Normativity in Perception

Edited by

Maxime Doyon
Université de Montréal, Canada

and

Thiemo Breyer
University of Cologne, Germany

palgrave
macmillan

Contents

Part IV Issues in Epistemology

3

Seeing Our World

Michael Madary

Introduction

We do not see *the* visual world. We see *our* visual world. That is, we see a world that we share with other humans, engaged in particular cultural practices. This claim is intended to be at odds with the way that many philosophers and scientists have traditionally thought about visual content. The traditional view has it that we see surfaces and shapes, colors and objects. I would like to defend an alternative. On this alternative, visual content includes social and rational norms. The mental operations that enable intelligent social interaction – often relegated to the inner and unconscious realm of cognition – can actually unfold in plain sight, as it were. These claims all lead to the main thesis of this chapter, which is that *visual content has a strong social element*. Call it thesis VCS.

VCS follows from a background claim about the general structure of visual experience. The background claim is as follows:

(AF) Visual experience has the general structure of anticipation and fulfillment.

In the next section I will sketch the main support for AF. Fully defending AF is far beyond the scope of one section of one chapter, but I will indicate sources where the reader can find more details. After motivating AF, I will, in section III, show that VCS follows naturally from AF. The argument in Section 2 is *a priori*. In Section 3, I will present a range of empirical results that show support for VCS. In the final section of the chapter, I will draw some lessons about mental architecture that follow from VCS. The main claim there will be that cognition is embedded in

the cycle of action and perception, not tucked away as a computational module taking input from perception and offering motor output.

1 Vision is a process of anticipation and fulfillment (AF)

AF can be defended both on *a priori* and empirical grounds. The fact that there is a convergence on AF from these two distinct – and sometimes conflicting – modes of investigation indicates a first point in favor of the truth of AF. I will first give the *a priori* argument and then turn to the evidence from psychology and neuroscience.

The *a priori* case for AF begins with what I take to be two undeniable properties of visual experience. Visual experience is perspectival, as Leibniz (Swoyer, 1995) and Hume (1777/1993) both noted. And all experience is temporal, as Kant emphasized in the second section of the *Transcendental Aesthetic* (1781/1787/1998). Due to the perspectival nature of visual experience, our experience of everyday properties is always *incomplete* at any particular time. In an elegant synthesis of these two features of visual experience, Husserl proposed that perceptual experience has the general structure of anticipation and fulfillment, or AF.

In visual perception, we are intentionally directed towards properties, such as shape and size, that can be seen from multiple perspectives.[1] This fact raises what is sometimes known as the problem of perceptual constancy: how do we perceive static properties despite the continuously changing appearance of those properties? Husserl's key proposal was that we experience these properties by implicitly anticipating how appearances should change as we move, or as the objects move.[2] As we move and take different perspectives on the object, we implicitly anticipate that appearances will change in a particular way. When things are going well, those anticipations are *fulfilled*, and new anticipations are immediately generated. When there is a misrepresentation of the environment, our anticipations may not be fulfilled. Importantly, the structure of anticipation and fulfillment incorporates the temporal nature of perception because anticipations are essentially future-directed.

There is little room for detail, but I should make four quick points about perceptual anticipations in order to illustrate my meaning and avoid misunderstanding.[3] First, anticipation has various degrees of determinacy (Husserl, 2001, p. 42; Siewert, 2005, p. 287). When we are looking at familiar objects, we are able to anticipate changes in appearances more determinately than when looking at unfamiliar objects. This first point will be important below. Second, anticipation is not limited to the hidden sides of objects. We can implicitly anticipate, for instance,

how a room will look as we explore it. Third, perceptual anticipation is usually not a deliberate mental act. Rather, it is an ongoing feature of all visual experience. Fourth, visual anticipations are continuously 'stirred up' (from Husserl's *erregen* in his 1900/1901/1993, VI §10) in the process of perception. What I mean with this fourth point is that we do not anticipate all of the possible appearances of an object at one time. There are a great many ways that objects can appear. The anticipations that are stirred up are constrained by the details of one's current viewing conditions.

Now that I have sketched the Husserlian *a priori* motivation for AF, I will sketch the empirical evidence for AF. Some of the empirical support for AF has been cited as support for the sensorimotor approach to vision (O'Regan and Noë, 2001; Noë, 2004). Here I will quickly mention three lines of evidence: selective rearing experiments, change blindness, and peripheral indeterminacy. In a well-known selective rearing experiment (Held and Hein, 1963), two kittens were harnessed together on a carousel from birth. The active kitten was allowed to explore the visual environment while the passive kitten was moved by the carousel. The movements of the passive kitten were determined by the movements of the active kitten via the carousel. Although both kittens received the same patterns of visual input, the passive kitten was left with permanent visual deficits. These results can be brought to support AF because they suggest that the passive kitten was never able to learn how visual appearances should change with self-generated movement. The passive kitten's ability to anticipate visual appearances never developed properly.

In the change blindness paradigm, subjects are presented with a visual image of a natural scene, and then given a quick distractor mask followed by a second image. The second image is identical to the first in all aspects except one, typically major, detail. Then the subject sees another distractor, followed by the first image again, and so on. The experimental paradigm is surprising because most subjects are unable to detect the change between the two images unless allowed to look at the alternating images for quite some time (O'Regan et al., 1999). There is some controversy about how to interpret the phenomenon of change blindness, but one plausible interpretation supports AF: change blindness suggests that we do not see the visual field in full detail. We are able to access the details of the visual environment because we continuously explore using head and body movements. This interpretation supports AF because it reveals the incomplete nature of visual perception. Repeated sampling through movement can be understood as the continuous process of anticipation and fulfillment. Similarly, the strong

indeterminacy of peripheral vision (Freeman and Simoncelli, 2011) is another feature of vision that suggests an ongoing process of anticipation and fulfillment. We do not simply open our eyes and see our world, we continuously sample our world in an ongoing process (Findlay and Gilchrist, 2003). The structure of this process is one of anticipation and fulfillment.

So far, I have given some empirical evidence from perceptual psychology, but there are also neuroscientific facts that support AF. The most obvious such fact is neurophysiological. If vision is anticipatory, one might expect to find massive feedback connectivity in the visual brain, which is exactly what has been found (Rockland and Van Hoesen, 1994). Apart from the neurophysiology, there is evidence from computational modeling. In a seminal work, Rajesh Rao and Dana Ballard (1999) designed an artificial neural network using anticipatory or predictive processing. After training the network on thousands of natural images, the neurons in the artificial network exhibited what is known as 'extra-classical' receptive-field effects. The extra-classical effects refer to neural responses to visual stimuli that have been recorded in vivo and are not easily accommodated by the traditional (or classical) feed-forward model of vision. Rao and Ballard created a feedback (anticipatory) model of vision, and their model captures both the classical and the extra-classical effects found in single cell recordings.

Rao and Ballard's work was at the beginning of a fruitful line of research using the predictive coding framework to make sense of way the visual brain works. For instance, Tai Sing Lee and David Mumford (2003) have applied the framework to a wide array of evidence from visual neuroscience, including a consideration of more brain regions than Rao and Ballard. More recently, Michael Spratling has applied the predictive coding framework to visual attention (2008), and to some of the details of neural response in area V1 (2012).[4]

To sum up, there is an *a priori* argument for AF, and there is empirical support for AF. There is a great deal that is left unexplored here,[5] but for the purposes of this chapter, let us assume AF in order to investigate the social content of vision.

2 Visual content has a strong social element

In this section I will develop the argument for VCS, which follows immediately from AF. First, I should note my intellectual debts: the idea that perceptual content is social is not a new one. It can be found in Husserl's posthumously published writings on intersubjectivity, as well

as in his well-known late work, *The Crisis of the European Sciences* (1970, §54b). Similarly, Merleau-Ponty emphasizes that the social element is ever-present in experience:

> We must therefore rediscover, after the natural world, the social world, not as an object or sum of objects, but as a permanent field or dimension of existence: I may well turn away from it, but not cease to be situated relatively to it. (1945/1962, p. 362)

The idea can also be found in contemporary work within the phenomenological tradition (Zahavi, 2005, chapter 6; Gallagher, 2005, chapter 9; Thompson, 2007, chapter 13).

What is new here, I hope, is that I am offering some details about *how* visual content is social. It is straightforward to talk about seeing colors and shapes, but how we see social content is not obvious. Although my own position differs slightly from hers, we can take Edith Stein's work on this topic (1913/1989) as a starting point. In her doctoral dissertation, supervized by Husserl, she applied Husserl's anticipation/fulfillment structure of perception to the way in which we empathize with others. She explains:

> The averted and interior sides of a spatial thing are co-given with its seen sides. In short, the whole thing is 'seen'. But, as we have already said, this givenness of the one side implies tendencies to advance to new givenness [...] The co-seeing of foreign fields of sensation [of other subjects; MM] also implies tendencies, but their primordial fulfillment is in principle excluded here [...] Empathic representation is the only fulfillment possible here. (1913/1989, p. 57)

Stein's suggestion, as I understand it, is that empathic awareness of the other is structurally similar to awareness of physical objects. The appearance of a table from a particular perspective serves as a fulfillment of my representation that there is a full table before me. The fulfillment involved in our perception of physical objects is the fulfillment of sensation, and is, in Stein's terminology, 'primordial'. The fulfillment involved in our empathic awareness of the other human being occurs through our perception of the other's animated body. Empathic fulfillment is not 'primordial', but it does share the general structural similarity with perceptual fulfillment.

Recently, Joel Smith (2010) has offered an account of the way in which the mental states of others can show up in visual experience. His account

also borrows from Husserl's notion of perceptual anticipation, and it turns out to be quite similar to Stein's results in some places, although he does not cite Stein as an influence. The way in which we perceive the mental states of others, Smith argues, is analogous to the way in which we perceive the hidden sides of objects through unfulfilled anticipations. He writes: 'Just as the rear aspect of the book is visually present without being visually presented, so another's misery is visually present even though only their frown is visually presented.' (2010, p. 739) He goes on to defend this idea against some objections by introducing a way to individuate perceptual states according to functional role.

Both Stein and Smith emphasize the way in which perceiving the mental states of others is similar to perceiving the hidden sides of objects. But it is important to be clear that Stein makes the additional point that experiencing mental states is disanalogous to experiencing hidden sides, due to the fact that the mental states and sides of objects have different modes of fulfillment – empathic and sensory fulfillment respectively.[6]

VCS shares similarities with the views that Stein and Smith have developed. All three of us rely on Husserl's notion of perceptual anticipation in order to identify social content in perception. But my view is stronger than theirs in the following way. I am claiming that the social content in perception is pervasive. On my view, the social element of visual content goes beyond the content involved with attributing mental states to others. The sociality of visual content is primary in development and it remains as an element of visual experience throughout maturity. The social aspect of perception is at play when we see each other, as Stein and Smith have argued, but it is also at play when we see our environment more generally: our homes and workplaces, our streets and sidewalks, our tools and food, and even the natural world.

If we accept that AF is true, then my claims about the social content of vision follow naturally. Visual anticipations are directed towards how the world might appear if we were to view it from different perspectives. Recall from section II above that visual anticipations have various degrees of determinacy. In more *familiar* environments, we will have more determinate anticipations. Similarly, the content of visual anticipations will be partly determined by previous experience, by how the world has appeared in the past. The influence of previous experience on AF content suggests that violations of anticipations would signal a departure from the *normal* course of appearances. Thus, familiarity and normality enter into perceptual content. Now, for creatures such as us, the familiar and the normal is social as a matter of fact. If familiarity and normality enter into the content of visual anticipation, and

if the familiar and the normal are social for us, then visual content is social.[7]

Here is the argument for VCS step by step:

1. (AF) Visual perception has the structure of an ongoing process of anticipation and fulfillment.
2. Visual anticipations have various degrees of determinacy.
3. The familiarity and normality of the visual scene enter into perceptual content through the cycle of anticipation and fulfillment.
4. Familiarity and normality are social properties for humans.

Conclusion: (VCS) Visual content has a strong social element (for humans).

The first two premises were quickly defended above in section II. The third premise was defended immediately above, prior to my presentation of the main argument step by step. Perhaps I should add some comments about my fourth claim: that familiarity and normality are social for creatures like us. I only mean to make the uncontroversial point that the human being is a social animal. When we interact with others, there are norms of behavior. Violations of those norms appear to us in the form of violations of perceptual anticipations. Adherence to these norms appears to us in the form of fulfillments of perceptual anticipations. When we use tools or move about our homes, workplaces, and towns, doing so normally fulfills perceptual anticipations – both our own and for others who might see us. Doing so abnormally violates perceptual anticipations. It is in these considerations that one can see how my view goes beyond that of Stein and Smith. I am suggesting that the social content of vision goes beyond the perception of other humans. My view is that there is a social element in the perception of other humans *as well as* in artifacts and our environment more generally. It should now be clear what I intend when I claim to illustrate how visual content is social. The key point is this: *social factors partly determine the content of visual anticipations*. The violation of visual anticipations can signal the violation of social norms. Thus, social norms show up for us at the level of visual content.[8]

There are three objections that I ought to address. First, one might ask about moments when we are alone, away from others, perhaps in pristine nature. In such situations, surely, one might say, there is no social content to what we see. I maintain that visual content remains social in such situations. My thinking behind this claim can be motivated by considering Merleau-Ponty once again: 'Just as nature finds its way into

the core of my personal life and becomes inextricably linked with it, so behavior patterns settle into that nature, being deposited in the form of a cultural world.' (1945/1962, p. 347) Even when one is alone, one still faces the choice of behaving in a way that would be socially normal or not. Behaving abnormally would bring about an unusual pattern of visual experiences that would, plausibly, alter the content of visual anticipations. When acting abnormally we may experience violations of anticipation, or a decrease in the degree of determinacy of anticipation. We can choose to act normally or abnormally, and perhaps no one will ever know what we choose, but the fact would remain that our visual experience would reflect whether we are acting in accordance with social norms, in accordance with the 'deposited cultural world', or in violation of social norms. The passage continues:

> Not only have I a physical world, not only do I live in the midst of earth, air and water, I have around me roads, plantations, villages, streets, churches, implements, a bell, a spoon, a pipe. Each of these objects is moulded to the human action which it serves. Each one spreads around it an atmosphere of humanity which may be determinate in a low degree, in the case of a few footmarks in the sand, or on the other hand highly determinate, if I go into every room from top to bottom of a house recently evacuated. (Ibid., pp. 347f.)

In other words, being alone does not remove the social nature of what we perceive. In the extreme case, when we are alone in untouched nature – not even footmarks in the sand – we have the unusual case of the lack of an 'atmosphere of humanity'. In such cases, we experience a kind of abnormality, the abnormality of having no sign of our conspecifics. Visual anticipations are disappointed because we normally anticipate traces of other humans in our perceptual environment. The disappointment of anticipation in such untouched environments may partly explain the thrill of being in such places.

The second objection is that humans are not necessarily social. We might consider a hermit or a feral child, or even a thought experiment in which a human child is artificially hatched and then nourished and raised by some impersonal machine. This objection reveals a possible limitation in the scope of my claim about visual content being social. It may not apply to all human beings. My claim above is that visual content is social because visual content includes the normal and the familiar, and the normal and familiar are social. There may be cases, however rare, in which the normal and familiar are not social. In such

cases, visual content may not be social. The *a priori* claim that does apply to all human beings would be my claim that the normal and familiar appear at the level of perceptual content (premise 3 above). VCS may not be a universal claim about all humans because the fourth premise above may not hold true for all humans.

Despite the fact that my claim is not universal to all humans, it is still a claim that departs from standard ways of thinking about visual content. Jesse Prinz, for example, argues that visual content is merely of the surfaces and shapes around us (2012, p. 52). Some cognitive neuroscientists might go further and include object recognition in the content of visual experience. None of these standard approaches include the possibility of social content in vision.

The third objection is as follows. One might insist that we must first see non-social visual properties in order to then make inferences about social content. Along this line of thought, one might conclude that non-social content is primary and that social content is somehow secondary or even non-perceptual. After all, we first must see an object as a hammer – see its shape and surfaces – before there is any concern about the normal or familiar way in which hammers are used. My reply to this objection is to return to the fact that perception is always perspectival. One consequence of this fact is that perception of properties is always incomplete. It is, strictly speaking, not correct to claim that we see a hammer *simpliciter*. Seeing a hammer in a normal orientation, or being used in a normal manner, stirs up relatively determinate anticipations, which give us a sense of the familiar. Seeing a hammer in an unusual orientation stirs up less determinate anticipations, reflecting the unfamiliarity of the perceptual scene. Once we reflect on the fact that there is no simple, non-perspectival view of an object, this objection loses its force.

Now that I have offered my general *a priori* motivation for social content in vision, I will turn to some of the empirical results that support my view.

3 Empirical support

My main argument for VCS is *a priori*, except for the uncontroversial observation that humans are social animals. In this section I wish to show that my position, which emphasizes the way in which social cognition can occur within visual perception, is also supported by a broad range of empirical evidence.

Let us begin with some observations about human development. It is well-known that newborns exhibit preferential viewing for faces

(Johnson et al., 1991). This strong preference suggests that visual content is primarily social from the very beginning of our lives outside of the womb.[9] Developmental psychology also supports my claim that the social content of vision is not limited to instances in which we are looking at other humans. In an experiment that illustrates this point, Krista Casler and colleagues (2009) taught two- and three-year-old children to use novel tools. Later, the children observed a puppet using those same tools. When the puppet used the tools in an atypical manner – that is, differently from the way in which the children were taught to use the tools – the children protested. For instance, they told the experimenter how the puppet was using the tool incorrectly, or they tried to intervene in order to teach the puppet the correct use for the tool. These results suggest that children naturally apply a kind of normativity to the use of artifacts. They presuppose that there is a proper way in which one ought to use a tool.

Children's normativity regarding tool use is a natural fit with VCS and AF. Perceiving an artifact stirs up anticipations about the ways in which the artifact will appear when one uses or handles that artifact in a normal manner. The content of these anticipations is largely determined by social factors, by the way in which we have seen others, and perhaps ourselves, use the artifact. The protest of the children is triggered by the violations of their anticipations about how the tools should appear when they are being used.

If VCS is correct, one might expect correlation between differences in social cognition, on the one hand, and differences in visual experience, on the other. There are at least two ways in which this expectation is met: in disorders of social cognition, and in cross-cultural comparisons. One of the main characteristics of autism is a disability in social cognition. If vision has a strong social element, then we should expect differences in visual experience for individuals with autism. The evidence supports this expectation. Many open questions remain as to the details, but it is becoming clear that there are differences in, for instance, facial processing (Behrmann et al., 2006) and motion perception (Kaiser and Shiffrar, 2009) for individuals with autism.

The second area in which we see a correlation between differences in social cognition and differences in visual experience is in cross-cultural comparisons. If vision has a strong social element, then we might expect that one's culture can have an impact on how one sees the world. Anecdotally, we might expect this impact to reveal itself in patterns of saccades as determined by common social interests. But experimental evidence also shows such an influence of culture on vision. Studies have

shown that East Asians and Westerners tend to differ in their 'cognitive styles', with East Asians being more holistic and Westerners being more analytic (Masuda and Nisbett, 2001). Recent work on this theme suggests that the differences occur in the allocation of resources for visual attention (Boduroglu et al., 2009), which offers further support for the tight connection between vision and social content. Our cultural background, it seems, partly determines our visual attention.[10]

Another well-documented source of evidence for the social element in visual content is something that readers can see for themselves. For over 30 years now, psychologists have known that some dynamic point-light displays are naturally perceived by us as walking human beings (Johansson, 1973). By fixing lights to various parts of a walking human, such as the hands, feet, hips, and head, psychologists create stimuli that are completely dark except for the little lights moving in a particular way. The surprising feature of these displays is that we cannot help but perceive the moving lights as a human walker. I urge readers to experience such displays for themselves.[11] Not only do we see the lights as humans, but we see more particular properties of the walker, including gender (Kozlowski and Cutting, 1977) and emotional state (Dittrich et al., 1996).

Consider how the point-light walkers can be described in terms of anticipation and fulfillment. We naturally spend a good bit of our visual lives watching other humans walking around. We need to do so in order to avoid collisions on busy sidewalks or in a stroll across campus. Thus, the motion of the point lights traces a pattern with which we are familiar. When we first see the moving lights, anticipations are stirred up based on previous (and common) experiences of others. As the lights move in a familiar pattern, those anticipations are fulfilled and we experience a walking human rather than a two-dimensional display of strangely moving lights.

There is a great deal of further evidence that can be cited as lending support to VCS. In the remainder of this section I will mention two related areas without entering into the details. These areas are social signals and cultural evolution. If VCS is true, then we would expect humans to use visual perception in order to receive, and eye movements in order to send, social information. There is a wealth of evidence indicating that we do so. In a survey article on this topic, Chris Frith (2008) covers the detection of mood from facial expressions and posture, gaze following as a way of understanding intentions, and the Chameleon effect (in which two interlocutors experience mutual respect and trust when they imitate each other's mannerisms). Building on some of the

same lines of evidence, Michael Tomasello (1999) has argued that *cultural evolution* explains how humans have developed advanced cognitive abilities in the relatively short period of six million years of evolutionary history that separates modern human from other great apes. Cultural evolution is driven by the human capacity to collaborate through shared intentionality. Crucially, shared intentionality is achieved in large part through the ongoing cycle of action and visual perception. For instance, infants learn one aspect of shared intentionality around one year of age by coordinating their gaze 'triadically' with their human interlocutor *and* with the object to which both of them are directed (Tomasello et al., 2005, p. 682).

There is a great deal here that must be left unexplored. In the final section I will draw some general lessons about mental architecture that are motivated by VCS.

4 Embedded rationality – the sandwich or the cycle?

If I am correct about VCS, then a general lesson about mental architecture follows. Borrowing from Susan Hurley (1998), we can distinguish two competing ways of understanding mental architecture. The classical 'sandwich', according to Hurley, treats perception as input and action as output. All of the heavy lifting is done by cognition, which is the real substance of the sandwich. Perception and action are just slices of bread holding it all together. Cognition, on this view, is often understood in terms of propositional attitudes, beliefs and desires, implemented in some kind of Fodorian language of thought. Hurley's alternative to the classical sandwich is a view in which perception and action are not mere input and output. Instead, they are interdependent in the form of an ongoing cycle on different temporal and spatial scales. So, which is it, the sandwich or the cycle?

Each model has strengths and weaknesses. As one might expect, the strength of the sandwich lies in modeling abstract cognitive tasks, and its weakness – according to its critics – lies in modeling tasks that involve ongoing skillful behavior, behavior that requires fine-grained coordination between action and perception. The cycle has the exact opposite profile. It excels in explaining ongoing skillful behavior, but has a weak spot when it comes to cognition, or rationality. Hurley was aware of this shortcoming, and offered an alternative conception of rationality: 'Rationality might instead emerge from a complex system of decentralized, higher-order relations of inhibition, facilitation, and coordination among different horizontal layers, each of which is dynamic

and environmentally situated.' (1998, p. 409) On Hurley's alternative, the mind is made of content-specific sensorimotor loops (ibid., p. 21). Rationality is not localized in a belief/desire processing engine, but rather emerges out of this vast layered network of loops. But a skeptic might respond to Hurley's view with the following question: how does higher-level rational content emerge out of a system of sensorimotor loops?

I think that VCS, and some of the themes discussed above, bring us slightly closer to an answer for the skeptic – at least for the case of social cognition. The best response to the skeptic's question is to deny the presupposition that perception is devoid of higher-level content. Leaving out a lot of explanation, one good reply to the skeptical question is to say that social cognition can emerge out of sensorimotor loops because social content is already present in visual experience.

The results about normativity in tool use for infants, cited above, are one way to illustrate Hurley's idea of rationality through sensorimotor loops. The children's perceptions of the objects become associated with various motor representations that involve using the object correctly. Seeing the object stirs up anticipations about how the object might be used. When it is used 'incorrectly', those anticipations are disappointed.

A closely related general line of evidence supporting Hurley's architecture suggests that *imitation* plays a large role for human intelligent behavior. Normal human perception of others performing an action may bring about an impulse to perform that same action, to imitate (Heyes, 2011; Belot et al., 2013). We act rationally by inhibiting these impulses, by using, in Patrick Haggard's terminology, our 'veto power' over the urge to act (Kühn et al., 2009). In some cases of brain injury (Lhermitte et al., 1986), patients lose the veto power and exhibit utilization behavior (using tools at inappropriate times) or imitation behavior (imitating conspecifics at inappropriate times). These considerations motivate an understanding of human action in which perception and action are tightly linked through sensorimotor loops. Seeing the artifact activates the motor routine to use the artifact; in my terms from above, seeing the artifact stirs up visual anticipations about how the artifact might appear when being used normally. Seeing the other's smile activates the motor routine to smile. Rationality is, in part, a matter of stopping – vetoing – these routines when appropriate.

To sum up, VCS can complement a mental architecture that focuses on the cycle of action and perception. The main contribution in that regard is that VCS offers some detail about how perceptual content can include

the kinds of things that are often relegated to the domain of cognition. One main advantage of the cycle is parsimony. The cycle seeks to explain intelligent behavior by appeal to the ongoing dynamics, and interplay between two obvious abilities: action and perception.[12] The sandwich posits a hidden third ability that is supposed to be doing all the work: cognition. Another main advantage of the cycle is that it finds support in 'veto' models of action from recent cognitive neuroscience.

Acknowledgments

I would like to thank Philippe Blouin, Thiemo Breyer, Maxime Doyon, Sascha Fink, Thomas Metzinger, and Lisa Quadt for helpful comments on this material. I also thank my audiences in Montréal, Osnabrück, and Zürich for their feedback. This research has been supported by the EC Project VERE, funded under the EU 7th Framework Program, Future and Emerging Technologies (Grant 257695).

Notes

1. This claim is widely accepted in contemporary philosophy of perception, but it would be denied by philosophers who deny that we represent any properties whatsoever in perception (Brewer, 2006; Martin, 2006; Travis, 2004).
2. See Madary (2012a) for a detailed discussion of Husserl on this topic and the way in which it relates to contemporary philosophy of perception.
3. For a full discussion, see Madary (2013).
4. For a discussion of the predictive coding framework from a philosophical perspective, see Clark (2013a) and Hohwy (2013).
5. For a discussion of the relationship between visual phenomenology, on the one hand, and predictive coding models on the other, see Madary (2012b), Clark (2013b), Seth (2014), and Madary (2014).
6. I thank Philippe Blouin for raising this point.
7. See Steinbock (1995, esp. pp. 132–37) for a discussion of Husserl's treatment of closely related themes.
8. My claim here is concerned with the way in which social content can occur at the level of conscious visual experience. But I should be clear that unconscious perceptual processing might include social content as well. I thank Lisa Quadt for raising this point.
9. Also highly relevant here is Colwyn Trevarthen's work on primary intersubjectivity (1979).
10. See Gallagher (this volume) for related themes.
11. An excellent demonstration can be found at www.biomotionlab.ca.
12. To be even more parsimonious, one could even look for a single computational strategy underlying both perception and action, thus putting pressure on the distinction itself. I will not defend this strategy, but for a step in this direction, see Clark (2013a), who is exploring themes from Karl Friston's work.

References

Behrmann, M., Thomas, C., Humphreys, K. (2006) 'Seeing it differently: visual processing in autism', *Trends in Cognitive Sciences*, 10, 258–64.

Belot, M., Crawford, V., Heyes, C. (2013) 'Players of "matching pennies" automatically imitate opponents' gestures against strong incentives', *Proceedings of the National Academy of Sciences*, 110, 2763–8.

Boduroglu, A., Shah, P., Nisbett, R. (2009) 'Cultural differences in allocation of attention in visual information processing', *Journal of Cross-Cultural Psychology*, 40, 349–60.

Brewer, B. (2006) 'Perception and content', *European Journal of Philosophy*, 14, 165–81.

Clark, A. (2013a) 'Whatever next? Predictive brains, situated agents, and the future of cognitive science', *Behavioral and Brain Sciences*, 36, 181–204.

Clark, A. (2013b) 'The many faces of precision', *Frontiers in Psychology*, 4, doi: 10.3389/fpsyg.2013.00270

Dittrich, W., Trosciano, T., Lea, S., Morgan, D. (1996) 'Perception of emotion from dynamic point-light displays represented in dance', *Perception*, 25, 727–38.

Findlay, J., Gilchrist, I. (2003) *Active Vision* (Oxford: Oxford University Press).

Freeman, J., Simoncelli, E. (2011) 'Metamers of the ventral stream', *Nature Neuroscience*, 14, 1195–1201.

Frith, C. (2008) 'Social cognition', *Philosophical Transactions of the Royal Society B*, 363, 2033–9.

Gallagher, S. (2005) *How the Body Shapes the Mind* (Oxford: Oxford University Press).

Held, R., Hein, A. (1963) 'Movement-produced stimulation in the development of visually guided behavior', *Journal of Comparative and Physiological Psychology*, 56, 872–6.

Heyes, C. (2011) 'Automatic imitation', *Psychological Bulletin*, 137, 463–83.

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

Hume, D. (1777/1993) *An Enquiry Concerning Human Understanding* (London: Hackett).

Hurley, S. (1998) *Consciousness in Action* (Cambridge, MA: Harvard University Press).

Husserl, E. (2001) *Analyses Concerning Passive and Active Synthesis*, trans. by A. Steinbock (Dordrecht: Kluwer).

Husserl, E. (1900/1901/1993) *Logische Untersuchungen* (Tübingen: Niemeyer).

Husserl, E. (1970) *The Crisis of European Sciences and Transcendental Phenomenology*, trans. by D. Carr (Evanston, IL: Northwestern University Press).

Johansson, G. (1973) 'Spatio-temporal differentiation and integration in visual motion perception', *Psychological Research*, 14, 201–11.

Johnson, M., Dziurawiec, S., Ellis, H., Morton, J. (1991) 'Newborns' preferential tracking of face-like stimuli and its subsequent decline', *Cognition*, 40, 1–19.

Kaiser, M., Shiffrar, M. (2009) 'The visual perception of motion by observers with autism spectrum disorders', *Psychonomic Bulletin & Review*, 16, 761–7.

Kant, I. (1781/1787/1998) *Kritik der reinen Vernunft* (Hamburg: Meiner).

Kasler, K., Terziyan, T., Greene, K. (2009) 'Toddlers view artifact function normatively' *Cognitive Development*, 24, 240–7.

Kozlowski, L., Cutting, J. (1977) 'Recognizing the sex of a walker from a dynamic point-light display', *Perception and Psychophysics*, 21, 575–80.

Kühn, S., Haggard, P., Brass, M. (2009) 'Intentional inhibition: how the "veto-area" exerts control', *Human Brain Mapping*, 9, 2834–43.

Lee, T., Mumford D. (2003) 'Hierarchical Bayesian inference in the visual cortex', *Journal of the Optical Society of America A*, 20, 1434–48.

Lhermitte, F., Pillon, B., Serdaru, M. (1986) 'Human autonomy and the frontal lobes. Part I: Imitation and utilization behavior: a neuropsychological study of 75 patients', *Annals of Neurology*, 19, 326–34.

Madary, M. (2012a) 'Husserl on perceptual constancy', *European Journal of Philosophy*, 20, 145–65.

Madary, M. (2012b) 'How would the world look if it looked as if it were encoded as an intertwined set of probability density distributions?', *Frontiers in Psychology*, 3, doi: 10.3389/fpsyg.2012.00419

Madary, M. (2013) 'Anticipation and variation in visual content', *Philosophical Studies*, 165, 335–47.

Madary, M. (2014) 'Perceptual presence without counterfactual richness', *Cognitive Neuroscience*, 5, 131–3.

Martin, M. (2006) 'On being alienated' in T. Gendler, J. Hawthorne (eds), *Perceptual Experience* (Oxford: Oxford University Press), pp. 354–410.

Masuda, T., Nisbett, R. (2001) 'Attending holistically vs. analytically: comparing the context sensitivity of Japanese and Americans', *Journal of Personality and Social Psychology*, 81, 922–34.

Merleau-Ponty, M. (1945/1962) *Phenomenology of Perception*, trans. by C. Smith (London: Routledge).

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

O'Regan, K., Rensink, R., Clark, J. (1999) 'Change-blindness as a result of "mudsplashes"', *Nature*, 398, 34.

O'Regan, K., Noë, A. (2001) 'A sensorimotor account of vision and visual consciousness', *Behavioral and Brain Sciences*, 24, 939–1031.

Prinz, J. (2012) *The Conscious Brain* (Oxford: Oxford University Press).

Rao, R.P.N., Ballard, D. (1999) 'Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects', *Nature Neuroscience*, 2, 79–87.

Rockland, K., Van Hoesen, G. (1994) 'Direct temporal-occipital feedback connections to striate cortex in the macaque monkey', *Cerebral Cortex*, 4, 300–13.

Seth, A. (2014) 'A predictive processing theory of sensorimotor contingencies', *Cognitive Neuroscience*, 5, 97–118.

Siewert, C. (2005) 'Attention and sensorimotor intentionality' in D.W. Smith, A. Thomasson (eds), *Phenomenology and Philosophy of Mind* (Oxford: Oxford University Press), pp. 270–94.

Smith, J. (2010) 'Seeing other people', *Philosophy and Phenomenological Research*, 81, 731–48.

Spratling, M. (2008) 'Predictive coding as a model of biased competition in visual attention', *Vision Research*, 48, 1391–1408.

Spratling, M. (2012) 'Predictive coding as a model of the V1 saliency map hypothesis', *Neural Networks*, 26, 7–28.

Stein, E. (1913/1989) *On the Problem of Empathy*, trans. by W. Stein (Washington, DC: Publications of the Institute of Carmelite Studies).

Steinbock, A. (1995) *Home and Beyond: Generative Phenomenology After Husserl* (Evanston, IL: Northwestern University Press).

Swoyer, C. (1995) 'Leibnizian expression', *Journal of the History of Philosophy*, 33, 65–99.
Thompson, E. (2007) *Mind in Life: Biology, Phenomenology, and the Sciences of Mind* (Cambridge, MA: Harvard University Press).
Tomasello, M. (1999) *The Cultural Origins of Human Cognition* (Cambridge, MA: Harvard University Press).
Tomasello, M., Carpenter, M., Call, J., Behne, T., Moll, H. (2005) 'Understanding and sharing intentions: the origins of cultural cognition', *Behavioral and Brain Sciences*, 28, 675–91.
Travis, C. (2004) 'The silence of the senses', *Mind*, 113, 57–94.
Trevarthen, C. (1979) 'Communication and cooperation in early infancy' in M. Bullowa (ed.), *Before Speech: The Beginning of Interpersonal Communication* (Cambridge: Cambridge University Press), 321–47.
Zahavi, D. (2005) *Subjectivity and Selfhood: Investigating the First-Person Perspective* (Cambridge, MA: MIT Press).