The Embedded View, its critics, and a radically non-representational solution

Author:

Thomas van Es, Centre for Philosophical Psychology, Universiteit Antwerpen, Belgium

Abstract

Whether perception involves the manipulation of representations is currently heavily debated. The Embedded View (EV) advanced by Nico Orlandi seeks a middle passage between representationalism and radical enactivism. In this paper I argue for a non-representational take on EV. I argue that this is the best way to resolve the objections EV has received from both representationalists and non-representationalists. I analyze this debate, and distinguish four sorts of objections: 1) the objection of the wrongfully cut middleman, 2) the argument against explanatory exclusionism, 3) the case for scientific benefits of representations, and 4) the charge of inconsistent ascription of representational status in EV. I argue that (1) the middleman was never cut in EV, and is controversial to boot, (2) *otherwise equal*, non-representational explanations have primacy over representational explanations, due to the lack of naturalistic grounds for representations and the unnecessarily ascribed cognitive load to the system. Further, I show that (3) puts the cart before the horse, and the arguments on offer are viciously circular. However, the final objection, (4) lays bare a deeper issue for EV. At the cost of giving up the middle position, however, the explanatory tools already available to EV can be shown to cover the work initially thought to require representation. I conclude that EV is best altered to be a non-representational theory of perception.

Keywords

Acknowledgements

1 Introduction

Whether perception involves the manipulation of representational content is currently heavily debated. A classic Marrian view sees perception as built up step-by-step from a 2D retinal

image into the full-blown 3D world we find ourselves in. Each individual step involves a separately produced representational image in which the brain detects relevant features to produce a more detailed, experientially rich picture (1982). There has been a large variety of theories that all regard representational content as central to perception (see Pitt 2017 for an overview).[1] Opposed to this, there is a tradition that rejects the idea that representational content plays any role in perception. Typically, non-representational approaches emphasize embodiment and environmental interaction, stressing that action and perception are intimately connected (Gibson 1979; Varela et al. 1991; Di Paolo et al. 2017; Hutto and Myin 2017).[2]

Recently, many theorists have tried to develop an approach that avoids either of these extremes. A popular strand here is prediction error minimization theory, which attempts to incorporate the intimate connections between action and perception into a fully representational framework (Hohwy, 2013; Clark, 2016). The Embedded View (EV) furthers this compromissory agenda by steering a middle course in the so-called 'representation wars'[3] (Orlandi 2014; see also 2012; 2013). Orlandi argues that perceptual processing is non-representational, but its perceptual products are representational. Yet EV has seen criticism from both sides for its limited use of representations. Where representationalists argue EV incorporates too few representations, non-representationalists argue there are too many.

In this paper, I argue that, to respond to the criticism from both representationalists and non-representationalists, EV is best off going fully non-representational. I will first set the stage by describing EV. Proceeding, I will then analyze the debate to show the perceived shortcomings from EV. I distinguish four such objections: 1) the argument that Orlandi leaves out the representational middleman; 2) the argument against *explanatory exclusionism*, according to which Orlandi seems to rule out representational readings prematurely; 3) the argument for the explanatory benefits of representations, according to which there may be benefits to invoking representations that have been overlooked, and 4) an inconsistency charge in EV's ascription of representational status. I argue that the first three objections can

---

[1] The textbook interpretation of Marr is representational, but it is up for debate whether this is the only interpretation (Orlandi, 2014, note 3).
[2] In both instances there is an incredible variation of theories, and though some certainly find themselves more in one camp than another, a certain degree of fluidity between these camps exists.
[3] I think the name 'representation wars' stems from Clark (2015).

be deflected, and so pose no threat to EV. Yet the fourth objection, the inconsistency argument, is successful and thus problematic for EV. Despite this, I argue that the tools and the solution are found in Orlandi's own work (2014; and 2012 respectively), hence EV can easily be amended to incorporate this criticism. Indeed, I will argue that, by taking the explanatory tools available to EV more seriously, the range of applicability of EV can be shown to cover those aspects of perception that Orlandi thinks require representation. If we buy into this solution, EV ceases to be the compromissory view Orlandi (2014) put forward and becomes wholly non-representational. In turn, it benefits in having resolved the raised objections.

2 The Embedded View Explained

In our processing of visual stimuli, we are sensitive to statistical regularities in the world (Mole and Zhao 2016, p. 366). This means that certain regularities that we have encountered before influence the way we engage with particular stimuli; particular changes in light intensity may for example indicate edges. Traditionally, it is thought that the information of the statistical regularities needs to be encoded in our brain to explain this sensitivity. For example, Marr argues that the brain encodes very specific statistical information regarding the occurrences of edges (Marr and Hildreth 1980, p. 202; Marr 1982). More recent proposals have reimagined the manner in which this encoding works. Still, the necessity of such encoding remains.[4] For example, currently popular prediction error minimization theories state that any statistical regularity we may be sensitive to must be, in some way or some form, encoded in causal-probabilistic models of the world (Gładziejewski 2016; see also Hohwy 2013; Clark 2016).

EV is not at odds with this sensitivity to statistical regularities. However, it challenges the encoding requirement. Orlandi argues for EV by way of an inference to the best explanation. Here she relies on a widely shared assumption that representations need only be invoked "when appeal to environmental and other conditions fails, or is not illuminating" (Orlandi, 2014 p. 7, see also Note 15; Pylyshyn 1984, p. 26; Brooks 1991; Burge 2010; Fodor 1987; Segal 1989; Van Gelder 1995). The tactic is thus to provide a non-representational

---

[4] Predictive processing (see Hohwy 2013 and Clark 2016 for some good introductory monographs) is one such theory, in which the rules are thought to be represented in the form of generative models. Some recent takes on the theory however, advocate a non-representational reading (Kirchhoff and Robertson 2018; Bruineberg et al. 2018; Hutto 2017)

alternative to standard representational explanations with at least equal explanatory power. With this in place, invoking representations is unnecessary. According to Orlandi, our visual system *relies* on the world's statistical regularities themselves rather than needing to encode those regularities internally. "Relying on a fact", then, "means acting in accordance with a fact, and with a corresponding principle, without representing either" (Orlandi 2014, p. 3; Orlandi 2012, 2013; see also Ramsey 2007). In which case, perceivers can rely on the facts in the world without having to representing those facts. Orlandi states:

> We can imagine (...) a connectionist network trained to detect something by being repeatedly exposed to it. Such training causes the network to display characteristic patterns of activation where low-level configurations are associated with high-level ones and *vice versa*. We can then think of a high-level state as 'checking' the pattern of activation at the lower, sensory level where this ultimately just means that the high-level state activates in a way that is more or less compatible with the lower-level pattern of activation. (Orlandi 2014, p. 88)

The idea here is that a connectionist network can become sensitive to a particular regularity. In more detail, the high-level states of the network become sensitive to the regularities at the lower, sensory level. This constitutes a certain bias in the system such that the higher level states will track the patterns of activation of lower level states, 'acting' in accordance with similar patterns, thus tracking, for example, edges. This, it should be noted, is a 'mere' tracking activity, and thus need not be thought of in representational terms. Indeed, there is nothing intrinsically representational about the reliable co-activation or covariation of two (parts of) systems (Hutto and Myin 2013). Further, Orlandi says:

> The transition between these two tracking states *is not regulated by an encoded assumption*. We can think of it as a mere associative transition. What is relevant is that there is a strong connection between states that track discontinuities in intensity and states that track edges. (2014 p. 153, emphasis added)

What this means is that the lower level states track the discontinuities in intensity, whereas the higher level states activate in accordance with particular patterns of lower level states, whilst tracking edges. Crucially, this associative transition does not require any encoding of the statistical regularities. "[T]he facts (...) explain why the device does what it does without being represented by the device" (Orlandi, 2012 p. 561). In explaining a particular behaviour, then, the facts themselves factor into our explanations, not representations of those facts.

Thus, we could certainly describe this particular network as *representing* the statistical regularities that it is sensitive to. *Prima facie*, at least, it is not conceptually incoherent. Adding this representational gloss, however, confers no explanatory advantage. Moreover, it comes at some cost. For it would ascribe more labor to our cognitive architecture than necessary. And it would lose some of the original ontological parsimony (Orlandi 2013, p. 739). Recall that, as we saw above, "representational notions are only needed when organic, functional, and environmental conditions are insufficient" (Orlandi 2014, p. 54). Here however, they suffice just fine.

One of the key issues with representational content is its origin (Hutto and Myin 2013; 2017). Broadly, the issue is as follows. Though representations have been defined in a multitude of ways, a core feature that distinguishes a representational relation from a merely covariational relation, is that they have content: they describe their target in a way that it may not be so (Travis 2004).[56] The hard problem of content lies in finding a naturalistic origin for content: a task that is, as of yet, unfulfilled (Hutto and Myin 2013, 2017).

Ideally, an alternative proposal, like EV, should avoid the problems it is supposed to resolve. If so, we can ask: what is the origin of the reliance relation? Orlandi speaks of the visual system as being *wired*, having features that "developed, and continue to develop under evolutionary and environmental pressure" (2014 p. 3). She argues that it would be perhaps more surprising if evolution had *not* wired us to be sensitive to, say, the particular changes in light intensity corresponding to edges (Orlandi 2014, p. 154).[7] However, while it is likely that the wiring of our visual system is a consequence of evolutionary pressures, it seems more is needed if we are to account for the full range of statistical regularities we can become sensitive to. For example, a lot of these sensitivities emerge only later in life, and some at timescales that evolutionary adaptation could only dream of! How then does EV account for this malleability of the sensitivities of the visual system? Or in other words: how do we account for the features that develop not only under evolutionary, but also under environmental pressures?

---

[5] This is not to say that content is all it takes for a relation to be representational, only that it is minimally necessary.

[6] This excludes deflationary notions of representation. In this paper, representations are minimally required to have content, as to not over-generate representations or ascribe them too liberally (see also Ramsey, 2007).

[7] Similar arguments can be found in the work of Järvilehto (1998) and Warren (2005).

Typically a wired system is conceived of as *hardwired* and incapable of adapting to the environment (Orlandi 2014, p. 146; see for example Rock 1983, p. 277). This need not be the case, however. Remember that a connectionist network is wired in the sense EV deems the visual processing system to be. Orlandi states:

> Networks are wired systems capable of adjusting to context. In the learning period, they come to process information in a new way without having any localizable structures that encode instructions. They come to be sensitive to the presence of mine echoes by simply adjusting their connections. Similarly, pigeons may stop pecking on a key as a result of extinction or counterconditioning without following any rules. (Orlandi 2014, p. 147)

She continues:

> The fact that they are adaptable in this way does not mean that they are rule-following. Rewiring due to inverting lenses, or to reinforcement and reward, amounts to changing the way a certain function is performed through a physical intervention. It constitutes an instance of malleability without being an instance of following (new) principles. (Orlandi 2014, p. 147)

This means that adapting a function to better fit the current environment does not entail an exchange of less fit rules for more adaptive rules. It does not entail any form of rule-following whatsoever. A functional change can do the trick of malleability just as well.

Above I have given a brief overview of how Orlandi envisions visual processing to be possible non-representationally, whilst also retaining sensitivity to statistical regularities. Orlandi also covers typical issues for non-representational accounts of perception, such as misperception, illusions, stability and constancy (Orlandi 2014, ch. 4) and multi-stability (Orlandi 2012). However, although Orlandi argues strongly for non-representational visual processing, she does think of the visual product as being representational. It is this concession to representationalism that puts her in the crossfire of the 'representation wars'.

Yet, one would think that if representations are not needed to account for the wide range of cases covered above, then what still remains unexplained? I will answer this question by identifying what Orlandi deems necessary for something to count as a representation. She discerns three distinct features of a representation: 1) standing in, 2), informing, and 3) guiding performance (Orlandi 2014, p. 9). Thus, representations need to "*stand in* for something other than themselves and in so doing they act as *mediators* within a system or a process" (p. 9). Representations also need to inform and although she concedes that the term 'information' is heavily disputed in representation contexts, Orlandi thinks that

representations inform, say or 'mean' something about the target (p. 9). Finally, representations need to be "causally active", "guiding the behavior of a system or organism, described in certain general terms, by standing in and informing" (p. 10).

My focus here will be on the first feature: standing in. For us to legitimately describe a system as containing representations that *stand in* for something, these representations need to be *decouplable* from what initially caused the representation. This means that the state should stand for its representational target (that which it represents) even in the *absence* of that target. If we take a portrait picture to be a representation, for example, the picture stands for the portrayed person even in the absence of the person. Of particular interest here is the term absence. For Orlandi, anything that does not directly impinge on our senses, should be deemed absent.

She uses two cases to exemplify this. First, when a rabbit is looking at a coyote, one may think that the coyote's reflected light is currently directly impinging on one's senses. Orlandi concedes that the *front* of the coyote is, but not a full coyote. For example, the coyote may even be partially occluded by a rock it is standing behind. Regardless, the rabbit reacts to the presence of a full coyote (Orlandi 2014, p. 127). Second, Orlandi claims that the same phenomenon occurs whenever we see, say, a cow. For although only the image of the frontside of a cow impinges on our senses, we still react to it, judge it and take it to be a full-blown cow. She argues that "this *taking* constitutes an early representational capacity. It is a capacity that involves an abstraction" (Orlandi 2014, p. 150). Representational capacities are thus twofold. The perspectively invisible (and thus absent) backsides of the objects we encounter require representation, since these aspects are abstracted from the direct stimuli. Further, to *take* the light reflected off a cow *as* a full-blown 3D cow also requires representation. The representation meets all three conditions: it *stands in* for the currently invisible backside, it *informs* about the cow, and it *guides our actions* in the sense that we judge it to be a full-blown cow (and, under this construction, this judgment guides us so that we may approach it to pet it etc.).

In sum, EV pulls visual processing apart from the visual percepts it produces. If EV is right, rather than needing to encode particular rules, we instead rely on the facts in the world. There is thus no need for the visual system to represent the statistical regularities it relies on. However, in the visual product there is the need to invoke representations. For despite being only exposed to a single side of an object at a time, we nonetheless take the stimulus to come

from a full-blown object. Further, we take the light reflected off of particular objects to be caused by objects of particular types. If so, then the visual product does require representational capacities.

3 The Embedded View Under Fire

Not loyal to either the representational or non-representational sides, EV has attracted criticism from both camps, creating an interesting bipartisan agreement that halfway is no way at all. The agreement extends beyond this point however. I distinguish four distinct objections put forward by opponents of EV, the first three explicitly in favour of a more thoroughly representationalist approach, the final amenable to theorists across the representational-non-representational divide. The first objection is that of the relevant middleman, and is specific to Bayesian brain theories. In Bayesian brain theories, priors are often identified with representations (Hohwy, 2013). The argument is that priors requiring further explanation does not mean they do not have any explanatory power so that they need to be excluded *tout court*. The second objection is against explanatory exclusionism: this argues against the notion that, if there are non-representationalist explanations, then their representationalist counterparts become obsolete. The third objection states that representations hold explanatory benefits for our scientific endeavours, and thus need to be included. The final, unifying objection is that Orlandi's application of representations is inconsistent. Below I will discuss these objections in more detail.

3.1 The relevant middleman

Rescorla argues that Orlandi ignores the middleman. He says: "[i]f we explain X by citing Y and then explain Y by citing Z, it hardly follows that Y is explanatorily irrelevant". He explains this with an example: "a physicist can at least partially explain the acceleration of some planet by citing the planet's mass and the net force acting on the planet, even if she does not explain why that net force arises" (Rescorla 2015). Put differently, even if Y requires further explanation, it may still serve in explaining X. Y's requiring further explanation does not diminish its explanatory power.

To be precise, Rescorla's point concerns Bayesian modeling theories of perception and cognition. An issue Orlandi brings up is that it is unclear where the priors, that is, the brain's anticipation of what it is about to encounter, come from (Orlandi 2014, p. 91-93). An

elaborate discussion of the issue of the origin of priors in Bayesian theories of perception is outside of the scope of this paper, but it is roughly as follows. According to Bayesian theories of perception, the brain attempts to reduce the uncertainty concerning the stimuli entering the system (X in Rescorla's example). An issue here is that, for any one stimulus, there is conceptually an infinite amount of possible causes (the underdetermination problem). To limit the hypothesis space, then, priors are introduced. These are (basic) assumptions about the way the world works that restrict the variety of possible causes of the stimuli (Y in Rescorla's example). Standardly, Bayesian brain theories are considered representational. The representations are thought to be instantiated in the form of priors (Hohwy 2013; Gładziejewski 2016).[8]

In this sense, the question about the origin of priors in Bayesian brain theories, though technically separate from, can be cast as a particular instantiation of the question about the origin of representations, specifically in the context of Bayesian brain theories. Indeed, if priors are thought to be representational in Bayesian brain theories, then, in that context, a discussion of the origin of priors in Bayesian brain theories is also a discussion of the origin of representations in Bayesian brain theories. Rescorla thinks that, due to this controversy, Orlandi excludes the priors (Y) because her environmental account (Z) would be needed to explain our visual system (X) anyway. Rescorla treats it as a form of cutting out the middleman.[9] Instead, he argues that the explanatory value of Y is not impacted by it requiring further explanation. Even more, he argues that "one can explain X by citing Y, without in turn explaining Y" (Rescorla, 2015). Representations thus need not be explained for them to be explanatorily useful, according to Rescorla. Put differently, even if the priors (and thus the representations) are to be explained further with Orlandi's embedded explanation, the priors themselves still have explanatory value.

3.2 Against explanatory exclusionism

EV is also accused of committing a form of *explanatory exclusionism* (Polger 2015, p. 345). Polger considers Orlandi to be choosing between competing explanations. He wonders: "[w]hy should we embrace an explanatory exclusion principle to the effect that the availability of functional or mechanistic explanations rules out the legitimacy of

---

[8] Discussion of Bayesian brain theories is outside the scope of this paper. Hohwy (2013) and Clark (2016) are two excellent monographs for an introduction into the most popular form of Bayesian brain theories.

[9] As we shall see below, I think this portrayal of Orlandi's position is incorrect.

representational explanations"? (Polger 2015, p. 345). Polger ponders rhetorically why representational explanations should be less preferable to functional or mechanistic explanations. Non-representational explanations are not, Polger thinks, by default preferable to representational explanations. The idea is that Orlandi would have to do more to show that her non-representational explanation is preferable to competing representational explanations. Indeed, there is no good reason to throw representational explanations out simply because there are non-representational accounts.

3.3 Representations with benefits

Another objection against EV is that invoking representations in our explanations of perceptual phenomena offers explanatory benefits. Leaving our explanation to merely "organic, functional, and environmental" conditions, so it is claimed, is insufficient. The addition of representations is supposed to make up for the difference. Rescorla argues that "Orlandi's analysis entrains a significant loss in explanatory power. If we adopt a realist perspective on priors", he thinks, "we can explain why various changes in environmental conditions yield various changes in the mapping from sensory stimulations to percepts: namely, because the priors change a certain way" (Rescorla 2015).

This again concerns Bayesian modeling theories of perception and cognition, where the priors refer to the brain's anticipation of what it is about to encounter. Orlandi *de-representationalized* this notion, thinking of such anticipation in functional terms, the training of a neural network. Rescorla then argues that priors add explanatory utility by exchanging a step in Orlandi's story. Rather than the environment influencing the neural network's training, it influences the priors. Both of these are thought to be the explanation for the change in our percepts in their respective theories. Rescorla thinks that this exchange, the realist position on the priors, adds explanatory value.

Polger (2015) puts forward a similar point. He thinks that "[r]epresentationalist explanations might be justified by their utility for explaining how those mechanisms of biases and constraints came to be, or what those systems have in common with other human perceptual systems and with physiologically distinctive visual systems in other creatures" (Polger 2015, p. 345). The addition of representations is thought to help both evolutionary and environmental explanations of the functional features described by Orlandi. The idea seems to be that, if we posit representational contents as that which is present in all our

perceptual channels (both our vision and hearing can be imagined as being representational) and also for such perceptual channels in non-human creatures, the shared representational contents can serve as a starting point for explaining what we have in common.

3.4 Inconsistency

The third and final objection has managed to unify both representationalists and non-representationalists alike. It is an inconsistency objection concerning Orlandi's ascriptions of representations in EV. As we have seen in Section 2, Orlandi understands visual processing as non-representational, but the perceptual product does require representations. However, if we were to consistently apply Orlandi's conditions for what counts as a representation, so it is objected, then a lot more features she has termed non-representational, would in fact count as representations. She thus applies her own conditions inconsistently. Put differently, her conditions for representational content actually lead to a proliferation of such representations.

Recall that Orlandi argues that representations are used in abstracting the image of a cow from mere electromagnetic radiation that impinges on our senses. Polger argues that

> "if this basic sort of abstraction is all that is required to count as a representation, then contrary to Orlandi, representation occurs at the earliest stages of visual processing … It seems to me that Orlandi has made it too easy for early visual states to count as representations." (Polger 2015, p. 344)

Despite Orlandi's care in the ascription of representations in visual processing, Polger argues that this frugality is not mirrored in explaining the visual product. Mole and Zhao (2016) argue along the same lines. In particular, they discuss one experiment that displays features of our visual system that, if we are to follow Orlandi's conditions, would be representational. Yet, when analyzed, all of the examples they offer are exactly of the sort that EV is supposed to de-representationalize, that is, they are examples in which the subject becomes sensitive to a newfound statistical regularity. Mole and Zhao portray these as counterexamples she may not have been aware of.[10] However, I claim these examples may be better understood as

---

[10] Mole and Zhao (2016) argue roughly that the subjects in the experiments have become sensitive to newfound statistical regularities in the environment, so that they could not have been hardwired. With a hidden assumption that systems can only become sensitive to new statistical regularities with representations, they conclude that representations thus *must* in some cases be used in visual processing. This bypasses Orlandi's explanation of malleability.

demonstrating the shortcomings in Orlandi's application of the stated conditions for representational content.

> Hutto and Myin raise a similar point. They state:

> Consider that Orlandi argues, as does REC[11], that positing sensitivities to statistical patterns of variation is all that is needed in order to explain perceptual processing and that there is no need or advantage in positing contentful representations in such explanations. Yet such statistical patterns are just as unavailable to directly inform perceptual processing as are the backsides of objects are to inform the final products of visual experiences. (Hutto and Myin 2017, p. 162)

Absence, or unavailability, is a central term in Orlandi's usage of representations. This is intended to cover the *decouplability* requirement as well as the *standing in* requirement for representational content. What is present and currently available in the environment, need not be represented internally. The 'original' may be used without making a 'copy'. Yet if this is all that it takes, Hutto and Myin argue, then what are we to make of the statistical regularities themselves? Surely, if the image of a cow requires representational abstraction from electromagnetic radiation, then how do the statistical regularities present themselves directly? Statistical regularities, of course, are never witnessed directly. Hutto and Myin thus argue that the representational content of statistical regularities stands or falls together with that of the backsides of objects. This too points out Orlandi's inconsistency in the ascription of representations. In sum, Orlandi's inconsistency in the ascription of representational content is consistently picked out by people from a wide variety of theoretical colors. As we will see below, it also is the most pertinent issue for EV.

4 The Embedded Fire Extinguisher and the Ember

In this section I will give a response to the above objections from the standpoint of EV. In the cases of answering the objections mentioned in Sections 3.1, 3.2, and 3.3, I will attempt to offer a rebuttal. I argue that the inconsistency issue of Section 3.4 proves to be trickier, although the explanatory tools available to EV provide the solution.

---

[11] REC stands for Radically Enactive Cognition, and is a non-representational approach to cognition proposed by Hutto and Myin (2017; see also 2013).

4.1 The controversial middleman

Rescorla objects that Orlandi cuts out a relevant explanatory mediator. Specifically, he argues we can explain X by citing Y, which in turn can be explained by citing Z. The appearance and necessity of Z, he argues, need not impair the explanatory utility of Y. He makes an analogy with a physicist, who explains the movement of celestial bodies (X) with net force (Y). Net force itself requires further explanation, perhaps by citing the mass of all physical bodies in the system (in the form of Z). Nonetheless, it can be used to explain X. Though I think the example is solid, it need not generalize to other cases. In particular, it need not apply to the use of representations to explain perceptual phenomena.

There are two relevant differences here. 1) Orlandi could be understood as cutting out Rescorla's imagined middleman, that is, if we can explain X by citing Z directly, then we do not need to first explain X by citing Y before finally turning to Z. For example, if we can explain perception by citing functional, organic mechanisms, then we need not explain perception by first citing priors, which can then be explained by citing functional mechanisms. If this is correct, then contrary to Rescorla, it is not the case that Y becomes explanatorily irrelevant because it is explained by Z. For the functional mechanisms do not explain the priors, but rather the phenomena themselves. In which case, Y and Z are thus *competing* explanations both intended to explain X. Y's representational explanation has thus never started her job as a middleman in EV; she thus could not have been fired as Rescorla seems to argue.

2) Recall the widely shared assumption that representational explanations are only necessary if we lack simpler, ontologically and computationally more parsimonious, explanations that we have seen in Section 2 (Orlandi 2013, 2014; Pylyshyn 1984, p. 26; Brooks 1991; Burge 2010; Fodor 1987; Segal 1989; Van Gelder 1995). It is thus not thought to be conceptually incoherent, or otherwise impossible for there to be a representational explanation in tandem with a functional, organic or environmental explanation. Instead, it is thought to be *unnecessary* for us to appeal to a term as ontologically and cognitively heavy as representations if we can show simpler explanations are sufficient. These considerations follow from Orlandi's inference to the best explanation strategy. She does not aim to show that representational explanations are incoherent or incompetent. Instead, she offers a better explanation, which at least in part relies on being a simpler explanation.

Further, at least to my knowledge, net force is an uncontroversial term in physics. This status allows it to be useful in explaining celestial movements. However, if net force were to turn out to be a fundamentally questionable concept, that is, one for which there is no naturalistic foundation (as of yet), then using it to explain celestial movement would not be warranted. Consider using the touch of Zeus to explain celestial movement. At least one of the reasons this explanation is not accepted is because the touch of Zeus has no naturalistic foundation. I propose that the same goes for representational contents. Consider that the project to naturalize content has so far proved unsuccessful, and increasingly more people are becoming pessimistic towards a resolution (Rosenberg 2015 and Hutto and Myin 2013; 2017 among others). Rescorla acknowledges this difficulty (2015), but fails to see why the lack of even a prospect of a naturalistic explanation makes the use of representational explanations inherently unappealing.

Note, however, that Rescorla's point only indirectly concerns representations. This is because the priors he discusses are typically conceptualized as representations in Bayesian brain theories (Hohwy 2013; Gładziejewski 2016). To recapitulate, to limit the hypothesis space for the brain's estimate of what stimuli it is about to encounter, priors are introduced. Here priors are continuously updated assumptions about the structure of the world. It is still unclear however, where priors originate from. Nonetheless, he argues, priors need not be explained in order to have explanatory force. Yet this seems highly problematic. We explain how the brain reduces uncertainty by reference to priors, of which the origins are unknown. It seems like we explain a mystery by a further mystery, merely moving the goalpost without actually cashing out an explanation. Indeed, it echoes the problem of invoking representations without explaining them. In both cases we attempt to use unexplained, quite possibly unexplainable concepts to explain the phenomena.[12]

4.2 Explanatory frugality

Polger (2015) wonders why we should exclude representational explanations merely due to the availability of functional or mechanistic alternatives. On the surface, this exclusion seems to presuppose the superiority of non-representational contenders from the offset. But remember that, as of yet, the naturalization of representational contents remains unsettled.

---

[12] See Clark (2016) for a proposed solution to the origin of priors, and see Hutto (2017) and Hutto and Myin (2017) for an objection to this.

Further, even among those that use representational explanations it is commonly thought that the ascription of representations is only necessary in the absence of simpler alternatives (Orlandi 2013, 2014; Pylyshyn 1984, p. 26; Burge 2010; Fodor 1987; Segal 1989). The only way out, some think, is to simply accept content as a basic metaphysical particle of existence (Shapiro 2014 p. 218). If non-representational explanations are explanatorily equal to representational explanations, then we should prefer the metaphysically lighter option (using Occam's time-tested razor), much like we do when rejecting the touch of Zeus as an explanation. Until the naturalization of content becomes a more promising area of research, equally or more powerful non-representational explanations should be preferred over representational ones. In this particular sense, non-representational explanations, otherwise equal, are superior to representational explanations. This, again, does not mean it is incoherent to add representations to a non-representational explanation. It is simply that doing so does not add explanatory value (as the non-representational explanation is at least equal in explanatory power to the representational explanation). Moreover, it unnecessarily adds a metaphysically unclear and cognitively heavy component to an otherwise simpler explanation. In short, it is not incoherent. But, it is not wise either.

4.3 Begging questions on explanatory benefits

Rescorla argued that we could gain explanatory benefits by adopting a realist stance on priors. This would allow us to explain *why* our percepts have changed. A realist's priors are, in a sense, quantifications of the neural network-like activity described by Orlandi (2014). Not only does the neural network change its wirings, but these wirings represent numerical statistical values distributed over a variety of possible outcomes. But does this offer more of an explanation than the functional approach? It is easy to see that EV's wiring explanation does the same work, by adapting Rescorla's sentence: "we can explain why various changes in environmental conditions yield various changes in the mapping from sensory stimulations to percepts: namely, because the [wirings] change a certain way" (2015, edited by the author).[13]

---

[13] This is not to discredit the empirical research programme of Bayesian brain theories of perception and cognition, which seems promising (see Hohwy, 2013; Clark, 2016). This is a point about how to interpret the priors used in this research. There is a difference between using Bayesian models as scientists to model and predict behaviour of animals, and a realist position concerning the models used as existing in the animal's head (or body). Rescorla (2015) favors an interpretation of the scientific literature that involves a realist position of these priors. Here I argue that doing so does not add any explanatory value.

There are two reasons why, in itself, a realist position on priors does not add any explanatory benefit. First, it is unclear what quantification of a biological process adds beyond numbers. These numbers may feature well in models of biofunctional mechanisms. But there is an important difference between using these numbers and models to describe or predict animal behavior and actually *ascribing* these numbers and models to the animal itself. Doing the latter prompts the question of how these models are instantiated in the brain, in what way the numerical probabilistic values are 'real', and even whether this requires a realist position about, say, abstract objects (and the problems that ensue). None of these questions are easily answered.

Second, even if we were to allow that there is an explanatory benefit, it is unclear how introducing a feature that is in itself fundamentally contested would help our explanatory efforts. This draws back on our earlier 'touch of Zeus'-style explanations. Until we have, at the very least, a promise of how the 'touch of Zeus' can be naturalized, naturalistic but otherwise equal explanations should take precedence. The same should go for representational contents.

Polger argues that representations can explain how our functional mechanisms came to be, and expose the similarities between similar systems within human organisms (think of different perceptual channels), but also similar perceptual channels of different organisms (such as a dog's vision). Yet it is unclear why representational explanations would do a better job at explaining how particular functions came to be than organic and functional explanations: the type more typically used in evolution-theoretic explanations. More importantly however, it seems circular. We first posit that our and non-humans' sensory channels share the use of representations, before using this posit to explain those similarities. Finally, we claim this supports a realist position of our earlier theoretical posits. With this we have come full circle, and viciously so.

4.4 The inconsistency and the non-representational solution

Polger (2015), Mole and Zhao (2016), Myin (2016) and Hutto and Myin (2017) have all argued that Orlandi's ascription of representations is inconsistent. Each of them has argued that at least one other feature of the visual processing system fulfills EV's conditions. In this section, I will focus in particular on Hutto and Myin's (2017) suggestion that the backsides of

objects and the perception of abstract properties such as COWness warrant the same representational status as the statistical regularities, but the solution extends to other cases.

Recall that Orlandi identifies three conditions for anything to be a representation: 1) standing in, 2), informing, and 3) guiding performance (Orlandi 2014, p. 9). A possible objection to Hutto and Myin's suggestion is then that our 'perception' of backsides of objects legitimately covers condition (1), whereas getting attuned to statistical regularities does not. The backsides are (to some extent) present in our visual percept (Orlandi 2014, p. 125). This requires, according to Orlandi, a representation to stand in for a backside that is not technically present, but does appear in our percept. The statistical regularities however, do not appear in the same way. There is only the brain's interactional attunement to regularities encountered in the world. There is thus nothing internally that 'stands in' for them. This would cause the statistical regularities to not meet condition (1). The backsides of objects do meet that condition.

The issue is whether this distinction between statistical regularities and the backsides of objects as being non-representational and representational respectively is justified. Both the statistical regularities and the backsides of objects are stimulus-free; they are absent in the sense discussed in Section 2. At this point, the ascription of representational status seems stipulative. Orlandi concedes that representational explanations for sensitivity to statistical regularities are possible. The reason why this feature of visual processing is de-representationalized, is because she provides a functional explanation that does not require representations. This is a metaphysically and cognitively less taxing way to explain the same phenomena, and should thus be preferred, given all that has been said above. Statistical regularities then do not have a stand-in, thus do not meet condition (1), and so are non-representational.

Orlandi (2014) does not discuss non-representational options for 'seeing the backsides', and thus makes it seem like the representational interpretation is the only route (perhaps much like how the sensitivity to statistical regularities may have seemed before her account). This, in the context of Orlandi (2014), makes it appear that she consistently applied her own conditions of favouring non-representational explanations when present. Interestingly, Orlandi (2012) has previously offered a non-representational solution to this issue. Here, our interaction with the rigidity of objects (and those objects thus having backsides) is explained through the same mechanism as other statistical regularities in the

world, like the co-occurrence of changes in light intensity and edges. After all, there is a robust statistical regularity so that the patterns of light pertaining to objects co-occur with these objects' rigidity. Indeed, it is a regular occurrence that the objects we encounter are 3D objects with front- and backsides. She says:

> "Objects in the world are typically rigid and the visual system can rely on this fact to produce a representation of objects directly from the retinal stimulation. Because the causes of retinal stimuli are typically rigid, we end up seeing the world that way" (Orlandi 2012, p. 560)

The rigidity of objects is thus explained by way of the exact same non-representational process as, for example, edges are. We do not have to invoke internal representations of the backsides of objects to explain our perception of rigid objects. Instead, we can rely on the robust statistical regularity that the objects we encounter are typically rigid. Any object that is rigid has a backside (even if they are not always the backsides that we have become attuned to, like a realistic cardboard cutout of a cow)[14], so if we see rigid objects, we 'see' the backsides to the extent necessary. Positing that we need representational content for the backsides of objects is superfluous.[15]

The further representational capacity concerns the abstraction involved in *taking* a particular pattern of electromagnetic radiation impinging on our senses as an image of a cow. It is now not difficult to see how her own proposed mechanism can be employed to explain this capacity to some extent. There is an important distinction that needs to be explicated here. I suggest that what is captured in the *taking* capacity by Orlandi is two distinct capacities. In saying that we *take* a particular pattern to be a cow, 1) there is a judgment that goes beyond perception in itself in which we *judge* the pattern to be a cow, and 2) there is the capacity to, by being exposed to that particular pattern, interact with the cow in the world we encounter it. (1), the judgment, is to be read in the same way we *take* a sarcastic comment to mean the opposite of what it explicitly says. This *taking* capacity is a lingual, socioculturally developed capacity that goes beyond the actual hearing of the comment itself. As such, this

---

[14] The possibility to make mistakes is not reason to doubt this explanation. Instead, it speaks in its favor. After all, it is because we are wired to see rigid objects when encountering particular patterns of stimulation, that the system sometimes fails and we become susceptible to illusions. Complete exposition of this is outside the scope of this paper, but see Orlandi (2014, ch. 4) for a more elaborate explanation, as well as similar non-representational, embedded explanations of misperception in general and multi-stability.

[15] See also Di Paolo et al. (2017) that explain the 'seeing of backsides' in terms of interactional relevance. We approach a particular object as having a backside when the backside is relevant to our current activity. A keyboard is not interacted with as having a backside currently not visible to the eye, unless we, say, pick it up. A teacup on the other hand, typically affords picking up and its backside will thus be 'seen' more regularly.

aspect of the *taking* capacity is not an explanandum in the category of perception, and with that it is outside the scope of EV, and this paper. Roughly, the idea is that language, or content-involving practice, is an extension of action in a sociocultural environment (further exposure of this idea is outside the scope of this paper, but see Moyal-Sharrock 2019; Hutto and Satne 2015; Hutto and Myin 2017 for positions along these lines).

Part (2) of the *taking* capacity, the capacity to interact with a cow in the world, by being exposed to a particular pattern of stimulation *is* part of the explanandum of EV, and can also be explained. The perception of a cow relies on the statistical regularities involved with the particular patterns of stimulation caused by cows, and the presence of cows. The perceptual system can become sensitive to these statistical regularities, and may rely on them in our perception of cows. These sensitivities go on to influence our interactional behaviour and perception much in the same way our sensitivity to edges does. Indeed, this explains our perception of cows in the same way the statistical regularities and interactional history explain edge detection or our perception of rigid objects. A primary difference is that a sensitivity to edges can plausibly be thought of as having developed by evolutionary pressures, whereas a sensitivity to cows is more likely to have developed ontogenetically due to environmental pressures. This allows EV to circumvent the inconsistency objection raised in Section 3.3. Given Orlandi's idea that representational explanations should only be instantiated if functional, organic or environmental explanations are insufficient, EV should thus adopt this non-representational approach.

What, then, is left of the compromissory position EV intended to take in the representation wars as put forward in Orlandi (2014)? EV started out as a view that held onto non-representationalism for visual processing, but argued for a representational view of the visual product. The solution for the inconsistency problem offered above de-representationalizes the two central features of vision that Orlandi thought required the ascription of representational capacities to be explained. As a consequence, EV has now gone fully non-representational and loses its status of a representationally compromissory view. Indeed, the view is now much closer to a fully non-representational view as defended by the likes of Hutto and Myin (2017).

Buying into this solution comes with a few further advantages. EV in its original form would still have to answer to the hard problem of content for visual products, that is, the problem of finding a naturalistic grounding for representational content, commonly agreed to

be minimally necessary for representation (Hutto and Myin 2013). Non-representational EV circumvents this hard problem of content by not positing representations in either perceptual processing nor perceptual products. This further increases the ascribed frugality of the brain's role as it need not trade in cognitively heavy representations. It also has ontological frugality. After all, the non-naturalistic status of representational contents is unlikely to change anytime soon (Rosenberg 2015; Shapiro 2014, p. 218; Hutto and Myin 2013, 2017). Further, not trading anymore in such unexplained posits allows this to remain metaphysically parsimonious.

Finally, it is worth noting that no new additions or similar *ad hoc* reasonings are added. Orlandi's body of work on EV already held the tools needed to fix the issue of inconsistency, while also broadening the explanatory range of its primary feature, namely, the outsourcing of cognitive tasks onto the environment.

5 Conclusion

In this paper I have assessed the debate around Orlandi's proposed Embedded View of vision. This view steers a middle course between representationalist and non-representationalist views by arguing that visual processing is non-representational, whereas its products are representational. Both representationalists and non-representationalists alike have challenged this view. I described four such challenges. I then offered rebuttals of three of those challenges, while also acknowledging that the fourth raises a significant problem. This problem concerned the inconsistency in Orlandi's ascription of representational contents. However, I showed that Orlandi has the tools necessary to fix this problem (and perhaps even part of the solution (see Orlandi 2012)). I conclude then that EV is best interpreted non-representational all the way down. As such, EV becomes a metaphysically and cognitively more frugal view, which, when placed within the context of the representation wars, is to its advantage.

Bibliography

Anderson, M. L. (2017) Of Bayes and Bullets: An Embodied, Situated, Targeting-Based Account of Predictive Processing. In Metzinger, T. and Wiese, W. (Eds.). *Philosophy and Predictive Processing: 4*. MIND Group

Brooks, R. (1991). Intelligence without representation. *Artificial Intelligence, 47*, 139–159.

Bruineberg, J., Kiverstein, J., Rietveld, E., (2018) The anticipating brain is not a scientist: the free-energy principle from an ecological-enactive perspective. *Synthese, 195*(5), 2417-2444

Burge, T. (2010). *Origins of objectivity*. Oxford University Press.

Clark, A. (2015). Predicting peace: The end of the representation wars. In Metzinger, T. and Windt, J. (Eds.). *Philosophy and Predictive Processing: 7*. MIND Group, 1–7

Clark, A. (2016) *Surfing Uncertainty: Prediction, Action and the Embodied Mind*. Oxford University Press.

Di Paolo, E., Buhrmann, T., Barandiaran, X., (2017) *Sensorimotor Life: An enactive proposal*, Oxford University Press.

Fodor, J. A. (1987). *Psychosemantics*. MIT Press.

Gibson, J. J. 1979. *The Ecological Approach to Visual Perception*. Houghton Mifflin.

Gładziejewski, P. (2016) Predictive Coding and Representationalism. *Synthese 193*: 559–582

Hohwy, J. (2013) *The Predictive Mind*. Oxford University Press

Hutto, D. D. (2017) Getting into predictive processing's great guessing game: Bootstrap heaven or hell? *Synthese, 195*(6), 2445-2458

Hutto, D. D. and Myin, E. (2013) *Radicalizing Enactivism: Basic Minds Without Content.* MIT Press.

Hutto, D. D. and Myin, E. (2017) *Evolving Enactivism: Basic Minds Meet Content*. MIT Press.

Hutto, D. D. and Satne, G. (2015) The Natural Origins of Content. *Philosophia 43*(3), 521-536

Järvilehto, T. (1998) The Theory of the Organism-Environment: System: I. Description of the Theory, *Integrative Physiological and Behavioral Science, 33*(4), 321-334.

Kirchhoff, M. D., and Robertson, I. (2018) Enactivism and predictive processing: a non-representational view, *Philosophical Explorations, 21*:2, 264-281

Marr, D. (1982). *Vision: a computational investigation into the human representation and processing of visual information.* Freeman.

Marr, D. and Hildreth, E. (1980). Theory of edge detection. *Proceedings of the Royal Society of London. Series B; Biological Sciences, 207*(1167): 187–217.

Mole, C. and Zhao, J. (2016) Vision and abstraction: an empirical refutation of Nico Orlandi's non-cognitivism. *Philosophical Psychology, 29*:3, 365-373

Moyal-Sharrock, D. (2019) From deed to word: gapless and kink-free enactivism. *Synthese,* 1-21

Myin, E. (2016) Perception as something we do. *Journal of Consciousness Studies* 23(5–6): 80–104.

Noë, A. (2004) *Action in Perception*. MIT Press

Noë, A. (2006) Experience of the World in Time. *Analysis* 66 (1), 26-32.

O'Regan, J. K., and Noë, A. (2001) A Sensorimotor Account of Vision and Visual Consciousness. *Behavioural and Brain Sciences, 24*, 939 –1031

Orlandi, N. (2012) Embedded seeing-as: Multi-stable visual perception without interpretation. *Philosophical Psychology, 25*(4), 555-573

Orlandi, N. (2013) Embedded Seeing: Vision in the Natural World. *Noûs 47*(4) 727–747

Orlandi, N. (2014) *The Innocent Eye: Vision is not a cognitive process*. Oxford University Publishing

Pitt, D. (2017) Mental Representation. In Zalta, E. N. (Ed.) *The Stanford Encyclopedia of Philosophy* (Spring 2017 Edition).

Polger, T. (2015) The Innocent Eye: Why Vision Is Not a Cognitive Process, by Nico Orlandi. [Review] *Analysis Reviews 75*(2), 343-345

Ramsey, W. M. (2007) *Representation Reconsidered*, Cambridge University Press.

Rock, I. (1983). *The logic of perception*. MIT Press.

Rosenberg, A. (2015) The Genealogy of Content or the Future of an Illusion, *Philosophia 43*(3), 537-547.

Segal, G. (1989). Seeing what is not there. *The Philosophical Review, 98*(2):189–214.

Shapiro, L. (2014) Radicalizing Enactivism: Basic Minds without Content, by Daniel D. Hutto and Erik Myin. [Review.] *Mind 123*(489), 213–220.

Travis, C. 2004. The silence of the senses. *Mind 113*(449): 57–94.

Van Gelder, T. (1995). What might cognition be, if not computation? *The Journal of Philosophy, 92 (*7), 345–381.

Warren, W. (2005). Direct perception: The view from here. *Philosophical Topics, 33* (1), 335-361.