

**Josefa Toribio**  
**ICREA-UB**

ICREA Research Professor at the University of Barcelona  
Departament de Lògica, Història i Filosofia de la Ciència  
Universitat de Barcelona  
Facultat de Filosofia  
Montalegre, 6-8, E-08001 Barcelona  
Phone: +34 934 037991 – Fax: +34 934 037980  
Email: jtoribio@icrea.cat

## **Visual experience: rich but impenetrable**

### **Abstract**

According to so-called “thin” views about the content of experience, we can only visually experience low-level features such as colour, shape, texture or motion. According to so-called “rich” views, we can also visually experience some high-level properties, such as being a pine tree or being threatening. One of the standard objections against rich views is that high-level properties can only be represented at the level of judgment. In this paper, I first challenge this objection by relying on some recent studies in social vision. Secondly, I tackle a different but related issue, namely, the idea that, if the content of experience is rich, then perception is cognitively penetrable. Against this thesis, I argue that the very same criteria that help us vindicate the truly sensory nature of our rich experiences speak against their being cognitively penetrable.

**Keywords:** visual experience; high-level properties; low-level properties; cognitive penetrability; social vision.

### **Acknowledgements**

Different versions of this material were presented at a variety of venues: the Central European University (Second Philosophy of Language and Mind Conference), Harvard University (Workshop on Cognitive Penetration), Helsinki Collegium for Advanced Studies (Dynamics of Active Perception Symposium) and the Swiss Center for the Affective Sciences (THUMOS: Geneva Research Group on Emotions, Values and Norms). I would like to thank the audiences at all these events for their questions and discussion, especially Santiago Echeverri, Fiona Macpherson and Susanna Siegel. I also thank Athanassios Raftopoulos and an anonymous reviewer for helpful comments on a previous version of this paper.

### **Funding**

Research for this paper was supported by the MINECO (Ministerio de Economía y Competitividad) via research grants MCINN FFI2011-26853 and PERSP CSD2009-0056 (CONSOLIDER INGENIO), and by AGAUR (Agència de Gestió d’Ajuts Universitaris i de Recerca) via research grant 2014-SGR-81.

## Visual experience: rich but impenetrable

### Abstract

According to so-called “thin” views about the content of experience, we can only visually experience low-level features such as colour, shape, texture or motion. According to so-called “rich” views, we can also visually experience some high-level properties, such as being a pine tree or being threatening. One of the standard objections against rich views is that high-level properties can only be represented at the level of judgment. In this paper, I first challenge this objection by relying on some recent studies in social vision. Secondly, I tackle a different but related issue, namely, the idea that, if the content of experience is rich, then perception is cognitively penetrable. Against this thesis, I argue that the very same criteria that help us vindicate the truly sensory nature of our rich experiences speak against their being cognitively penetrable.

### 1. Thin vs. rich views.

According to so-called “thin” views about the content of experience, we can only visually experience low-level features such as colour, shape, texture or motion. By contrast, according to “rich” views, we can also visually represent high-level properties. Properties that count as high-level when discussing the richness of the content of visual experience include, but are not limited to: natural kinds (e.g., being a pine tree), artificial kinds (e.g., being a bicycle), dispositional (e.g., being breakable), emotional (e.g., being sad), moral (e.g., being virtuous), and aesthetic (e.g., being beautiful) properties. The core claim of the rich view is that at least some of these high-level properties are part of the sensory phenomenology of visual experience, i.e., that we can indeed see and not just think that, for instance, Mary is sad or that the tree in front of us is a pine tree. Advocates of thin views do not deny that things can visually *seem* to us as having high-level properties. They deny that these *seemings*—the way things appear to us in perception—have any phenomenology of their own (e.g. Tye, 1995) or any sensory, as opposed to cognitive, phenomenology (e.g. Lyons, 2005). It may seem to us as though we could, for instance, see the sadness on other people’s faces, but the distinctive phenomenal character of such a mental state is, according to thin view theorists, nothing but the result of some cognitive, i.e., non-perceptual, element affecting the way in which the typical low-level visual properties involved in the expression of sadness appear to us.

One prominent and detailed argument in favour of rich views is Siegel's (2006, 2010) phenomenal contrast argument. Here is her pine tree scenario (Siegel, 2006: 491):

Suppose you have never seen a pine tree before, and are hired to cut down all the pine trees in a grove containing trees of many different sorts. Someone points out to you which trees are pine trees. Some weeks pass, and your disposition to distinguish the pine trees from the others improves. Eventually, you can spot the pine trees immediately... Gaining this recognitional disposition is reflected in a phenomenological difference between the [overall] ... experiences you had before and after the recognitional disposition was fully developed.<sup>1</sup>

To get the argument off the ground, Siegel makes the plausible assumption that there is a phenomenological difference between the overall experiences one is in before and after acquiring certain recognitional capacities, such as the disposition to recognise pine trees—let's call them O1 and O2 respectively. The argument is then set out in three premises. First, if there is a phenomenological difference between O1 and O2, then the specific visual experiences—E1 and E2—that are parts of O1 and O2 differ in phenomenology. Second, if E1 and E2 differ in phenomenology, then E1 and E2 have different content. Third, if E1 and E2 have different content, the difference consists in the fact that E2 represents the high-level property of being a pine tree. Conclusion: in some visual experiences, some high-level properties are represented.

Siegel's argument is an argument to the best explanation, as she proceeds by discussing and rejecting a variety of alternative hypotheses for each of the premises. Predictably, these alternative explanations and some of the standard objections against rich theories overlap. One of the most important objections, which Siegel addresses as part of her defence of the first premise, is that the difference in phenomenology between O1 and O2 is not the result of a difference in the sensory phenomenology of E1 and E2. Instead, the objection goes, the difference in phenomenology between O1

---

<sup>1</sup> *Visual* experiences are part of the *overall* experiences one is in before and after acquiring the pine tree recognitional capacity. These overall experiences include cognitive states, such as one's beliefs and desires, emotional states of all sorts and perhaps experiences from other sense modalities. I have used 'overall experiences' in place of 'visual experiences', which is the expression Siegel originally uses, to make this distinction clearer.

and O2 is the result of a judgment or some other cognitive episode occurring downstream visual consciousness.<sup>2</sup> Judgments are here taken as episodes that occur later in information processing time and that subsume visually represented low-level properties under a cognitive, as opposed to a perceptual, representation.<sup>3</sup> This alternative explanation is often formulated by appealing to states of seeming, which are different from perceptual experiences and could be considered as perceptual judgments (see e.g. Brogaard, 2013). The second premise in Siegel's argument reflects her commitment to intentionalism. According to this version of intentionalism, the phenomenology of an experience is determined by the properties it represents, i.e., by its content. Hence, if there is a difference in the phenomenology of E1 and E2, their content is different. Some of the objections to this claim appeal to the role of attention. Here is a classic formulation. There is indeed a phenomenological difference between E1 and E2, but it does not arise from a difference in the properties visually experienced. It is instead the result of a difference in how we distribute perceptual attention (see e.g., Price, 2009). Finally, some alternative explanations to the third and last premise in Siegel's argument appeal to gestalt shifts that would allegedly track appearance-types instead of high-level properties (see e.g. Nanay, 2011).

Of course, all three premises of the argument have to be true for the conclusion to follow. However, I am here solely concerned with the first premise. For the purposes of this paper, I take the truth of the other two premises for granted and thus ignore both the role played by attention and alternative explanations involving gestalt shifts. This allows me to re-formulate the first premise in terms that already appeal to the experience of high-level properties. The main issue to be addressed here is whether or not the phenomenology of experiences that seem to involve high-level properties is truly sensory—as opposed to being the result of perceptual judgments occurring downstream of visual consciousness. The considerations I offer in what follows thus

---

<sup>2</sup> Siegel discusses not just judgments, but also cognitive states such as dwelling on a belief, entertaining a hunch or intuition, and entertaining a proposition without committing to its truth. In addition, she considers and rejects another alternative: that the difference between O1 and O2 is due to some background phenomenology, whose effect on how the world is presented to us in perception is similar to the effect of e.g. moods. In what follows, I focus only on the contrast between experience and (perceptual) judgment.

<sup>3</sup> Perceptual judgments, in particular, are often taken to be a type of hybrid representation, with a perceptual and a cognitive component. Raftopoulos (2011), for instance, characterizes perceptual judgments as hybrid visual/conceptual constructs. See below.

provide support for rich views only inasmuch as we grant that the rich sensory phenomenology here vindicated is not the result of attentional factors or gestalt shifts that do not involve the representation of high-level properties. Even so, the exercise is worth pursuing, as it ultimately targets a central issue in philosophy and cognitive science: how to characterize the subtle divide between perception and cognition.

The paper is organized as follows. Firstly, I argue in favour of the view that some of our visual representations of high-level properties have a genuine sensory phenomenology. I do so by relying on some recent work in social vision, both with regard to the perception of emotion (e.g. Adams & Kveraga, 2015; Weisbuch & Adams, 2012) and the perception of animacy (Scholl & Gao, 2013). My general strategy is to review findings in these fields so as to isolate a set of conditions that the representation of a certain type of high-level properties—socially salient properties—has to meet to be considered genuinely perceptual. Section 2 will introduce some basic ideas about social vision. In Section 3, I review some findings about perception of emotion. In Section 4, I examine empirical research in the perception of animacy that will be decisive to finally offer such a set of conditions.

Secondly, in Section 5, I tackle a different, but related issue, namely the view that, if the content of visual experience is rich, then visual experience is cognitively penetrable. Although there is a variety of characterizations in the philosophical market, cognitive penetrability is usually portrayed as the nomological possibility that two subjects (or the same subject at different times or in different counterfactual circumstances) could have different perceptual experiences as the result of differences in other cognitive (including emotional) background states, while sharing the same proximal stimulus and attending to the same distal stimuli under the same external conditions (cf. Siegel, 2012; Macpherson, 2012).<sup>4</sup> Susanna Siegel explicitly endorses a version of the above conditional. If we can visually experience high-level properties, she claims, then, even if early vision is cognitively impenetrable, in the sense defended by Fodor (1983) and Pylyshyn (1999), visual experience is cognitively penetrable by prior cognitive, including affective, states (Siegel, 2006,

---

<sup>4</sup> See e.g. Stokes (2013; 2015) and Machery (2015) for a good discussion of the concept of cognitive penetration and some different proposals.

2010).<sup>5</sup> Adams and collaborators go a step further by defending the cognitive penetrability of early vision as a way of accounting for the functional interaction of an array of social cues, including, but not limited to, gender, race and emotion, thus endorsing an empirical version of the conditional (Adams & Kveraga, 2015). In this article, I remain neutral on the empirical issue of whether or not visual experience and/or early vision processing, is indeed cognitively penetrable. I argue, however, that the very same set of criteria that helps establishing the genuine visual nature of our representations of some high-level properties makes the transition from rich views to the cognitive penetrability thesis problematic.

Throughout the paper, I make two assumptions, which I take to be neutral between rich and thin theories. First, I assume representationalism, i.e., the view that mental states and perceptual experiences have content, i.e., that they represent the world as being a particular way. Second, and perhaps more controversially, I assume that there is an interesting distinction between experiences and perceptual judgments, even if perceptual judgments lack some of the paradigmatic characteristics of full-fledged, all-things-considered judgments—characteristics like, for instance, being the result of some sort of deliberation based on evidence or background knowledge. A representation at the level of judgment—even at the level of perceptual judgment—is governed by standards of rationality that differ from those governing perceptual representations in the following sense: perceptual judgments cannot go against your knowledge without thus signalling irrationality. There is no irrationality, however, in perceiving what is known to be false.

This second assumption may be considered somewhat controversial for, at least, the following reason. One may plausibly maintain that there is no real divide between perceptual judgments and perceptual experiences. Perceptual judgments, it could be argued, are more like perceptual experiences with regard to the way they enter into rational relations with the subject's propositional attitudes and general background knowledge. After all, when making a perceptual judgment, non-inferentially and on the basis of perceptual experience alone, one typically judges that things are as they

---

<sup>5</sup> Parker Crutchfield (2012) argues, conversely, that if visual experience is cognitively penetrable, then we can visually represent high-level properties. Although this conditional also strikes me as problematic, for similar reasons to the ones I will mention in Section 5, this is a topic for another paper.

seem.<sup>6</sup> This is, of course, right. However, when there are discrepancies between how things seem to be and some piece of background knowledge, perceptual judgments may be suspended, modified or abandoned—and such weighing of evidence and reasons will come out at the level of all-things-considered judgments. By contrast, perceptual experiences typically remain the same and cannot be thus suspended, modified or abandoned, even in the face of contradictory background knowledge: just consider perceptual illusions. Despite the similarity between perceptual experiences and perceptual judgments, the former thus have a resilience to rational scrutiny than the latter lack.<sup>7</sup> Even if perceptual judgments are considered hybrid constructs, with a sensory and a conceptual component, there is still a clear difference between them and perceptual experiences, since, unlike perceptual judgments, perceptual experiences are a type of event with just sensory phenomenology. The discussion that follows could thus be formulated as targeting the question of whether our seeming to experience high-level properties is the result of an event with a hybrid sensory/conceptual component or a purely sensory one.

## 2. Social vision

We form impressions of other people in a flash. The mere look on a face speaks volumes. It may speak of tenderness, or contempt, or hostility. It may trigger in us a feeling of sympathy, or rejection, or fear. Nonverbal, specifically visual communication is ubiquitous in our social interactions. Our social behaviour is often guided by the meaning we extract from visual cues, which not only help us interpret other people's feelings, intentions and emotions but also, and importantly, help us predict what their likely behaviour is going to be, thus oiling the wheels of our social life. This predictive element is part and parcel of a research programme in social vision: the functional forecast model of emotion expression processing (Weisbuch and Adams 2012). This is a model about the visual processing of social and emotional

---

<sup>6</sup> These considerations gain greater weight within so-called seemings-internalism with regard to epistemic justification (see e.g. McGrath, 2013). Their dialectical import in the present discussion is, however, negligible.

<sup>7</sup> It is important here that the form of the relevant judgment should be “to judge that *p*”, as opposed to “to judge that it seems to be as if *p*”.

cues. The model characterizes our behaviour as a way of responding to what we *see* on the faces of other people with regard to their desires, emotions and intentions.

The functional forecast model of emotion expression processing is a research programme fuelled by the conviction that social perception and vision science should walk together in the pursuit of a rich understanding of the cognitive underpinnings of our on-going social interactions. The hypothesis behind this model is that our visual system has evolved in such a way so as to quickly interpret the social visual cues that convey information about our fellow humans' emotions and intentions. This fast visual interpretation of social cues makes adaptive sense, the model's advocates claim, because it allows us to anticipate other people's most likely behaviour.

Two more features help characterize this version of the social vision programme. First, the functional forecast model relies heavily on a Gibsonian model of perception, where the notion of shared social affordance occupies a central place. The notion of affordance, originally introduced by Gibson (1979), captures the relation between certain environmental features and a subject's abilities to act upon them. So-called grasping-like affordances, for instance, refer, in particular, to the relation between the features of an observed, graspable object and the particular motor abilities that such a potential quality of the object calls for in the observer. The sight of a doorknob affords, according to this view, a reaching-for-pulling motor action to a suitably endowed subject. The notion of shared social affordance expands on the kind of environmental features located at one end of the affordance relationship to include other people's actions and their mental states. Central to this account is the idea that our representation of other people's emotional facial expressions tells us a lot about their most likely immediate behaviour, and this informs our own.

The functional forecast model of emotion expression processing is, secondly, characterized by the view that the fast and automatic impressions that we form of other people's emotions through vision have an influence in the way other socially relevant visual information is processed. Advocates of this model maintain that the perception of, for instance, an angry look on someone's face affects the speed at which we recognize the gender of the person displaying such an emotion. We are much quicker at spotting anger on men's than on women's faces. This is taken to



show that our brain has, over time, developed specific structures that enable fast and accurate detection of properties that indicate the presence of dangerous situations in our environment—and encountering angry men is apparently one of such dangerous situations. Quickly seeing anger on male’s faces has prompted adaptive behaviours in the past. As a result of this, the visual system has evolved to be especially sensitive to the instantiation of such high-level properties, which have proved to be particularly relevant for our fitness. The suggestion is that “the visual system may be calibrated to respond to shared social affordances of compound social cues at the very earliest stages of perceptual processing” (Adams & Kveraga 2015, p. 10). Adams and collaborators contend that “visual perception, and especially social visual perception, is deeply permeated by the observer’s internal state, prior knowledge, and context, and as such, is highly cognitively penetrable” (Adams & Kveraga, 2015, pp. 3–4).

I will come back to the issue of the cognitive penetrability of perception vis-à-vis the visual representations of socially salient high-level properties in the last Section of the paper. In the next two Sections, I discuss whether the functional forecast model of emotion expression processing as well as some other studies in social vision provide empirical support to rich views of visual content.

### **3. Expressing emotions: sensitivity vs. experience**

Facial expressions of emotions and other standard social visual cues, such as gender or race, have traditionally been considered distinct sources of information processed by independent sub-modules. Many of the studies behind the functional forecast model of emotion expression show, however, that they functionally interact with each other at a very early stage of visual information processing so as to signal the same behavioural affordance (Adams & Kveraga, 2015). Adams and collaborators report top-down influences impacting functional integration of social cues at a *very early* stage of visual information processing. But, how early is very early? And how relevant are these results on visual processing for clarifying issues pertaining to the *phenomenology* of visual experience? Much of this research seems to show that the visual system is *sensitive* to properties such as being angry or being threatening. But the interesting question, the one that concerns me here, is whether or not data in this

field give support to the idea that we can visually *experience* such high-level properties.

In order to answer these questions, it would help to distinguish between the pre-attentive and the attentive pathways involved in the processing of visual information. The first type of visual information processing takes place outside conscious awareness. Visual information processing in so-called early vision is pre-attentive in this sense. It is usually characterized in terms of a feed-forward sweep, transmitted through the ventral pathway, and it is taken to last around just 120 milliseconds after stimulus onset (see e.g. Raftopoulos, 2009, p. 33).<sup>8</sup>

When discussing the pre-attentive processing of social visual cues, Adams and Kveraga refer to studies showing the activation of the superior colliculus (SC), which itself has projections to the koniocellular (K) layers of the lateral geniculate nucleus in the presence of threat cues. All these areas of the brain seem to be sensitive to the presence of threat cues and their activation is linked to the activation of typically emotion-related areas, such as the amygdala. Adams and Kveraga hypothesize a convergence of pathways onto the dorsal stream, which is characteristically involved in action-related vision. In discussing this “very early integration” of social visual cues, they claim (Adams & Kveraga, 2015, p. 13):

... magnocellular projections to association areas in the brain (e.g., orbitofrontal cortex) quickly project back to visual centers, helping guide even simple object recognition (Kveraga et al. 2007), and all within the first ~200 ms of visual input (Bar et al. 2006). Thus, we must consider the importance of top-down integration of contextual meaning when guiding and organizing social visual processing. In this way, compound cues can still be quickly

---

<sup>8</sup> The work of Milner and Goodale (1995) provides evidence supporting a ‘dual stream’ model of the human visual system. On the one hand, the dorsal stream seems to provide information for the guidance of skilled visuo-motor action. Dorsal processing is also unconscious. Unlike its ventral cousin, however, it is in all probability entirely outside the scope of attention. On the other hand, the ventral stream subserves conscious perceptual judgement. The case of visual agnosics illustrates how the visual system can unconsciously process information exclusively aimed at guiding skilled sensorimotor behaviour without the subject’s recognition of any the objects involved. These subjects do not seem to have conscious visual experience of the shape and orientation of objects at all, yet they can, in forced choice conditions, engage in action-oriented tasks with objects in ways that match control groups of normal-sighted subjects.

integrated in a functionally meaningful way.

The most relevant studies about this type of pre-attentive processing involve exposure to subliminal stimuli. For instance, negative emotional responses to emotion expressions are observed in black and white subjects when they are subliminally exposed to expressions of joy on the other race's members' face (Weisbuch & Ambady, 2008). In a different study involving blindsight patients, researchers report that their response to bodily expressed emotions presented in their impaired visual field mimics the emotional valence of the unseen body stimuli, so that, for instance, an unseen bodily expression of happiness is matched by a facial expression with greater zygomaticus activity, i.e., an expression of joy, while a greater corrugator activity is observed as a response to unseen bodily fearful expressions (Tamietto and de Gelder, 2008). These studies suggest that, even when visual awareness is absent, emotional bodily expressions are still processed.

All findings involving this pre-attentive pathway indicate that we are able to detect and process emotional social affordances quickly and efficiently. That we are thus sensitive has a clear evolutionary explanation, as the processing of the lower-level properties, which constitute them, would not have the same evolutionary advantage on their own. The detection of emotional social affordances, and not the detection of the low-level factors that constitute them, prepares us to anticipate forthcoming physical danger in our environment. However, it is not at all clear that we could talk here about visual representations and, a fortiori, about representations of high-level properties.<sup>9</sup> The central issue about sensory phenomenology still remains unresolved by evidence involving pre-attentive pathways. Could it be that we do not visually experience high-level properties, even if we are visually sensitive to them? In particular, could it be that our seeming to experience them is the result of a perceptual judgment? To answer these questions, we need to focus on mental events that do involve attentive pathways.

---

<sup>9</sup> It is also not clear that the kind of top-down integration illustrated by these studies could in fact be considered a case of cognitive penetration. Although this is a topic for the final Section, the following seems true: even if the kind of contextual influence Adams and Kveraga discuss has a top-down effect, the timing suggests that it will be due to attentional mechanisms, which exert their influence at roughly the time the authors mention, i.e., after what is typically considered early vision processing time.

Processing of information through attentive pathways involves conscious attention. With attention comes a fine-tuning of the expectations that the visual perception of emotion expressions can generate (Weisbuch & Adams, 2012). The influence of the expression of emotion in gender recognition is one of the better-known studies in this regard. Hess et al. (2009) suggest, for instance, that subjects are more likely to perceive androgynous faces with angry expressions as male. Androgynous faces exhibiting joy or fear are, by contrast, more likely to be perceived as female. In a different study, subjects are shown a combination of male and female faces expressing emotions such as joy, sadness or anger, but also displaying neutral, emotion-free expressions. The task is, again, gender recognition. In this set-up, Hess and collaborators (2009) measured the time it takes subjects to identify males and females. What they found is that female faces expressing anger take the longest to be identified (Hess et al. 2009).

Adam & Kveraga (2015) rely on this type of studies to support the idea that we visually experience socially relevant properties such as gender and emotion. If they are right, our visual *experience* of emotions, instantiated in facial expressions, will be precisely what helps the gender categorization task—or what disrupts it, as the second study shows. We should not forget, however, that the subjects in these experiments, like in any experiment involving categorical identification, respond to the stimuli only after (post-attentive) identification of the relevant properties. It thus becomes rather difficult to assess whether they really visually experience high-level properties, such as anger, as opposed to judging them to be there.<sup>10</sup>

---

<sup>10</sup> A recent study by Firestone and Scholl (2015) questions the classic paper by Levin and Banaji (2006), in which they defend the influence of racial categories on the perception of lightness. Levin and Banaji (2006) report that when looking at faces with exactly the same luminance, Black faces appeared consistently darker than White faces, thus allegedly demonstrating the influence of relatively abstract concepts, such as race, on the perception of lightness. Firestone and Scholl (2015) have designed some experiments in which the images of Black and White faces are blurred to rule out racial recognition. The subjects in these experiments could not perceive the race of the faces and even explicitly judged the faces to be of the same race. Yet, the results about lightness are exactly the same as in Levin and Banaji's (2006) studies. Firestone and Scholl conclude that the difference in perceived lightness is hence not due to a top-down influence from high-level representations about race, but just the result of subtle bottom-up elements within visual processing (Firestone & Scholl, 2015, p. 694).

These results illustrate the complex intricacies involved in assessing the occurrence of cognitive penetration *per se*, but tell us very little about whether or not the content of our experiences is rich and also about the alleged support that the truth of rich theories lends to the cognitive penetrability thesis, which are my only two concerns in this paper. As I will argue in Section 5, the very same reasons that allow us to establish the truly perceptual nature of the representation of some high-level properties speak against the idea that such representations are the result of cognitive penetration.

In interpreting this kind of results, advocates of the functional forecast model of emotion expression processing appeal to the high speed of visual information processing and to selective activation in different brain regions to vindicate the essentially visual nature of the target representations (Weisbuch & Adams, 2012). That, for instance, it takes us much longer to recognize female faces which express anger is interpreted as the result of a cognitive adjustment that takes place partly due to the overriding perceptual influence of the emotional properties on display. In the end, the fact that both pre-attentive and attention driven response pathways are automatic, effortless and unintentional routes is taken to be the crucial factor for distinguishing between visual and post-perceptual events. It is important to keep in mind, however, that, these criteria would be decisive only if the relevant candidate as a post-perceptual event type were judgments, i.e., all-things-considered judgments. Adding the qualification of ‘perceptual’ to ‘judgments’ is no trivial matter. Paradigmatically, (all-things-considered) judgments are conscious events deliberately formed based on evidence or as a result of reasoning. No one thinks of *perceptual* judgments in these terms. Perceptual judgments often—almost always—occur without us realizing it. They also tend to be effortless, automatic and unintentional. Furthermore, other post-perceptual processes that have nothing to do with vision share these properties. Semantic priming, for instance, is typically effortless, automatic and unintentional. Meeting *just* these conditions thus falls short of drawing a principled divide between experience and perceptual judgment.

The general conclusion that we can draw from the discussion of these results is thus the following: *our brain* seems to have evolved to quickly detect certain socially relevant, high-level properties. Sensitivity to high-level properties seems warranted due to the creation, over our evolutionary history, of highly specialized mechanisms. *We* also seem to be quite fast at detecting socially relevant, high-level properties, but the automatic, effortless and unintentional nature of these responses is not enough to plausibly vindicate the truly sensory nature of the phenomenology of these representations—even if these characteristics can be used as a starting point for developing a final set of criteria. In order to secure the view that the phenomenology

that accompanies our representation of certain high-level properties is truly sensory phenomenology, we need something else.

#### **4. The irresistibility of the stimulus criterion**

Even if the studies on gender recognition described above offer no conclusive evidence for the truth of rich theories, they put us on the right track of what will turn out to be a decisive factor for drawing the fine line that separates experiences and perceptual judgments. When considering the experiments run by Hess et al. (2009), I mentioned that one way of interpreting the results is to think that the subjects seem to be irresistibly caught up in the representation of the emotions displayed by facial expressions, and that this overpowering visual engagement accounts for their task performance. I will now argue that this particular feature is the relevant additional criterion we are searching for, since it brings us closer to the notion of irresistibility with which we experience perceptual illusions.

The idea is not new. It has been recently developed by Brian Scholl and collaborators while working on the visual representation of some other socially relevant high-level properties such as animacy or, as they also call it, intentionality (see e.g. Scholl & Tremoulet 2000; Scholl & Gao 2013). They rely on this type of considerations when arguing for the genuine visual nature of our representations of this property. The displays these vision scientists use to investigate the representation of animacy are much simpler than those involving expression of emotion and involve only geometric shapes. In one of the studies (Gao, McCarthy & Scholl, 2010), subjects watch a group of oriented dart shapes and a green circle moving around on a computer screen. The motion of the darts is totally random and independent of the green disc's motion. However, when the dart orientation was adjusted so as to always be pointing directly at the disc, subjects would report the random movement of the darts as chasing, which is taken to be a case of *perceived* animacy (Scholl & Gao, 2013, p. 213).<sup>11</sup>

---

<sup>11</sup> See <http://perception.research.yale.edu/Animacy-Wolfpack/Animacy-Wolfpack-BasicDemo-Pointing.mov> for online demonstrations.

The representation of animacy when viewing these simple movies is “fairly fast, automatic, irresistible and highly stimulus-driven”, Scholl and Tremoulet (2000, p. 299) claim. This is what makes the representation of this property a genuine visual event, as opposed to a post-perceptual one, such as a perceptual judgment. This kind of irresistibility is, I contend, the key factor we need to add to the set of relevant criteria identified above. The representation of animacy is mandatory and irresistible: “mandatory in the way that most visual illusions are” (Scholl and Tremoulet 2000, p. 306). Studies on animacy have also an advantage over gender categorization studies in defining this type of irresistibility. The blatantly artificial nature of the kinematics in these movies makes it easier to appreciate that subjects obviously *know* that the geometric figures on the computer screen are not animate. Yet, as Scholl and collaborators acknowledge, subjects cannot help but perceive animacy given certain conditions. In other words, these studies reproduce a situation in which sensory phenomenology and background knowledge clash. If, under these conditions, sensory phenomenology trumps cognition, as it does, then we have good reasons to think that the content of the visual experience involves the high-level property represented—that the phenomenology of this visual representation is genuine sensory phenomenology.<sup>12</sup>

The idea that we can visually experience some high-level properties thus get reinforced by evidence coming from vision science, but only inasmuch as what I now would like to call ‘the irresistibility of the stimulus criterion’ is met, i.e., only inasmuch as we cannot help but experience such properties—even in the face of conflicting background knowledge. This additional constraint reduces the kind of high-level properties that we could justifiably include in the set of truly perceivable, since this kind of irresistibility seems to be the result of some hardwired or highly specialized perceptual processing—the kind that matters for meeting our biological and social needs.

Advocates of thin views might, however, object that meeting the irresistibility of the stimulus criterion only shows that the processes that retrieve high-level properties are hardwired or specialized, not that their output is phenomenologically noticeable. It

---

<sup>12</sup> I explicitly narrow down the scope of the relevant notion of cognition to that of perceptual judgments further down in this Section.

may be true that our visual system is hardwired or specialized enough to deliver representations of socially and biologically relevant high-level properties, in the same way in which hardwired and specialized visual structures yield representations of depth or motion. Yet, it could be argued, this does not entail that such representations have any phenomenology of their own. It may very well be that our seeming to experience high-level properties is the result of our visually representing low-level properties while, at the same time, inferring the presence of high-level properties due to the cognitive element provided by perceptual judgments.<sup>13</sup>

In order to address this objection, let us go back for a moment to the work on animacy carried out by Scholl and collaborators. First, it is important to notice that the motivation behind their research is, without a doubt, to support the idea that, in talking about the perception of animacy, we are not just speaking loosely or metaphorically. Their experiments are designed to test for the truly perceptual nature of our representations of animacy. The hypothesis they set up to prove is that (Scholl & Gao, 2013, p. 202):

... vision itself may traffic in animacy as well as physical features such as shape and orientation. This is an exciting possibility, suggesting (as with other forms of social perception) that the purpose of vision is (in part) not only to recover the physical structure of the local environment, but also to recover its causal and social structure.

The experiments are set up in such a way so as to check whether or not the representation of animacy falls on the perceptual side of the perception / cognition divide by also making sure that lower-level visual properties cannot be invoked in the explanation of the results.<sup>14</sup> Although early studies involved subjects just watching simple geometric shapes moving around on a computer display, as I described above, the subjects have a much more active role in more recent experiments. They have to complete a task. The display in the relevant experiments is packed with shapes that move very rapidly. There is a chasing shape, referred to as the wolf, which pursues another shape: the “sheep”. The sheep is the only figure that is salient and can be

---

<sup>13</sup> I thank Professor Raftopoulos for pressing me on this point.

<sup>14</sup> Lower-level properties such as motion trajectories, rotational motion or degrees of correlation between the chasing and target shapes.



controlled by the subjects. The task is to move the sheep around to avoid being touched by the wolf. This could be done because the sheep moves faster than the wolf. The details get somewhat technical and they vary in different studies (see e.g. Gao et al., 2009), but the key point is that the only way to detect the wolf is by its spatiotemporal behaviour, since the shape of the wolf is identical to other shapes on the display. Only the sheep is clearly different. In this way, the task is truly a chasing detection task, as there is no other feature through which the wolf could be identified, and as I said earlier, chasing detection is taken to be a form of *perceived* animacy. In each trial, either the subject-controlled sheep escapes from the wolf after a certain time, i.e., it moves away in a random direction as soon as the wolf gets close to it, or is caught—touched—by the wolf. The overall visuomotor performance is measured in terms of the percentage of successful escapes. Crucially, Scholl and Gao claim: “this ability to detect chasing could not be explained by appeal to any lower-level form of perception such as correlated motion or proximity” (Scholl & Gao, 2013, p. 211).<sup>15</sup>

In a different set of studies, Gao and Scholl (2011) examine explicitly the possible influence of judgment on the perception of animacy by randomly interrupting the wolf’s chasing behaviour and alternating it with periods of random movement. The task remains the same: to avoid the wolf. In each trial, the relative percentage of chasing and random motion varied from 0 to 100%. When measuring performance, they found that both small and large percentages of random motion led to comparatively successful performance (although for different reasons: in the former case, the wolf was easier to detect and avoid. In the latter case, the wolf’s chasing behaviour was simply less efficient). At the same time, when the percentage of random motion stayed in the middle range, i.e., when the wolf was still relatively easy to detect, the performance was dramatically compromised.<sup>16</sup> What these results show, according to Scholl and Gao, is that the perceived animacy cannot be due to a conscious *cognitive* event, simply because subjects not only do not know but also (and this is particularly relevant if we thought that the cognitive event might be a perceptual judgment) because they cannot even tell the difference between small and

---

<sup>15</sup> For an online demonstration of one of the already performed trials see <http://perception.research.yale.edu/Animacy-Wolfpack/Animacy-Wolfpack-Game-Pointing-NoCheating.mov>. See also Gao & Scholl, 2011.

<sup>16</sup> For online demonstrations of all studies on interrupting chasing see <http://www.yale.edu/perception/Brian/demos/animacy-ChasingTemporal.html>.

medium percentages of random motion—the difference between, e.g., 20% and 40%. Hence, it does not make sense to think that the subjects' performance is the result of a judgment, not even a perceptual judgment, about animacy, because in order to assume that, we would have to think that the subjects have information that is not even consciously available to them (Scholl & Gao, 2013, p. 213). They conclude (Scholl & Gao, 2013, p. 214):

The point of reviewing these studies here is to note that chasing detection (as a form of perceived animacy) is influenced in systematic ways by rather subtle display parameters, in the form of a psychophysical function (and in ways that do not seem readily explainable by appeal to higher-level judgment) ... [These studies] ... also support a social vision interpretation of perceived chasing in an even more direct way. Beyond the compelling phenomenology of the displays themselves, the data reported in these studies are all measures of visuomotor *performance* rather than explicit reports or ratings. This is an important distinction since overt decisions about what should and should not count as animacy can directly influence reports and ratings ... but have no way to directly influence visual performance of the type studied here.

These results help reinforce the relevance of the irresistibility of the stimulus criterion as a good litmus test for dissociating perception from cognition with regard to certain high-level properties, such as animacy. They do so by showing, on the one hand, that animacy is truly perceived, and not the result of a perceptual judgment and, on the other, that the phenomenal character of this type of visual experiences cannot be explained in terms of the phenomenology of some other low-level properties.

## **5. Cognitive penetrability vis-à-vis rich views**

So far, I have argued that vision science provides empirical support for the truth of rich views, at least with regard to a certain kind of evolutionarily and developmentally relevant high-level properties. I urge, following Scholl and collaborators, that in order to settle the issue of whether we visually perceive high-level properties, as opposed to seeming to visually experience them as a result of a perceptual judgment, some important conditions have to be met: the response pathways have to be automatic,

effortless, unintentional, and perceptually irresistible, in the sense explained above. As we have just seen, the last crucial condition is satisfied only if the relevant high-level properties present themselves to us in experience with the kind of irresistibility that comes from the processes that bring them about being “strongly and directly controlled by specific features of the visual stimulus itself” (Scholl & Gao, 2013, p. 204).

In this last Section, I focus on the relationship between the truth of rich theories and the cognitive penetrability thesis. It is not necessary for my purposes to enter into a discussion about how to exactly characterize the cognitive penetrability view. The account provided in the first Section captures all its important features. To wit, if perception is cognitively penetrable, then it is nomologically possible that there will be changes in a subject’s perceptual experience as a result—not of changes in attention, not of changes in sensory organs or distal stimuli—but as a result of differences in the subject’s background states. We have seen that social vision scientists such as Adams and collaborators take the way in which we seem to represent social visual cues to yield adaptive behavioural responses as evidence for the claim that the subject’s background cognitive states, knowledge, and context influence visual perception—including very early stages of visual information processing. A little more conservatively, but also more engaged with experience as opposed to processing, philosophers like Susanna Siegel (2006, 2010) endorse the view that, if we can visually experience some high-level properties, then visual experience is cognitively penetrable—even if early vision is encapsulated.

It requires only a moment’s reflection to see that the satisfaction of the irresistibility of the stimulus criterion in no way ensures the cognitive penetrability claim. Quite the opposite, that our experiences of high-level properties have a genuine sensory phenomenology seems to be best vindicated—perhaps could only be vindicated—by assuming that the processes responsible for this kind of representations are encapsulated from cognition. Encapsulation guarantees the mandatory, irresistible, fast, automatic and highly stimulus driven nature of the visual representations of those high-level properties of which it makes sense to say that we truly visually represent them. By the irresistibility of the stimulus criterion, the stronger the evidence in favour of the claim that we visually experience some high-

level properties, the less plausible it is that our so experiencing them is the result of the cognitive penetrability of perception.

Interestingly, I take Siegel's (2006) hologram argument to implicitly invoke such a criterion. The argument is put forward precisely to give support to the thesis that the phenomenological difference between the specific experiences a subject has before and after learning to recognize pine trees is genuinely sensory. Here it is (Siegel, 2006, p. 494):

Suppose that you're an expert pine-spotter looking at some pine trees in the forest. Then someone tells you that the forest has been replaced by an elaborate hologram, causing you to cease to dwell on the belief that you're looking at a familiar tree. If an event such as (ii)(d) were what contributed to the phenomenological change before and after acquiring the disposition to recognize pine trees, then we would expect your acceptance of the hologram story to make the hologram look as the forest looked to you before you knew how to recognize pine trees. But intuitively, the hologram could look exactly the same as the forest looked to you after you became an expert. So the familiarity with pine trees does not seem to have its phenomenological effects at the level of belief.

In this argument, the illusory factor takes the form of a hologram. The presence of the hologram nicely illustrates a situation in which the subject's recognitional response is automatic, fast, unintentional and, importantly, irresistibly caught up in certain features of the stimulus—despite clear and immediate evidence to the contrary. The type of cognitive event Siegel chooses for her discussion is the event of dwelling on a belief with the content of “*that kind of tree is familiar*” (this is what (ii)(d) refers to in the quote above), but the same could be said about other cognitive events, including perceptual judgments, which could be formulated as, for instance, “*that is a pine tree*”. On this scenario, sensory phenomenology trumps the effect of cognitive events.

The case is interesting because, as I said, Siegel also, and independently, endorses the claim that the truth of rich theories supports the cognitive penetrability thesis. Yet, the natural conclusion from the hologram argument seems to be quite the opposite.

When Siegel relies on the hologram scenario to vindicate the sensory phenomenology of the expert's visual representations (of, as it turned out, pine trees)<sup>17</sup>, she needs to envision a setting where perceptual judgments (“*that tree* is familiar” or “*that* is a pine tree”) or any other kind of background information shaped into the form of a cognitive event stand on hold. This is surely the point of describing the situation as one in which the subject *knows* in advance that she has a hologram in front of her. The subject knows that what she is looking at is an illusion and knows, for that reason, that not only are there no pine trees in front of her, but no trees at all. The plausibility of the idea that the hologram would look to the expert exactly as the forest did after learning depends precisely on the expert's recognitional disposition being irresistibly caught up in the nuances of the stimuli. The thought experiment argument thus works only inasmuch as it can be taken as a case in which there is no top-down influence from some background explicit cognitive representation on the processing of visual information, i.e., only inasmuch as the experience is not cognitively penetrated.

This appeal to modular perceptual processing as guarantee of genuine rich visual content is bound to strike some as implausible because modularity is usually wedded to innateness and hence the prospects for a successful explanation of development and learning may seem dim. However, contrary to one of the tenets of the traditional modularity thesis (see e.g. Fodor 1983), modules need not be innately specified. There is plenty of neuroscientific evidence that suggests, as it seems plausible, that for many of these computational modules, environmental triggers are necessary, with some developing “from ‘scratch’ over time, based on experience” (Scholl & Tremoulet 2000, p. 306).<sup>18</sup> Modularity places restrictions on the flow of information into and from the modules—their outputs feed into, but not from, cognitive processes underlying higher cognitive faculties. Yet, there are no restrictions on how the information is treated within the module: low-level sensory plasticity and constant adjustment and re-organization of neural connectivity within modules is standard. The encapsulated plasticity of this model accounts nicely for the development of

---

<sup>17</sup> Of course, we would have to put aside alternative explanations in terms of, for instance, attentional or gestalt shifts to plausibly claim that the experience is an experience of pine trees. Siegel does that in arguing for the truth of premises two and three of her phenomenal contrast argument. I remind the reader that I am here taking for granted the truth of these other two premises.

<sup>18</sup> A similar idea is behind Pylyshyn's notion of compiled transducer, i.e., post-perceptual processes that become encapsulated with enough time and repetition (Pylyshyn, 1984; 1999, p. 360).

recognitional abilities of the type illustrated by Siegel's pine tree scenario while upholding the sensory nature of the phenomenology of our visual representations.

There is, of course, another way of satisfying the irresistibility of the stimulus criterion, and hence another way of accounting for the genuine sensory nature of the phenomenology of our visual representations of some high-level properties: to show that the mechanisms responsible for the processing of the relevant information are innate. This is the preferred route for those vision scientists who pioneered the research into the perception of causality, like Michotte (1963), and also for some contemporary researchers working on this topic (see e.g. Leslie 1984, 1994). On this view, the irresistibility of the stimulus criterion is satisfied because the mechanisms responsible for the visual representation of high-level properties are regarded as hardwired. Standard evolutionary constraints account for the development of the relevant neural structures.<sup>19</sup> Although I do not find this route as attractive as the plastic encapsulated model defended by Scholl and collaborators, it does illustrate, once again, how problematic it is to move freely from the truth of rich theories to the truth of the cognitive penetrability thesis.

As I said at the beginning of the paper, I have only engaged with one of the multiple issues that the defence of rich theories of content raises: to show that we do visually represent high-level properties, and do not just seem to represent them as the result of a perceptual judgment. I have relied on empirical research to do so. What I have called the irresistibility of the stimulus criterion has proven to be crucial for distinguishing visual representations from perceptual judgments. I have also argued, against some standard claims to the contrary, that satisfaction of this criterion breaks the assumed connection between rich theories and the cognitive penetrability thesis. It appears that we can safeguard genuine sensory phenomenology for our visual representations of some high-level properties only by making the mechanisms responsible for such representations encapsulated.

---

<sup>19</sup> Interestingly, at places, Adams and collaborators seem to be following this route in the way they present their research, as if the visual system's capacity to represent complex emotional features was just the result of evolution making it more sensitive to them and better wired to other regions of the brain. Although this is highly speculative, I wonder whether their insistence on taking the results of their research in social vision as confirmation of the cognitive penetrability thesis is just a case of conflation between the philosophical thesis, as characterized here, and a visual information processing model, such as predictive coding, where top-down influences are pervasive (see e.g. Hohwy, 2013).

## References

- Adams, R. & Kveraga, K. (2015). Social vision: Functional forecasting and the integration of compound social cues. *Review of Philosophy and Psychology*. DOI 10.1007/s13164-015-0256-1
- Bar, M., Kassam, K. S. Ghuman, A.S., Boshyan, J., Schmid, A. M., Dale, A. M., et al. (2006). Top-down facilitation of visual recognition. *Proceedings of the National Academy of Sciences*, 103, 449–454.
- Brogaard, B. (2013). Do we perceive natural kind properties? *Philosophical Studies*, 162(1): 35–42.
- Crutchfield, P. (2012). Representing high-Level properties in perceptual experience, *Philosophical Psychology*, 25(2): 279–294.
- Firestone, C. & Scholl, B. (2015). Can you experience ‘top-down’ effects on perception? The case of race categories and perceived lightness. *Psychonomic Bulletin and Review*, 22(3): 694–700.
- Fodor, J. (1983). *Modularity of mind*. Cambridge, MA: MIT Press.
- Gao, T., McCarthy, G., & Scholl, B. J. (2010). The wolfpack effect: Perception of animacy irresistibly influences interactive behavior. *Psychological Science*, 21, 1845–1853.
- Gao, T., Newman, G. E., & Scholl, B. J. (2009). The psychophysics of chasing: A case study in the perception of animacy. *Cognitive Psychology*, 59, 154–179.
- Gao, T., & Scholl, B. J. (2011). Chasing vs. stalking: Interrupting the perception of animacy. *Journal of Experimental Psychology: Human Perception and Performance*, 37, 669–684.
- Gibson, J. J. (1979). *The ecological approach to visual perception*, Boston: Houghton-Mifflin.
- Hess, U., Adams, R.B. Jr., Grammer, K., & Kleck, R.E. (2009). Sex and emotion expression: Are angry women more like men? *Journal of Vision*, 9, 1–8.
- Hohwy, J. (2013). *The predictive mind*. Oxford: OUP.
- Kveraga, K., Boshyan, J. & Bar, M. (2007). Magnocellular projections as the trigger of top-down facilitation in recognition. *Journal of Neuroscience* 27, 13232–13240.
- Leslie, A.M. (1984). Spatiotemporal continuity and the perception of causality in infants. *Perception*, 13, 287–305

- Leslie, A.M. (1994). ToMM, ToBy, and agency: core architecture and domain specificity. In L. Hirschfield & S. Gelman (Eds.), *Mapping the Mind: Domain Specificity in Cognition and Culture* (pp. 119–148). Cambridge: Cambridge University Press.
- Levin, D. T., & Banaji, M. R. (2006). Distortions in the perceived lightness of faces: The role of race categories. *Journal of Experimental Psychology: General*, 135, 501–512.
- Lyons, J. (2005). Perceptual belief and nonexperiential looks. *Philosophical Perspectives*, 19, 237–256.
- Machery, E. (2015). Cognitive penetrability: a no-progress report. In J. Zeimbekis & A. Raftopoulos (2015). *The cognitive penetrability of perception. New philosophical perspectives*. New York: OUP.
- Macpherson, F. (2012). Cognitive penetration of colour experience: Rethinking the issue in light of an indirect mechanism. *Philosophy and Phenomenological Research*, 84(1), 24–62.
- McGrath, M. (2013). Siegel and the impact for epistemological internalism. *Philosophical Studies*, 162, 723–732.
- Michotte, A. (1963). *The perception of causality*. Oxford: Basic Books.
- Milner, A. D. & Goodale, M. A. (1995). *The visual brain in action*. Oxford: OUP.
- Nanay, B. (2011). Do we see apples as eatable? *Pacific Philosophical Quarterly*, 93(3): 305–322.
- Price, R. (2009). Aspect-switching and visual phenomenal character. *The Philosophical Quarterly*, 59(236), 508–518.
- Pylyshyn, Z. (1984). *Computation and cognition: Toward a foundation for cognitive science*. Cambridge, MA: MIT Press.
- Pylyshyn, Z. (1999). Is vision continuous with cognition? The case for cognitive impenetrability of visual perception. *Behavioral and Brain Sciences*, 22, 341–365.
- Raftopoulos, A. (2009). *Cognition and perception. How do psychology and neural science inform philosophy?* Cambridge, MA: MIT Press.
- Raftopoulos, A. (2011). Late vision: processes and epistemic status. *Frontiers in Psychology*, 2, 17–28.
- Scholl, B. J. & Tremoulet, P. D. (2000). Perceptual causality and animacy. *Trends in Cognitive Science*, 4(8), 299–309.



- Scholl, B. J., & Gao, T. (2013). Perceiving animacy and intentionality: Visual processing or higher-level judgment? In M. D. Rutherford & V. A. Kuhlmeier (Eds.), *Social perception: Detection and interpretation of animacy, agency, and intention* (pp. 197–230). Cambridge, MA: MIT Press.
- Siegel, S. (2006). Which properties are represented in perception? In T. S. Gendler & J. Hawthorne (Eds.), *Perceptual experience* (pp. 481–503). New York: OUP.
- Siegel, S. (2010). *The content of visual experience*. New York: OUP.
- Siegel, S. (2012). Cognitive penetrability and perceptual justification. *Noûs*, 46(2), 201–222.
- Stokes, D. (2013). Cognitive penetrability of perception. *Philosophy Compass*, 8(7), 646–663.
- Stokes, D. (2015). Towards a consequentialist understanding of cognitive penetration. In J. Zeimbekis & A. Raftopoulos (2015). *The cognitive penetrability of perception. New philosophical perspectives*. New York: OUP.
- Tamietto, M. & de Gelder, B. (2008). Emotional contagion for unseen bodily expressions: Evidence from facial EMG. In *Proceedings from the 8<sup>th</sup> International Conference on Automatic Face and Gesture Recognition* (pp. 1-5), Amsterdam: IEEE.
- Tye, M. (1995). *Ten problems of consciousness*. Cambridge: MA: MIT Press.
- Weisbuch, M. & Adams, R. B. (2012). The functional forecast model of emotion expression. *Social and Personality Psychology Compass*, 6/7, 499–514.
- Weisbuch, M., & Ambady, A. (2008). Affective divergence: Automatic responses to others' emotions depend on group membership. *Journal of Personality and Social Psychology*, 95, 1063–1079.