

Contents lists available at [ScienceDirect](http://ScienceDirect.com)

Consciousness and Cognition

journal homepage: www.elsevier.com/locate/concog

Encapsulated social perception of emotional expressions



Joulia Smortchkova

RUB, Universitätsstraße 150, 44801 Bochum, Germany

ARTICLE INFO

Article history:

Received 8 July 2016

Revised 3 September 2016

Accepted 6 September 2016

Available online 12 September 2016

Keywords:

Social perception

Encapsulation

Rich content view

Emotion perception

ABSTRACT

In this paper I argue that the detection of emotional expressions is, in its early stages, informationally encapsulated. I clarify and defend such a view via the appeal to data from social perception on the visual processing of faces, bodies, facial and bodily expressions. Encapsulated social perception might exist alongside processes that are cognitively penetrated, and that have to do with recognition and categorization, and play a central evolutionary function in preparing early and rapid responses to the emotional stimuli.

© 2016 Elsevier Inc. All rights reserved.

1. Introduction

Two debates are currently at the foreground in philosophy of perception: the debate about encapsulation and the debate about the reach of perceptual content ([Hawley & Macpherson, 2011](#); [Zeimbekis & Raftopoulos, 2015](#)). A special case of the debate about the reach of perceptual content is the debate about *social perception*.

Social perception ([Rutherford & Kuhlmeier, 2013](#)) includes cases when perception is attuned to properties of other individuals, properties that are (in some sense) socially relevant – for example, being a goal-directed action, being an agent, being a fellow human being and so on. One particular case of social perception is emotion perception: seeing others' emotional facial and bodily expressions. Seeing emotional expressions is of obvious relevance to social cognition, because it puts the viewing subject into contact with information about the mental states of her fellow human beings. This information needs to be rapidly and automatically processed in order to allow the subject to produce reactions on the fly, depending on the emotional state of the other (run if the other is expressing fear or anger, for example).

Elsewhere, I have argued that we can perceive agents as agents, as opposed to as objects ([Murez & Smortchkova, 2014](#)), and that we can perceive emotional expressions without mindreading them ([Smortchkova, 2016](#)). In this paper I clarify the difference between two stages of social perception and argue that the detection of emotional expressions is, in its early stages, *informationally encapsulated*. This implies that there are two forms of social perception: one early and encapsulated, and the other late and possibly cognitively influenced. While most discussions of social perception have focused on the latter, the possibility that the former also exists has yet to be given proper consideration.

The central question of this paper is thus how encapsulated social emotion perception is possible. In order to answer this question, I start from a narrow definition of encapsulation, due to Pylyshyn, and argue that some social perception is encapsulated, in the relevant sense. There are already further arguments published supporting the idea of direct perception of emotions, e.g. [Marchi and Newen \(2015\)](#), which especially allow for cognitive penetration to be involved in the process of recognizing a basic emotion. What is missing so far is a detailed analysis of the process of social perception leading to

E-mail address: joulia.smortchkova@rub.de

<http://dx.doi.org/10.1016/j.concog.2016.09.006>

1053-8100/© 2016 Elsevier Inc. All rights reserved.

the recognition of an emotion. I argue that we need to distinguish two stages of social perception, encapsulated social perception and cognitively penetrated or cognitively modified social perception.

Note that the two forms of social perception are not mutually exclusive. Encapsulated social perception might exist alongside processes that are cognitively penetrated, and that have to do with recognition and categorization. The hypothesis that social perception has a dual nature, comprising both encapsulated social vision, and (sometimes) cognitively penetrated social visual recognition, is compatible with the view that I defend in this paper. A view that is incompatible, on the other hand, is the view that emotion recognition needs always to be cognitively shaped by concepts (this is the proposal suggested by Gendron, Lindquist, Barsalou, & Barrett, 2012 who argue that conceptual knowledge shapes the initial processing of emotions).

The discussion will unfold as follows. In Section 2 I briefly introduce the debate on the reach of perception. In Section 3, I introduce Pylyshyn's notion of encapsulation, and defend it from some objections. In Section 4, employing Pylyshyn's notion of encapsulation, I argue that there is encapsulated social emotion perception, which is a version of the rich content view. Finally, in Section 5, I reply to potential objections, distinguish my view from neighboring ones, and draw some consequences concerning the role of encapsulated social perception within social cognition more generally.

2. The reach of perceptual content

While there is widespread agreement that low-level properties (such as colors, shapes and orientations) are represented in perceptual content, and computed by early visual processes, it is controversial whether high-level properties (broadly those properties that are not obviously sensory, for example causation, meaning, or emotional expressions) are also part of perceptual content (Hawley & Macpherson, 2011). According to the *poor content view* only low-level properties can be represented in perception; according to the *rich content view* also high-level properties can be represented in perception. There is no uncontroversial way to draw the line between the two sorts of properties, and an array of intermediate positions on the issue are possible.

The debate about the reach of perceptual content is primarily a debate about the reach of conscious perceptual experience (Siegel, 2011) or phenomenal content (Briscoe in Zeimbekis & Raftopoulos, 2015). The debate, however, can also be framed as concerning the contents of perception *tout court*, conscious and non conscious, and the properties that can, consciously or non consciously, be represented in perceptual contents. Indeed, one question that is asked in the debate is whether perceptual content is restricted only to low-level sensory properties, or whether it can also include higher-level properties. Thus, I prefer not to restrict the debate to conscious experience, but focus on the properties that can be represented in perception (see also Burge, 2010). Indeed, the debate about the reach of perception is properly understood to be about which properties can enter into perceptual content. *De facto* this leaves open the possibility that both conscious and unconscious perceptual representations are concerned. Unconscious perceptual representations might play a central role in social perception by guiding fast reactions in response to perceptual stimuli, without conscious access to the contents of the representations (see Section 5).

Therefore, we can distinguish between two versions of the debate: a debate about the reach of perceptual conscious experience and a debate about the reach of perceptual content *tout court*, conscious and non conscious. I will be concerned with the question of perceptual content in general and the properties that are represented by it (this opens the possibility that there might be a disconnection between the outputs of early vision and the conscious contents of perception, an issue to which I return in Section 5). For this reason, I will freely appeal to experimental evidence that taps both into the subject's conscious visual experiences and into the subject's unconscious perceptual representations.

If one adopts an extreme poor content view, the only properties represented in perception are those whose processing results from direct stimulation of the sensory organs (Lyons, 2011). Bayne (who is himself a partisan of the rich content view) introduces what may be the mainstream view:

Proponents of what I shall call the conservative view hold that the phenomenal character of visual experience is exhausted by the representation of low-level properties – color, shape, spatial location, motion, and so on. Conservatives give similar accounts of other perceptual modalities: the phenomenal character of audition is exhausted by the representation of volume, pitch, timbre, and so on; the phenomenal character of gustation is exhausted by the representation of sweetness, sourness, and so on. The phenomenal world of the conservative is an austere one.

[Bayne in Hawley & Macpherson, 2011, p. 16]

Tye, who also endorses the poor content view, writes: “Thereby, it seems plausible to suppose, they [the output representations of visual processing] represent those features, they become sensations of edges, ridges, colors, shapes, and so on. Likewise for the other senses.” (Tye, 1995, p. 103). This kind of view is similar to a poor content theory that states that the only properties that enter perceptual visual content are those represented by Marr's 2 and ½ -D sketch (Marr, 1982): shape, color, spatial disposition, and movement, but not depth; for example, Prinz claims that the upper limit for perceived conscious states are attended 2 and ½ -D representations (2006a) (also Raftopoulos, 2009).

According to a richer approach, 3-D properties are also represented, even if they are not in plain view, such as the occluded part of a cup and depth-properties in general. These representations are complex and Marr says they are in a format that is available for recognition (Marr, 1982, Chapter 5). This approach is intermediate between the poor and the rich content view.

As a way to introduce the rich content approach, its proponents often use certain intuitive examples that appeal to a phenomenal contrast between an experience in which some high-level property is present and an experience in which it is missing (Siegel, 2011). Among *rich content positions*, an array of options is available. Current candidates for high-level properties are causation (Butterfill, 2009; Ducasse, 1967); visual objects (Scholl & Leslie, 1999); agency and actions (Gao, Newman, & Scholl, 2009); natural or artificial kinds (being a face, being a tomato, being a duck, being a coin, being a clock) (Siegel, 2011); affordances (Nanay, 2011). Emotional expressions are examples of high-level properties: they are not primary and secondary sensory properties. Hence, to claim that emotions are perceptually represented is to endorse a version of the rich content view.

3. Encapsulated early vision

Cognitive penetrability is, roughly, the claim that one's cognitive states (for example, beliefs and desires) influence what one sees. The influence can be at the level of conscious outputs, i.e., perceptual experiences or at the level of the processes involved, some of which are unconscious (Pylyshyn, 2003; Stokes, 2013). The influence can also be synchronic or diachronic: it can happen at a given moment depending on the current cognitive states of the subject, or it can be the result of a slow change over time due to cognitive factors. In the latter case, however, rather than 'cognitive penetrability' strictly understood, it is more accurate to speak of 'perceptual learning' (Pylyshyn, 2003).

Cognitive impenetrability or *encapsulation* (I use the terms interchangeably) is the opposing claim that perception (or a part of perception) is immune to such influences. Two central issues in the debate surrounding cognitive impenetrability are: firstly, how many notions of cognitive penetrability there are, and secondly, what kind of empirical evidence is best suited to prove the existence (or nonexistence) of cognitive penetrability. Here, I will not rehearse these debates, extended overviews of which can be found in (Firestone & Scholl, 2015; Vetter & Newen, 2014; Zeimbekis & Raftopoulos, 2015).

In this paper, I will adopt the following definition of synchronic cognitive penetrability, due to Pylyshyn:

[A] system is cognitively penetrable when the function it computes is sensitive, in a semantically coherent way, to the organism's goals and beliefs, that is, it can be altered in a way that bears some logical relation to what the person knows (...). Note that changes produced by shaping basic sensors, say by attenuating or enhancing the output of certain feature detectors (perhaps through focal attention), do not count as cognitive penetration because they do not alter the contents of perceptions in a way that is logically connected to the contents of beliefs, expectations, values, and so on, regardless of how the latter are arrived at.

[Pylyshyn, 1999, p. 343]

Three things are important in this definition: first, penetrability is defined functionally; second, the crucial feature of penetrability is the notion of "semantic coherence", or "logical dependence"; third, it is claimed that penetrability is the mark of cognitive processes (such as decision making), while impenetrability is the mark of perceptual (or rather early visual) processes.

In defining early vision, one has to be able to distinguish it from later visual processing and cognitive processing. There are two definitions of early vision: one that is *functional* (Pylyshyn, 2003) and another that is based on *the neural correlates* of early vision and appeals to activations of primary visual areas (Raftopoulos, 2009). The latter approach distinguishes early vision from other kinds of processing via their relative timings. While I prefer to work with the functional definition, I will nevertheless appeal to neural evidence in order to study properties of functionally defined early vision. The appeal to neural evidence is common in the encapsulation debate, given that it is plausible that there is some (relatively) systematic correspondences between functional and neural levels of description. For example, Fodor – a functionalist – appeals to neural data in order to argue in favor of the modularity of perception (Fodor, 1983).

According to Pylyshyn, functionally defined early vision is encapsulated:

[...] I will argue that visual perception, in the everyday sense of the term, does indeed merge seamlessly with reasoning and other aspects of cognition. But the everyday sense of the term is too broad to be of scientific value [...]. I will argue that within the broad category of what we call "vision" is a highly complex information processing system, which some have called "early vision", that functions independently of what we believe. This system individuates, or picks out, objects in a scene and computes the spatial layout of visible surfaces and the 3D shape of the objects in the scene."

[Pylyshyn, 2003, p. 51]

This does not mean that there are no influences from cognition on perception, but that these influences can only occur at some *particular stages*:

Our hypothesis is that cognition intervenes in determining the nature of perception at only two loci. In other words, the influence of cognition upon vision is constrained in how and where it can operate. These two loci are: (a) in the allocation of attention to certain locations or certain properties prior to the operation of early vision (...). (b) in the decisions involved in recognizing and identifying patterns after the operation of early vision. Such a stage may (or in some cases must) access background knowledge as it pertains to the interpretation of a particular stimulus."

[Pylyshyn, 1999, p. 343]

To be more precise, Pylyshyn allows some top-down processing as part of early vision, while distinguishing this top-down influence from cognitive penetrability by cognitive states such as beliefs, desires and goals:

The early vision system is a significant part of vision proper [...] (i.e., it involves the computation of most specifically-visual properties, including 3-D shape descriptions). [...] Many of these computations involve what is called top-down processing [...] What this means is that the interpretation of parts of a stimulus may depend upon the joint (or even prior) interpretation of other parts of the stimulus, resulting in global-to-local influences such as those studied by Gestalt Psychologists. Because of this some local vision-specific memory may also be embodied in early vision..

[Pylyshyn, 1999, p. 343–344]

To show cognitive penetrability with respect to early vision functionally defined one would need to demonstrate that (semantically coherent) information from higher-level cognitive processes influences early vision's elaboration of visual appearances in on-going tasks (and not only top-down influence such as filling-in).

Recently, data from neurophysiology have been used to challenge Pylyshyn's claims (reviews can be found in [Ogilvie & Carruthers, 2015](#); [Vetter & Newen, 2014](#)). While I agree that this is a promising strategy, I nevertheless do not think that these challenges are fully successful, once one considers the precise scope of Pylyshyn's claims – which a brief consideration of these challenges should help to clarify.

Take, for instance, the line of evidence that comes from neurophysiological effects of color terms on categorical perception. This evidence shows very early activations in visual processing in response to the presence of color terms ([Thierry, Athanasopoulos, Wiggert, Dering, & Kuipers, 2009](#) for Greek speakers; [Winawer et al., 2007](#) for Russian speakers; [Mo, Xu, Kay, & Tan, 2011](#) for Mandarin speakers), and is used to argue that there is cognitive penetrability of color terms on the perception of colors ([Ogilvie & Carruthers, 2015](#))

The first worry I have with this line of argument is that learning a linguistic term takes time and is a case of *diachronic* (and not *synchronic*) influence. Most diachronic changes of visual processing, however, are accepted by the defenders of Pylyshyn-style encapsulation ([Pylyshyn, 2003](#); [Raftopoulos, 2009](#)): they argue only against *synchronic* influences, when perception is influenced by the subject's current goals and desires. Diachronic cases of cognitive penetrability (such as learning a color term in infancy and using it throughout adult life) are cases of perceptual learning. After perceptual learning, both post-perceptual and perceptual modifications take place, so it is not surprising to see changes in perceptual processing with time ([Pylyshyn, 2003, pp. 86–88](#)).

Note that all the relevant studies show the need for some time for the effects to take place, and interestingly, there are no studies on the effects of color term acquisition in the case of adults who learn a new language with a new color term, but only studies that show the disappearance of the effect when adults learn a new language that does not have the color term ([Thierry et al., 2009](#)).

Much stronger evidence in favor of cognitive penetrability would be provided if there were cases where learning a new color term, such as the Russian голубой (goluboy) for light blue, changed the activity in early visual processing *instantaneously*, showing that the acquisition of a color term makes one see things differently. But such evidence is lacking.

Moreover, changes in visual processing due to one's ability to categorize similar colors differently does not seem to show that beliefs and goals influence perception. Classifying a shade of blue as синий (sinji) or голубой (goluboy) in a certain linguistic community seems hardly to be a case of beliefs influencing perception depending on the organism's *needs and goals* at the time of the task. Classifications are *stable* in a certain linguistic community. Cognitive influence on encapsulated vision would have to be task congruent and depend on the needs of the organism for the influence to be a clear-cut case of cognitive penetration (that is congruent with the selected hypothesis).

The second worry is a very general one. To interpret neurophysiological data one needs to know the exact mapping between neural areas and psychological abilities – a knowledge that we largely lack so far (see [Firestone & Scholl, 2015](#) for discussion). The presence of feedback connections is not clearly linked to the existence of cognitive processes that depend on, or are consistent with, the subject's current goals and desires, and not facilitation processes for the categorization and elaboration of the stimulus (see also Section 4).

This does not mean that all perception is encapsulated – processes involved in recognition are plausibly influenced by background knowledge, but it does suggest that early stages of perception are not dependent on the subject's goals, beliefs and desires. The examples above concern color perception, but a similar claim can be made for social perception as well.

4. Encapsulated social perception

Is there a connection between encapsulation and the reach of perceptual content? To claim that early vision is cognitively impenetrable seems *prima facie* a rather conservative view of perception. For instance, Vetter and Newen write:

He [Pylyshyn] conceptualizes early visual cortex as an encapsulated module which receives possibly attentionally modulated responses from the eye as input and outputs shapes, size, colors and other typically "early" visual features.

[[Vetter & Newen, 2014, p. 65](#)]

Pylyshyn, however, is neutral on the kind of properties that are computed by early vision:

Early vision depends on how the visual system was wired up by evolution as well as on biological, chemical, and physical principles, and on the incoming patterns of light, but on little else. It would not be a great exaggeration to say that early vision – the part of visual processing that is prior to access to general knowledge – computes just about everything that might be called a “visual appearance” of the world except the identities and names of the objects.

[Pylyshyn, 2003, p. 51]

Concerning the output representations of early vision, Pylyshyn writes:

The precise nature of the output in specific cases is an empirical issue that we cannot prejudge. There is a great deal that is unknown about the output, for example, whether it encodes nonvisual properties, such as causal relations, or primitive affective properties, like “dangerous”. In principle, the early-vision system could encode any property whose identification does not require accessing general memory, and in particular that does not require inference from general knowledge.

[Pylyshyn, 2003, p. 136]

This passage clearly leaves open the possibility of rich encapsulated early vision. The aim of this section is to provide an actual case of rich encapsulated vision, in the framework of social perception.

In this section I will argue that there is a part of vision (functionally defined early vision, or more probably a functional subset of early vision) that encodes some social properties (in the case at hand, emotional expressions) without on-line access to general knowledge and central inferences. Note that this does not mean that further processing of the emotional stimuli is immune from cognitive influences, but it supports the existence of encapsulated social perception that is immune to background knowledge.

The claim that we can see emotional expressions is not entirely new. Usually, however, these claims focus on further processing of the emotional stimuli, involved in recognition, and not on the very early mechanisms of social perception. For example, Butterfill (2015) has argued that perception of emotional expressions might be a case of *categorical perception*. In his paper, however, he is not concerned with questions related to cognitive penetrability, and in particular his position is perfectly compatible with beliefs and other cognitive states influencing the *categorization* of the emotional stimuli.

Marchi and Newen (2015) introduce a proposal in some respects similar to Butterfill’s, in the sense that they are concerned with emotion categorization (they call it ‘emotion recognition’) and they claim that ‘cognitive penetrability [...] shapes our perception of socially relevant information.’ (Marchi & Newen, 2015, p. 4). Later on, however, they clarify that, in their view, it is emotion *recognition* that is shaped by cognitive penetrability.

Both Butterfill, and Marchi and Newen, do not consider the possibility that, even if some instances of social perception might be cognitively influenced (in the case of emotion *recognition*, for instance), the early visual processing of the emotional stimuli might not be.

I think that two aspects of social perception need to be distinguished: one aspect is an early, initial processing of emotional stimuli that happens “on the fly” and is encapsulated; the second aspect concerns further processing – such as categorization and/or recognition of the emotional stimulus – which is probably influenced by background knowledge. The role of the former would be to prepare fast reactions to the emotional stimulus, necessary for the survival of the subject.

The data on cognitive influence on emotion perception Marchi and Newen appeal to concern *recognition* and not early visual processing of emotions: in an experiment the reading of emotional texts make the subjects classify a face as angry instead of as fearful because of the influence of the background story. Independently, they also appeal to research by Bar (2003, 2009) in favor of the quick activation of higher level cognitive areas before a visual stimulus is recognized, to neurophysiological data on V1 and V2 activations in response to illusory contours, and on feedback connections in object (and not emotion) perception in order to establish that “[cognitive penetrability] is physiologically possible” (Marchi & Newen, 2015, p. 3).

Crucially, however, there is no connection between the first set of evidence (on background knowledge changing recognition) and the latter set of evidence (quick involvement of higher level areas in recognition), so the inference that the latter, neurophysiological data, might support or add physiological plausibility to the first set of data has not yet been established.

First of all, Bar’s proposal concerns cases where a quick response is necessary, and this is not the case of the story experiment. But there is a stronger objection: in Bar’s model there is a projection of a partially analyzed representation (such as a blurred image) from early visual areas to the prefrontal cortex (Bar, 2003). In the case of a fearful-angry face, since the two emotions share more perceptual similarities between them, as in respect to a happy face, it might be that the initial visual stimulus is already analyzed as being one emotion, but the story introduces an alternative categorization. In order to claim that there is cognitive penetrability for the early stage, one would need to show that there is no independent initial elaboration of the visual stimulus, something that even Bar’s research does not show.

It might well be that in *recognizing* a facial expression as an expression of happiness, one needs background knowledge and stored information (this seems a very sensible proposal). And it might also be that recognition is a much earlier and faster process than normally thought, but the *initial elaboration of the emotional stimulus* might not need background knowledge at all.

Thus, Marchi and Newen’s paper cannot be used to support the more radical proposal that even the initial elaboration of the emotional stimulus depends on background beliefs. For this reason, we can, even in the case of social perception, distinguish between encapsulated early and unencapsulated later portions of the process.

Before I further elaborate this view, I need to dispel a *prima facie* worry. Indeed, one could argue that it is obvious that emotional expressions are not perceived because they do not impact on the retina. ‘Dangerous’ or ‘fearful’ cannot be visually represented because they do not reflect light (see also McNeill, 2012).

This objection identifies perceivable with sensory. It is, however, at odds with our best scientific theories of vision: even in its early stages, vision is dedicated to retrieving information that is in some sense, absent, that is information that does not directly impact the retina. For instance, in constructing Marr’s 2 and ½ sketch, visual computations build up information about prospectual properties of objects that are not reflected on the retina (that only receives a ‘flat’ projection) (Marr, 1982). Other examples are when vision represents the backside of objects, allowing us to grasp a cup as a three dimensional object, or when it interprets a square overlapping with another square as being two objects, or when absent features, such as the sides of a Kanisza triangle, elicit neural activation at the early stages of visual processing (Stanley & Rubin, 2003).¹

This does not mean, however, that higher-order properties are on the same processing level as sensory features. There might be an asymmetric dependency between the two: a fearful face is a configuration of low-level sensory properties while at the same time not reducible to this configuration. But this is just an extension of extant models of object perception to social perception: A Gestalt configuration is a set of low-level properties differently grouped, that asymmetrically depends on those low level properties. Gestalt properties by themselves do not impact the retina, but, according to commonsense in psychology of perception, they are part of our perceptual experience (Palmer, 1999). Similarly, “fearful” and “dangerous” can be perceived, despite not being purely sensory.

The hypothesis proposed in this section is that there is an encapsulated processing of high-level properties in vision that is independent from the subject’s semantic knowledge. Crucially, this hypothesis is supported by experimental evidence and compatible with Pylyshyn’s notion of encapsulation. This support is given by the following evidence: social perception stimuli are processed before attention is allocated, thus eliminating cognitive penetrability via attention, they arise before categorization and recognition, and do not show any semantic dependence on cognitive states.

The first source of evidence comes from well-known face and body-specific perceptual processes. Evidence on faces comes from the discovery of the FFA (fusiform face area) (Kanwisher, McDermott, & Chun, 1997) and while there is disagreement on which properties this area exactly correlates with, this evidence is supported by the existence of neurons responding selectively to faces at 90–100 ms (Kiani, Esteky, & Tanaka, 2005). Evidence on body perception converges on the existence of areas selectively activated by the visual presence of bodies (Servos, Osu, Santi, & Kawato, 2002). This is initial and weak evidence since these effects could be due to some low-level properties rather than to face and body perception *per se*. Note, however, that the presence of *selectively* responding neurons at stages of perception reliable correlated with early vision points toward the idea that faces *do* elicit the response.

The second source of evidence comes from attentional effects: faces (Gliga, Elsabbagh, Andravizou, & Johnson, 2009), bodily biological motion (Shi, Weng, He, & Jiang, 2010), and intentional actions (Neufeld, Brown, Lee-Grimm, Newen, & Brüne, 2016) are powerful attention-grabbers. This means that they strongly impact attention distribution: for example, in inattention blindness experiments, participants better detect changes in face (Ro, Russell, & Lavie, 2001) and body (Downing, Bray, Rogers, & Childs, 2004) stimuli, compared to objects. There is also evidence of rapid visual attention allocation for fearful bodies and facial expressions (Jessen & Kotz, 2011).

The appeal to empirical evidence based on attention needs to be clarified: attentional effects can be either bottom-up (attentional guidance by external factors, independently from the subject’s will, based on the saliency of certain stimuli) or top-down (voluntarily direction of attention to a stimulus based on the subject’s goals and knowledge) (Katsuki & Constantinidis, 2013). All the effects cited in the previous paragraph are examples of *bottom-up* attention. The fact that some social stimuli attract attention might mean that these high level properties are processed *before* attention is allocated, and are thus bottom-up attention grabbers. Similar groups of low-level properties (for example inverted faces) do not have such an effect on attention. If stimuli matched for low level features, but with and without the high level social feature have different effects on attention (one attracts attention, the other does not) then it is a sign that the high-level property is represented. Moreover, since they attract attention and are pre-attentionally processed, this helps to rule out indirect, top-down attentional, cognitive influence on (this part of) social perception.

The third source of evidence comes from two studies that show a very fast and automatic processing of bodily emotional expressions: one study found that seeing fearful body language rapidly freezes the observer’s motor cortex (Borgomaneri, Gazzola, & Avenanti, 2015); the other study shows the presence of different stages of the involvement of motor cortex during perception of emotional body language (Borgomaneri, Gazzola, & Avenanti, 2014).

In the 2015 experiment, TMS (transcranial magnetic stimulation) was applied to healthy subjects whose task was to observe pictures of actors in different kinds of bodily poses (the face was concealed). These body poses could be either emotional (happy or fearful) or neutral. TMS was applied to the right M1 in a first set of subjects and to the left M1 in a second set of subjects. The subjects also had to perform a task, i.e., answer the question “what did you see?” with a forced-choice verbal reply. Forced choices were happy, fear and neutral. Concerning neurophysiological data, the authors tested M1 excitability in the time frame between 100 and 125 ms. They found a very early modulation of the motor cortex in reaction to emotional stimuli, motor suppression was stronger in seeing fearful bodies than in seeing happy and neutral bodies, and stronger in

¹ Sometimes this effect is presented as a case of cognitive penetrability. But in this case what is the background belief? That triangles have three sides? These effects, just as the Müller-Lyer effect, seem to be *independent* from a subject’s beliefs and knowledge.

seeing happy bodies than in seeing neutral bodies. They interpret this as the sign of a temporary reduction of motor preparation in order to give the subject the possibility to monitor the environment and the dangerous stimulus.

The 2014 experiment was concerned with distinguishing between different stages of motor cortex involvement in emotion perception. The authors found that when participants observed happy, fearful and neutral bodies there was a reduction of activity in the right motor cortex at already 150 ms. This reduction was correlated with the subjects reporting distress. Moreover, applying TMS at 150 ms to the right M1 disrupted recognition of emotions (but not to the left M1).

These experiments show that there is an automatic, fast, early motor activation in response to a visual emotional stimulus; and that motor involvement plays a role in the processing of the stimulus, because the interference with motor cortex interferes with a behavioral recognition task. The early processing of the emotional stimulus does not seem to be modulated by top-down attention and it plays a role in directing attention from the bottom-up. Thus, these experiments provide some support for the hypothesis that in vision, emotional expressions play a role independent from conceptualization at very early stages. As a consequence, it is plausible that there is a part of vision that encodes some high-level properties and that is nevertheless encapsulated.

This is not to say that these experiments show that we perceive specific emotions. It might turn out that the properties that are perceived early are not 'happy, fearful, and neutral' but 'potential threat and no threat' (i.e. the presence or absence of danger). Spelling out the exact kind of social property that can be perceptually represented early is a matter of future research.

Furthermore, this extension fits nicely with recent evidence on the impact of emotion stimuli on early vision (Ferneynough, Kim, Phelps, & Carrasco, 2013). If social perception is part of the encapsulated visual module, then free flow of information *within the module* is to be expected between emotion perception and early visual processing.

Given the facts about early activation, attentional direction and independence from the subjects' goals, beliefs and desires, this processing of the emotional stimulus, which gives rise to an early motor preparation, is a good candidate for being the encapsulated part of social perception. A noteworthy conclusion is that to endorse the rich content view (social perception) does not imply giving up encapsulation. However, a careful distinction needs to be made between full recognition of an emotion, and early processing of the emotional stimulus, which provides a first gist, ready to be further analyzed, categorized, and recognized.

5. Potential objections and conclusion

I now consider some potential objections to the hypothesis sketched above. The shared denominator of all these objections is that they claim that there might already have to be cognitive penetrability in place for the observed effect to obtain, or that the effect is not a case of genuine early processing.

The first objection appeals to the fact that there is very early involvement of the motor system, and that therefore the early visual processing is not *pure*. This objection can be spelled out in two ways. The first way is to say that the involvement of the motor system influences the recognition of the visual stimulus: according to one model of object recognition, which describes the possible time course of object recognition, such recognition arises at around 150–300 ms (Johnson & Olshausen, 2003), and the motor system contributes to the pre-150 ms object recognition. The second way is to say that all motor involvement is a case of cognitive penetrability, whether it participates or not in object recognition. Therefore, by definition, visual processing of the emotional expression is not encapsulated. First of all, note that, following Pylyshyn's approach, the encapsulated part is what happens *before recognition* in the very early stages of processing. In reality, the central issue is rather whether the motor system can be coupled with early vision without this constituting cognitive penetrability.

There is nothing in the original definition of encapsulation to prevent this possibility. On the contrary, "[i]n certain cases the module should be viewed as a visuomotor system, rather than as a strictly visual system." (Pylyshyn, 2003, p. 127). This means that interactions between the motor system and vision is compatible with encapsulation.

Moreover, the subjects did not formulate any *intention* before the activation of their motor system (contrary to cases discussed in Wu, 2013). The motor preparation, which is congruent with the display, is triggered by the emotional stimuli. An interesting hypothesis is whether this motor preparation helps in later processing such as the recognition of the stimulus.

The second objection appeals to two facts: first, that emotion perception is a case of categorical perception and second, to data about processing effects in categorical perception that suggest cognitive penetrability. As we saw, there are data that seem to suggest the dependence of the categorization effect on the presence of a color term in a language. Since emotion perception, like color perception, is a case of categorical perception (see Butterfill, 2015; Marchi & Newen, 2015) then we have *prima facie* evidence for believing that emotion perception is cognitively penetrated as well.

The reply to such an objection depends on whether all emotion perception is categorical, even in its early stages. This in turn depends on whether the outputs of early visual processing are the representations that are categorized or not. In other words, the question is whether the outputs of early vision coincide with our visual conscious experiences ready to be categorized and classified.

Pylyshyn connects early vision's representations to MOT (multiple object tracking), where the tracking occurs within the module. So the representations elaborated by early vision are uncategorized objects that are not yet full-blown objects, but "proto-objects" (Pylyshyn, 2007). Proto-objects are tracked pre-attentively and need not be conscious. They are arguably

represented differently from our common-sense consciously perceived objects (see Murez, Smortchkova, & Strickland, in prep.). If the outputs of early vision *are not* the contents of perception that are categorized, but the categorized contents are the result of further processing, then categorical perception is not a real challenge to encapsulated emotion perception, because it is not categorical.²

The third objection challenges the background definitions of this paper. I have relied throughout this paper on classic definitions of cognitive penetrability, due to Pylyshyn. According to this objection, in so doing, I am using a notion of cognitive (im)penetrability that is too restrictive, and that should consequently be abandoned, because there is a variety of different notions of cognitive penetrability (Vetter & Newen, 2014). This objection threatens to beg the question: the crux here is to show that there is cognitive penetrability even at the early stages of processing of the emotional stimulus.

My aim in this paper was to start from a clear notion of encapsulation, due to Pylyshyn, and show that it can be extended to social perception. Whether there are other sorts of cognitive penetrability is not an issue that I need to discuss here. Emotional stimuli attract attention *bottom-up* and in turn influence some motor and behavioral responses that are stimulus-congruent, but in order to do so, they must be first perceptually processed *pre-attentively*. One would need to show that even this pre-attentive perceptual processing of the emotional stimulus at an early stage is cognitively shaped – something that has not yet been demonstrated.

One's favorite notion of cognitive (im)penetrability depends on many background assumptions and overall theoretical considerations, in particular concerning how one draws the perception/cognition divide. For instance, Marchi and Newen (2015) claim that if encapsulation proves to be false, this will blur the distinction between perception and cognition. Others (Ogilvie & Carruthers, 2015) think that the rebuttal of encapsulation does not erase the distinction between perception and cognition, it only shows that encapsulation is not the right criterion for drawing the line.

My claim is that a part of vision being encapsulated subserves evolutionary functions, such as the ability to detect threats on the fly, independently from what one thinks or wants. Given that early social perception is encapsulated, the phenomenon of *pareidolia* (seeing familiar patterns – such as faces – where there aren't any), that is sometimes presented as a case of cognitive penetrability (Marchi & Newen, 2015) is, on the contrary, a case in favor of encapsulation. It is not because I believe that there is a face on Mars that I see a face on Mars, but I believe that there is a face on Mars because I cannot stop my perceptual system from processing the face-like stimulus. Thus, while it is always possible to argue with the traditional conception of cognitive penetrability I have been assuming, the onus of proposing a better definition rests on the shoulders of my opponents.

While other philosophers and cognitive scientists have introduced views similar to the one presented here, the present account differs considerably from them. First of all, I do not appeal to cognitive penetrability to argue in favor of the rich content view (contrary to Crutchfield, 2011 who appeals to synchronic cognitive penetrability), not even as a case of diachronic perceptual learning (as in Siegel, 2011). The reason is that encapsulated social perception is not similar to a case when one learns, e.g., to recognize pine trees: it is an evolutionary ancient mechanism, parts of which are plausibly innate (see Section 4). This does not mean that *all* social perception is fixed: just as in the case of colors there is some learning involved for the categorization of color categories, similarly for cases of emotion perception, there might be perceptual learning involved.

Secondly, while the view presented here is close in spirit to Scholl and Gao (2013) and Firestone and Scholl (2015), I think that their notion of encapsulation is too strong: in their view, the processes that gives rise to *conscious perceptual experiences* is encapsulated, and social perception belongs to high-level vision. They thus seem to assume that the outputs of encapsulated vision coincide with conscious perceptual experiences. In my account I distinguish between the outputs of early vision and conscious perceptual contents, and argue that the former (but not the latter) are encapsulated and rich. The reason is that I distinguish between early social perception and high-level social perception. The form of encapsulated social perception that I defend is modest, but this does not mean that it is unimportant.

6. Conclusion

To conclude, in this paper I presented a case for the extension of the reach of perceptual content within the framework of encapsulated vision and I have formulated the hypothesis that there is a precise sense in which one can justifiably talk of *encapsulated social perception*.

What are the consequences of such a view? First of all, this proposal is empirically assessable and is based on current experimental data. It might turn out that some parts of perception are cognitively penetrated, for example during the processing occurring between early vision's outputs and recognition, but this does not, in and of itself, challenge the proposal that the encapsulated part of perception can be extended. Secondly, this challenges the assumption that if only a portion of perception is encapsulated then modularity is not "an interesting organizing principle" (Prinz, 2006b, p. 34). Indeed, encapsulated social properties, besides being crucial for survival, also have an immediate bearing on how humans process the world around them, for instance how they detect threatening or friendly stimuli.

² One might also want to distinguish categorical perception (the warping of a continuous stimulus into two distinct and discrete categories) from recognition and categorization of emotions. The first is in most cases independent from beliefs and goals, even if susceptible of diachronic modifications due to perceptual learning (Section 2). The latter is easily prey of cognitive influences, as Marchi and Newen (2015) show in their paper.

Finally, it is worth stressing that keeping encapsulation does not mean giving up on the idea that perception is connected to the social world. Quite on the contrary, it actually means doing even better justice to the paramount importance of the social world within our lived experience: some parts of perception are already social *before, and without*, social influence.

Acknowledgments

This work has been supported by the Volkswagen Foundation project “Situated Cognition. Perceiving the World and Understanding other minds” led by Professor Tobias Schlicht. I am grateful to Albert Newen and Francesco Marchi for helpful discussions, to Michael Murez for his invaluable help and support in writing this paper, to the audience of the conference “Cognitive Penetration and Predictive Coding” in Bochum for their questions on a first version of this paper, and to two anonymous reviewers for their insightful comments.

References

- Bar, M. (2003). A cortical mechanism for triggering top-down facilitation in visual object recognition. *Journal of cognitive neuroscience*, 15(4), 600–609.
- Bar, M. (2009). The proactive brain: Memory for predictions. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 364(1521), 1235–1243.
- Borgomaneri, S., Gazzola, V., & Avenanti, A. (2014). Temporal dynamics of motor cortex excitability during perception of natural emotional scenes. *Social Cognitive and Affective Neuroscience*, 9(10), 1451–1457.
- Borgomaneri, S., Gazzola, V., & Avenanti, A. (2015). Transcranial magnetic stimulation reveals two functionally distinct stages of motor cortex involvement during perception of emotional body language. *Brain Structure and Function*, 220(5), 2765–2781.
- Burge, T. (2010). *Origins of objectivity*. Oxford University Press.
- Butterfill, S. A. (2009). Seeing causings and hearing gestures. *Philosophical Quarterly*, 59(236), 405–428.
- Butterfill, S. A. (2015). Perceiving expressions of emotion: What evidence could bear on questions about perceptual experience of mental states? *Consciousness and Cognition*, 36, 438–451.
- Crutchfield, P. (2011). Representing high-level properties in perceptual experience. *Philosophical Psychology*, 25(2), 279–294.
- Downing, P. E., Bray, D., Rogers, J., & Childs, C. (2004). Bodies capture attention when nothing is expected. *Cognition*, 93(1), B27–B38.
- Ducasse, C. J. (1967). How literally causation is perceivable. *Philosophy and Phenomenological Research*, 28(December), 271–273.
- Ferneyhough, E., Kim, M. K., Phelps, E. A., & Carrasco, M. (2013). Anxiety modulates the effects of emotion and attention on early vision. *Cognition & Emotion*, 27(1), 166–176.
- Firestone, C., & Scholl, B. J. (2015). Cognition does not affect perception: Evaluating the evidence for “top-down” effects. *Behavioral and Brain Sciences*, 1–77.
- Fodor, J. (1983). *Modularity of mind*. MIT Press.
- Gao, T., Newman, G. E., & Scholl, B. J. (2009). The psychophysics of chasing: A case study in the perception of animacy. *Cognitive psychology*, 59(2), 154–179.
- Gendron, M., Lindquist, K., Barsalou, L., & Barrett, L. F. (2012). Emotion words shape emotion percepts. *Emotion*, 12, 314–325.
- Gluga, T., Elsabbagh, M., Andravizou, A., & Johnson, M. (2009). Faces attract infants' attention in complex displays. *Infancy*, 14(5), 550–562.
- Hawley, K., & Macpherson, F. (2011). *The admissible contents of experience*. John Wiley & Sons.
- Jessen, S., & Kotz, S. A. (2011). The temporal dynamics of processing emotions from vocal, facial, and bodily expressions. *Neuroimage*, 58(2), 665–674.
- Johnson, J. S., & Olshausen, B. A. (2003). Timecourse of neural signatures of object recognition. *Journal of Vision*, 3(7), 4.
- Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The fusiform face area: A module in human extrastriate cortex specialized for face perception. *The Journal of Neuroscience*, 17(11), 4302–4311.
- Katsuki, F., & Constantinidis, C. (2013). Bottom-up and top-down attention different processes and overlapping neural systems. *The Neuroscientist*, 1073858413514136.
- Kiani, R., Esteky, H., & Tanaka, K. (2005). Differences in onset latency of macaque inferotemporal neural responses to primate and non-primate faces. *Journal of Neurophysiology*, 94(2), 1587–1596.
- Lyons, J. (2011). Circularity, reliability, and the cognitive penetrability of perception. *Philosophical Issues*, 21(1), 289–311.
- Marchi, F., & Newen, A. (2015). Cognitive penetrability and emotion recognition in human facial expressions. *Frontiers in Psychology*, 6.
- Marr, D. (1982). *Vision: A computational approach*. Freeman.
- McNeill, William E. S. (2012). On seeing that someone is angry. *European Journal of Philosophy*, 20(4), 575–597.
- Mo, L., Xu, G., Kay, P., & Tan, L.-H. (2011). Electrophysiological evidence for the left-lateralized effect of language on preattentive categorical perception of color. *Proceedings of the National Academy of Sciences*, 108(34), 14026–14030.
- Murez, M., Smortchkova, J., & Strickland, B., (in preparation). The Mental Files Theory of Singular Thought: A Psychological Perspective.
- Murez, M., & Smortchkova, J. (2014). Singular thought: Object-files, person-files, and the sortal PERSON. *Topics in Cognitive Science*, 6(4), 632–646.
- Nanay, B. (2011). Do we see apples as edible? *Pacific Philosophical Quarterly*, 92(3), 305–322.
- Neufeld, E., Brown, E. C., Lee-Grimm, S. I., Newen, A., & Brüne, M. (2016). Intentional action processing results from automatic bottom-up attention: An EEG-investigation into the Social Relevance Hypothesis using hypnosis. *Consciousness and Cognition*, 42, 101–112.
- Ogilvie, R., & Carruthers, P. (2015). Opening up vision: The case against encapsulation. *Review of Philosophy and Psychology*.
- Palmer, S. E. (1999). *Vision science: Photons to phenomenology*. MIT press.
- Prinz, J. J. (2006a). Beyond appearances: The content of sensation and perception. In T. Gendler & J. Hawthorne (Eds.), *Perceptual experience* (pp. 434–460). Oxford University Press.
- Prinz, J. J. (2006b). Is the mind really modular? In R. J. Stainton (Ed.), *Contemporary debates in cognitive science* (pp. 22–36). Blackwell.
- Pylshyn, Z. W. (1999). Is vision continuous with cognition? The case for cognitive impenetrability of visual perception. *Behavioral and Brain Sciences*, 22(3), 341–365.
- Pylshyn, Z. W. (2003). *Seeing and visualizing: It's not what you think*. A Bradford Book.
- Pylshyn, Z. W. (2007). *Things and places: How the mind connects with the world*. The MIT Press.
- Raftopoulos, A. (2009). *Cognition and perception: How do psychology and neural science inform philosophy?* MIT Press.
- Ro, T., Russell, C., & Lavie, N. (2001). Changing faces: A detection advantage in the flicker paradigm. *Psychological Science*, 12(1), 94–99.
- Rutherford, M. D., & Kuhlmeier, V. A. (2013). *Social perception: Detection and interpretation of animacy, agency, and intention*. MIT Press.
- Scholl, B. J., & Gao, T. (2013). Perceiving animacy and intentionality: Visual processing or higher-level judgment. *Social perception: Detection and interpretation of animacy, agency, and intention*, 4629.
- Scholl, B. J., & Leslie, A. M. (1999). Explaining the infant's object concept: Beyond the perception/cognition dichotomy. *What is Cognitive Science*, 26–73.
- Servos, P., Osu, R., Santi, A., & Kawato, M. (2002). The neural substrates of biological motion perception: An fMRI study. *Cerebral Cortex*, 12(7), 772–782.
- Shi, J., Weng, X., He, S., & Jiang, Y. (2010). Biological motion cues trigger reflexive attentional orienting. *Cognition*, 117(3), 348–354.
- Siegel, S. (2011). *The contents of visual experience*. Oxford University Press.
- Smortchkova, J. (2016). Seeing emotions without mindreading them. *Phenomenology and the Cognitive Sciences*, 1–19.

- Stanley, D. A., & Rubin, N. (2003). fMRI activation in response to illusory contours and salient regions in the human lateral occipital complex. *Neuron*, 37(2), 323–331.
- Stokes, D. (2013). Cognitive penetrability of perception. *Philosophy Compass*, 8(7), 646–663.
- Thierry, G., Athanasopoulos, P., Wiggert, A., Dering, B., & Kuipers, J.-R. (2009). Unconscious effects of language-specific terminology on preattentive color perception. *Proceedings of the National Academy of Sciences*, 106(11), 4567–4570.
- Tye, M. (1995). *Ten problems of consciousness: A representational theory of the phenomenal mind* (Vol. 282). MIT Press.
- Vetter, P., & Newen, A. (2014). Varieties of cognitive penetration in visual perception. *Consciousness and Cognition*, 27, 62–75.
- Winawer, J., Witthoft, N., Frank, M. C., Wu, L., Wade, A. R., & Boroditsky, L. (2007). Russian blues reveal effects of language on color discrimination. *Proceedings of the National Academy of Sciences*, 104(19), 7780–7785.
- Wu, W. (2013). Visual spatial constancy and modularity: Does intention penetrate vision? *Philosophical Studies*, 165(2), 647–669.
- Zeimbekis, J., & Raftopoulos, A. (2015). *The cognitive penetrability of perception: New philosophical perspectives*. Oxford University Press.