal suggests that both the possession of a success pattern and a constancy mechanism are necessary and sufficient for a state to consist of accuracy conditions and to be capable of misrepresentation. On one hand, he argues that the success pattern condition establishes the accuracy conditions of the state, as the representational state is used to guide behavior in accordance with how the world actually is. It also accommodates the possibility of error, since if the state does not correspond to reality, behavior based on it will fail. On the other hand, the constancy mechanism would maintain a reliable representation of the environment, ensuring accuracy conditions, while also accounting for the possibility of error due to failure of the constancy mechanism.

However, Souza Filho’s account lacks a clear justification for why both criteria are necessary to guarantee accuracy conditions and the possibility of misrepresentation. Specifically, the success pattern alone seems enough for both conditions. In other words, his proposal appears to accommodate the idea that what allows accuracy and error is the system’s use of an internal state as a proxy for reality. Yet, he simultaneously suggests that another condition is necessary to account for them. Nevertheless, although the constancy mechanism may help to stabilize the state, it does not seem strictly necessary for an error to occur, nor is it necessary to guarantee the accuracy conditions.

The need to establish a second criterion for representations arises from the problem of liberality: based solely on the condition of success pattern, representations could be attributed to organisms that are clearly non-representational. This is why Souza Filho introduces a second criterion. However, the minimal criteria for representations must be those that are both sufficient and necessary to explain and accommodate accuracy conditions and misrepresentations, as these are the defining elements of representations. And Souza

Filho’s theory seems not to need the constancy mechanism criterion to accommodate either the accuracy conditions or the possibility of error. That is, the need to restrict the attribution of representations cannot be the only reason for establishing a second criterion.

In conclusion, Souza Filho’s dual proposal attempts to establish two minimal conditions for the attribution of representations but does not provide a clear justification for why both are truly necessary. On the contrary, the satisfaction of one of the minimal conditions seems sufficient. Thus, the constancy mechanism is not strictly necessary for a system to misrepresent, nor is it necessary for accuracy conditions, making the requirement for both criteria to ascribe representations seem somewhat arbitrary.

By contrast, adopting an evolutionary approach provides a biological grounding for understanding representations. Through this perspective, we gain empirical reasons rather than relying on explanations that may appear arbitrary. That is, the evolutionary approach can explain accuracy conditions based on adaptability and the natural selection of representations according to their ability to guide efficient behaviors. Likewise, an evolutionary approach can accommodate the possibility of misrepresentation as a natural by-product of the evolution of representations, since a representational system must balance accuracy and efficiency. Error is possible in the search for this balance.

Although an evolutionary approach appears to offer clear advantages by providing empirical explanations, we must first ask whether a biological grounding for representations can truly be established. In this paper, I will argue that the biological grounding of basic representations is the presence of a neural system, i.e., groups of neurons that form a simple neural network. Such a neural system was necessary for basic representations to emerge during the evolutionary process. Furthermore, the possession of such a neural system serves as the tipping point between representational and non-representational organisms.

### 3. Decoupled Representations

In the last section, I outlined the general notion of representations. However, there are many different types of representations, including decoupled representations. Put simply, decoupled representations have the function of identifying and collecting information, while leaving open what the information will be used for. According to the general notion of decoupled representations, they serve to track features in the environment and may trigger different behaviors as appropriate, or no behavior at all.

Although various definitions of decoupled representations can be found in the specialized literature, this article will focus on the theory defended by Sterelny (2003), as it is one of the most developed accounts of decoupled representations. Sterelny differentiates between coupled and decoupled representations. Coupled representations are internal states that directly and automatically connect environmental features to specific behaviors. When an organism relies on coupled representations, the same environmental input consistently triggers the same behavior. In this case, the behavior is tightly coupled to the input. By contrast, decoupled representations are distinguished by their capacity to give rise to multiple possible actions or behaviors following the same input. In Sterelny’s schematic terms, decoupled representations are “internal cognitive states which (a) function to track features

of the environment, and (b) are not tightly coupled functionally to specific types of response” (Sterelny, 2003, p. 49).

Sterelny further explains the evolution of decoupled representations, noting that they are belief-like and have evolved from broad-banded responses. An organism with broad-banded responses can trigger different behaviors depending on the context, providing a range of available behavioral options. Thus, decoupled representations satisfy the response breadth criterion, which states that a representation is decoupled if it can trigger different behaviors in different contexts.

To illustrate these concepts, consider a bird that is used to facing aerial predators. The shadow of a predator flying overhead creates an internal state that tracks the predator’s presence. A bird with coupled representations reacts to the shadow in a fixed manner, namely, by flying to nearby bushes to hide. In this case, the bird’s response is rigid and automatic: every time it sees a shadow, it flies to the bushes, regardless of the context. This stimulus-response pattern is tightly linked to the input.

In contrast, a bird with decoupled representations processes the shadow more flexibly. The internal state still tracks the presence of a potential predator, but it does not always trigger the same behavior. Depending on the context, the bird might fly to nearby bushes, stay still and camouflage itself, or fly up to confirm whether the shadow is indeed a predator. In this way, decoupled representations enable broad-banded responses, as the internal representation of the shadow as a potential predator can lead to a variety of context-sensitive behaviors.

As this example illustrates, response breadth refers to having a wide range of possible responses, which contrasts with the concept of narrow-banded response. Narrow-banded responses involve specific and rigidly determined behaviors in reaction to detected information. Broad-banded responses, on the other hand, allow information about the environment to trigger different behaviors depending on the situation. This flexibility is foundational for being belief-like, as Sterelny explains: “such a breadth of response is important for belief-like representation is not tied to specific actions and hence evolves from broad-banded response” (Sterelny, 2003, p. 55).

Thus, Sterelny’s account demonstrates that decoupled representations are belief-like, meaning they can guide different behaviors depending on the context rather than being selected to trigger one specific behavior. Furthermore, this belief-like quality distinguishes decoupled representations from goal-like representations. Unlike goal-like representations, decoupled representations are a record of the world that are “not functionally tied to specific action” (Sterelny, 2003, p. 51). This means that decoupled representations can (potentially) influence a variety of actions.

#### **4. Evolution of Nervous System and Learning**

As claimed above, I will adopt an evolutionary approach to address the question of where different kinds of representations begin. However, it is worth noting that Sterelny has also attempted to establish an evolutionary perspective on this issue. In the first part of his book, Sterelny discusses the evolution of animal cognition, focusing on the emergence of

decoupled representations as a critical stage in the evolution of cognitive abilities. Decoupled representations are considered especially significant because they appear necessary for certain higher-order cognitive skills. If their importance lies in this connection to advanced cognition, then, as I previously stated, it is essential to pinpoint the time of their emergence in evolutionary history. Sterelny attempts to do this by arguing that decoupled representations arose relatively late, coinciding with great apes' capacity for social surveillance of their own cognitive states. However, this suggestion may be incorrect, as it seems more plausible that decoupled representations emerged earlier, as I will argue later in this paper.

To answer the question of where different kinds of representations begin, I propose that we should study the most basic biological mechanisms that enabled both basic and decoupled representations during evolutionary history. Moreover, identifying such specific features would allow us to determine which organisms possess each type of representation.

To develop this argument, I will draw on the theory proposed by Ginsburg and Jablonka (2019) regarding the evolution of the nervous system (NS) in relation to complex cognitive abilities. In their book *The Evolution of the Sensitive Soul*, Ginsburg and Jablonka seek to uncover the biological basis of consciousness. To this end, they explore the evolutionary trajectory from minimal neural system to complex central nervous system (CNS), analyzing how these stages correspond to different forms of learning, from nonassociative (NAL) to unlimited associative learning (UAL).

Although their work focuses primarily on the evolution of consciousness, Ginsburg and Jablonka identify three types of representations and map them to different stages in the evolution of the nervous system and learning. I will adopt their framework to map both basic and decoupled representations, as these are understood in philosophy and as defined earlier in this paper.

#### **4.1. From Neural System to Central Nervous System: From Nonassociative Learning to Unlimited Associative Learning**

Ginsburg and Jablonka describe the evolutionary development from individual neurons to centralized nervous systems as a response to increasing demands for sensory integration, motor coordination, and adaptive behavior. This progression is tied to the evolution of learning mechanisms, which also became more complex, transitioning from nonassociative to associative learning.

Neurons arose from cellular precursors with sensory and motor functions, which specialized into distinct sensory and motor cells. The connecting cells between them evolved into true neurons through mechanisms such as synapse development, allowing efficient information transmission (communication). Early nervous systems, as seen in organisms like cnidarians, consisted of diffuse nerve nets capable of basic sensory and motor coordination.

Over time, neural networks became specialized and hierarchical, driving the centralization of nervous systems to meet the demands of multicellular complexity. This centralization enabled faster and more efficient coordination, as well as improved processing and integration of sensory information and control of body functions.

Nonassociative learning, involving habituation and sensitization, was linked to simple neural networks, and allowed organisms to modify their responses to repetitive stimuli. These simple neural networks supported basic forms of memory and adaptation through changes in the strength of synaptic connections of patterns of neural activity. These abilities enabled organisms to optimize their interactions with the environment, conserving energy by ignoring irrelevant stimuli or enhancing responses to significant ones.

As neural systems grew more sophisticated, associative learning emerged (AL). This form of learning, based on synaptic plasticity, enabled organisms to connect sensory inputs with behavioral outputs. Associative learning allowed for more effective adaptation to the environment, as repeated experiences strengthened or weakened neural pathways.

The evolution of centralized nervous systems marked a shift toward more complex organizational and functional capabilities. Centralization supported hierarchical levels of information processing and control mechanisms, which simultaneously provided additional layers that supported predictive relationships between stimuli and actions. These mechanisms enabled adaptive behaviors that were not dependent on hardwired reflex or immediate environmental inputs. This evolution established connections between stimuli, actions and consequences, introducing associative learning in two forms: limited and unlimited associative learning (LAL and UAL, respectively).

Limited associative learning is the process by which “spontaneous and stochastic exploratory activities and preexisting simple reflex reactions can be combined, reinforced, and recalled” (Ginsburg and Jablonka, 2019, p. 295). In other words, it combines exploratory activities and reflexive reactions to create anticipatory responses, enabling organisms to predict and adapt to the environment more effectively. It encompasses classical and operant conditioning, allowing organisms to associate specific actions with outcomes.

Unlimited associative learning, in contrast, enables the formation of novel and compound associations, such as second-order conditioning, through complex neural architectures. In other words, UAL is characterized by its generativity: it can form practically limitless associations among stimuli and reinforced actions. These capabilities require hierarchical processing, memory systems, flexible reinforcement mechanisms, and integration units.

The transition from limited to unlimited associative learning occurred as neural complexity increased, allowing organisms to integrate compound or sequential stimuli and perform more sophisticated behaviors. More precisely, “distinct, hierarchical, bidirectionally integrating neural circuits, which support motor-sensory-motor integration and new memory and value systems must have evolved to enable open-ended AL” (Ginsburg and Jablonka, 2019, p. 345). The development of UAL provided significant adaptive advantages, enabling organisms to adjust behaviors based on past experiences and expand their ecological niches.

Unlimited associative learning required complex central nervous systems with the following specific features: first, hierarchical neural processing was required to organize sensory and motor inputs into compound patterns. Furthermore, memory systems retain traces of learned associations for future use. Thirdly, flexible reinforcement systems enabled dynamic adaptation and reinforcement of learning based on prediction errors, allowing the system to detect discrepancies between expected and received outcomes. Finally, integration units allowed the combination of sensory and motor information for adaptive behavior.

The possession of UAL also facilitated new cognitive abilities, such as the ability to retrieve and reconstruct learned patterns (even when the stimuli are incomplete), giving rise to pattern completion and memory retrieval. UAL also involves discriminative flexibility, enabling an organism to distinguish and respond to novel compound stimuli. Moreover, it fosters cumulative learning, constantly improving behaviors based on past experiences. This adaptability provided a crucial evolutionary advantage in complex and dynamic environments.

The evolution of neural systems into centralized nervous systems facilitated increasingly sophisticated learning mechanisms. The link between nervous system complexity and associative learning underscores the adaptive significance of these changes, culminating in the development of unlimited associative learning. Furthermore, Ginsburg and Jablonka define three types of representations and map them to different stages in the evolution of nervous system complexity. This mapping will be the topic of the next section.

## 4.2. Mapping of Representations

Ginsburg and Jablonka define three types of representations that can be distinguished within their neurobiological framework: namely, neural (nonmental) representations, neural-mental representations, and neural-mental-symbolic representations. I will now analyze how they are defined and how they are mapped to different stages in the evolution of the nervous system, relating to the learning abilities that each stage supports.

Before discussing the three types of representations defined by Ginsburg and Jablonka, it is important to note that their account does not attribute any form of representation to the most basic nervous systems. In contrast, they argue that basic nervous systems are only capable of causal and functional correspondences, such as reflexes.

According to Ginsburg and Jablonka, representations are associated with limited associative learning, which in turn is linked to simple centralized nervous systems. Specifically, neural representations are illustrated by an LAL reaction, where the organisms relate a simple conditioned stimulus to an unconditioned stimulus. To facilitate this, a simple centralized nervous system is required, as demonstrated in the previous section. Secondly, the authors define neural-mental representations as those that allow for composite and dynamic percepts, as they model the relationships between stimuli and responses in an integrated manner. It is worthwhile to notice that the authors define mental states as conscious, so neural-mental representations are conscious states. That is, mental representations are subjectively experienced configurations that integrate information across sensory modalities and contextual reinforcements, enabling a primate form of self-awareness.

According to Ginsburg and Jablonka, another defining feature of mental representations is that they are intentional: they are *about* something in the external environment. To support this type of representations, the organism must not only be capable of information processing but also of generating predictions and updating internal models. Thus, mental representations require a higher level of nervous system complexity. More precisely, mental representations are supported by the dynamic architecture of UAL, as it enables hierarchical information processing, multimodal integration of sensory stimuli and motor response, and the use of memory systems that allow the updating and storage of composite patterns.

Therefore, mental representations arise when evolution fosters higher plasticity and canalization “so that different inputs can be categorized and processed by the representing system and lead to the same output (canalization), and the same inputs can be processed in different ways within the system to yield different outputs (plasticity)” (Ginsburg and Jablonka, 2019, p. 371).

Finally, neural-mental-symbolic representations are the most complex type of representations, as they involve symbolic cognition and abstract reasoning, which are based on the foundations established by UAL. These representations are related to encoding information symbolically, requiring abstract and linguistic thought abilities. Some additional information about symbolic representations is provided by Ginsburg and Jablonka, such as their relation to cultural and linguistic systems. However, this paper is not concerned with any of the abilities related to symbolic representations as defined by Ginsburg and Jablonka, so I will not delve into them further. That is, in this article, I will not make any statements about abilities such as abstract reasoning, symbolic cognition, or linguistic thinking.

To summarize, Ginsburg and Jablonka map neural nonmental representations to basic centralized nervous systems that are complex enough to enable LAL. On the other hand, mental representations require the abilities supported by the dynamic architecture of UAL and, as explained in the previous section, necessitate the formation of a complex CNS. However, neither the definition of neural representations nor the definition of mental representations used by Ginsburg and Jablonka matches the definitions of basic and decoupled representations to which I have subscribed in this article.

## 5. Biological Grounding of Representations

As mentioned throughout the paper, my main aim is to establish how both basic and decoupled representations emerged during the evolutionary process and, therefore, to identify the biological grounding that enabled them. These foundations can also serve as a demarcation criterion to answer the question of where basic and decoupled representations begin. To achieve this goal, I will use the biological explanation of Ginsburg and Jablonka, as outlined above regarding the evolution of the nervous system, and I will map basic and decoupled representations to different stages of complexity.

### 5.1. Basic Representations

Regarding basic representations, I argue that an organism qualifies as representational as soon as it possesses a basic neural system. That is, basic representations emerged during the evolutionary process when a neural system was formed. In this way, a neural system serves as the biological grounding for basic representations and, thus, as the demarcation element between representational and non-representational organisms.

To back this argument, I will demonstrate how this claim is supported by the evolutionary theory of Ginsburg and Jablonka. Moreover, the neural system as a minimal condition for representations aligns with the basic features that philosophers typically require from a complete account of representational status. Let’s first recall the basic notion of representations. A representation is an internal state that represents the world to trigger behavior, and

it can be correct/incorrect or fulfilled/unfulfilled according to its content. In other words, the basic function of a representation is to match behavior to circumstances.

This idea aligns with the necessity of a neuronal system for representations, as it introduces a form of internal information processing that is not present in purely reactive systems. This is consistent with the requirements that Souza Filho imposes for an organism's behavioral explanation to qualify as intentional success explanation (as opposed to a causal-sequence or non-intentional success explanation). A neural system, unlike tropistic systems, not only transmits a fixed response to a given stimulus but also integrates, processes, and modulates information before generating an action. That is, the stimulus generates an internal state that not only reacts but also stores information, allowing the internal state to be modulated based on, for example, new experiences.

Thus, a neural system enables the same sensory signal to trigger different responses depending on the context, rather than merely eliciting a fixed, unmodifiable reaction. As such, the account defended here cannot be explained by causal-sequence or non-intentional success explanations. The neural activity it describes cannot be reduced to a causal sequence or an adaptive biological function that does not involve representations, since a neural system introduces an internal processing of information, making it necessary to postulate an intentional success explanation.

My statement that a neural system is needed to possess basic representations is also supported by the evolutionary explanation of Ginsburg and Jablonka. On the one hand, they claim that simple neural networks enable organisms to optimize their interactions with the environment. Indeed, representations serve to trigger behaviors that are appropriate in each situation. On the other hand, they argue that neural patterns map simple relationships between stimuli and responses. In fact, neural patterns are the basis of representations that encode information about environmental conditions and trigger responses accordingly.

It is worth noting that Ginsburg and Jablonka claim that LAL evolved in moving bilateral animals that possess brains. However, an interesting study recently carried out by Bielecki et al. (2023) shows that box jellyfish (*Tripedalia Cystophora*) do possess associative learning, even though they lack a brain. This suggests that the preconditions for associative learning set out by Ginsburg and Jablonka may be incorrect. In this paper I will make a distinction that is not clear in their account. Specifically, they claim that associative learning is only possible with a properly formed brain. In contrast, given the recent findings of Bielecki et al., it seems that what is required for associative learning is a primitive or basic nervous system, such as that possessed by box jellyfish. Therefore, in this paper I defend that basic learning could be enabled by a decentralized nervous system, while unlimited associative learning is supported by a complex central nervous system capable of hierarchical processing, among other functions.

Moreover, this recent scientific evidence not only contrasts with Ginsburg and Jablonka's views on the relationship between the nervous system and learning but also suggests that the authors may be mistaken in not attributing representations to the most basic neural systems. Specifically, the latest studies indicate that representation can exist in simple nervous systems, without the need for a centralized nervous system.

As we have seen, Ginsburg and Jablonka argue that the most primitive neural systems allow only causal and functional correspondences, i.e., direct reflex responses to stimuli without complex processing, and that representations only emerge when the nervous system is sufficiently sophisticated. However, we cannot assert that every simple nervous system operates solely through reflexes. As the recent scientific evidence by Bielecki et al. demonstrates, organisms with basic nervous systems can learn and remember associations, which cannot be explained solely by simple reflexes.

For instance, experiments show that box jellyfish can learn from past experience, such that they not only respond automatically to stimuli but also use previous information to adjust their behavior. This illustrates that these organisms can generate internal states that guide behavior, that is, organisms with basic neural systems can form internal representations of their environment.

The associative learning observed in box jellyfish supports the idea that basic neural systems can also qualify as representational because they distinguish a mere reflex from a representation. A reflex, on one hand, is a fixed, unmodifiable response to a stimulus. In contrast, representations are internal states that store information about the environment and adapt the behavior of organisms. Therefore, associative learning involves representations because it stores information about the environment (i.e., if an organism can learn that a given stimulus is linked to a particular consequence, it provides evidence of a representation) and allows behavior to be modified based on past experience (indicating flexible adaptation rather than an automatic reaction).

In sum, the ability of organisms with basic neural systems to engage in associative learning supports the thesis presented in this article: representations are already present at this stage, without the need for a centralized nervous system. This stands in contrast to the theory of Ginsburg and Jablonka.

Further developing and recovering the idea of misrepresentation as a fundamental and defining feature of representations (that is, the notion that a state qualifies as representational only if it is capable of misrepresenting the world), this account accommodates the possibility of misrepresentation. For example, if an organism smells a predator, it represents the presence of a predator. If, after such a representation, it triggers the avoidance behavioral response, the representation is correct if the predator is, in fact, nearby. By contrast, the organism would have misrepresented the environment if the predator is not present, and the avoidance behavior would serve no purpose. This is what distinguishes a simple cause-effect (non-representational) channel between input and output, i.e., the possibility of misrepresentation. If the external circumstances are correctly reported by the internal state, the behavior will succeed in its goal; however, if the behavior is incorrectly supported by the internal state, then the behavior won't be appropriate. In other words, the representation would be incorrect or unfulfilled if its content does not match (or represent) the real environmental conditions.

To put it another way, and using the definitions outlined previously, my account, which establishes a neural system as the minimal condition for basic representations, fulfills Souza Filho's success pattern condition. According to him, a state is representational if a success pattern exists in the organism's behavior, and the organism uses the state as a proxy for the

presence of an external feature (e.g., a predator) to achieve the desired result. That is, when a neural pattern encodes the information and represents a predator in the environment, using this representation as an indicator of the predator, the organisms will trigger the avoidance behavior to achieve the desired result of avoiding the predator. If there is indeed a predator in the environment, the success pattern will occur in the organism's avoidance behavior. If there is no predator, the organisms would have misrepresented the world, and the success pattern would not occur.

In claiming that my neural system criterion satisfies the success pattern condition, one could argue that I must confront the same issue that Souza Filho sought to address: the success pattern condition alone sets the boundary of intentionality too low, allowing systems that are clearly non-representational to be treated as if they were. This is why Souza Filho introduces the constancy mechanism as a second condition. That is, it could be argued that my account also requires an additional constraining condition.

That is, as for the criticism of liberality, it could be argued that my proposal is too liberal by not imposing additional conditions, as Souza Filho's constancy mechanism does. However, postulating a neural system as a minimal condition for possessing representations by itself overcomes the problem of liberality that Souza Filho faces: establishing the possession of a neural system as a minimal condition for the attribution of representations already imposes a sufficient constraint, which prevents this account from being considered too liberal. Requiring the possession of a neural system is neither a trivial nor overly broad condition, since it excludes a large number of systems that are clearly not representational. Many of the organisms that more liberal accounts might accept as representational, e.g., plants, bacteria or any living organism without a neural system (Price, 2001; Artiga, 2016; Ganson, 2020), are automatically excluded from the category of representational systems. While Souza Filho's success pattern criterion might apply to overly simplistic systems, my proposal avoids this issue by requiring a concrete biological basis that restricts the attribution of representations.

Moreover, this account does not require the success pattern condition as a separate criterion, as the presence of a neural system already automatically possesses a success pattern by itself. It is not necessary to first demonstrate that an organism's behavior follows a success pattern in order to then include the neural system as a condition. On the contrary, the presence of a neural system already automatically implies the existence of a success pattern. This is because a neural system is not only a biological structure but also functions to coordinate the organism's behavior successfully in response to its environment. Thus, evolution has selected nervous systems that enable the generation of success patterns in behavior, ensuring adaptive responses to environmental stimuli. In other words, it is unnecessary to postulate two separate conditions, as possessing a neural system inherently implies that the organism exhibits a success behavioral pattern.

Thus, this account does not require additional conditions, such as the constancy mechanism, since the possession of a neural system already establishes a sufficient basis for representation. On the one hand, it filters out overly simple systems, thereby restricting the attribution of representations. On the other hand, it guarantees a pattern of behavioral success, as the neural system's function is to enable adaptive responses to the environment.

Finally, it does not rely on arbitrary definitions, as it is based on a concrete biological structure and its evolutionary function, rather than imposing conditions that might be considered more abstract.

To sum up, I am claiming that when a basic neural system appears during the evolutionary process, basic representations emerge. That is, the possession of a neural system is the minimal condition that an organism must fulfill in order to qualify as representational, and thus, it is the demarcation element between representational and non-representational organisms. This assertion satisfies the philosophical requirement for a complete account of representational status and is also justified by the biological/evolutionary theory. In other words, this account provides us with empirical reasons to map basic representation to the early stage when a neural system appears in evolutionary time, as empirical reasons to distinguish and define basic representations as has been done in this paper.

## 5.2. Decoupled Representations

Decoupled representations, understood as those whose main goal is to identify and collect information, are not directly tied to any specific response. In contrast, they can trigger different behaviors or no behavior at all. However, although they may not trigger any response, the ability to trigger different responses after the same input or after different ones must be explained. To do so, I argue that the ability to contrast present and past situations, calculate different available and possible behaviors, and choose among them according to the consequences they will entail, enables the ability to trigger different behaviors in the same environmental situations. More precisely, I assert that off-line simulation ability is the key feature that enables these capacities. In other words, off-line simulation is the defining feature of decoupled representations and the main ability required to enable UAL.

To possess this ability, decoupled representations require a specific architecture that supports hierarchical and open-ended processing. Thus, I argue that decoupled representations must be mapped to a stage of evolution where the nervous system has become complex enough to support such processing, which in turn enables off-line simulation. According to Ginsburg and Jablonka, a complex central nervous system is required to hold up these abilities.

Even if we can identify some differences between decoupled representations, as defined here, and mental representations as understood by Ginsburg and Jablonka, it is important to note that both types of representations require the same abilities, as I will show. That is, I claim that, despite their differences, the evolutionary timing of decoupled and mental representations can be the same: both types of representations emerge at the same point in the evolutionary process, as both require a complex central nervous system.

Regarding the differences, Ginsburg and Jablonka define mental representations as conscious states or subjectively experienced states that form the basis for self-awareness. I have already clarified in this paper that I am not committed to such features. Therefore, the definition of decoupled representations that I support here is not concerned with subjectivity or related features.

However, aside from this difference, decoupled and mental representations can be equated. As outlined earlier, both types of representations require a complex CNS, so it can be assumed

that they appeared at the same time in the evolutionary process. I will now demonstrate that the architectural and functional requirements for both types of representations are the same, justifying the idea that they emerged simultaneously.

Many of the capabilities that were required for the emergence of mental representations, as explained above, are also necessary for decoupled representations. First, a central nervous system is required for decoupled representations to emerge, as it enables hierarchical and open-ended processing. According to Ginsburg and Jablonka, the evolution of the nervous system toward centralization facilitates mechanisms that enable adaptive behaviors that are not dependent on hardwired reflexes or immediate environmental inputs. This is a fundamental requirement for decoupled representations: if they can trigger different behaviors after the same input, the same behavior after different inputs, or trigger no behavior at all, it indicates that the behaviors based on decoupled representations are not directly coupled to environmental conditions.

Furthermore, decoupled representations must be open-ended, as they are not directly tied to any specific behavior. As the basic definition of decoupled representations suggests, they lack a definite or predetermined outcome, meaning the information they carry can evolve, change or continue in multiple directions. Thus, the absence of a fixed answer or a specific way of being completed is one of the defining features of decoupled representations, allowing for flexibility and multiple possibilities. According to Ginsburg and Jablonka, such an open-ended ability emerged with the evolution of distinct, hierarchical, bidirectionally integrating neural circuits that support motor-sensory-motor integration and new memory systems. In other words, open-ended processing appeared during the transition from limited to unlimited associative learning. This provides empirical support for the claim that decoupled representations emerge only after a complex central nervous system has formed.

Hierarchical and open-ended processing are, at the same time, what enables the off-line simulation ability, which has been proposed as a defining feature of decoupled representations and has also been discussed as part of the capabilities of mental representations as defined by Ginsburg and Jablonka. To trigger different behaviors after the same input, decoupled representations rely on past experiences and associations, which are remembered by the organisms, and confront these with present environmental situations to calculate different possible behaviors to address the current situation. The organism then selects one of these possible behaviors based on error predictions and learning from the consequences of past behaviors in similar situations.

In other words, off-line simulation is enabled by the architecture of the central nervous system and, therefore, can be found in organisms that possess both mental and decoupled representations. Furthermore, off-line simulation is one of the abilities required for unlimited associative learning. As Ginsburg and Jablonka have shown, mental representation emerged from a complex CNS and enabled UAL. Now, I will demonstrate why decoupled representations can also be understood as emerging from a CNS, thus as enabling UAL.

Ginsburg and Jablonka outline the requirements that give rise to UAL, and they claim that mental representations enable it. That is, they show that mental representations are involved in UAL. Since the same abilities are required for both mental and decoupled representations, we can infer that decoupled representations, as defined here, are also involved in

UAL. I will argue this further to justify that mental and decoupled representations emerged at the same evolutionary moment.

Decoupled representations are involved in UAL abilities insofar as they enable a flexible reinforcement system, one of the requirements for UAL that Ginsburg and Jablonka define as the ability to dynamically adapt and reinforce learning based on prediction errors. This is also linked to the ability to form associations that can be updated as the organism encounters new experiences, which is another requirement for unlimited associative learning. In other words, UAL requires off-line simulation ability, because this general ability allows for predicting errors and updating those associations based on new experiences.

In this way, decoupled representations can be mapped to the evolution of the complexity of the nervous system, emerging at the same stage where a centralized nervous system is in place. At this point, the organisms are complex enough to support hierarchical and open-ended processing, which enables off-line simulation. Furthermore, decoupled representations are involved in unlimited associative learning to the extent that they possess these abilities. Thus, the biological grounding of decoupled representations is a central nervous system, and this can serve as a demarcation element: what distinguishes decoupled representational from non-decoupled representational systems is the possession of a central nervous system.

## 6. Conclusion: where representations begin

In this paper, I have answered the question of where both basic and decoupled representations begin from an evolutionary perspective. Specifically, I have mapped the evolution of the neural system toward a central nervous system, identifying in different stages of this evolutionary process where basic and decoupled representations emerge. On the one hand, I have argued that a basic neuronal system is sufficient to enable basic representations, understood as an internal state that matches environmental conditions with the appropriate behavior. On the other hand, I have claimed that a complex central nervous system is required for decoupled representations to emerge, as it enables hierarchical and open-ended processing, which in turn gives rise to off-line simulation, that is, the defining feature of decoupled representations.

Moreover, I have argued that mapping different kinds of representations in the evolutionary process helps us answer the question of where representation begins. By doing so, we can identify the biological grounding for each type of representation. If a neuronal system is required for basic representations to appear, the demarcation element between basic representational and non-representational systems is the possession of a neuronal system. Similarly, a central nervous system complex enough to support off-line simulation is the tipping point between organisms that possess decoupled representations and those that do not, since this complexity enables the emergence of decoupled representations in the evolutionary process.

Thus, mapping both basic and decoupled representations at different stages of the evolution of the nervous system provides empirical evidence to distinguish different types of representations and establish the minimal conditions from which they evolved. In this way, we can use this biological grounding as the minimal condition required for an organism to

qualify as representational in terms of both basic and decoupled representations. An organism qualifies as representational as soon as it possesses a neuronal system that matches the environmental conditions with behavioral output. Regarding decoupled representations, an organism possesses them if it is complex enough to have a central nervous system that supports off-line simulation abilities.

The biological grounding from which each type of representation emerged represents the minimal conditions from them, and thus, this condition can help us answer the question of where both basic and decoupled representations begin. The simplest organism that possesses basic representations is the one with the most basic neural system that supports the matching between environmental conditions and behavior. Similarly, the simplest organism that possesses decoupled representations is the one with the most basic central nervous system that supports off-line simulation abilities.

## References

- Artiga, M. (2021). Strong Liberal Representationalism. *Phenomenology and the Cognitive Sciences*. <https://doi.org/10.1007/s11097-020-09720-z>.
- Artiga, M. (2016). Liberal representationalism: a deflationist defense. *Dialectica*, 70, 407–430. <https://doi.org/10.1111/1746-8361.12152>.
- Bielecki, J., Dam Nielsen, S. K., Nachman, G., & Garm, A. (2023). Associative learning in the box jellyfish *Tripedalia cystophora*. *Current Biology: CB*, 33(19), 4150–4159.e5. <https://doi.org/10.1016/j.cub.2023.08.056>.
- Burge, T. (2010). *Origins of objectivity*. Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780199581405.001.0001>.
- Butlin, P. (2021). Directive Content. *Pacific Philosophical Quarterly*, 102: 2–26. <https://doi.org/10.1111/papq.12328>.
- Dretske, F. (1981). *Knowledge and the Flow of Information*. Cambridge, MA: MIT Press.
- Fodor, J. A. (1987). *Psychosemantics: The Problem of Meaning in the Philosophy of Mind*. Cambridge, MA: MIT Press, Bradford Books.
- Ganson, T. (2020). A role for representations in inflexible behavior. *Biology and Philosophy*, 35( 37). <https://doi.org/10.1007/s10539-020-09756-0>.
- Ginsburg, S., & Jablonka, E. (2019). *The evolution of the sensitive soul: Learning and the origins of consciousness*. MIT Press.
- Grice, P. (1957). Meaning. *The Philosophical Review*, 66, 377–388.
- Millikan, R. (1984). *Language, Thought and Other Biological Categories*. Cambridge, MA: MIT Press.
- Millikan, R. (2004). *Varieties of Meaning*. Cambridge, Mass: MIT Press.
- Papineau, D. (1984). Representation and Explanation. *Philosophy of Science*, 51: 550–72.
- Price, C. (2001). *Functions in Mind: A Theory of Intentional Content*. Oxford: Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780199242009.001.0001>.
- Ramsey, W. (2007). *Representation reconsidered*. Cambridge University Press. <https://doi.org/10.1017/CB09780511597954>.
- Schulte, P. (2024). *Toward a radical gradualism about representation*. Paper draft.
- Shea, N. (2018). *Representation in Cognitive Science*. Oxford: Oxford University Press.
- Souza Filho, S. (2022). A dual proposal of minimal conditions for intentionality. *Synthese*, 200(115), 1–22. <https://doi.org/10.1007/s11229-022-03546-5>.
- Sterelny, K. (1995). Basic minds. *Philosophical Perspectives*, 9, 251–270. <https://doi.org/10.2307/2214221>.
- Sterelny, K. (2003). *Thought in a hostile world: The evolution of human cognition*. Blackwell Publishing.

## PRESENTATION AND CONTENT

Ivan V. Ivanov\*

**Abstract:** It is commonly thought that an intentionalist account of perceptual experience accounts for the representational aspects of experiences better than the alternative, disjunctivist account. The contrary is argued here. Disjunctivism can account better for the person-level representational contents of experiences, in light of the distinctive features of those contents.

**Keywords:** perceptual content, phenomenal content, naïve realism, disjunctivism, intentionalism

What sort of representational content do perceptual experiences have? Might they even lack contents?<sup>1</sup> If experiences were representational states, as according to the intentionalist view of experience (Crane 2002; Tye 2000; Pautz 2010; Speaks 2015; Schellenberg 2018), the nature of an experience would delineate a specific representational content for the experience. But if not, the issue would be complicated. Experiential disjunctivists, in particular (Campbell 2002; Martin 2004; Johnston 2004; Brewer 2011; Langsam 2011), have argued that perceptual experiences are not fundamentally representational, but, instead, fit into two fundamentally different kinds. Those could be called presentations and projections. For example, a visual experience would consist either in a primitive relation to the item seen on the occasion, or in a state that, despite not being a presentation of anything, would be subjectively indistinguishable from the presentation of an item of some sort. The distinction concerns the sheer phenomenal character of experiences. Presentations would have a relational, world-involving phenomenal character, while projections would have a phenomenal character of a different sort. The proposed disjunctivist account of visual experience would generalize with some modifications to the other senses.

Based on the disjunctive account of experience, some (Travis 2004; Brewer 2006; Raleigh 2013) have argued, further, that experiences lack contents in the sense of specific and intrinsic accuracy conditions that are appreciable by the subject of experience. In response, it has been proposed that, regardless of whether one is a disjunctivist or not, it must be granted that experiences have such contents, since experiences can be non-arbitrarily associated with specific and non-trivial accuracy conditions (Siegel 2010; Schellenberg 2011; Giananti 2019). But there is a residual impression that such accuracy conditions would be a matter of artifice (Pautz 2009; Crane 2013). For some purposes, it would be suitable to ascribe an existential content to an experience, for others: a singular content. Similarly, for some purposes it would

---

<sup>1</sup>The contents in question would determine intrinsic, determinate and substantive accuracy conditions that would be conveyed to the subject (see Travis 2004: 1; Siegel 2011: Ch. 2). In this sense, the thesis is stronger than the one that experiences lack conceptual content, but consistent with the view that experiences are either partly grounded by or involve representations at the subpersonal level.

\*School of Politics and Law, Capital Normal University, Beijing, China; iivanov@warwickgrad.net; ORCID ID: <https://orcid.org/0000-0002-9445-7241>

©2025 Ivan V. Ivanov. This article is distributed under a Creative Commons CC BY-NC 4.0 license.

be suitable to regard the experience as representing blueness, for others: a disjunction of colors, for yet others: an appearance property corresponding to blueness. And in the absence of a commitment to a representationalist account of the nature of perceptual experiences, there seems to be no reason why any content ascription should be deemed truer or deeper than the rest. So there is a residual sense in which robust representational contents go with intentionalism about perception, and are in the least unnecessary under disjunctivism.

It should be noted that some disjunctivists (Logue 2012b; Kennedy 2013) have posited both a relational and a representational aspect in the nature of successful perceptual experiences in order to better account for some aspects of hallucinations in terms of the remnant representational aspect. On such views, perceptions would have both relational phenomenal character and robust representational contents.<sup>2</sup> But two essential aspects would mean that the nature of veridical perceptual experiences is disunified, and also threatens to render the putative relational aspect explanatorily redundant.<sup>3</sup> By contrast, I propose that disjunctivists can and should provide a robust account of perceptual content otherwise than on the basis of a representational aspect in the nature of perceptual experiences. In the core cases, this would be purely due to the relational aspect. In the rest, this would be due to the experience's subjective indiscriminability from a core case.<sup>4</sup>

The proposed account of experiential content is not merely the best one available to the disjunctivist, but turns out to be preferable to the account of experiential content that an intentionalist could offer. This is because, as will be argued, only the disjunctivist can explain certain distinctive features of experiential contents. Specifically, they can make better sense: i. of the residual allure of pluralism about experiential content; ii. of the distinctive restrictions on experiential contents (the Laws of Appearance). And so the point of the discussion is not the old one (Logue 2014; Ivanov 2017) that even the proponent of disjunctivism could allow that experiences have intrinsic representational contents (and so that their view is not subject to objections based on the plausibility of such contents), but that disjunctivism helps account for such contents better.

---

<sup>2</sup>Such views are to be distinguished from earlier attempts to reconcile relationalism and representationalism, on which only the representational aspect was deemed relevant to the explanation of the phenomenal character of perceptions (see Byrne & Logue 2008; McLaughlin 2010; Schellenberg 2010). Instead, the relational aspect is supposed to fully account for the phenomenal character of perception, and the representational aspect: for something else, e.g. the rational and concept-grounding significance of hallucinations.

<sup>3</sup>I am well aware that the authors have argued that each essential aspect is needed to account for a different aspect of experience (see also Mehta 2021). But the resulting accounts of veridical perceptual experience are treading a fine line between incoherence and exceptional complexity. For example, Mehta must argue that the relational and representational essential aspect ground two different sorts of experiential phenomenal character. For an assessment of the explanatory efficacy of views of this sort, see Niikawa (2019).

<sup>4</sup>Martina (2022) makes some progress in this regard, specifically: in accounting the intentional directedness of perceptual experiences to particulars. The account on offer here is more systematic and ambitious insofar as it aims to simultaneously account for the directedness and accuracy conditions of perceptual experiences, and to do so in a way that is superior to the one available under intentionalism.

## 1. Perceptual Phenomenal Content

Perceptions, at least paradigmatically, are states with phenomenal character (in other words, they are experiences).<sup>5</sup> When you see an oval-shaped object, there is something it is like for you. At least superficially, hallucinating based on the same neural activity, or seeing a round object in some distorting context, also producing the same type of neural activity, would for the subject be exactly like the perception. We can call the aspect of phenomenal character shared by such experiences an oval\* aspect.<sup>6</sup>

Perceptual experiences of specific sorts also tend to cause beliefs and judgements of corresponding specific sorts in attentive perceivers lacking reasons to doubt the accuracy of their senses. It is plausible that the phenomenal type of an experience at least partly determines the attributive aspects of the contents of the world-directed attitudes (judgements, beliefs) that would normally be caused by it. For example, an experience with a round\* phenomenal character would normally cause an attribution of roundness. This might be because that aspect of the experience would intrinsically determine a content representing roundness, or because it would do so in a subject possessing shape-recognition capacities. Such content may ordinarily be expressed by stating that something looks to be round to the subject of that experience. This much is uncontroversial and should be acknowledged by any theorist of perception. But I will argue for something stronger: experiences with a round\* aspect represent roundness intrinsically.

Pautz (2010) has given a series of seemingly theory-neutral arguments for that conclusion. For example, it seems that a subject could learn the concept ROUND, purely based on having hallucinatory experiences, disconnected from instantiations of roundness, insofar as the experiences feature such a phenomenal aspect.<sup>7</sup>

In response, it has been said that this argument begs the question against those who would deny that the nature of experiences should be fundamentally characterized in terms of a relation to propositional contents (Alford-Duguid & Arsenault 2019). On the envisioned alternative (Travis 2007; Brewer 2011) experiences might relate us merely to particularities (objects, events, and tropes), with generalities (i.e. property-types) entering at the level of thought or of the conceptually penetrated aspects of perception.

It is a difficult question whether Pautz's arguments are question-begging or not, and I will not pursue it. My strategy in arguing that experiences have intrinsic representational contents is different from his. The view that perceptions present us solely with particulars is supposed to be an articulation of the naïve realist account of perception. Below, I propose

---

<sup>5</sup>I intend to remain neutral on the debate whether perception must be phenomenally conscious, although from what follows it would become clear that if there were unconscious perception, it must take a very different form from conscious perception (not being a presentation of the world, it would not be subject to the constraints I argue necessarily apply to perception qua presentation).

<sup>6</sup>Note that I am not supposing that (and, in fact, deny) that Oval\* gives the most determinate classification of the phenomenal character of the matching experiences.

<sup>7</sup>Of course, many would deny this possibility, precisely because they have very strong theoretical commitments regarding what it takes to represent a property. But off-hand, it seems that you can learn the concept merely by having an indistinguishable experience from the one you would when looking at a round thing. This would be an even more pervasive phenomenon, if, as many suppose, there are no perfectly round things.

what I take to be the conservative version of naïve realism: that is, the version that departs minimally from common sense<sup>8</sup>

On it, the experiences involved in perception in ideal circumstances would be transparent and detailed presentations not just of the particulars perceived, but also: of their intrinsic characteristics. I then show that this view of the phenomenal character of optimal perceptions has significant implications both for how the phenomenal character of suboptimal perceptions (including illusions) should be understood, and also for what experiences which would introspectively match a perception of either type might say about the world.

## 2. The Limits of Presentation

Recall that the disjunctivist takes a perceptual experience to either consist in the perception of an object and its character or in a state that is merely subjectively indistinguishable from a perception of a specific type (cf. Martin 2004). The respective view of perception has been called naïve realism.

Specifying naïve realism has made it clear that what disjunctivists have in mind is a thesis about the phenomenal character of the experience (see Logue 2012a; Soteriou 2016; Langsam 2017). Consciously perceiving would be a matter of a special, relational type of consciousness, constituted by the perceived object. I call such consciousness *presentation*. By contrast, whatever nature the consciousness involved in a perceptual hallucination has would be hidden from the subject (otherwise they would be able to tell, based on introspection, that they are not perceiving). Thus the naïve realist is at once a *phenomenal* and a *metaphysical* disjunctivist about perceptual experiences.

A presentation would necessarily be luminous: in a good case, you would always be able to tell that your experience is a relation to an object purely by introspecting your experience. This would also entail that in a good case one would have access to conclusive evidence for the existence of an object, and in some cases: for its being in some condition. Thus the naïve realist also manages to earn epistemological disjunctivism: in a good case one has a better reason to affirm the existence of an object than in a bad case.

As it turns out, specifying the constitutive requirements for the presentation of an object paves the way for deriving experiential content. But the former is an interesting and challenging issue in its own. What does it take for one to be presented with an external particular?

Most naïve realists have been happy to allow for some latitude in the circumstances under which an object could get presented. For example, these days most allow that perceptual illusions are cases in which an object would not just be perceived, but would be presented to the subject, despite significant distortion (French & Phillips 2020). This has created some

---

<sup>8</sup>There are more revisionary forms of naïve realism, to which the argument could not be applied. Some of them would be compatible with the view that experiences lack intrinsic representational contents. But it is unclear what sort of advantages such views would have. In fact, the most prominent revisionary versions (Beck 2019; Mehta 2022) have been designed with the explicit intention to be hospitable to intrinsic representational contents. From that perspective, the departure from common sense, which those forms of naïve realism would require, seems unmotivated.

worries: is this not stretching the notion of presentation beyond its limits? (Pautz 2020a: Ch. V).

Here is one way of articulating the worry. For an item to be presented in an experience, would not at least some of its intrinsic characteristics also have to be presented? This seems to be part of the common-sense understanding of what it takes to be presented with an object (see Johnston 1992: 257). But if that's the case, it seems that only cases of veridical perception of at least some of the object's intrinsic properties could involve the presentation of what gets perceived. This would result in a strange overall view of perception, on which most cases of perception would be fundamentally different from an extremely few elite cases involving presentation (Smith 2010).

I believe the right response to this problem is to allow that even a perceptual illusion would involve the presentation of the relevant intrinsic condition of an object (cf. Martin 2010, French & Phillips 2020). But, *prima facie*, at least, the proposal is implausible. For example, in what sense is it right to say that I am presented with the blueness of a bead (illuminated with pink light) when I experience the bead as black?

Moreover, the proposal seems unprincipled, threatening to result in too much latitude for presentation. For example, Logue (2012) has suggested that the hueish aspects of our visual appearances might amount to distorted presentations of corresponding types of reflectances. But this would strain the ordinary notion of presentation. Could features of the subject's perceptual system play such an overwhelming role in determining the phenomenal character of color experiences without also managing to occlude the relevant intrinsic features of objects?

On the customary alternative, classically favored by naïve realists (Campbell 1993; Kalderon 2008; Allen 2016), color experiences would present *sui generis* properties of the surfaces of objects, rather than their reflectances. I submit that such a primitivist view of color is required if color experiences are to be understood as presentations of the colors of objects.<sup>9</sup>

Similarly, it has recently been argued that most hallucinations are misdescribed as such and are actually presentations of the intrinsic features of some non-standard objects (Raleigh 2014; Ali 2018). But is it plausible that if my brain were removed from my head and envatted, the residual subject would be presented with properties of the vat machinery, aspects of the computer program running them, or with one's brain processes? Neither option seems plausible. On the other hand, there seems to be no principled way to rule the options out, if one is willing to allow that distortion of the character of an object is compatible with presentation of the relevant characteristic.

So naïve realists are faced with a dilemma. They should not make the notion of presentation too restrictive, so as to rule in only fully veridical perceptions. But once they loosen it, they risk overshooting, distorting the notion beyond recognition. Is there a principled middle ground?

---

<sup>9</sup>This may be seen as a large cost for the view, but as it will be shown, the classical version of naïve realism also has significant explanatory advantages over other views of perception, which would be lost under more moderate versions of naïve realism.

### 3. What Goes into Presentation

Let us first take a look at the cases clearly fit to involve presentation of the object perceived.

Why is describing a case of normal perception of a round object as a case in which the object is presented plausible? Because it is also plausible that the roundness of an object (some of its intrinsic conditions) is presented. This was the crux of Johnston's observation. But why is this plausible? Johnston also provides a hint.<sup>10</sup> In the described case, it seems that not just the roundness, but what it is to be round is also presented in your experience. This is why you can easily tell what sort of object you are looking at: the symmetry, distinctive of roundness, is also plausibly presented in your experience, giving you a basis on which to distinguish the property from all others.

So suppose that in the optimal cases the presentation of an instance of a property would require the presentation of the property's nature. Why might one want to allow that instances of properties could also be presented in other cases? Consider why we might want to say that one sees the actual color of an orange ball even in conditions of poor illumination. You might say something like this:

Although the nature of the color is not presented in as much detail as in daylight, some of it (e.g. the brightness of the orange shade) is still presented. This is why turning on the lights does not seem to present you with a new property: it merely fills in the nature of the experienced color further (cf. Kalderon 2011).

So we can allow that an instance of a property could get presented in circumstances in which its nature is not presented as fully as normally. But note that this is based on evidence that some of its nature would still be presented. The underlying thought seems to be that if none of the nature of the property were present, it would make no sense to say that the property-instance itself is present. In such cases, the instance might stand in a robust connection to the experience. But it would not be experienced. Someone might push back and argue that, strictly speaking, only an aspect of the orange shade would get presented in poor illumination. But although it is natural to say this, I believe it is mistaken. Compare: strictly speaking I am not presented with a 3-D object, but just with its facing surface. I think the right position on this issue is that I am presented with both. Especially for naïve realists, the inevitable partiality of perception does not screen off wholes from being objects of perception, because they take perceptual phenomenal character to have depth.<sup>11</sup>

I will now argue that at least some cases of perceptual illusion could be understood similarly: as featuring a suboptimal presentation of the relevant property-instance.

---

<sup>10</sup>“Part of my pleasure in seeing color is . . . the pleasure of having access to the natures of those features of external things which are the colors” (Johnston 1992: 257).

<sup>11</sup>It should be noted that there might be cases in which it is right to say that only an aspect of the relevant property is presented. But not all cases will be like that. Compare: how much of a wall must I be presented with to be presented with the wall? In some cases, I would only be presented with one or several bricks, while in others, I would also be presented with the wall. But just as with the case of the wall it seems unlikely that we can specify a non-trivial sufficient condition for the presentation of the wall, we should expect that same when it comes to intrinsic features and their aspects.

In the dark, one cannot experience most of the nature of a shade. This seems to be merely a matter of low-grade resolution. But a presentation could feature not just low-grade resolution, but some distortion of the presented essential aspects. So an illusion might involve both a loss of essential detail presented, and translucency in the presentation of the remaining essential detail. For example, the black appearance of the bead might be due to: i. the pink light screening off everything to the nature of the bead's blueness, except its determinate lightness, and ii. affecting the manner of presentation of the lightness (making it appear slightly darker).

Here is one reason to buy into this view of perceptual illusions. In some cases, such as the Muller-Lyer illusion, the effects of gradually altering the perceptual context are better described in terms of constancy than in terms of variation. In the case at hand, by gradually erasing the hashes, it becomes clear that the actual sizes of both lines were experienced all along, with one's subsequent experiences gradually correcting the original slight distortion generated by the presence of the hashes (cf. Brewer 2006).

The account of perceptual illusions as involving distorted presentation of some essential aspect of an intrinsic characteristic of the perceived item helps explain why naïve realism need not be confined to the optimal cases of perception. The naïve realist can offer the same sort of explanation of the latter and of cases of perceptual illusion. Their account of illusion also helps explain why reflectances could not be presented in a hueish guise. There would be nothing qualitative in the nature of reflectances which could be presented distortedly in this manner. The situation is fundamentally different from that of the blue bead's looking black.<sup>12</sup>

Note that the proposed account of both optimal and suboptimal presentation of an object ultimately appeals to the presentation of some of the nature of an intrinsic characteristic of the item. So if this putative constraint on the presentation of an item is genuine, and the naïve realist account of perception is correct, all perception must involve the presentation of some of the nature of some of the intrinsic properties of the particular perceived, along with it.

## 4. Presentation and Representation

In this section, I will outline the connection between perceptual appearances and perceptual contents under the elaborated form of naïve realism. From the proposed account of what the presentation of an item must involve, it follows that perceptions do not just present objects/events, property-instances and aspects of their intrinsic nature, but also represent the objects/events as instantiating specific properties.

---

<sup>12</sup>A reviewer notes that such an account leaves many cases of mismatch as candidates for distorted presentation of the relevant feature. For example, would a colorblind person be presented with blueness when looking at one and having a greenish experience? I do not see why not. As long as there is a plausible story to be told about which aspects of the property get presented along with it, it is not out of the question that we have a genuine case of presentation. Note also that the account is considerably more restrictive than the recently proposed naïve realist accounts of illusion.

Supposing perception were a matter of presentation, external particulars would never be presented simply. They would always be accompanied by some of their intrinsic properties, and those, in turn, by some of their essential aspects. In short, presentation would involve qualitative inheritance (Kalderon 2017; Logue 2017). But precisely because of this, all presentations would have intrinsic representational contents.

This could be illustrated most clearly when it comes to a case of optimal presentation. In cases of this sort, the instance of the property would be presented along with the full nature of the property. So in ideal circumstances, when one sees a circular object, for example, one would not simply be presented with its roundness, but also with the nature of roundness. Based on one's experience of the object's roundness, not only would the subject be able to distinguish the object from distinctly shaped one, but this would be done on the basis of the trope's being represented as round.

Generalizing, we have the first principle linking presentation and representation:

- (i) The optimal way of being presented with an instance of F-ness would also present the full nature of F-ness: @. So supposing the perception of an F object, O in ideal circumstances would involve an optimal presentation of O's F-ness, the perception of O's F-ness would represent O as being F, by virtue of representing F as being essentially @.

For example, if there is a perfectly round object, when presented in an optimal way it would be represented as perfectly round. And an object shaded a maximally determinate shade 325 of Red would be represented as being that shade.

Most cases of perception would involve suboptimal presentation. However, even when the presentation of the relevant intrinsic characteristics would be suboptimal, they would still be represented in some (specific) general manner.

In cases featuring merely degraded resolution, without distortion, the experience would either omit or abstract from aspects of the nature of the instance. In such cases, due to the way it would appear, an instance of F would be represented as X (a specific determinable of F). For example, the perfectly round object would be represented as oval, and the determinate shade would be represented merely as crimson or as dark red.

In cases featuring distortion, translucent features of the experience would affect the character of the appearance of the instance. Perceptual appearances in such cases would be a matter of falsified concreteness. For example, in the blue bead's case, the truth in the appearance would be the color's darkness, and the falsity would be the specific degree of darkness. In such cases, instead of being represented as X (very dark), F would be represented as G (maximally dark), i.e. a close determinate in the same determinable family (dark).

Based on these examples, we can derive a second, wider-ranging principle:

- (ii) A suboptimal way of being presented with an instance of F-ness would represent some aspects of F's nature:  $\lambda(\alpha)$ . So supposing the perception of an F object, O, would involve the suboptimal presentation of O's F-ness, depending on whether the presentation of the F-ness is merely low-resolution or features distortion, the perception of O's F-ness would represent O either as X (a determinable of F) or as

G (a distinct close determinate of X), respectively, in virtue of F's being represented either as  $\lambda(\alpha)$ , or as  $\beta$ .

Under the assumption that perception is a matter of presentation, both principles turn out to cover all perceptions.

Perceptual experiences exceed perceptions. I take most perceptual hallucinations to be correctly called thus (implying an absence of perception). However, as Martin suggests (2004: section 3), such experiences are at least superficially defined in terms of their subjective indiscriminability from perceptions. But given that perceptions would have representational contents intrinsically, a state subjectively indiscriminable from a perception would also have content intrinsically:

Take a hallucination H and a perception P, from which H is subjectively indiscriminable. Suppose that P would involve the presentation of O and of some of its intrinsic character (O's F-ness). Given the specific circumstances of perception, some aspects of the nature of F-ness would be presented along with the F-ness. This would be either with or without distortion. As the perception would represent O's F-ness as being either @,  $\lambda(\alpha)$ , or as  $\beta$ , and thus O as being F or X or G, the hallucinatory experience would represent that something is either F or X or G on the same basis. So the corresponding hallucinatory experience, featuring no presentation of anything, would still represent the instantiation of a specific property. This would be due purely to its being subjectively indiscriminable from the presentation of some aspects of the nature of an instance, in some manner.

Note that to get this result nothing further needs to be presupposed about hallucinations (certainly not that they have a representational nature, or a representational aspect to their nature). Often, the negative epistemic conception of hallucination, on which hallucinatory consciousness would consist in nothing beyond an inability to tell one is not perceiving an object, has been found faulty due to an alleged difficulty it has accounting for the fact that hallucinatory experiences represent specific properties (Pautz 2010; Mehta 2024: Ch. VI). In response, some proponents of the epistemic conception have denied that hallucinatory experiences represent specific properties (Brewer 2011; Alford-Duguid forthcoming). The alternative response offered here is that an experience can represent a specific property merely due to its subjective indiscriminability from another.<sup>13</sup>

## 5. The Advantages of Presentation-based Representationalism

In this section, I will argue that disjunctivism not only can account for experiential contents, but can do so better than intentionalism. Basically, this is because the intentionalist has nothing more basic in terms of which to account for certain puzzling features of experiential contents. By contrast, for the disjunctivist, experiential contents would derive from the more fundamental nature of perceptual experiences.

---

<sup>13</sup>Martina (2024) puts forward a similar proposal concerning the intentional directedness of hallucinatory experiences. This might be due purely to their subjective indiscriminability from corresponding types of perceptions. My proposal is more ambitious, in that I derive specific accuracy conditions per hallucinatory experience with a specific phenomenal character, purely on the basis of its subjective indiscriminability from a type of perception.

a. the contents of hallucinations

An intentionalist aims to account for the matching phenomenal character perception and a matching hallucination in terms of the presence of the same semantic property. Assuming hallucinations lacked objects, they could not have singular contents. On the other hand, it is plausible that successful perceptual experiences would not be adequately characterized merely in terms of existential contents (Soteriou 2000). The natural thought is then to take both experiences to consist in something intermediate: a relation to a gappy or a demonstrative content (Tye 2007; Burge 2010; Schellenberg 2010). But, first, it is dubious that gappy contents are real contents (Tye 2014). Second, it is difficult to model experiential contents on indexical contents (Gersel 2019). Experiential contents could not be token-reflexive. Moreover, if perceptions made subjectively available only the character-like aspects of the respective contents, it becomes difficult to make sense of the distinctive rationalizing role of perceptual experiences.

A disjunctivist need not go for any of those options.<sup>14</sup> First, an experiential disjunctivist could adopt content disjunctivism. They could take perceptions to have singular contents, while taking hallucinations to have contents of a different sort. Moreover, they need not take the latter to have existential contents. Instead, they might propose a *sui generis* type of demonstrative content, distinctive to hallucinations. The hallucination would represent demonstratively, not in virtue of representing itself, but by virtue of being indiscriminable from a case of presentation.

b. pictorial content

The proposed account has another virtue: it offers a fresh take on what it means for perceptual contents to be pictorial. Many take this to be a matter of the format in which perceptual contents are encoded. But this will not do.

As Pautz has proposed, there seem to be distinctive constraints on the contents of experiences. For example, an experience could not represent that something is both round and square, or both to one's left and to one's right. By contrast, the scenario can be freely formulated in thought. This is an example of what can be called an Exclusion Law of Appearance (Speaks 2024). In addition, an experience can represent something as red only insofar as it represents it as extended and with a certain hue. By contrast, I can think of a red object without thinking of it as extended or as reddish. This is an example of what can be called an Inclusion Law of Appearance (Ibid.).

Intentionalists encounter problems with accounting for the Laws of Appearance's ostensible necessity. They either have to take it at face value and take it as a basic fact, or explain it away. Both options face serious problems (Pautz 2020b). Taking pictoriality to be a matter of representational format fits with the second strategy and inherits its problems.

By contrast, the disjunctivist can both respect and account for the ostensible necessity of Laws of Appearance, precisely because they take the phenomenal character of experiences to

---

<sup>14</sup>Above I have addressed a similar worry for the naïve realist that during distortion only parts of or surrogates for the intrinsic aspects would be subjectively available, and argued that in most such cases the intrinsic aspects would themselves be subjectively available. Moreover, accounting for the distorted appearance of the intrinsic aspects need not appeal to anything semantic: there will just be more and less transparent sorts of presentation (see French & Phillips 2020: 5.1).

be more basic than the corresponding phenomenal contents. The phenomenal character of a perceptual experience would either be a matter of presentation or be constrained by the range of objects of possible presentation via introspective indistinguishability. Since presentation is qualitative inheritance, the necessary consistency and determinacy of the inherited external reality would constrain both what scenarios optimal and suboptimal presentations could represent. Supposing hallucinatory experiences must match types of perceptions, they also could not represent any wilder scenarios.<sup>15</sup>

Disjunctivism offers a very neat explanation of pictorial content: it is content fully grounded in presentation. By contrast, propositional content is content not fully grounded in presentation.<sup>16</sup> Consequently, the contents of thoughts need not adhere as closely to the character of external reality as the contents of experiences.

## FUNDING

This paper is part of the project “A Metaphysics for Naïve Realism”, funded by China’s Ministry of Education Humanities and Social Science Research Grant scheme (project 24YJA-720002).

## References

- Alford-Duguid, D. (forthcoming). Hallucination without sensible qualities. In Ori Beck & Farid Masrour (eds.), *The Relational View of Perception: New Essays*. Routledge.
- Alford-Duguid, Dominic & Arsenault, Michael (2017). On the explanatory power of hallucination. *Synthese* 194(5): 1765-85.
- Ali, R. (2018). Does hallucinating involve perceiving? *Philosophical Studies* 175(3): 601-27.
- Allen, K. (2016). *A Naïve Realist Theory of Colour*. Oxford University Press.
- Beck, O. (2019). Rethinking naive realism. *Philosophical Studies* 176(3): 607-33.
- Brewer, B. (2006). Perception and content. *European Journal of Philosophy* 14(2):165-81.
- Brewer, B. (2011). *Perception and Its Objects*. Oxford University Press.
- Burge, T. (2010). *Origins of Objectivity*. Oxford University Press.
- Byrne, A. & Logue, H. (2008). Either/or. In Adrian Haddock & Fiona Macpherson, *Disjunctivism: Perception, Action, Knowledge*. Oxford University Press.
- Campbell, J. (1993). A simple view of colour. In John Haldane & Crispin Wright (eds.), *Reality, Representation, and Projection*. Oxford University Press.
- Campbell, J. (2002). *Reference and Consciousness*. Oxford University Press.
- Crane, T. (2002). The intentional structure of consciousness. In Aleksandar Jokic & Quentin Smith (eds.), *Consciousness: New Philosophical Perspectives*. Oxford University Press.
- Crane, T. (2013). The given. In Joseph K. Schear (ed.), *Mind, Reason, and Being-in-the-world: The McDowell-Dreyfus Debate*. Routledge.
- French, C. & Phillips, I. (2020). Austerity and illusion. *Philosophers’ Imprint* 20(15): 1-19.
- Gersel, J. (2019). Please mind the gappy content. *Philosophical Studies* 176(1): 219-39.
- Giananti, A. (2019). The way things look: a defence of content. *Review of Philosophy and Psychology* 10(3): 541-62.

<sup>15</sup>I develop this argument in Ivanov (forthcoming).

<sup>16</sup>This leaves it open that at least some propositional contents may be partly constitutively dependent on relations to presentations.

- Ivanov, I. V. (2017). Property-awareness and representation. *Topoi* 36(2): 331-42.
- Ivanov, I. V. (forthcoming). Naïve realism and the Laws of Appearance. *Australasian Journal of Philosophy*.
- Johnston, M. (1992). How to speak of the colors. *Philosophical Studies* 68(3): 221-63.
- Johnston, M. (2004). The obscure object of hallucination. *Philosophical Studies* 120(1-3): 113-83.
- Kalderon, M. E. (2008). Metamerism, constancy, and knowing which. *Mind* 117(468): 549-85.
- Kalderon, M. E. (2011). Color Illusion. *Noûs* 45(4): 751-75.
- Kalderon, M. E. (2017). *Sympathy in Perception*. Cambridge: Cambridge University Press.
- Kennedy, M. (2013). Explanation across good and bad experiential cases. In Fiona Macpherson & Dimitris Platchias (eds.), *Hallucination: Philosophy and Psychology*. Cambridge, MA: MIT Press.
- Langsam, H. (2017). The intuitive case for naïve realism. *Philosophical Explorations* 20(1): 106-22.
- Langsam, H. L. (2011). *The Wonder of Consciousness: Understanding the Mind Through Philosophical Reflection*. MIT Press.
- Logue, H. (2012a). Why naive realism? *Proceedings of the Aristotelian Society* 112(2pt2): 211-37.
- Logue, H. (2012b). What should the naïve realist say about total hallucinations? *Philosophical Perspectives* 26 (1): 173-99.
- Logue, H. (2014). Experiential content and naive realism: a reconciliation. In Berit Brogaard (ed.), *Does Perception Have Content?* Oxford University Press.
- Logue, H. (2017). Are perceptual experiences just representations? , in Bence Nanay (ed.), *Current Controversies in Philosophy of Perception*. Routledge.
- Martin, M. G. F. (2010). What's in a look? In Bence Nanay (ed.), *Perceiving the World*. Oxford University Press.
- Martin, M. G. F. (2004). The limits of self-awareness. *Philosophical Studies* 120(1-3): 37-89.
- Martina, G. (2022). Phenomenally-grounded intentionality for naïve realists. *Phenomenology and Mind* 22(22):138-48.
- McLaughlin, B. P. (2010). The representational vs. the relational view of visual Experience. *Royal Institute of Philosophy Supplement* 67: 239-262.
- Mehta, N. (2021). The fragmentation of phenomenal character. *Philosophy and Phenomenological Research* 104(1): 209-31.
- Mehta, N. (2022). Naïve realism with many fundamental kinds. *Acta Analytica* 37(2):197-218.
- Mehta, N. (2024). *A Pluralist Theory of Perception*. Cambridge, Massachusetts: The MIT Press.
- Niikawa, T. (2019). Classification of disjunctivism about the phenomenology of visual experience. *Journal of Philosophical Research* 44: 89-110.
- Pautz, A. (2009). What are the contents of experiences. *Philosophical Quarterly* 59(236): 483-507.
- Pautz, A. (2010). Why explain visual experience in terms of content? In Bence Nanay (ed.), *Perceiving the World*. New York: Oxford University Press.
- Pautz, A. (2020a). *Perception*. New York, NY: Routledge.
- Pautz, A. (2020b). The puzzle of the laws of appearance. *Philosophical Issues* 30(1): 257-72.
- Raleigh, T. (2013). Phenomenology without representation. *European Journal of Philosophy* 23(4): 1209-37.
- Raleigh, T. (2014). A New Approach to 'Perfect' Hallucinations. *Journal of Consciousness Studies* 21(11-12):81-110.
- Schellenberg, S. (2010). The particularity and phenomenology of perceptual experience. *Philosophical Studies* 149(1): 19-48.
- Schellenberg, S. (2011). Perceptual content defended. *Noûs* 45(4): 714-50.
- Schellenberg, S. (2018). *The Unity of Perception: Content, Consciousness, Evidence*. Oxford University Press.
- Siegel, S. (2010). *The Contents of Visual Experience*. Oxford University Press.
- Smith, A. D. (2010). Disjunctivism and illusion. *Philosophy and Phenomenological Research* 80(2): 384-410.
- Soteriou, M. (2000). The particularity of visual perception. *European Journal of Philosophy* 8(2): 173-89.
- Speaks, J. (2024). Pautz on the laws of appearance, internalism, and color realism. *Inquiry*, 67(8): 2271-82.

Ivan V. Ivanov

- Speaks, J. (2015). *The Phenomenal and the Representational*. Oxford University Press UK.
- Travis, C. (2004). The silence of the senses. *Mind*, 113(449): 57-94.
- Travis, C. (2007). Reason's reach. *European Journal of Philosophy* 15(2): 225–48.
- Tye, M. (2000). *Consciousness, Color, and Content*. MIT Press.
- Tye, M. (2007). Intentionalism and the argument from no common content. *Philosophical Perspectives* 21(1):589-613.
- Tye, M. (2014). What is the content of a hallucinatory experience? In Berit Brogaard (ed.), *Does Perception Have Content?* Oxford University Press.

## ROSENTHAL'S REPRESENTATIONALISM: IS CONSCIOUSNESS A HIGHER ORDER THOUGHT?

Ksenija Puškarić\*

**Abstract:** In this paper I discuss Rosenthal's theory of consciousness. Being in a conscious state, according to his view, consists in having an appropriate higher order mental representation – a higher order thought (HOT) about a lower order non-conscious state. I will explain the key points of his theory and discuss several problems, such as the status of qualia in explaining the phenomenal character of sensory experience, versus the role of the HOT content accompanying such states. A version of this problem also underlies the tension in reconciling Rosenthal's view that all mental states are intentional, with the view that sensory states are non-intentional, exhibiting qualitative properties. Finally, I raise a dilemma concerning the nature of transitive consciousness attributed to HOTs. If transitive consciousness is factive, and thoughts are not, how can all HOTs constitute transitive awareness?

**Keywords:** representationalism, consciousness, qualia, Rosenthal, non-conscious mentality

### 1. Introduction

The view of a sunset and the sound of a piano differ in how they are felt, or in *what it is like* to experience them, as Nagel said. They are both conscious sensory experiences, and such experiences always have a phenomenal aspect. This aspect presents the so-called hard problem of consciousness. As the knowledge argument (Jackson, 1982) famously suggests, one can know everything about the brain and still sensibly ask: *how* does such and such brain state give rise to any kind of 'feel'? The so-called explanatory gap is a real threat for naturalistic approaches that try to reduce consciousness to a physical state.

In this paper I want to explore a representational theory of consciousness put forward by D. Rosenthal in a series of papers. Instead of direct reduction of consciousness to brain states he proposes an *indirect* approach: first by reducing consciousness to non-conscious states, and then, reducing non-conscious states to physical states. This way, he attempts to avoid the explanatory gap and preserve the naturalistic view. In other words, Rosenthal's view is that consciousness is a *relational* property of mental states. Since not all mental states are conscious – he believes – it is possible to give an informative and non-circular explanation of consciousness, namely what conscious states have that unconscious states lack. The core of his theory is that a mental state is conscious when one is conscious of oneself being in that state. Being conscious of a mental state consists in having an appropriate higher order mental representation – a higher order thought (HOT) about a lower order non-conscious state.

Despite its initial appeal, many philosophers have argued that because of its complexity, the HOT theory loses plausibility. However, sometimes complex phenomena might demand a complex explanation – one could argue – if such is the only good explanation available. Led

\*Institute for Interdisciplinary Studies, Seton Hall University, 400 S Orange Ave, South Orange, NJ, USA; email: ksenija.puskaric@gmail.com; ORCID ID: <https://orcid.org/0009-0008-5378-2751>

©2025 Ksenija Puškarić. This article is distributed under a Creative Commons CC BY-NC 4.0 license.

by that principle, firstly I will present Rosenthal's argument to the conclusion that the only tenable view is that consciousness is a relational property of mental states, and briefly sketch his own reductive strategy. In the second part of this paper, I will focus on his argument for the actuality of non-conscious states. Since the explanation of consciousness – in Rosenthal's view – is based on non-conscious states, it is important for him to establish that there are no philosophical nor common-sense obstacles to non-conscious mentality. In the third part, I will briefly introduce Rosenthal's terminology which he uses to analyze the puzzle of consciousness. Then I will lay out the key characteristics of the HOT theory. Finally, in the fourth part I will discuss the problems of Rosenthal's theory regarding qualia, the explanatory power of HOTs, and the dilemma of whether HOT and transitive consciousness are interchangeable.

## 2. Consciousness as a relational property

Let me briefly reconstruct my understanding of Rosenthal's (1990, 1997) line of thought which led him to the view that the most plausible theory of consciousness lies within the realm of a representational theory.

- (1) We either explain consciousness in: a) non-mental or b) mental terms.
- (2) If a) then the direct reduction to non-mental terms leads to the (problem of) explanatory gap. Therefore, this strategy is not acceptable.
- (3) If b), then consciousness is a property of mental states.
- (4) Properties are intrinsic or relational.
- (5) If all mental states are conscious, then consciousness is an intrinsic property of mental states.
- (6) If consciousness is an intrinsic property of mental states, the explanation of consciousness is circular.
- (7) Therefore, it is not plausible that all mental states are conscious.
- (8)  $\therefore$  Therefore, consciousness must be a relational property of mental states. (from 3,4,7)

Let me consider the first premise in more detail. According to Rosenthal (1990, 1997), there are three strategies available in explaining consciousness:

- (1) explaining consciousness in non-mental terms (physicalism)
- (2) explaining consciousness in mental terms: (2a) mental & conscious states (Cartesian view), or (2b) mental & non-conscious (Rosenthal).

The strategy 1) promotes reduction of consciousness to neurological or functional states. According to Rosenthal, direct reduction is not likely to be successful because there seems to be an unbridgeable gap between consciousness and physical/functional states. Jackson's (1982) knowledge argument, as already mentioned, articulates a strong intuition that neurology/functionalism cannot answer the question of *how* a phenomenal state, say pain, could be a specific activation of neurons or instantiation of certain causal roles. Something crucial seems to be left out of the physicalist's story. Consequently, the best conclusion one can make

from the explanatory gap is the following: our inadequacy of understanding consciousness in physical/functional terms points to a corresponding gap in the world. Namely, consciousness is not identical to a brain/functional state. As Rosenthal himself points out we won't get a satisfactory answer to the question of how this or that neurological state gives rise to a 'feel'. That is his ground for rejecting *direct* reduction of consciousness to physical/functional states. Thus, the first strategy is not acceptable. This leaves him with a more plausible mental-to-mental reduction, which implies that consciousness is a property of mental states. Given the obvious premise 4), to the effect that 'properties are intrinsic or relational', there are two further possibilities.

First 2a), consciousness might be an *intrinsic property* of mental states; one might attempt to explain consciousness in mental terms and claim that all mental states are conscious. According to the Cartesian view, all mental states are conscious, i.e. the mind is transparent. To put it into other words, Rosenthal suggests that in the Cartesian view, consciousness is an inseparable property of mental states – it is an intrinsic property. Sensory states and bodily sensations especially count in favor of this view because it is difficult to see how one might be in pain without having a feeling of pain. The occurrence of non-conscious pains or itches seems extremely implausible. As Kripke (1972) pointed out, being in pain is feeling pain. However plausible this view might appear to be, Rosenthal (1990, p. 12) claims it faces a serious difficulty. If all mental states are conscious and we want to explain consciousness as being such and such a property of mental states, which are all conscious, then our explanation would appeal to consciousness and thereby it would be circular and non-informative. Therefore, to paraphrase premise 6), Rosenthal believes the Cartesian view of consciousness is unacceptable on the grounds of explanatory circularity.

A possible way of responding to Rosenthal's objection would consist in drawing the distinction between 'explanation' and 'definition'. Perhaps Descartes could define consciousness in terms of 'mental states which all have the property of being conscious'. Such a definition may be circular but not necessarily vicious, insofar as the explanation is not circular. Take for example the following circular definition: something is red iff it looks red. We pick out 'red' via experience of red, but 'red' is also a property of such and such a surface of the object. So, in the case of color the explanation is not circular although the definition is. Why not argue that an explanation like that is possible for consciousness?

Nevertheless, even if the Cartesian view of consciousness can escape from the problem of circularity there is another reason against this view. Rosenthal indirectly argues that the Cartesian view is simply false (in the next section I shall explain this point in more detail); there are non-conscious mental states, and our commonsense intuitions strongly support this view. Therefore, Rosenthal thinks that consciousness is not likely to be an intrinsic property. The only plausible option left is that consciousness is a relational property of mental states.

Rosenthal's reductive approach is a strategy which promises to circumvent both problems mentioned: the explanatory gap and circularity, while maintaining the spirit of the naturalistic approach to consciousness. Consciousness can be explained in terms of mental states, but of those states that are themselves not conscious. Furthermore, those non-conscious mental states can be reduced to physical/functional states. This way, Rosenthal proposes to extend the explanatory chain of consciousness-explanantia by introducing an intermediate between conscious and physical, non-conscious mental states.

To illustrate the strategy (“ $\rightarrow$ ” stands for “is explained by”):

Conscious states  $\rightarrow$  intentional/not-conscious states  $\rightarrow$  physical/functional states

Naturalistic reduction of consciousness proceeds in two steps:

1) reduction of consciousness to intentional but non-conscious states

This step of reduction concerns the hard problem of consciousness which consists in explaining what makes a mental state conscious. Rosenthal’s theory explains consciousness in terms of non-conscious states.

2) not-conscious states to neurological/functional states

Rosenthal argues that the question of how to explain having mental representation of the world in terms of physical/functional states is the problem of intentionality and not the problem of consciousness *per se*. (That is why I will focus on discussing the first stage of reduction.)

Next, Rosenthal must convince us that non-conscious states are not postulated *ad hoc*, just because the HOT theory needs them. As we have already seen, there is an intuition which is strongly against the possibility of non-conscious states. That is why Rosenthal must explain why we have such strong yet wrong intuitions and furthermore provide reasons for accepting non-conscious mentality. In the next section I will consider arguments in favor of those claims.

### 3. Non-conscious mental states

Some philosophers, Descartes for example, thought that the mind is transparent, that one is aware of every mental state one has. Especially for sensory states or bodily sensations like pain or hunger, it doesn’t seem intuitively plausible that one can have them while not being aware of them and not knowing what it is like to have them. Rosenthal’s (1990, p. 20) answer to these strong points is that not knowing that we are in a mental state doesn’t show that non-conscious mental states don’t exist. For non-conscious pains for example, there is no what it’s like to be in pain, and consequently, there is no awareness of feeling pain. Rosenthal (1986, p. 346) acknowledges that it is hard to see how there could be non-conscious states that could be *qualitative*. The difference between the experiences of redness and greenness is constituted by the distinct feeling of seeing red and green. Namely, we distinguish mental states of seeing red and green only if they are conscious. But there is no what it is like to see red or green if the experiences are not conscious. Therefore, one could argue, contrary to Rosenthal, there is no quality unless the sensation is conscious. Rosenthal (1990b) believes that even though we are not able to distinguish sensory states when not conscious, it doesn’t follow that there is no qualitative difference among them. Although we fix the referent of ‘redness’ on the ‘feel’ it still doesn’t show that not conscious states lack this quality. The way of fixing the referent need not coincide with an essential property of the referent. Rosenthal claims that both conscious and not conscious mental states of seeing red and green have qualitative aspects. However, in the latter case we are not aware of the difference. Not being aware of X doesn’t mean that there’s no X. Rosenthal concludes that there is no reason to suppose that sensory quality cannot occur unless the state is conscious.

Besides arguments which are supposed to show that there are no philosophical obstacles to the actuality of not-conscious mental states, Rosenthal (1990a,1990c) further gives examples from our everyday life which show that there is a commonsense intuition in favor of his view. Non-conscious mental states are the best explanation of some situations people encounter in everyday situations. Here are some such situations. First, it seems that emotions are good candidates for being non-conscious. Sometimes we are not aware of being depressed and we realize that when someone points that out to us: 'you seem to be depressed'. However, such examples seem to demonstrate only that one feels depressed after realizing it, but it doesn't show that the mental state of depression has a qualitative aspect while being unconscious. Next, consider the 'absent driver' case. Sometimes, a driver who is quite familiar with the road shifts while driving his attention from driving itself to thinking about how to, for example, spend his vacation. Suddenly, after ten minutes the driver realizes he wasn't aware of what he was doing or what he saw on the road at that time. However, even though he didn't pay attention to driving he didn't hit some trees or another car. That is best explained by the fact that he saw what was happening on the road. His sensitivity in behavior (breaking, turning the wheel, etc.) is better explained by assuming that he had visual information rather than merely reacting purely by reflex. Rosenthal argues that non-conscious perception is the best explanation of such behavior. Alternatively, one could argue that the driver had conscious perceptual experience, but he lacks memory of it within those 10 minutes due to the lack of attention. Even if in this case this strategy could work, Rosenthal thinks that lack of memory is not a plausible explanation for various phenomena psychologists have pointed out. For example, subliminal perception, peripheral vision or blindsight. Blindsight patients who are not able to tell anything about objects in the 'blind half' of their visual field, and report that they don't see them, are still able to act pretty much as if they have visual information. What blindsight patients lack, Rosenthal (1990a, 2006) argues, is phenomenal consciousness, the distinctive feel. Still, some information about objects came through blindsighter's visual system. They can form beliefs about the size and the color of objects; they are able to behave based on this information, and so on. In Rosenthal's view, blindsighters have non-conscious perception. Therefore, Rosenthal concludes, there are empirical reasons in favor of the actuality of non-conscious mental states: those states are intuitive after all, and there are no philosophical obstacles to the claim that consciousness is a relational property of mental states.

To summarize, Rosenthal's points for supporting the thesis that not all mental states are conscious are the following:

- (i) there are no philosophical reasons to reject non-conscious mentality.
- (ii) common sense is open to non-conscious mentality.
- (iii) some scientific cases of perception are best explained by appeal to non-conscious mental states.

In the next section I will focus on his explanation of phenomenal consciousness, conscious mental states that involve a distinct quality of "what is like" to be in such states.

#### 4. Consciousness explained

According to Rosenthal (1994), two basic uses of the term “consciousness” must be distinguished: intransitive and transitive consciousness.

##### intransitive consciousness (“being conscious”)

This term refers to consciousness as a property of something. For example, we can say that a boxer having been knocked out is becoming conscious without saying anything about the content of his consciousness. Here, there is no reference to a direct object of consciousness. In this sense, we prescribe consciousness to creatures or mental states. A creature is intransitively conscious when it is awake and can have some information from the environment. Whereas creature wakefulness is not problematic to define, consciousness as a property of mental states, which is called phenomenal consciousness or ‘what is it like’, doesn’t seem to be that easy endeavor.

##### transitive consciousness (“being conscious of”)

To be transitively conscious is to be conscious of something. Here, we refer to the object of consciousness. So, the boxer can be conscious of the fact that he is losing the match or conscious of pain. There are two ways in which a creature can be transitively conscious: thinking about something or by perceiving. For example, one can be conscious of a table by having a thought about it or perception of it. What matters is that the object is in some way represented in our minds. It is important to emphasize that transitive consciousness is not consciousness in a phenomenal sense; it is awareness which need not itself be a conscious state. In order to be conscious of something, it is sufficient to be in a non-conscious mental state which represents the object. Now, the next question is: what is sufficient for intransitive or phenomenal consciousness? According to Rosenthal, (1994, p. 359) *a mental state is (intransitively) conscious when a creature is (transitively) conscious of it*, e.g. watching a sunset is a conscious mental state when a creature is conscious of watching the sunset. If a creature isn’t conscious of that state, perception remains on an unconscious level, i.e. there is no “what is it like”.

Rosenthal (1994) claims that when we analyze the relation between transitive and intransitive consciousness, we shall find the solution to the puzzle of state consciousness. The relation between the two kinds of consciousness consists in the following:

- (i) Intransitive  $\rightarrow$  Transitive
- (ii) Not (Transitive  $\rightarrow$  Intransitive)

Being in a conscious state implies that there is transitive awareness. However, the converse doesn’t hold. Being aware of the rabbit, say by perceiving it, is not sufficient for being in a conscious state. Namely, one can have non-conscious perceptions.

(“O” stands for “object”)

Being transitively conscious = being aware of O

Being aware of O doesn’t imply being aware of being aware of O.

Therefore, not-(transitive  $\rightarrow$  intransitive)

Rosenthal explains consciousness in terms of 'consciousness of.' Since 'consciousness' is the explanandum and it figures in the *explanans*, the question is whether that makes the explanation circular. The answer is negative – we can understand 'consciousness of' without understanding what phenomenal consciousness is. The *explanans*, therefore, does not presuppose the *explanandum*.

Given that consciousness is a relational property of mental states, sometimes the right relation is obtained, sometimes not, and that yields the difference between non-conscious and conscious states. Now, the question is what kind of appropriate relation or awareness gives rise to consciousness. Consider the following example Rosenthal (1990, p. 27) gives: "...suppose that I myself notice my angry behavior and infer that I'm angry but again experience no conscious anger." One can be non-consciously angry and then – by observing one's own behavior – infer that one is angry. However, Rosenthal claims, we wouldn't assume that anger and thereby became conscious. These kinds of examples show there is a need to introduce an additional condition for intransitive consciousness. A mental state is conscious iff one is noninferentially and nonobservationally conscious of it. But how exactly are we conscious of our mental states?

There are two ways in which a creature can be transitively conscious of mental representations: by thinking or perceiving (1990, p. 32). Take for example an object of one's desire. According to the *higher order experience theory* (defended by Armstrong and Lycan) a creature can be conscious of a desire by perceiving it with some sort of "inner eye". Rosenthal (1990a, 1989) rejects this view because sensory perceiving involves a quality, and we are not aware of any higher order qualities while reflecting upon our experiences. He suggested something different: a mental state is conscious when one has a higher order thought (HOT) about it, which must be a noninferential and not observationally based thought. As it has been emphasized, transitive consciousness need not imply a state of being intransitively conscious. Thus, HOT need not itself be a conscious thought<sup>1</sup>.

HOT is a current thought that oneself is in a particular mental state. For example, when I look at a sunset there is something it is like to be in the state due to me having a thought about *myself watching the sunset*. But what kind of thought is HOT? How does it give rise to phenomenal consciousness? It seems that we can think about feeling depressed, but usually, that won't make us feel depressed. Rosenthal (2000, p. 233) agrees as Rey and Balog pointed out: if consciousness consists in having a HOT about a mental state, then consciousness would be "too cheap", since we can think about almost anything. Rosenthal (1989) recognizes that entertaining a HOT is easy, but thinking is not all about its content. The *how* we think matters here. Consider the following example:

It's easy to have thoughts about Santa Claus, but not so easy to form an occurrent, assertoric thought that Santa Claus is here in this room – really to believe it, as we might say. It's not the content that makes it hard, but the assertoric mental attitude. (Rosenthal, 2000, p. 233)

HOT is thus not any kind of thought, akin to imagining. HOT is an *assertoric* thought. What matters here is *how* the thought refers to a mental state. When we think about

---

<sup>1</sup>We are aware of a HOT only when we introspect our mental states already containing it. That is, only when we have a HOT about another HOT, do we have a conscious HOT about conscious experience, e.g. of watching a sunset.

watching the sunset, the perception is conscious only if we think about it *as if* we really have such a perception. However, it is possible to be conscious of a mental state by having different non-assertoric mental attitudes such as doubt, desire, hope etc. For example, we can be conscious of a desire by doubting whether the desire is moral. However, Rosenthal (1989) holds that doubt itself doesn't make us transitively conscious. Even if we doubt that desire is moral or not, we still think about it as an actual desire. Higher order doubt makes us transitively conscious insofar as it already assumes an assertoric thought that we have a desire. Therefore, a mental state is conscious only if the creature is conscious of it by having an *assertoric* HOT about it.

Now let me briefly address the issue of HOT theory's explanatory power. "What is it like" to be in a mental state is determined by the content of a higher order mental representation. Rosenthal clearly makes this point:

So what concepts we have makes a difference to what it's like for us to be in various sensory states. We can best explain why concepts make a difference to what it's like for us to be in sensory states by reference to the HOTs these concepts figure in. HOTs presumably result in it's being like something for one to be in various sensory states. (Rosenthal, 1993, p. 362)

Concepts which figure in HOTs are 'responsible' for the quality of our experience. The advantage of HOT theory is that it can explain how different conceptual resources reflect upon our experience by giving various examples. Consider the wine taster example by Rosenthal (1991, pp. 33-4). A person who has fine-grained concepts of wines will consequently have a richer phenomenal experience of a particular wine than an ordinary taster. The degree of the richness of our experience is thus determined by conceptual resources. However, we have experiences of color and sounds which outstrip our conceptual capacities. If our experience is determined by concepts, then how can HOT theory explain phenomenality in those cases? Rosenthal (1989, 2005) is aware of our limited conceptual capacities, but he believes that it doesn't pose a problem for the HOT theory. In cases where concepts can not "keep up" with experience, HOTs refer to lower order states demonstratively. For example, the ordinary wine taster might have a HOT with a content, 'I feel *this* taste'. However, at this point, difficult questions for the HOT theory begin to emerge.

## 5. Problems for the HOT theory

Rosenthal (2005, p. 237) allows HOTs to misrepresent their object: "The HOT model allows for a HOT to occur unaccompanied by its notional target, but that occurrence might result in nothing, so far as consciousness is concerned". A famous example of a HOT-misrepresentation comes from a common experience at the dentist. When he drills teeth, people report feeling pain although they are under anesthesia. Experiencing fear while being conscious of pain could be interpreted as follows: there is something that is like being in pain for them although the "real" feeling is fear. According to the HOT theory, fear is wrongly represented as pain. In normal situations, the first order mental state (e.g. pain) causes a HOT and thereby determines the content of the HOT. Pain is a conscious state when it is accompanied with a thought about oneself being in pain, i.e. the content of the HOT determines the "feel". "If I am conscious of myself [via a HOT] as being in a P state, it's phenomenologically as though I'm in such a state whether or not I am" (Rosenthal, 2004, p. 35; cf. p. 41). Furthermore, he says, "Since HOTs make one conscious of oneself as being

in a particular state, what it's like for one to be in a state is a function of how one's HOT represents that state" (Rosenthal, 2002, p. 658). But in the dentist's office there is no first order pain, just a HOT about pain. In the case of misrepresentation, fear perhaps causes a thought about pain because the first to second order connection is interrupted. How that happens remains somewhat mysterious.

Rosenthal claims that a mental state is conscious if a creature is conscious of it, and satisfies a further condition, i.e. is conscious of it noninferentially and nonobservationally:

... it may well be that much mental processing mediates between our conscious mental states and our being transitively conscious of those states, despite its seeming, from a subjective point of view, that no mediation occurs. (Rosenthal, 1994, p. 359)

Rosenthal's condition excludes the inference that we are aware of, but he doesn't exclude the inference which we are *not* aware of. So, it is possible that a non-conscious inference makes a state conscious. It seems that our conscious experience could then be some kind of non-conscious self-interpretation of our own states, and that is a dangerous territory. The social or cultural background, for example, may affect the way we interpret ourselves and what we feel. Two people with the same (type of a) lower order mental state from different cultures could have different conscious experiences due to the different interpretations of their respective lower order states. Hence, the role that the lower mental states play in explaining 'what is it like' would be subordinate to the active role of interpretation, which opens Pandora's box of appearance-reality contrasts within the mental itself. It seems that Rosenthal's account is not immune to this kind of problem.

Next, in Rosenthal's account, it is unclear (at best) what the status of qualia is in explaining the phenomenal character of sensory experience, versus the role of the HOT content accompanying such states. Let us begin with the concept of qualia.

Phenomenal consciousness is usually characterized in terms of 'what is it like' to be in a mental state. Most philosophers agree that for sensory states, emotions, moods and so on there is always 'what is it like' to be in those states, and Rosenthal aligns with such a view:

As already noted, the qualitative properties that belong to our bodily and perceptual sensations are the distinguishing properties of those sensations. They are not only the properties by means of which we distinguish introspectively among sensations of various types; they are the properties in virtue of which the sensations actually differ from one another. Conscious sensations exhibit characteristic qualitative character; for every type of conscious sensation, there is something it's like to have that sensation consciously. (Rosenthal, 1993, p. 362)

It is important to note that Rosenthal here claims that the qualitative aspect of mental states (qualia) is primarily a property of sensory states. Thus, pain differs from the taste of chocolate because of the lower order state that is qualitative in such-and-such way. But how is that consistent with the claim that HOTs are solely responsible for the 'feel'? "Since HOTs make one conscious of oneself as being in a particular state, what it's like for one to be in a state is a function of how one's HOT represents that state" (Rosenthal, 2002, p. 658). It is unclear how much qualia and/or HOT explains phenomenal experience.

Rosenthal seems somewhat enigmatic when it comes to the further point, namely, whether he would accept qualia as non-intentional states. According to Tim Crane's (2001) definition, sensations lack intentional content, but instead have qualitative properties. These

properties are called *qualia*. Crane (2001) contrasts sensations with propositional attitudes, which clearly exhibit intentionality. Rosenthal (1990a, in p. 46), on the other hand, endorses the view that all mental states are intentional: “To be a conscious state implies being a mental state, and mental states are all states of being conscious of something or other.” To be conscious of is to be in a mental state which is directed upon an object, and thus, it has a relational structure<sup>2</sup>. Generally, directedness upon an object is taken to be a condition for a mental state being intentional/ representational<sup>3</sup>. According to this interpretation, supported by Rosenthal’s claim that “mental states are all states of being conscious of something or other”, it follows that all mental states have relational structure. But if *all* mental states are intentional then how could sensory states be non-intentional exhibiting qualia? The question is how it is possible for Rosenthal to sustain both claims:

- (i) qualia are qualitative properties of sensory states
- (ii) all mental states (including sensory states) are intentional

Perhaps it’s too early to claim inconsistency in the HOT theory. When one has a toothache, one is aware of one’s tooth. Since pain(-state) is directed upon the tooth, pain is an intentional state. The content of pain explains the feeling of pain. But we can abstract from the totality of the toothache-state and focus upon its distinct *sharpness*. We thus obtain a “sharpness qualia”. Its instantiation then partly explains the phenomenal character of a toothache. The main idea here is that all mental states are intentional, and in addition, *some* sensory states have non-intentional properties.

Now, the question is: could Rosenthal accept this compromise? I think not, because he explicitly says that every sensory state is qualitative. On the other hand, “weak intentionalism” postulates qualia only for some states, those for which intentional content cannot explain the totality of the phenomenal experience (e.g. directedness of the pain cannot itself explain its feeling of sharpness). Alternatively, qualia might be an intrinsic property of sensory states. If we further suppose that Rosenthal could take qualia to be a higher level non- intentional property, we will get the following result: qualia are non-intentional intrinsic properties of all sensory states. Rosenthal might adopt the view according to which all sensory states are representational, and they all have non-representational properties which ‘latch onto’ sensory states. Is this combination plausible? I think that such a ‘latching’ property of a sensory state would be redundant in explaining ‘what it is like’ to be in such states. That brings us to the problem already mentioned – the content of a HOT is necessary and sufficient for subjective, phenomenal experience. Rosenthal says that the HOT targets qualia: “...when we’re conscious of being in those states, those qualities seem somehow to light up.” (Rosenthal, 1993, p. 362). Qualia then could indirectly contribute to ‘what is it like’ by being an object of a HOT. But why not instead claim that a HOT targets just lower

---

<sup>2</sup>Tim Crane also thinks that one can talk about intentionality in terms of transitive consciousness: “Transitive consciousness seems at first sight to be a form of intentionality: when one is conscious of the bus approaching, the approaching bus is the intentional object of your consciousness. It is natural to conclude that since claims of the form ‘X is conscious of Y’ are ways of talking about the intentionality of conscious states ...” (Crane, 2001, p. 78)

<sup>3</sup>“Since, according to intentionalism, all mental states have directedness and aspectual shape, then all mental states have an intentional object and intentional content.” (Crane, 2001, p. 28)

order purely representational mental states? There are two options available in explaining the phenomenology of sensory experiences:

- (1) Either a HOT is targeting lower order representational states (LORS), or
- (2) HOT is targeting qualia-states

If it is possible to maintain an equally plausible, but ontologically less costly explanation of our experience by adopting option 1), then there is no need to adopt option 2) which introduces an additional kind of entity. Would Rosenthal's theory lose some explanatory strength without the concept of qualia? I don't think so, if the content of a HOT determines and explains 'what it is like' to be in a mental state. Rosenthal would have to show that there is an explanatory advantage of option 2) over 1). To further complicate the dilemma, Rosenthal says that having a HOT about a non-existent mental state or having a HOT about an actual mental state wouldn't make a difference in the subjective experience:

HOTs can presumably occur in the absence of the states they purport to be about. What would that be like subjectively? Since HOT makes the difference whether there is or isn't something it's like to be in a mental state, it may be that having a HOT without a relevant state is, something at least, subjectively indistinguishable from having both. (Rosenthal, 1997, p. 744)

If, in some cases, there are no lower mental states then there are no qualia either, and yet there is a phenomenal experience. What explains the phenomenal experience in isolated cases arguably explains it in all cases. Consequently, there would not be a subjective difference between the two options mentioned: 1) either HOT is targeting LORS or 2) HOT is targeting a quale state. In other words, if there is no *explanandum*, there is no need for quale *explanans*.

Finally, there is another problem, which I will present in the form of a dilemma for the HOT theory. The first horn of the dilemma can be put in the following argument:

- (1) By definition, HOT is thought.
- (2) "...having a thought about something suffices for one to be transitively conscious of that thing..." (Rosenthal, 1993, p. 360)
- (3) "...since being conscious of something is factive, we cannot be conscious of something that is not there." (Rosenthal, 1997, p. 744)

Then (4): For every HOT, there is a lower order state which HOT is about. (from 1, 2, 3)

Therefore, (5) HOT cannot be about non-existent lower order mental states. (from 4)

The second horn of the dilemma consists in the following claims:

1\*) One can think about things which do not exist. One can think about unicorns, Napoleon, and so on. That means thinking about an object doesn't entail that the object exists. And Rosenthal (1997, p. 741.) is explicit about this point: "...since perceiving is factive and thinking is not".

2\*) Therefore, HOT can be about non-existent lower order states which Rosenthal (1997, p. 744) clearly allows by saying "HOTs can presumably occur in the absence of the states they purport to be about..."

Ultimately, the question is: which premise can Rosenthal deny? Each seems quite important, and most seem independently plausible. Together, they land him in an apparent contradiction.

## 6. Conclusion

The core of Rosenthal's theory is that a mental state is conscious when one is conscious of it. The condition he adds is that transitive consciousness has to be non-inferential. I pointed out that Rosenthal's account of inference is taken in the sense of "inference we are aware of", which leaves an open route for the claim that consciousness is a product of un conscious self-interpretation. Such a view carries well-known risks. Secondly, I argued that if Rosenthal were to accept a quale as a non-intentional sensory state, he would have to give an additional argument to reconcile it with the claim that all mental states are intentional. Further, if he were to accept the 'latching' qualia thesis, these properties would be redundant in explaining 'what it is like' to have sensory experience, since a HOT which targets lower order intentional states can do that explanatory work. Thirdly, I presented a dilemma concerning the question of whether HOTs can be about non-existent lower order states. The following points in Rosenthal's account seem to be puzzling. First, he says that all thoughts are making us transitively conscious, and that being 'conscious of' is factive. But then, it follows that having a HOT implies existence of its object, i.e. a lower order mental state. On the other hand, thinking is not factive. Thinking that p doesn't entail that p, (and thinking about x does not entail that there is x, since one can think about non-existent things). Consequently, HOTs could be about non-existent lower order states, and that's clearly endorsed by Rosenthal. All these claims are plausible, but more distinctions must be made to avoid an apparent contradiction. Bottom line: Rosenthal's HOT theory is defended in a plausible way. However, its complexity carries unresolved problems that keep emerging.

## References

- Block, N. (1995). On a confusion about a function of consciousness. *Behavioral and Brain sciences* 18, 227-287.
- Carruthers, P. (2000). *Phenomenal consciousness*. Cambridge University Press.
- Crane, T. (2001). *The Elements of Mind*. Oxford, Oxford University Press.
- Jackson, F. (1982). Epiphenomenal Qualia. *Philosophical Quarterly* 32: 127-136.
- Kripke, Saul A. (1972/2001). *Naming and Necessity*. Cambridge, Mass.: Harvard University Press.
- Rosenthal, D. M. (1986). Two concepts of consciousness. *Philosophical Studies* 49 (May): 329-59.
- Rosenthal, D.M. Thinking that one thinks, Report No. 11/1989 of the Research Group on 'Mind and Brain', Perspectives on Theoretical Psychology and the Philosophy of Mind, University of Bielefeld, Germany.
- Rosenthal, D.M. A theory of consciousness, Report No. 40/1990a. of the Research Group on 'Mind and Brain', Perspectives on Theoretical Psychology and the Philosophy of Mind, University of Bielefeld, Germany.
- Rosenthal, D.M. The independence of consciousness and sensory quality, Report No. 5./1990b. of the Research Group on 'Mind and Brain' at the center of Interdisciplinary Research, University of Bielefeld, Germany.
- Rosenthal, D.M. Why are verbally expressed thoughts conscious?, Report No. 32/1990c. of the Research Group on 'Mind and Brain', Perspectives on Theoretical Psychology and the Philosophy of Mind, University of Bielefeld, Germany.

- Rosenthal, D.M. (1991). The Independence of Consciousness and Sensory Quality. *Philosophical Issues* 1: 15–36.
- Rosenthal, D. M. (1993). State consciousness and transitive consciousness. *Consciousness and Cognition* 2 (3):355-63.
- Rosenthal, D. M. (1997). A theory of consciousness. In Ned Block, Owen Flanagan & Guven Guzeldere (eds.), *The Nature of Consciousness: Philosophical Debates*. MIT Press.
- Rosenthal, D. M. (2000). Metacognition and higher-order thoughts. *Consciousness and Cognition* 9 (2):231-242.
- Rosenthal, D.M. (2001). Consciousness and sensation: Philosophical aspects. In Neil J. Smelser & Paul B. Baltes (eds.), *International Encyclopedia of the Social and Behavioral Sciences*. Elsevier.
- Rosenthal, D.M. (2002). How Many Kinds of Consciousness? *Consciousness and Cognition* 11: 653-65.
- Rosenthal, D.M. (2004). Varieties of Higher-Order Theory. In *Higher-Order Theories of Consciousness*, ed. Rocco J. Gennaro. Amsterdam: John Benjamins Publishing.
- Rosenthal, D.M. (2009). Concepts and definitions of consciousness. In P W. Banks (ed.), *Encyclopedia of Consciousness: A - L*. Elsevier.
- Williamson, T. (2000). *Knowledge and its Limits*. Oxford: Oxford University Press.

## REVISITING SEARLE'S REFUTATION OF SENSE DATUM THEORY OF PERCEPTION

S. Sreenish\*

**Abstract:** John R. Searle criticized the sense-datum theory of perception, particularly A. J. Ayer's version of it (Searle, 2015). According to Searle, (i) Ayer subscribes to the view that physical objects are not directly perceived but mind-dependent sense-data (Searle, 2015). Following this, Searle says, (ii) Ayer committed the conclusion of the "Bad Argument": "any argument that attempts to treat the perceptual experience as an actual or possible object of experience" (Searle, 2015). This paper revisits (i) and (ii). It will show that while (i) and (ii) are compatible with the sense-datum theory in general, it has some reservations when it applies to Ayer's version of it. Hence, Searle's refutation of Ayer warrants reconsideration.

**Keywords:** Searle, Ayer, Perception, Sense Data, Object

### 1. Introduction

Searle criticized the sense-datum theory of perception in general, Ayer's in particular (Searle, 2015).<sup>1</sup> According to Searle, Ayer repeated the same mistake committed by Locke and Berkeley while discussing perception. For Locke and Berkeley, it is the ideas or impressions that are directly perceived. Searle says (i) Ayer subscribes to the view that physical objects are not directly perceived but sense-data (Searle 2015: 85-7). Following this, he says, (ii) Ayer committed the conclusion of the "Bad Argument": "any argument that attempts to treat the perceptual experience as an actual or possible object of experience" (Searle 2015: 29).<sup>2</sup>

The expression 'Bad Argument' is a technical term that Searle often used in his book *Seeing Things as They Are: A Theory of Perception*. He says Bad argument is a fallacy that arises when one commits to the view that "we never directly perceive objects and states of affairs in the world, but directly perceive only our subjective experiences" (Searle 2015: 11). This subjective experience is sometimes called ideas, or impressions, or sense data. Sense datum theory states that we are only directly aware of sense data. In this sense, they committed to the Bad Argument.

In contrast to sense-datum theory, Searle argues that physical objects are directly perceived. He maintains direct realism, whereas sense-datum theory maintains indirect realism. According to the former, perception is a direct presentation of mind-independent objects, whereas the latter views we perceive directly mind-dependent items (sense data). This paper will revisit (i) and (ii). We will show that although (i) and (ii) make sense regarding

<sup>1</sup>For a staunch criticism of Ayer's sense-datum theory, see Austin (1962).

<sup>2</sup>Searle identifies perceptual experience and sense-datum as one and the same.

sense-datum theory in general, they do not fit with Ayer. In addition, we will show some other inconsistencies in Ayer's sense-datum theory.

This paper contains four sections. In the first section, we will discuss Searle's refutation of sense-datum theory in connection with Ayer. In the second section, we will examine the plausibility of (i) and (ii) concerning sense-datum theory, in general. In the third section, we will examine the plausibility of (i) and (ii) concerning Ayer, in particular. In the fourth section, we will show some inconsistencies in Ayer's account.

## 2. Searle's Criticism of Ayer

In this section, I will discuss Searle's refutation against Ayer's sense-datum theory of perception. The former aims to establish the case that the latter subscribes to (i) and thereby commits to (ii).

Searle cites the passage below from Ayer's famous work *The Foundations of Empirical Knowledge* as an instance to convey both (i) and (ii):

Nevertheless, even in the case where what we see is not the real quality of a material thing, it is supposed that we are still seeing something, and that it is convenient to give this a name. And it is for this purpose that philosophers have recourse to the term 'sense-datum.' By using it they are able to give what seems to them a satisfactory answer to the question. What is the object of which we are directly aware, in perception, if it is not part of any material thing? Thus, when a man sees a mirage in the desert, he is not thereby perceiving any material thing, for the oasis which he thinks he is perceiving does not exist. At the same time, it is argued, his experience is not an experience of nothing, it has a definite content. Accordingly, it is said that he is experiencing sense-data, which are similar in character to what he would be experiencing if he were seeing a real oasis, but are delusive in the sense that the material thing which they appear to present is not actually there. (Ayer 1940: 4)

For Ayer, the 'experience is *not* an experience of nothing' because it has content. Searle says Ayer's mistake lies in the fact that the latter thinks if an experience has definite content, then it must have a definite object. Accordingly, Ayer views that since no physical object is involved, it is legitimate to call the mind-dependent sense datum (oasis) the object of experience.

Searle criticizes Ayer, saying that viewing the content of perceptual experience as necessarily implying the object is a mistake. Searle maintains that the experience has definite content does not imply that it has a definite object (Searle 2015: 86-7). Illusory experiences and hallucinatory experiences are cases where the experiences have definite content but no relevant object. He says, "...it is the Bad Argument to suppose there must be a sense-datum as an object" (Searle 2015: 182).

According to Searle, the content of perceptual experience is the way things are organized in our perceptual experience. Content determines the way the object is given to our experience. Thus, the content of every perceptual experience sets certain conditions: (1) the presence and features of the apparent object (let us say oasis) that appears in the perception must be the cause of the perceptual experience, (2) the apparent object must possess the properties it appears to have, (3) the experience must be experienced as caused by the apparent object. A perceptual experience is veridical if all the above conditions (1, 2, 3) are fulfilled. Only a veridical perception can have the relevant object. None of these conditions

are fulfilled in the above-mentioned mirage case. Once the perceptual experience has no object to satisfy the conditions set by the content, then the experience in question is an experience of nothing in the intentional sense of perception (Searle 2015: 87). It is a non-veridical case. Now, Searle raises a legitimate question: How does a non-veridical experience have an object?

Searle thinks that the problem mentioned above manifested in Ayer can be analyzed in the light of the Intentionality Theory of Perception (ITP) and offers a solution. ITP makes a case for understanding immediate (direct) perception in two ways by making a distinction between the constitutive sense of perception and intentional sense of perception (Searle 2015: 83). In the constitutive case, what we immediately perceive is ontologically subjective experience (Searle 2015: 81). In the intentional sense of perception, the object of perception is the ontologically objective object (Searle 2015: 69). In both cases, we have categorically different objects of experience. Given the distinction, Searle interprets Ayer in the following manner: the object of the constitutive sense of perception refers to an ‘oasis,’ and the object of the intentional sense of perception refers to ‘nothing’ since no mind-independent object is presented. Searle says one cannot be reducible or identified with the other.

Searle argues that Ayer confused the constitutive sense of perception with the intentional sense of perception. This confusion led him to conclude that a sense datum is directly perceived. Searle complains that Ayer fails to acknowledge the crucial distinction between these two senses of perception and mistakenly reduces the intentional sense of perception into the constitutive sense of perception. As a result, he commits to the view that only subjective experiences (sense datum) are immediately perceived. Hence, Searle claims that Ayer committed the fallacy of the ‘Bad argument’ (Searle 2015: Chapter 1). These are the grounds for Searle to uphold (i) and (ii).

### 3. Sense Datum Theory

This section aims to examine whether (i) and (ii) fit with sense-datum theory in general. The sense-datum theory is one of the central discussions in the earlier twentieth-century analytic philosophy. In fact, sense-datum theory has its roots in representationalism (Locke) and phenomenalism (Berkeley). In connection with representationalism, the sense-datum theory of perception defends indirect realism. The sense-datum theory is the best-known theory that defends indirect realism (Chediak 2016: 520). Similarly, in favor of phenomenalism, they often argued that “material objects are some kind of ‘construction’ out of sense-data” (Crane 2001: 132). However, sense datum as a thesis began with Moore (1903) and Russell (1912, 1914), then moved to C. D. Broad (1923) and then H. H Price (1932). The term ‘sense data’ was first introduced in philosophy by G. E. Moore, and it was meant to stand for whatever it is ‘given or presented to the senses’ in acts of perception (Crane 2001:132). According to Moore, sense-datum is the immediate object of perception. Moreover, he opines that sense-datum is, to some extent, identifiable with physical objects (Ayer 1945: 293-5). Hence, by perceiving sense datum (directly), we perceive physical objects (directly). However, Moore’s account does not get much appreciation from the latter proponents of sense-datum theory. Although there were continuous disputes over the nature of sense data among the adherents of sense-datum theory, they all agree that the subject is directly and immediately aware of

sense-datum in perception. Perception of sense datum involves no other perception. In other words, sense data is the immediate object of perception.

### 3.1. The Road to Sense-datum Theory

As we said earlier, sense-datum theory is a variant of traditional representationalism and phenomenalism. However, the theory builds on the solid material supplied by the argument from illusion and the argument from hallucination. It is generally argued that both these assumptions put us on the road to sense-data theory.<sup>3</sup>

#### 3.1.1. Argument from Illusion

The argument from illusion suggests that 'when one is subject to an illusion, then one is experientially aware of something's having a quality, F, which the real object perceived does not have (Crane 2006: 132). Brewer commented:

[T]he argument from illusion is supposed to establish that, at least in cases of illusion, and therefore also in all cases of experience subjectively indistinguishable from some possible illusion, including those of genuine perception, such objects must be mind-dependent ideas, sense-data, or whatever, rather than mind-independent things themselves. (Brewer 2006, 'Perception and Contents,': 168)

Argument from illusion shows that physical objects appear other than they really are. The fact that physical objects appear to the subject other than they really are leads to the view that something else is directly perceived. It follows that, in illusory experience, one is not directly perceiving the physical object, but instead encounters a mind-dependent entity. The argument from illusion extends the same analysis to veridical perception, i.e., in perception, too, one is directly aware of mind-dependent items, not physical objects. In short, the argument from illusion suggests that what we are directly aware of is not physical objects.<sup>4</sup>

#### 3.1.2. Argument from Hallucination

The Argument from hallucination is another version of the argument from illusion. Both have the same template. In hallucinatory experiences (total), one does not perceive physical objects, although the subject experiences something. Since hallucinatory experience lacks a physical object, the object involved in the hallucinatory experience must be a mind-dependent (experience-dependent) item.<sup>5</sup> The argument from hallucination extends the same analysis to perception by arguing that the structure of perception is the same as that of a hallucination (Schellenberg 2010: 28).

For instance, consider Subject S's perception of object O as F at t1 and the hallucinatory experience of O as F at t2 are phenomenally indistinguishable. It is argued that 'phenomenally indistinguishable perceptual states are mental states of the same type' (Crane

<sup>3</sup>For more on argument from illusion and argument from hallucination, see H. Robinson (1994) and A. D. Smith (2002).

<sup>4</sup>For a recent defense of the invalidity of argument from illusion, see C. French & L. Walters (2018).

<sup>5</sup>See Moran (2019) for a discussion on the difference between experience-dependent and mind-dependent.

2001: 133).<sup>6</sup> The presence and absence of the actual object perceived have no effect on the phenomenal experiences of the subject. If the hallucinatory experience is phenomenally the same as the indistinguishable perceptual experience (veridical) and if two phenomenally indistinguishable states belong to the same mental type or kind, then the subject is aware of a mind-dependent item, not a mind-independent object. The crux of the argument from hallucination is that ‘if one’s experience is of the same type when he or she really seeing an object is same as hallucinating, and experiences of the same type have the same immediate objects, then the immediate object of the experience in the case of the genuine perception cannot be a physical object either’ (Crane 2001: 133). It follows that we directly perceive mind-dependent things.

The central arguments demonstrated by the argument from illusion and the argument from hallucination can sketched in the following manner:

Step 1: In both the veridical case and in the non-veridical case (illusion and hallucination), there is a common element—a qualitative subjective experience going on in the visual system.

Step 2: Since the common element is qualitatively identical in both cases, whatever analysis we give of one, we must give of the other case.

Step 3: In both the veridical case and the non-veridical case, the subject is aware of something or conscious of something or sees something.

Step 4: But in the non-veridical case, the perceived item cannot be a material object; therefore, it must be a subjective mind-dependent entity.

Step 5: But by step 2, we have to give the same analysis for both cases (veridical and non-veridical). Therefore, in the veridical case, as in the non-veridical case, the subject sees only the mind-dependent entity (Searle 2015: 30).

So, the argument from illusion and the argument from hallucination together establish the case that in perception, we are directly aware of mind-dependent items, not mind-independent objects.<sup>7</sup> Taking advantage of both theses, the proponents of sense-datum theory conveniently named the mind-dependent item as ‘sense datum.’ They viewed that we directly perceive (veridical and non-veridical case) sense data and perceive physical objects (if at all perceiving) indirectly.<sup>8</sup>

So far, we discussed sense-datum theory in general. We have seen that in the light of argument from illusion and argument from hallucination, sense-datum theory subscribes (i). According to Searle, since sense-datum theory subscribed to (i), it is fair to say they

---

<sup>6</sup>Some “deny that phenomenally indistinguishable mental states are always mental states of the same type” (Crane 2001: 134).

<sup>7</sup>For more on the argument from illusion and argument from hallucination, see A. D. Smith (2002); H. Robinson (1994).

<sup>8</sup>Although sense datum theorists, in general, deny that physical objects are not the immediate object of perception, the adherents are “not suggesting that one never perceives physical objects...” (Stoljar 2009: 4). According to them, sense data is directly given to perceptual experience does not mean that physical objects are imperceivable. Sense data is the object of direct perception (immediate object of perception), whereas physical objects are perceived by perceiving sense-data. Thus, physical objects are objects of indirect perception (nonimmediate objects of perception).

committed to (ii). In the next section, we will discuss Ayer's account of sense datum and examine whether Ayer subscribes to (i) the way ordinary sense-datum theorists do.

## 4. Sense-datum Theory-Ayer

A. J. Ayer (1940; 1945; 1956) is one of the significant proponents of sense-datum theory in the later phase. His use of the term 'sense data' differs from his predecessors. He maintains a significant distance away from the ordinary sense-datum theorists (Russell, Broad, Price). Because the ordinary sense-datum theory views sense data as the object of direct perception (immediate object of perception), whereas physical objects are indirectly perceived (nonimmediate object of perception).<sup>9</sup> Ayer rejected Moore's view and said, along with Price, "no sense-datum can possibly be identical with any physical entity" (Ayer 1945: 300-1). However, on the one hand, he agrees with Moore by saying that physical objects are directly perceivable (Ayer 1940: 60-1), and on the other hand, he totally agrees with others that sense-data is the object of which we are directly aware in perception (Ayer 1940: 4-5).<sup>10</sup>

### 4.1. A Shift to Language Analysis

Ayer acknowledges that the conclusion of the argument from illusion and the argument from hallucination is illuminating (Ayer 1940: 26-7; 1945; 1956: 94).<sup>11</sup> He is of the view that both supply potential materials to view that perception involves something other than physical objects, namely sense data. However, he subscribed to the conclusion that material things are not directly perceived but only from the linguistic perspective, not from the ordinary perspective. He says:

The main proposition which the argument from illusion, as I have interpreted it, was supposed to establish was that what we see, or otherwise directly experience, are never material things, but only sense-data. This conclusion I accept, but not as a proposition in the ordinary sense (Ayer 1940: 26-7).

Ayer invokes a linguistic analysis in the propositions of perceptual experience. He says there are two ways perception of objects can be categorized. In the ordinary sense, when we say S perceives X, S presupposes that X exists. But the argument from illusion and the argument from hallucination conveyed that there are cases where X need not exist, but still, S perceives X. Here arises the problem. Ayer offers a linguistic solution to this problem. He discusses two different languages in the analysis of the propositions of perceptual experience. One is physical object language, and the other is sense datum language.

#### 4.1.1. Physical Object Language

Physical object language is a form of language used to talk about things that objectively exist. Ayer says that in terms of physical object language, one can legitimately say that physical objects are often directly perceived. He says:

<sup>9</sup>Moore is exceptional. According to Moore, sense data and physical objects are immediate objects of perception.

<sup>10</sup>See Griffiths (1960) for more on Ayer's account of perception.

<sup>11</sup>However, his focus rests on argument from illusion. Hatfield comments that "Ayer (1940: Ch. 1) found that the Argument from Illusion supports choosing sense-data language" (Hatfield 2021).

It is indeed legitimate to use the phrase “direct observation” in such a way that things like houses and trees and stones can properly be said to be directly observable, and this usage can perfectly well be made to cover the case of delusive as well as veridical perceptions, provided that it is allowed that what is “directly observed” may not in fact exist, and that it may not really have the properties that it appears to have. (Ayer 1940: 234)

According to physical object language, the proposition that S perceives object O as F entails the existence of O and the existence of F-property in O. However, the argument from illusion and hallucination refutes the propositions. Because there can be cases where O need not exist (hallucination), and there can be cases where O lacks the F-property it appears to have (illusion). Ayer says this is a genuine challenge, and the solution is developing a new language to analyze the propositions of perceptual experience.

#### 4.1.2. Sense Datum Language

Ayer proposed sense datum language to avoid the shortcomings of the physical object language. He says:

It is useful for us to have a terminology that enables us to refer to the contents of our experiences independently of the material things that they are taken to present and this the sense- datum language provides. (Ayer 1940: 26)

The function of physical object language is to talk about the mind-independent object involved in perception. Its concern is restricted to the existent, not to appearance. Ayer says appearance is the way things seem to the subject. They are mind-dependent. He considered appearances as nothing but ‘sense data.’ They directly presented the subject’s experience. There can be appearances without having relevant physical objects (like the mirage case noted in the first section). An object may appear to the subject in a particular way, but it does not entail that the object is necessarily like the way it appears. In the experience of a material object, the object may not have the properties it appears to have, or the object in question may not even exist. Since appearance occurs without depending on physical objects, physical object language is insufficient to explain appearance. Because physical object language relies not on appearance but on physical object. Considering this, Ayer proposes sense datum language to explain how things *seem* or appear to the subject in a perceptual situation. He says that in terms of sense-datum language, one can legitimately say sense-data is directly perceived. He says:

But I have shown that it is also legitimate to use the phrase “direct observation” in such a way that it is only what is designated by the term “sense-datum”, or some equivalent term, that can be said to be directly observable, and that it is this usage that, for my present purpose, is to be preferred. (Ayer 1940: 234)

So, it is clear that Ayer’s introduction of sense datum does not deny that physical objects are directly perceived. Instead, it suggests that there is a sense (linguistic) in which we can legitimately hold that sense datum is directly perceived. The proposition that sense datum is directly perceived does not have any commitment to say that physical objects exist. Because “[t]he proposition that a sense-datum is veridically sensed does not entail that any material thing is veridically perceived” (Ayer 1940: 235). Thus, sense datum language, unlike physical object language, is consistent with the truth of the argument from illusion, and the argument from hallucination is not denied. Moreover, Ayer says that the

physical object language allows us to view that whatever is perceived may or may not exist or whatever appears may or may not really be there in the objects. Thus, physical object language cannot assure us whether or not we perceive things as they really are. It cannot assure whether things we perceive really exist or not. On the other hand, in the case of sense-datum language, such confusion does not arise.

According to sense datum language, the object we are aware of, sense data, “really have the properties that it appears to have” (Ayer 1940: 235). There is no sense in saying the properties appear to sense data does not exist. There is no way one can sense a sense datum without actually sensing it. Considering this, Ayer prefers sense datum language over physical object language for the analysis of the propositions of perceptual experience.<sup>12</sup>

Ayer held that sense data is conceptually prior to physical objects. He says:

...sense-datum language is logically prior *to physical object language*, and this is shown by the fact that, while referring to sense-data is not necessarily a way of referring to physical objects, referring to physical objects is necessarily a way of referring to sense-data. (Ayer 1945: 312)<sup>13</sup>

It suggests that sense-data is prior to physical objects as far as the linguistic analysis of the propositions of perceptual experience is concerned. Given any perceptual experience, let us say it necessarily involves sense data, whereas physical objects are contingent because the experience could be either an illusion or a hallucination (Ayer 1940: 230).

So far, we have discussed Ayer's distinction between physical object language and sense datum language in analyzing the propositions of perceptual experience. According to the former, it is convenient to say physical objects are directly perceived, whereas according to the latter, sense datum is directly perceived. The distinction between physical object language and sense datum language ultimately aims to legitimize the consistency of the propositions that (a) ‘physical objects are perceived directly’ and (b) ‘things that appear to us may not exist or may appear other than what they are.’ However, one may be confused about what exactly the distinction implies without unpacking the notion of sense datum itself.

#### 4.1.3. Sense Data

Ayer's use of the expression sense datum has a considerable distinction from his predecessors. For Ayer, unlike ordinary sense-datum theorists, “the existence of sense data is not factual but linguistic convention” (G. Hatfield 2013: 966). He employed sense-datum as a linguistic tool to examine the propositions of perceptual experience (Ayer 1945: 311). In his project, sense-data becomes a linguistic entity rather than a metaphysical unit. Thus, his notion of ‘sense-datum’ is ontologically neutral.<sup>14</sup> It does not entail existence. Hence, sense datum involves no metaphysical commitments.

<sup>12</sup>Ayer says “‘sense-datum language’ is, in a certain sense, more comprehensive than the ‘physical-object language’” (Ayer 1945: 312).

<sup>13</sup>Italics mine.

<sup>14</sup>It will be fruitful to revisit L. S. Carrier's *linguistic* analysis of perception to make sense of Ayer's ontological neutrality towards sense data. Carrier employed linguistic analysis to statements of immediate and mediate perception (Carrier 1969). He argues that explaining perception from the linguistic perspective secures neutral commitment to the existence of the ‘object’ of perception.

As long as sense datum is ontologically neutral, the propositions of sense datum language do not posit any ontological commitment to the existence (real) of what is directly observed (sense data). For instance, the proposition that ‘S directly perceives sense datum R’ does not entail the question of the existence of R. Because R is not a real entity but a linguistic entity. He thinks conceiving sense data as a linguistic entity secures it from the traditional ontological problem associated with sense data.

The difference between Ayer and the other proponents of sense-datum theory is that the former uses sense-datum as a tool to explain the proposition of perceptual experience, whereas the latter uses it to explain the nature of perceptual experience itself. The former has no metaphysical commitment, whereas the latter does. When Ayer says that sense datum is the only immediate object of perception, he is making a considerable departure from physical object language or ordinary language to sense datum language (Ayer 1956: 94). Such distinction is not maintained (at least not explicit to us) in the case of ordinary sense-datum theory.

In this section, we have discussed sense-datum theory in connection with Ayer. However, one can see some fatal inconsistencies involved in his approach. In the next section, we will discuss some of them.

## 5. Objections to Ayer

There is no exaggeration in saying that Ayer’s version is a new wave to the sense-datum theory of perception. However, his account suffers from multiple issues.<sup>15</sup>

### 5.1. Is the Sense Datum Really Unreal?

Ayer says that sense data are unreal. They are unreal in the sense that they are metaphysically neutral. However, this is not conveyed. Things fall under the category of either metaphysically specific or metaphysically neutral (Crane 2001). Something that falls under the former category must be substantive in nature. They are substantive because they must have substantive properties. For example, physical objects have substantive properties. They exist in space and time and must fall under the law of physics. Hence, they are metaphysically specific. Similarly, abstract entities like numbers also have substantive nature. They must not exist in space and time, and so on. Hence, they are also metaphysically specific. On the other hand, metaphysically neutral entities lack substantive properties.<sup>16</sup> Such entities have neither mental nor physical existence. Considering this, it is difficult to conceive how sense datum fits into the category of metaphysically neutral since the sense datum shares the property of being colored or shaped or being sensed. As we said, metaphysically neutral things do not share any properties, whereas according to Ayer, a sense datum has the properties it appears to have. Thus, whether sense data is metaphysically neutral or metaphysically specific is not conveyed. As long as it is not conveyed, the question ‘Is sense data really unreal?’ remains sound.

---

<sup>15</sup>Some of the prominent critics of sense datum theory in recent times are Martin 2000, 2002; Crane 2013.

<sup>16</sup>In Crane’s (2001) view, Intentional objects are metaphysically neutral entities. They lack substantive properties.

## 5.2. Does Sense Data have Properties it Appears to have?

Ayer says a sense datum has the property it appears to have (Ayer 1945: 303-8). They are often sensed as colored or shaped. He says:

When I look at the coin obliquely, the sense-datum that I am sensing is, it is said, in itself elliptical, at the same time I know that if I were looking at the coin vertically, my sense-datum would be circular, and I take this as evidence that the coin really is circular in shape, accordingly, I subconsciously correct my elliptical sense-datum so as to bring it into accord with the "real" shape of the coin. (Ayer 1940: 119)

It says sense datum is sensed, and the sensed sense datum is shaped. However, this does not fit with Ayer's overall project. Something which is shaped and sensed must exist. It must exist in space. However, it is clear that sense data do not exist the way physical objects exist (see Huemer 2001: 149–68).<sup>17</sup> So, it is fair to ask where exactly they exist. If at all a sense datum exists, it must have mental existence. But Ayer denied any mental existence to sense data. Another option left to Ayer is to identify sense data with brain states (Russell 1927: 383; O'Shaughnessy 2003: 186). However, this option does not seem plausible, because the brain state does not have the colored or shaped property that the sense datum perceptually appears to have. If the sense datum is shaped or colored, then the sense datum is not a brain state. Moreover, Ayer altogether rejects the view that sense datum is a brain state (Ayer 1940: 61–65, 76–78).<sup>18</sup> Hence, it is difficult to conceive how the sense data has the properties it appears to have provided sense data have no existence.<sup>19</sup>

## 5.3. Is the Sense Datum Really a Linguistic Entity?

For Ayer, sense datum is not a metaphysical entity but a linguistic unit. However, this view is not consistent with his position that a sense-datum has the property it appears to have. If a sense datum is shaped or colored and sensed, it cannot be called a mere linguistic entity. Linguistic entities can only be grasped, not sensed. A Kantian would say linguistic entities belong to the category of understanding, not to sensibility. Hence, it is fair to say that things that are sensed are not linguistic but metaphysical.

In this section, we have pointed out a few inconsistencies in Ayer's sense-datum theory. If the above-noticed inconsistencies are worth considering, then it is fair to demand a reevaluation of Ayer's account.

## 6. Conclusion

Searle argued that sense-datum theory subscribes to (i): physical objects are not directly perceived but only sense data in perception. In section 2, we examined this charge concerning sense-datum theory in general and found that Searle's criticism fits with the sense-datum theory. Followed by the argument from illusion and the argument from hallucination, they

---

<sup>17</sup>For instance Ayer says "sense-data are not states of mind, we must not then look round for some other container for them, such as the brain or "the psycho-cerebral compound." For it is already a mistake to suppose that they can be phases of any substance, or anywhere in physical space, at all" (Ayer 1940: 78).

<sup>18</sup>See footnote no 29. See also Hatfield (2021).

<sup>19</sup>M. Huemer (2004) comments that O'Shaughnessy (2003: 186) seems to be rejecting the traditional doctrine that sense data literally have the features that perceptually appear to us.

conclude that the object of direct perception is mind-dependent item, not mind-independent items. They call this mind-dependent item sense data. According to the ordinary sense-datum theory, a sense datum is a real entity involved in perceptual experience. They are metaphysical units with relevant properties. They think that sense data is the medium by which we perceive the physical world. Thus, physical objects are indirectly perceived.

On the other hand, in the third section, we have noticed a significant departure from the ordinary sense-datum theory. Ayer, unlike the ordinary sense-datum theory, uses sense datum as a linguistic entity, not as a metaphysical entity. His analysis of perception is considerably linguistic, not empirical. His aim is to analyze the propositions of perceptual experience. In this linguistic analysis, he finds that both the argument from illusion and the argument from hallucination posit certain limits to the ordinary language (physical object language) of perceptual experience. To overcome the limitations of ordinary language and to fit with the truth of the argument from illusion and the argument from hallucination, he proposed sense datum language. He says one can legitimately hold that in perception, sense data is directly perceived. However, this does not deny that physical objects are not directly perceived. Within the boundary of ordinary language (physical object language), one can legitimately view that physical objects like tables, chairs, etc., are directly perceived, whereas, within the boundary of sense-datum language, one can say sense data is directly perceived. However, Ayer prefers sense datum language over physical object language in analyzing the propositions of perceptual experience because it is more “comprehensive” than physical object language (Ayer 1945).

It is worth reemphasizing that there is a significant transition from ordinary sense-datum theorists to Ayer. The latter focuses on the linguistic analysis of perception and makes it explicit that the existence of sense data is a linguistic convention. For him, sense data exists neither in the mind nor in the physical world. He thinks this move has the virtue of avoiding unnecessary ontological commitments. He also thinks that maintaining a sharp distinction between sense datum language and physical object language could legitimize that physical objects and sense datum are directly perceived.

Now, we are in a position to make judgments over (i) and (ii). As far as physical object language is concerned, it is not appropriate to attribute (i) to Ayer. Thus, he is not committed to (ii). However, (i) is legitimately attributed to Ayer if we consider the sense datum language. In that sense, Ayer committed to (ii). However, the original motivation behind the sense-datum theory is missing in Ayer’s account. It is worth examining whether the introduction of sense-datum language really serves the purpose that ordinary sense-datum theorists raised. Moreover, there are some fatal inconsistencies involved in Ayer’s approach. If those inconsistencies are really considerable, then it is fair to demand a reevaluation of Ayer’s account.

## References

- Austin, J. L. (1962). *Sense and Sensibilia*. Oxford University Press.
- Ayer, A. J. (1940). *The Foundations of Empirical Knowledge*. London: Macmillan.
- Ayer, A. J. (1945). The terminology of sense-data, *Mind* 54(216), 289-312.
- Ayer, A. J. (1956). *The Problem of Knowledge*. Middlesex: Penguin.

- Brewer, B. (2006). Perception and Content, *European Journal of Philosophy* 14(2), 165-81.
- Broad, C. D. (1923). *Scientific Thought*. London: Routledge & Kegan Paul.
- Carrier, L. S. (1969). Immediate and Mediate perception. *The Journal of Philosophy*, 66(13), 391-403.
- Chediak, K. (2016). Intentionalism and the Problem of the Content of Perception, *Revista Portuguesa de Filosofia*, 519-529.
- Crane, T. (2001). *Elements of Mind*. Oxford University Press.
- Crane, T. (2006). Is There a Perceptual Relation, in T. S. Gendler & J. Hawthorne (eds.), *Perceptual Experience*. Oxford University Press, 126-46.
- Crane, T. (2013). *The Objects of Thought*. Oxford University Press.
- French, C. & Walters, L. (2018). The invalidity of the argument from illusion, *American Philosophical Quarterly*, 55(4), 357-364.
- Griffiths, A. P. (1960). Ayer on perception, *Mind*, 69 (276), 486-498.
- Hatfield, G. (2013). Perception and Sense Data, in M. Beaney (eds.), *The Oxford handbook of the history of analytic philosophy*. Oxford University Press, 948-974.
- Hatfield, G. (2021). Sense Data, *The Stanford Encyclopedia of Philosophy* (Fall 2021 Edition), Edward N. Zalta (ed.), URL=<https://plato.stanford.edu/archives/fall2021/entries/sense-data>.
- Huemer, M. (2001). *Skepticism and the Veil of Perception*. Rowman & Littlefield.
- Martin, M. G. F. (2000). Beyond dispute: Sense-data, intentionality and the mind-body problem, in T. Crane and S. Patterson (eds.), *History of the mind-body problem*. Routledge, 195-231.
- Martin, M. G. F. (2002). The transparency of experience, *Mind & language*, 17(4), 376-425.
- Moore, G. E. (1903). The refutation of idealism, *Mind*, 12(48), 433-453.
- Moore, G. E. (1913-14). The Status of Sense-Data, *Proceedings of the Aristotelian Society*, 14: 355-380. doi:10.1093/aristotelian/14.1.355
- Moran, A. P. (2019). *Austinian Disjunctivism Defended: A Presentational Theory of Visual Experience* (Doctoral dissertation).
- O'Shaughnessy, B. (2003). Sense Data, in B. Smith (eds), *John Searle*. Cambridge: Cambridge University Press, 169-88.
- Price, H. H. (1932). *Perception*. London: Methuen.
- Robinson, H. (1994). *Perception*. London: Routledge.
- Russell, B. (1912). *The Problems of Philosophy*. Oxford: Oxford University Press.
- Russell, B. (1914). The relation of sense-data to physics, *Scientia*, 16(16).
- Russell, B. (1927). *The Analysis of Matter*. London: Kegan Paul, Trench, Trubner.
- Schellenberg, S. (2010). The particularity and phenomenology of perceptual experience, *Philosophical Studies*, 149, 19-48.
- Searle, J. R. (2015). *Seeing Things As They Are: A Theory of Perception*. Oxford University Press.
- Smith, A. D. (2002). *The Problem of Perception*. Cambridge MA: Harvard University Press.
- Stoljar, D. (2009). Perception, in Shand J. (ed.), *Central Issues of Philosophy*. Blackwell Publishing Ltd: UK, pp 51-67.

## UNDERSTANDING ‘REPRESENTATION-ESCHEWING’ APPROACHES: AN ‘ANCIENT’ VIEW

Hari Narayanan V\*

Yogeshwar Joshi\*\*

**Abstract:** The paper argues that the fundamental claim of approaches to cognition that are ‘representation-eschewing,’ such as the enactive one, must have been more intuitively plausible in some cultures where the self is not understood in terms of an inner entity that can feel alienated or separated from the world. Taking cues from ancient literature, the paper demonstrates that even a simple act of perception is understood differently in some ancient cultures. It was more like an activity spread across eyes and objects and not simply an internal event receiving input from the world. This involves understanding oneself as continuous with the rest of existence. Accordingly, there was hardly any need to posit discrete representations inside that stand for objects outside. This suggests that approaches to cognition that are representation-eschewing may appear intuitively implausible in our culture because of certain inveterate but culturally constructed ways of understanding oneself as separate from the world.

**Keywords:** representation, enactivism, perception, self, literacy

### 1. Introduction

The notion of representation, understood as inner states that bear content or stand in for certain things, plays a central role in the study of cognition. It is said that a substantive assumption of cognitive science is that cognition is computation over representations (Von Eckardt 1995). One influential way to understand contents is in terms of causal-informational theories, according to which there exists a causal relation between a representation and what it represents (Dretske 1981, Fodor 1992). A representation becomes a mental representation when it helps the entity possessing it to exhibit intelligent behavior towards something (Rupert 2011). That is to say, when we see a tree and exhibit intelligent behavior towards it (e.g. standing under its shade when it is sunny), it is because we make use of the representation of a tree. As Robert Rupert (2018) points out, the positing of representations need not be in the service of vindicating folk psychology but as a scientific tool to explain intelligent behavior. That is to say, though representations are not observable, they need to be posited to explain cognitive activities just like many unobservable things are posited in the course of scientific inquiry.

Approaches to the study of cognition have been divided along the lines of the importance they give to the notion of representation in their explanatory framework. Roughly a distinction can be made between representation-rich and representation-eschewing approaches<sup>1</sup>

---

<sup>1</sup>The literature refers to representation-hungry domains of cognition. Approaches that consider cognition to be primarily involving such domains are here called representation-rich.

\*School of Liberal Arts, IIT Jodhpur, India; email: hari@iitj.ac.in; ORCID ID:

<https://orcid.org/0000-0002-3280-8632>

\*\*School of Liberal Arts, IIT Jodhpur India; email: p22hs005@iitj.ac.in; ORCID ID:

<https://orcid.org/0000-0003-1930-0823>

(Clark and Toribio 1994, Degenaar and Myin 2014, Zahnoun 2021). The former tends to consider representations to be essential as far as cognitive processes are concerned. For instance, perception involves constructing and using mental representations of objects and their properties. Similar is the case with, say, abstract reasoning. Representation-rich approaches hold that all cognitive processes require mental representations understood as inner “vehicles of specific contents” (Clark and Toribio 1994). This view is often treated as the default view with which cognition is studied.

On the other hand, representation-eschewing approaches tend to shun the use of representations. It need not amount to completely doing away with the notion of representation altogether, though radical enactivism attempts to do so, but ‘the cognitive’ is not seen as essentially involving representations. It is said that non-representational approach is one feature of what is called 4E (embodied, embedded, enactive and extended) cognitive science (Alexander 2025). Many cognitive processes customarily taken to be representation-hungry, such as abstract thinking, are seen as capable of being explained without invoking the notion of representation. It is even claimed that the ‘hunger’ for representation can be satisfied without appealing to inner representations but by reconceptualizing cognition as enaction (Kiverstein and Rietveld 2018). The central claim is that cognitive processes are to be understood primarily in terms of the survival of the organism, actively engaged with an environment. The organism is in continual effort to regulate its activity to continue its existence. In this regard, it is coupled with the environment, even while maintaining its autonomy, so as to receive whatever is required from the environment to maintain its integrity. It is systems like computers that require internal representations because they do not have such a relation with the environment. Higher cognitive processes such as reasoning are continuous with the basic ones, and therefore, it is claimed that they, too, can be explained without invoking detailed inner representations. There are attempts to explain even the act of remembering without positing inner stored memories (Hutto 2024).

The paper is an attempt to argue that representation-eschewing approaches to cognition must have been intuitively more plausible when processes like perception and concepts like self are understood in a particular way. This is done by discussing some accounts of perception, both ancient and contemporary. The claim is that when the self is not seen as an inner entity, the agent of actions, e.g., seeing is not seen as located inside, then the need for positing internal representations to explain cognitive processes becomes minimal. To support this claim, the paper additionally attempts to explain how the inner existence of the self becomes the default way of understanding the self as a result of several processes, such as literacy. As a result, representation-eschewing approaches may become more intuitively implausible and this may underlie the difficulty in appreciating, say, enactive approaches to cognition. We shall not argue for the corollary of this claim that when the agent is seen as an inner entity with a sharp boundary between the inner and the outer, then the approach to the study of cognition will become representation-rich.<sup>2</sup> Further, the purpose is not to defend any particular approach to cognition, such as the enactive one, but only to argue that their approaches to representation may appear to be implausible under a particular way of looking at phenomena like perception and self.

---

<sup>2</sup>We are thankful to an anonymous reviewer who suggested distinguishing between these two claims.

The aim of the paper is only to show that positing inner representations appears to be less compelling in certain ways of looking at perception and self. Though we use phrases like representation-eschewing approaches, our claim is consistent with positions that accommodate representations in certain areas of cognition. Further, the present claim does not presuppose that representation is posited in the service of defending folk psychology. Representation may be a scientific posit used to explain intelligent behavior but the claim made in the paper is that the very way we intuitively understand a phenomenon can shape the kind of questions that are asked in the scientific study of it.

The paper is structured as follows. The next section discusses some ancient accounts of perception, which are understood more in terms of an ongoing activity spread across objects, eyes, and organisms. In the third section, we shall show how this account is similar to some contemporary accounts of perception. In the last section, we attempt to demonstrate how the ancient accounts lost their plausibility in view of the gradual alienation of humans from the rest of existence and contend that as long as the self is understood in terms of an inner entity, enactive or representational-eschewing accounts would appear *prima facie* implausible. Thus, the paper argues that when a sharp distinction is not made between the inner and the outer and the self is not essentially seen as an inner entity, then it would be possible to appreciate the plausibility of approaches that do not heavily rely on the notion of inner representation.

## 2. An Ancient Account of Perception and the Self

It has been pointed out that vision was understood by the Greeks in a radically different way from our current default way of understanding it. In the writings of Homer, there are several verbs to refer to sight (Snell 1953). For instance, when it is written that a snake is ‘seeing one’ with the word *δέρκεσται* (*Derkesthai*), it refers to mainly the gleam of the eyes that can be noticed by someone else. That is to say, the reference is not to the internal experience of seeing but the relation between the eyes and the object. Another word was used to refer to visual attitudes, such as looking inquisitively. Similarly, there was another word to refer to seeing something bright. When Homer writes of an eagle as ‘looking sharply,’ Snell notes that it does not refer to the capacity of the visual organ, but to the beams of the eagle’s eye, which can penetrate any object. Here again, what we see is a connection between the organism and an object. There are similar expressions, meaning ‘seeing with joy’ or ‘seeing with fear,’ and such expressions refer to the emotions being determined by the object and not a quality of sight. It is again the relation that matters. The claim that Snell makes is that there was no single word to refer to sight as such, but many verbs refer to different types of seeing. The words referred to relations with things or quality of experience or to the quality of objects but not to an abstract thing called sight. In the words of Iain McGilchrist, “sight had not been abstracted yet from its context within the lived world . . . not yet thought of as unidirectional, detached, dead: not yet *observation*” (2010: 266). Subsequently, the Greek language developed a noun to stand for ‘sight,’ and the notion of a spectator, a person standing away and looking, gained currency. But this was absent in the works of Homer.

This lack of an exact word corresponding to vision can be understood in the larger context of how entities like body and mind were understood in Homer’s works. There was

no single word standing for body or mind (Snell 1953). The word ‘psyche’ was used, but it referred to many things, including breath and bodily parts. When a reference is made to a person’s thoughts, the word *thumos* was used but, ‘*thumos*’ is understood not as an immaterial thing but as something that has flesh-like bodily parts. That means, from the modern standpoint, we cannot say whether it is a mental word or a physical word. The same is the case with words like *menos*, *aion*, etc. The point is made by Michael Clarke when he says that “the Homeric man does not have a mind rather his thought and consciousness are as inseparable a part of his bodily life as are movement and metabolism” (1999: 115). That means all kinds of activities, whether it be called mental or physical in our times, were seen as continuous with one another.

In the words of Hermann Fränkel, “In the *Iliad*, then, man is completely a part of his world. He does not confront an outside world with a different inner selfhood, but is interpenetrated by the whole, just as he on his part by his action and suffering penetrates the total event” (1975: 80). Christopher Gill makes a similar point when he says that during the time of Homer, the sense of self was bound up with the ‘communal dialogue’ (Gill 1999). He points out that there is no equivalent to ‘personality’ or ‘self’ in ancient Greek, and the ancient understanding of self, Gill argues, cannot be understood by modern notions of the self in terms of an inner entity, forming the center of will, action, thinking, etc. What is required, in Gill’s words, is an ‘objective-participant model’ and not a ‘subjective-individualist’ model to understand the Homeric notion of self.

When there is no inner self, vision is not something that happens internally but is an ongoing process in which sense organs, the brain, and the objects outside all participate. Thus, the words referring to visual experience can refer to a multitude of things, including the quality of objects or the quality of the person who undergoes experience. In other words, visual experience is not an inner happening and vision is not the capacity of a person.

It is interesting to note that some accounts of indigenous<sup>3</sup> consciousness, too, reflect such an understanding of the self. For instance, Richard Sorenson (1998) holds that the way indigenous people looked at the world can be called liminal consciousness in contrast with the modern way of looking at the world. In the words of Sorenson, “Feeling and awareness are focused on at-the-moment, point-blank sensory experience – as if the nub of life lay within that complex flux of collective sentient immediacy” (ibid : 3). This suggests that attention is mostly around whatever takes place in the present flux and there is no strict compartmentalisation of experience. Everyone participates in the flux, and there is hardly any room for sharp demarcation between the inner world and the outer world.

Sorenson further points out the number of abstract concepts was much less in indigenous societies and even the names of individuals were not binding. This suggests that the identity of a person did not derive from a name. This is expected in a situation where abstraction is given less prominence and life is understood to unfold in real concreteness. Further, it is pointed out that individual and collective elements are expected to co-exist in a person in indigenous societies.

Further, it is important to note that non-verbal attention is predominant in liminal consciousness. Even after the mastery of oral language, sensory contact with the world, such

---

<sup>3</sup>The word indigenous may be contested. It is used here primarily in the sense of pre-literal.

as tactile sensation, is the primary way of being in the world. When symbols and abstraction are less, it is expected that the sensory mode retains the dominant mode of experience.

Regarding the Dinkas of Sudan, it is claimed that they did not have the concept of inner life (Lienhardt 1961). There was no word corresponding to belief, desire, etc., in their parlance. The identity of a person is always understood in fluid terms and behaviour is influenced by a multitude of factors. Such a state of affairs is expected when, in the words of Rebecca Adamson (2008), “the environment is perceived as a sensate, conscious entity suffused with spiritual powers through which the human understanding is only realized in perfect humility before the sacred whole.” What this indicates is the lack of a sharp separation between the conscious subject and objects in the world because all that exists is understood in terms of the integral conscious whole.

There are some accounts of perception, even in our times, that resonate with the ‘ancient view’ discussed in this section. In the next section, we discuss two such accounts. Subsequently, we shall discuss why the intuitive plausibility of such accounts got gradually diminished.

### 3. Perception: Some Contemporary Accounts

Maurice Merleau-Ponty, who famously identified the self with the body, looked at perception in terms of participation. It is the living body, integrally connected with the world around, that forms the real subject. This position, in the words of David Abram, “rejuvenate[s] the organic basis of our thoughts and our intelligence” (2017: 47). Merleau-Ponty contends that whatever human intellect is capable of doing, the moorings always lie in sensory experience.

Perception, according to him, involves both receptivity and creativity in dealing with the world. The contours of the world are always uncertain, and one needs to adjust constantly to whatever one comes across. It is this requirement that makes the perceptual process both receptive and creative. This clearly shows the two-way nature of perception. Merleau-Ponty writes, “I give ear or look in the expectation of sensation, and suddenly the sensible takes possession of my ear or my gaze, and I surrender a part of my body, even my whole body, to this particular manner of vibrating and filling space known as blue or red” (2013: 246). The sensible world is active and alive, and he even uses terms like ‘beckoning’ and ‘responds’ to describe the active nature of the objects. Perception is even called “a coition of my body with things” (ibid: 373). Both the perceiver and the perceived have the role of acting as well as receiving.

It is important to note that what is described is a matter of direct experience and, therefore, pre-reflective. It is the way the ‘sensing body’ experiences things. Sensing, in the words of Merleau-Ponty, is a “vital communication with the world which makes it present as a familiar setting of our life” (ibid: 61). The basic unit of sensation is a whole to be understood in terms of a figure appearing in contrast with a background. This account of perception becomes clear when we understand that the body is always something that is oriented to the world. It is “intention incarnate.” It is even claimed that the body’s situation in the world is as organically related as that of a “heart in the organism” (ibid: 235).

One can find a similar account of perception in the sensory-motor accounts of consciousness propounded by Alva Noë and Kevin O’Regan (O’Regan and Noë, 2001; Noë, 2004, Noë 2021). Their approach falls within the broader field of enactivism in cognitive science. The enactive approach in cognitive science considers the interaction between an organism and the environment to be the basis on which other cognitive structures emerge and looks at perception as a form of action (Varela et al. 1991). Enactivism is of different types, and we shall discuss only sensory-motor enactivism (Ward et al. 2017). It tries to explain the perceptual experience in terms of the interaction between the body and the environment, eschewing the need for the brain to build a detailed model because the world is present to us and can be accessed directly. As per this approach, called actionism by Noë, perception is basically an act of exploring the environment. For instance, vision is not the result of forming a detailed picture of the world inside but an act of exploring the world, making use of the knowledge we have about how our bodily movements affect sensation.

Actionism challenges the view that perception is primarily a matter of receiving some data from the world and then forming a detailed picture of the object in the brain. Much of the existing studies on vision are focussed on figuring out how the brain can form a detailed picture, which, in turn, gives rise to visual experience. Actionism tries to challenge this picture by holding that there is no requirement for the brain to form any detailed picture as the world is “its own model” and is available to us at any time. That is to say, the vivid experience of the visual field is not the result of brain processing to complete whatever fragmented information we get through senses but the fact that we have got constant access to the world.

To consider perception as a form of action is not just to go beyond the dichotomy of perception and action but also to conceptualize cognition as spanning from the head to the environment. The study of vision often assumes it to be taking place in the head. That the sense data we get through our eyes is incomplete is beyond doubt (O’Regan 1992, 2011). First of all, a retinal image is two-dimensional, but we see the world in three dimensions. Similarly, there are many gaps in the visual system, such as blind spots and poor peripheral vision. But despite all these, our experience of the visual world appears quite vivid. The traditional way to explain this is by holding that the brain fills in the missing information and builds a detailed picture of the world. This conclusion appears to be ineluctable because, after all, it is “I” who sees, and the “I” is an inner entity. Then, the inner process constitutes cognition, and the role of the rest of the world is limited to input and output.

Actionism challenges the view that the brain needs to form a complete picture of the world. As Alva Noë (2009) puts it, to think that the brain produces experience is to treat it like, say, digestion. A more appropriate metaphor to understand consciousness, according to Noë, is that of dance or of playing a musical instrument. It is an action or a process where multiple factors are involved and not a thing or product as such.

Actionism holds that our perception depends upon our knowledge of changes in sensation as a result of movement, and this underlies the vivid impression of a continuous visual field. For instance, when we see a house, we do not see its backside, but this does not prevent us from the feeling of seeing the whole house because we know that we have ready access to it. Similarly, we know that small movements of the eyes can result in seeing the periphery, and this feeling of easy access can generate the impression that the whole visual field is equally

vivid. The point here is that there is no need for the brain to fill in gaps or form a detailed picture because the world is out there at any time. What is called a retinal image is not a picture in the usual sense, so there is no need to make that image a better or complete picture.

Further cases such as inattentional blindness and change blindness demonstrate that we do not see all that there is but only what matters to us. Inattentional blindness is a case of not seeing certain things in the visual field simply because our attention is focussed on certain other things. Change blindness is a matter of failure to notice small changes in a scene, although they can be noticed if we look for them. These phenomena are quite common and show that we do not have any detailed representation of all that is there, but see those things that matter to us. This implies that seeing is a matter of exploring and gaining information based on our needs. It is, as per actionism, a way of relating to the world and not a matter of forming a picture in the head. In other words, seeing is a matter of the world and its objects being present to us. Even if all seeing does not necessarily involve explicit movement or action, when we see something, the relationship is always oriented toward action.

It is important to note that while doing away with the need for the brain to construct any detailed representation of whatever we see, the view does not cast aside the brain as entirely irrelevant. It certainly plays a key role, just like a musical instrument is important, but is not the sole locus of a musical performance. The brain, using implicit knowledge of relations between movements and stimulation (called sensorimotor contingencies), is certainly essential in coordinating our actions, including the enactment of perception. What matters for perception is the link between incoming information and bodily movements and not the existence of any inner picture or mysterious quality, often referred to in the literature as *qualia*. Experience is ultimately realised in the “active life of the skilful animal” (Noë 2004: 226).

This way of looking at perception clearly resonates with the ancient account discussed in the previous section. If the words for vision in Homer meant a connection between an organism and the objects in the environment, then it implies that seeing is not primarily an inner occurrence but a matter of engaging with the world. Similarly, when sight is not abstracted from its context, vision is something that happens between the organism and the environment.

The purpose of the present paper is not to defend actionism or any related attempt as the right account of vision. There have been voices from within camps sympathetic to enactivism that are critical of the approach of Noë. Andy Clark, a leading exponent of embodied approaches, considers the sensory-motor approach as “taking one step too far” (2011: 177). Ian Robertson (2023) holds that Noë’s account cannot be accepted with respect to the position that skilled action involves understanding. The present purpose is only to show that representation-eschewing approaches look plausible when the inner world or the self are understood in a different manner, and this way of looking was common in some cultures. This implies that if actionism or a similar position currently appears intuitively implausible, that can be due to the ingrained ways of understanding ourselves in terms of an inner entity, making us oblivious to how we are inextricably embedded in the environment.

Our discussion of Merleau-Ponty’s account of perception suggested that the actual nature of perception can be captured only in its pre-reflective moment. The moment conceptualisation comes in, perception may no longer be seen in terms of a reciprocal encounter because then, it can happen that perception is understood in terms of the activity of the inner self receiving data from external objects. The point is that when conceptualisation starts, it often takes place in the agentive mode in which it is the ‘I’, often understood in terms of an inner entity, that sees. This may lead to the view that what is central to seeing, or for matter any kind of perception, is an inner process and as a result, so much effort is spent on figuring out how the brain can manage this thing called perception.

However, our discussion in the first section indicated that the ‘I’ is not always understood as an inner entity. In that case, even if ‘I see’, vision need not be predominantly inner. Just like the self is distributed, so can agency be. That means the issue is not one of agency of the ‘I’ as such, but how the ‘I’ is understood.<sup>4</sup> An enactive account of self can be in the form of ‘I-ing’ or ‘selfing’ (Thompson 2015). Self is not seen essentially as a fixed entity or the subject that looks at the object but more as a process. As Mark Johnson puts it, “what philosophers call ‘subjects’ and ‘objects’ (persons and things) are abstractions from the interactive process of our experience of a meaningful self-in-a-world” (Johnson 2007: 20). It means that the idea of self as an entity opposed to the world of objects is the result of abstraction from basic experience. Thus, when cognition itself is understood in terms of an enaction, the self is no longer an abstract entity but something that unfolds during the enaction. Then still, the “I” may be seeing, but that “I” is not an inner entity clearly demarcated from the world but something that spans over the body and the environment. This is certainly similar to the man of the *Iliad*, discussed earlier, who “does not confront an outside world with a different inner selfhood.”

To conceive of oneself as an inner entity is to abstract it or separate it from the larger context. This, in turn, engenders plenty of avenues to conceptualize different kinds of narratives. Narrative or remembering self (Strawson 2008, Kahneman 2011) is predominant in our times, in which the self is understood as an entity rooted in the past that looks towards the future through the present. That is to say, the temporal division of past and future is central to the narrative self, and this in turn, gives rise to an understanding of the self that stands away from the flux in the present. This point can be elucidated further by looking into the possibility of the emergence of an inner, separative self-conception under certain conditions such as literacy. In the next section we try to show that the abstraction from the ongoing experience into fixed subjects and objects and the entrenchment of the subjective-individualist model of the self can be due to various factors such as literacy.

#### 4. The ‘Literate’ Worldview and the Entrenchment of the Inner World

Orality and literacy are distinct in profound ways, though they appear to go together intimately in our culture. Given the relatively recent advent of written language, for a long time in human history, orality was the foremost medium of acquiring and transmitting knowledge. There have been attempts to highlight the unique aspects of oral cultures in

---

<sup>4</sup>Latour (2014) even speaks of non-human agency.

terms of their way of acquiring, transmitting, and preserving knowledge, which is distinct from literate cultures that rely on written scripts. This is often termed as the orality-literacy divide or the “great divide” (Havelock 1963; 1982; 1986, McLuhan 1962, Goody 1977; 1987, Ong 2002, Olson & Torrance 1991, Abram 2017). It is claimed that with the advent of writing systems, particularly with the invention of the ‘alphabet’, certain fundamental changes happened, which resulted in the transition of human civilization from oral culture to the one dominated by literacy. The shift from orality to literacy entailed significant implications for human consciousness, cognition, and its relation with the surrounding sensorial world. Alphabetic literacy, especially the Greek alphabet, is argued to be the cause of Greek enlightenment, which is said to have aided Western civilization in transcending the so-called limitations of the “oral mind” (Havelock 1963; 1982, Halverson 1992).

#### **4.1. The Nature of the Alphabet**

Archaeological evidence indicates that pre-writing systems, and the earliest pictographic scripts of pre-historic humans, evince their profound entanglement with the ‘more than human world’. As Abram (2017) notes, the survival and subsistence of our pre-historic hunter-gatherers and tribals depended on their ability to correctly interpret cues in the form of scrawls, tracks, and trails left in nature by other wild creatures. To ‘read’ such signs found in nature required a high level of sensorial alertness and awareness of the enviroing lifeworld.

This entanglement with the natural world was visible in the earliest writing systems developed by humans. When it comes to writing systems that humans have invented, it is claimed that as early as 27000 BCE, humans were carving stylized, intentional marks on animal bones (Rhodes 2023). Pictograms or ideograms, developed independently in different cultures such as Mesoamerica, Mesopotamia, Egypt and China, consisted of characters that picture the visible thing by referring to some quality commonly associated with the pictured entity.

Since such ostensive ideograms were insufficient to capture the entirety of meaningful discourse, more techniques were employed. Further, pictographic systems were unwieldy as they consisted of thousands of characters to refer to various terms in language. Rebuses were used to refer to terms with no clear visual representation (Abram 2017: 98). Attempts to refer to the sounds of human speech resulted in a phonetic script that relies on transcribing the sound of the spoken voice rather than its external referent or meaning. The discovery that every syllable in the language is composed of what is presently called consonants and vowels led to the system of alphabets to represent those elements. This drastically reduced the number of characters in the writing system and led to the modern writing systems.

With the widespread use of the alphabet, the gap between direct sensorial reality and human communicative systems drastically widened (Abram 2017: 100). If the pictographic system represented visual objects, the alphabets refer to only the sounds produced by the human voice. The imagistic logic that connected humans with nature was left behind with the advent of the alphabet, although it continued in certain measures in pictographic systems of Chinese script, Egyptian hieroglyphs, and the Mayan system (Tedlock 1993).

Based on such observations, Abram (2017: 100-101) argues that the invention of writing led to a progressive dissociation of humans from their surrounding environment. The senses of pre-historic ancestors slowly started getting tuned to the world of written scripts. The earlier written scripts, as noted earlier, retained some connection with the surrounding world as the pictographic systems resonated and aided humans in recalling various aspects of the surrounding animate world. With the advent of the alphabet, the meaning of the linguistic terms moved beyond the reference to the enveloping lifeworld. This entailed a shift in attentional landscape from sensorial reality to the shape of utterance invoked by the written word. In alphabetic writing, the abstraction of linguistic terms became prominent (Ong 2002, Goody 1977). This change is evident in the contrast between the oral character of Homeric epics and Plato's dialogues. In Homeric epics, the animate and the participatory nature was salient as many of the gods in the epics are the very forces of nature, whereas, in Plato's dialogues, one can clearly see an emphasis on the contemplation of the abstract world of unchanging ideas and essences away from the sensorial terrain. Socrates' visibly dismissive attitude towards nature as 'nothing to learn from' in *Phaedrus* peculiarly shows the mark of a mind that is more attuned to the world of ideas than the world of the sensorial reality (Abram 2017: 102-104, Havelock 1963; 1986).

This point needs to be elucidated further to show how literacy might have contributed to the richness of the inner world or the sense of a separative self. In oral cultures, the acoustic medium was not conceived as a phenomenon different from the person who used it. However, literate cultures could objectify the medium as the medium is different from the person using it. That means whatever is written down can be looked at again as if one were looking at an object. With alphabetic writing, it became possible to stand at a distance and objectively think about nature and the existence of written thoughts. As a result, the power of reflexivity strengthened. As Abram (2017: 110-111) states, earlier qualities like 'justice' were understood in the context of their actual occurrence and not as abstract entities.

But with alphabetic writing, it became normal to look at them as entities independent of people and situations. For instance, Socrates' concern with unchanging eternal essences of 'justice', 'virtue', 'courage' and so on, rather than its fleeting occurrences in the real world, shows the impact of the power of reflexivity borne by alphabetic writing. The introduction of the very distinction between the word and what it 'stands for' was arguably the result of this kind of writing. In the oral setting, the word participated in the actual phenomenon, in its very embeddedness. Thus 'justice', 'virtue', 'courage' etc. were conceived as they unfolded in real-life situations. The thorough abstraction possible in the literate mind led to the enhanced reflective ability within human consciousness, which was called *psyche* by Socrates. The psyche may be appropriately called the "literate intellect" (ibid: 113). This *psyche* enabled intellectual contemplation away from the ordinary sensory world towards the world of ideas and forms, which was considered immortal and essential.

Thus, it may be reasonably inferred that the widespread acceptance of alphabetic writing enabled the practice of looking at one's ideas abstracted from real-world occurrences, and this, in turn, generated the notion of the self as an inner entity that possesses and uses such ideas. In fact, the self itself can be seen as an idea removed from direct bodily embeddedness in the environment. As a result, the 'objective-participant' model of the self gave way to the 'subjective-individualist' model. Thus, it is claimed that training in alphabetic reading and

writing fosters a reflexive sense of self as an autonomous, independent, unchanging entity away from sensuous surroundings. Then, ‘sight’ becomes abstracted from its real-world occurrence and treated as something that happens to the inner self, and the ability to see needs to be explained mainly in terms of inner processes. This amounts to a significant transformation in how humans looked at themselves and the world. In a nutshell, it may be said that the invention of writing, in turn, reinvented humans, or as Ong (2002: 77) stated, “More than any other single invention, writing transformed human consciousness.”

## 4.2. Storage and Knowledge

The magnitude of the shift can be gauged further when we consider the fundamental differences in the way oral and written language are acquired, transmitted, preserved, and received (Horowitz & Newman 1964). The ability to collect and store information is effectively actualized when human thought is written down. In the absence of such mechanisms, different ways are employed for preserving information. The totality of the oral community’s knowledge of its identity, history, culture, and ways of life is often remembered and transmitted through folklore, stories, myths, narratives, riddles, poems, and songs. Language in oral settings acts as a performative tool to effectively communicate and store information in the collective memory; therefore, knowledge is fluid and dynamic. Oral cultures learn information embedded in a situation through listening, repeating, and participating in situ, and the emphasis is on correspondence between the narrated events and human experience. In literate cultures, language is understood to be representing thought. The thought is then preserved in written format and stored in a large set of data that can be readily referred to, and the emphasis is on neutral, objective facts understood to be bereft of the subjective point of view. Thus, in literate culture, people learn through abstract thoughts represented in written scripts through books and diverse automated mediums that depend on written scripts (Ong 2002, Sherwani et al. 2009: 40).

Based on these practices of knowledge-building and preserving, certain distinctive characteristics of the oral and literate culture emerge (Ong 2002: 33-48). The verbalized and the acquired knowledge in oral cultures remain close to the concrete reality as it appears in human experience. In this way, knowledge in oral cultures is tied to the human lifeworld. It is situational in nature. Literate cultures, on the other hand, employ analytic means and structure knowledge at a distance from lived reality by way of abstraction. They rely on syllogistic and formal logic and define concrete objects in abstract terms (Goody 1977). As spoken words are susceptible to being forgotten or misunderstood, oral cultures rely on mechanisms of redundancy and repetition. These enable the speaker to maintain continuity and aid the hearer to stay on track with the speaker. In literate cultures, authors can organize their thoughts at their own pace and then put them down in written format. In such conditions, there is always a way for readers to refer to the earlier content; thus, the redundancy and repetition mechanisms are greatly reduced.

Moreover, in oral cultures, since the speech often functions as a performative tool, the speakers tend to invoke feelings of emotional resonance with the audience. In other words, the affective component is strong in oral cultures. Oral reciters use rhetoric techniques such as humour, exaggerations, puns, metaphors etc. The speaker and the audience participate together in meaning-making. In the literate culture, the accuracy of the message is given

more importance. Thus, the informational and logical content are valued mainly rather than its affective tone. The tools of reasoning and evidence are used to augment the accuracy of the message, which is objectively construed. Oral cultures tend to forget memories that are useless and obsolete and thus are often described as homeostatic (Ong 2002, Sherwani 2009: 40-42, Yeganeh 2022: 566-568). Oral cultures do not hesitate to use the same word in different contexts carrying different meanings. As a result, the vocabulary remains restricted. On the other hand, literary cultures are accumulative as archaic words from the past are stored and defined. Dictionaries, therefore, become particularly useful. Literate cultures emphasize precision in the meaning of the word. In oral culture, since knowledge is not stored in any objective fashion, it is fluid. It keeps changing over a period of time. In literate culture, knowledge remains solidly precise in written letters. Furthermore, in oral cultures, knowledge is stored in the collective memory of the people. Thus, the importance of communal learning is manifold. However, in literate cultures, individual learning is emphasized as learning is looked at as an impersonal and neutral process that can be done based on reference to recorded texts. Finally, due to its participatory nature and emphasis on communal learning, in the oral culture, the produced knowledge is construed as subjective. In contrast, in literate cultures, due to the emphasis on individualistic learning, the produced knowledge is often construed as objective and high on impersonal content, or the ‘knower’ is seen as separated from the ‘known’ (Ong 2002: 31-73, Goody 1977, Yeganeh 2022: 565-569, Mianowski 2011: 127-139).

These contrasts between orality and literacy seem to imply almost different ways of being in the world. Both the nature of the alphabet and the way the knowledge is stored in the written medium contribute to this difference. When abstraction from the living world becomes the norm and meaning is treated as precise, then a clear division between the inner and the outer, or between the self and the world, becomes the default way of looking at things. This is what is meant by saying that the ‘knower’ is separated from the ‘known’. When the “knower” and the “known” becomes sharply separated the knower is reified into an inner entity that appears to be separated from all that is known. This underlies the subjective-individualist model of understanding the self in contrast with the objective-participant model.

Subsequent technological developments arguably strengthened this trend. Printing makes the world of the written word much easier to produce and transmit. As Elizabeth Eisenstein (1980) argues, a new idea of selfhood emerged with the printing press because of the possibility of keeping one’s ideas fixed forever. She even says that printing is the first invention which created controversy regarding who invented it first. That is to say, claims of individualism became stronger with the written word, accentuated by the invention of the printing press. In the words of Neil Postman (2000), the printing press ‘altered human consciousness’ itself. To quote Eisenstein, “It encouraged a sharper division between private worlds and public affairs” (Eisenstein 1980: 133).

Further, the notion of agency became more potent with the series of such ‘disruptive technologies.’ All of them, apart from making human life comfortable, required us to engage in more and more activity. Consider electricity. Millions of people must be engaged throughout the day to ensure it is always available. So is the case with many other life

transforming inventions. Thus, the whole techno-economic system requires humans to remain active; therefore, agency becomes a central feature of humans. All this strengthens the trend to conceive of oneself as distinct from the world, giving rise to what is called a thick psychological boundary with the world (Albahari 2006). The human-nature dichotomy became the default way of understanding, and nature is seen as the ground upon which to exercise human agency.

The present conception of nature implicitly assumes it to be without humans, but this understanding is of recent origin. As G. E. R Lloyd (1991) puts it, nature is something that is ‘invented.’ Bruno Latour (2013) argues that it is modern humans who tend to think in terms of there being a nature outside of us, consisting of so many things that exist intrinsically. Apart from the technological changes discussed above philosophical positions have also contributed to the emergence of this world view. For instance, as per the corpuscular world-view, matter is ultimately minute particles without any end or purpose, and the Cartesian view denies consciousness to anything other than humans (Martin and Baressi 2006). When the non-human world is deemed unconscious and without any purpose, a strong foundation gets built to bifurcate humans and nature. Further, this pushes consciousness in humans to an ‘inner’ realm because the body is material and therefore not conscious. Thus, the essence of a human being is understood to be inner, the locus of thinking and reasoning. This can be understood in terms of the ‘subjective-individualist’ model referred to earlier. When the essence of the self is understood in terms of an inner entity, all that one does, whether it be seeing, thinking, or reasoning, needs to be explained in terms of internal processes. This, in turn, may generate the need to posit a complex representational system that can carry out cognitive processes internally.

The above discussion suggests that literacy and the subsequent technological developments have the impact of firmly etching the sense of the self as an inner entity separated from the world. With this kind of understanding of the self, seeing becomes an inner activity because the self that sees is primarily an inner entity. The eyes become just a tool of the self, and the object is independent and simply seen by the subject. It must be then that the Greek language developed a noun to stand for ‘sight,’ and the notion of a spectator, a person standing away and looking, gained currency. When ‘sight’ had not been abstracted from living experience, vision was not seen as an internal affair, and therefore, there was no formidable problem with how the brain accomplishes seeing. But in our times, when the self is generally understood to be internal and all that one does is largely conceptualised as an internal activity, the ‘problem’ arises as to how the brain can accomplish the task of seeing based on scant data received by the senses. This, in turn, may give rise to the assumption that the brain constructs detailed inner representations and our knowledge of the world through such representations, making the ancient and enactive approaches discussed in this paper appear implausible. This line of thinking does not assume that representations are posited only to defend folk psychology. As Robert Rupert (2018) points out, representation may be a posit in the scientific explanation of behavior; therefore, it is on par with other posits like electrons, etc.<sup>5</sup> Further, he argues that critics of the notion of representation expect it to be in tune with folk psychological understanding. Our claim is only that when the self is understood as an inner entity, cognitive processes are also treated as essentially

---

<sup>5</sup>We are thankful to an anonymous reviewer for drawing attention to this point.

internal, and therefore, such an understanding of the self makes representation-eschewing approaches look implausible. This is not to be understood as a critique of the notion of representation or expecting it to be aligned with folk psychological commitments. Of course, what is intuitively plausible may turn out to be scientifically untenable. At the same time, if the shift from an objective-participant model of understanding the self to the subjective-individualist model was due to cultural factors and not because of any ‘scientific’ knowledge we gained about the nature of the self, then it is reasonable to caution that the intuitive implausibility of any approach to cognition arising out of that shift needs to be taken with a pinch of salt. Further, it is generally accepted that any scientific study makes assumptions about the nature of the phenomenon that is being studied, and this constrains the kind of questions asked and the methods adopted. Hence, it is important to take seriously the approaches to cognition that are rooted in a different way of understanding the self.

## 5. Conclusion

Approaches to the study of cognition that are representation-eschewing must have been intuitively much more plausible in some ancient modes of living. It is reasonable to hold that with the widespread adoption of literacy and subsequent developments, humans started becoming estranged from the living world or ‘sensorial reality’ and began feeling inner life to be the seat of oneself. This can be called the shift from the objective-participant model of understanding the self to the subjective-individualist model. This might have given rise to the notion that our activities, such as perception, need to be explained internally because the self is essentially an inner entity. When this picture of ourselves as inner entities is overthrown, avenues get opened to appreciate representation-eschewing approaches to cognition, such as the enactive one.

## References

- Abram, D. (2017). *The Spell of the Sensuous: Perception and Language in a More-Than-Human world*. USA: Vintage Books.
- Adamson, R. (2008). ‘First Nations’ Survival and the Future of the Earth’ in M. K. Nelson (ed.), *Original Instructions: Indigenous Teachings for a Sustainable Future*, 27–35. Rochester: Bear & Company.
- Albahari, M. 2006. *Analytical Buddhism*. London: Palgrave Macmillan.
- Alexander, C. (2025). What is 4E cognitive science?. *Phenomenology and the Cognitive Sciences*. <https://doi.org/10.1007/s11097-025-10055-w>.
- Clark, A. (2011). *Supersizing the Mind*. OUP.
- Clarke, M. (1999). *Flesh and Spirit in the Songs of Homer*. Oxford: OUP.
- Clark, A., & Toribio, J. (1994). Doing without representing? *Synthese*, 101, 401-431.
- Degenaar, J., & Myin, E. (2014). Representation-hunger reconsidered. *Synthese*, 191, 3639-3648.
- Dretske, F. (1981). *Knowledge and the Flow of Information*. USA: MIT.
- Eisenstein, E. (1980). *The Printing Press as an Agent of Change*. Cambridge University Press.
- Fodor, J. A. (1992). *A Theory of Content and Other Essays*. The MIT press.
- Fränkel, H. (1975). *Early Greek Poetry and Philosophy: A History of Greek Epic, Lyric, and Prose to the Middle of the Fifth Century*. New York: Harcourt Brace Jovanovich.
- Gill, C. (1996). *Personality in Greek Epic, Tragedy and Philosophy*. Clarendon Press: Oxford.
- Goody, J. (1977). *The Domestication of the Savage Mind*. UK: Cambridge University Press.

- Goody, J. (1987). *The Interface between the Written and the Oral*. Cambridge: Cambridge University Press.
- Halverson, J. (1992). Havelock on Greek Orality and Literacy. *Journal of the History of Ideas*, 53(1), pp. 148–163. <https://doi.org/10.2307/2709915>
- Havelock, E. (1963). *Preface to Plato*. Cambridge: Belknap Press of Harvard University Press.
- Havelock, E. (1982). *The Literate Revolution in Greece & Its Cultural Consequences*. New Jersey: Princeton University Press.
- Havelock, E. (1986). *The Muse Learns to Write: Reflections on Orality and Literacy From Antiquity to the Present*. New Haven: Yale University Press.
- Horowitz, M. W., & Newman, J. B. (1964). Spoken and written expression: An experimental analysis. *The Journal of Abnormal and Social Psychology*, 68(6), 640–647. <https://doi.org/10.1037/h0048589>
- Hutto, D. D. (2024). Acts of Remembering: From a Radically Enactive Point of View. *Psychoanalytic Inquiry*, 1-19.
- Johnson, M. (2007). *The Meaning of the Body: Aesthetics of Human Understanding*. Chicago: The University of Chicago Press.
- Kahneman, D. (2011). *Thinking, Fast and Slow*. Penguin Books.
- Kiverstein, J. D., & Rietveld, E. (2018). Reconceiving representation-hungry cognition: an ecological-enactive proposal. *Adaptive Behavior*, 26(4), 147-163.
- Latour, B. (2013). *An Inquiry into Modes of Existence: An Anthropology of the Moderns*. USA: Harvard University Press.
- Latour, B. (2014). Agency at the time of the Anthropocene. *New Literary History*, 45(1), 1–18.
- Lienhardt, G. (1961). *Divinity and Experience: The Religion of the Dinka*. Oxford: Clarendon Press.
- Lloyd, G. E. R. (1991). “The invention of nature.” In: *Methods and Problems in Greek Science: Selected Papers*, 417–434. Cambridge: Cambridge University Press.
- Martin, R., and Barresi, J. (2006). *The Rise and Fall of Soul and Self: An Intellectual History of Personal Identity*. New York: Columbia University Press.
- McGilchrist, Iain. (2010). *The Master and His Emissary: The Divided Brain and the Making of the Western World*. New Haven: Yale University Press.
- McLuhan, M. (1962). *The Gutenberg Galaxy: The Making of Typographic Man*. Canada: University of Toronto Press.
- Mianowski, J. (2011). Oral and literate cultures – Where do the boundaries overlap?” In Zdzisław & Wąsik (Eds.), *Languages in Contact*, pp.127-139. ISBN 978-83-60097-09-0.
- Noë, A. (2021). The enactive approach: A briefer statement, with some remarks on “radical enactivism.” *Phenomenology and the Cognitive Sciences*, 20(5), 957–970.
- Noë, A. (2004). *Action in Perception*. USA: MIT Press.
- Noë, A. (2009). *Out of Our Heads*. Hill and Wang.
- O'Regan, J. K. (1992). Solving the “real” mysteries of visual perception: The world as an outside memory. *Canadian Journal of Psychology/Revue canadienne de psychologie*, 46(3), 461-488.
- O'Regan, J.K. (2011). *Why red doesn't sound like a bell: understanding the feel of consciousness*. New York: OUP.
- O'Regan, J. K., & Noë, A. (2001). A sensorimotor account of vision and visual consciousness. *Behavioral and Brain Sciences*, 24(5), 939–973.
- Olson, D., & Torrance, N. (Eds). (1991). *Literacy and Orality*. Cambridge, UK, and New York: Cambridge University Press.
- Ong, W. (2002). *Orality and Literacy: The Technologizing of the Word*. USA: Routledge.
- Merleau-Ponty, M. (2013). *Phenomenology of Perception*. Tr. by Colin Smith. Routledge.
- Postman, N. (2000). *Building a Bridge to the 18th Century: How the Past Can Improve our Future*. New York: Vintage.
- Rhodes, M. (2023, July 2023). From Orality to Literacy: A Global History of Writing. *DIG: A History Podcast*. <https://digpodcast.org/2023/07/23/global-history-of-writing/>
- Robertson, I. (2023). The unbearable rightness of seeing? Conceptualism, enactivism, and skilled engagement. *Synthese* 202, 173. <https://doi.org/10.1007/s11229-023-04385-y>

- Rupert, R. D. (2011). Embodiment, consciousness, and the massively representational mind. *Philosophical Topics*, 99-120.
- Rupert, R. D. (2018). Representation and mental representation. *Philosophical Explorations*, 21(2), 204-225.
- Sherwani, J., Ali, N., Rose, C., & Rosenfield, R. (2009). “Orality-grounded HCID: understanding the oral user.” *Information Technologies and International Development*, 5 (4), pp. 37-49.
- Snell, B. (1953). *The Discovery of the Mind: The Greek Origins of European Thought*. New York: Harper and Row.
- Sorenson, R. (1998). “Preconquest Consciousness.” In Wautischer (Ed). *Tribal Epistemologies: Essays in the Philosophy of Anthropology*. Ashgate, 79-116.
- Strawson, G. (2008). “Against Narrativity” in *Real Materialism and Other Essays*. Oxford: Clarendon Press. pp. 428–452.
- Tedlock, D. (1993). *Breath on the Mirror: Mythic Voices & Visions of the Living Maya*. San Francisco: HarperCollins.
- Thompson, E. (2015). *Waking, dreaming, being: Self and consciousness in neuroscience, meditation, and philosophy*. Columbia University Press.
- Varela, F. J., Thompson, E., & Rosch, E. (1991). *The Embodied Mind*. Cambridge: MIT Press.
- Von Eckardt, B. (1995). *What is cognitive science?*. USA: MIT press.
- Ward, D., Silverman, D. & Villalobos, M. (2017). Introduction: The Varieties of Enactivism. *Topoi*, 36, 365–375. <https://doi.org/10.1007/s11245-017-9484-6>
- Yeganeh, H. (2022). Orality, literacy, and the “great divide” in cultural values. *International Journal of Sociology and Social Policy*, 42(5/6), pp. 564-582. <https://doi.org/10.1108/IJSSP-04-2021-0088>
- Zahnoun, F. (2021). On representation hungry cognition (and why we should stop feeding it). *Synthese*, 198 (Suppl 1), 267-284.