The predictive mind and the experience of visual art work

Ladislav Kesner*

Department of Art History, Masaryk University, Brno, Czech Republic

*Correspondence:
Ladislav Kesner, Department of Art History, Masaryk University, Arne Nováka 1, Brno, Czech Republic
e-mail: kesner@mail.muni.cz

Among the main challenges of the predictive brain/mind concept is how to link prediction at the neural level to prediction at the cognitive-psychological level and finding conceptually robust and empirically verifiable ways to harness this theoretical framework toward explaining higher-order mental and cognitive phenomena, including the subjective experience of aesthetic and symbolic forms. Building on the tentative prediction error account of visual art, this article extends the application of the predictive coding framework to the visual arts. It does so by linking this theoretical discussion to a subjective, phenomenological account of how a work of art is experienced. In order to engage more deeply with a work of art, viewers must be able to tune or adapt their prediction mechanism to recognize art as a specific class of objects whose ontological nature defies predictability, and they must be able to sustain a productive flow of predictions from low-level sensory, recognitional to abstract semantic, conceptual, and affective inferences. The affective component of the process of predictive error optimization that occurs when a viewer enters into dialog with a painting is constituted both by activating the affective affordances within the image and by the affective consequences of prediction error minimization itself. The predictive coding framework also has implications for the problem of the culturality of vision. A person's mindset, which determines what top–down expectations and predictions are generated, is co-constituted by culture-relative skills and knowledge, which form hyperpriors that operate in the perception of art.

Keywords: predictive coding, predictive error minimization, art perception, art experience, affective affordance, reward in art, culturality of vision

INTRODUCTION

The old notion of perception as unconscious, knowledge-driven inference (Helmholz, 1860/1962) or hypothesis testing (Gregory, 1980), which asserts that the brain actively anticipates upcoming sensory input rather than passively registering it, has now been recast in the terms of contemporary neuroscience, and has recently undergone an unprecedented revitalization. It has been linked to the idea of the Bayesian brain – a probability machine that constantly makes predictions about the world and then updates them based on what it senses. According to predictive coding model of perceptual inference, subjects try to infer the causes of their sensations based on multi-level generative models of the world (Rao and Ballard, 1999; Kersten et al., 2004; Bar, 2009; Friston, 2009; Rauss and Pourtois, 2013). Predictions (priors) about the probable cause of sensory input, generated in higher levels of processing hierarchy, are continuously updated by prediction errors which code mismatches between expected and actual data. Recently, some theorists have further extended the predictive coding framework (or predictive error minimization, PEM) from *brain* to *mind*, applying it to a variety of cognitive mechanisms beyond perception itself (Hohwy, 2013). According to philosopher Clark (2013) the Bayesian approach constitutes the "grand unified theory of mind as perception, action and attention are all in the same business of reducing sensory prediction error resulting with our exchanges with environment" (Clark, 2013, p. 21). As Friston (2013, p. 1330) notes, the higher-order aspects of inference in the brain represent the frontiers of theoretical neurobiology. The main challenge is to link prediction at the neural level with prediction at the cognitive-psychological level and to find conceptually robust and empirically verifiable ways to harness this theoretical framework toward explaining higher-order mental and cognitive phenomena, including the subjective experience of aesthetic and symbolic forms. A step in this direction has been taken in the recently proposed "tentative prediction error account of visual art" (TPEA; Van de Cruys and Wagemans, 2011).

It is the aim of this article to further extend the application of the predictive coding framework to the visual arts. To do this in a productive way requires moving from the rather abstract level of theory formulation to a more detailed discussion of particular instances of response to a work of art, that is, it requires testing the theoretical model against a specific case study. Consequently, the article is organized into three sections. First, I shall articulate some objections to and problems with the current formulation of the prediction error account of art perception. Based on these observations, the second section presents a case study of an encounter with a particular painting in order to expand upon some key some aspects of predictive coding in visual art. I shall focus especially on the problem of the emotional response to a work of art within the PEM framework. Finally, I shall point out further implications of this theoretical model for the question of the social and cultural determination of vision.

KEY CONCEPTS

Bayesian brain

The notion that perceptual impression results exhaustively from neuronal groups computing probability (P) of a hypothesis (H) given data (D) according to Bayes rule for P(H| D):

P(H|D) = P(H) × P(D|H)/P(D)

The left-hand part of the equation corresponds to *posterior probability*, that is probability that the hypothesis is true given the new data.

P(D|H) refers to the *likelihood of compatibility* of new data with prior hypothesis and

P(H) or *prior* denotes one's prior expectations about the probability of the hypothesis and drives the interpretation of sensory data.

Predictive coding

According to hierarchical predictive coding framework of perceptual inference, brain does not passively register sensory input but actively anticipates it. Neuronal representations at the higher levels of processing hierarchy generate "conditional expectations of perceptual causes" or predictions (priors), based on information in memory and context, about the probable cause of sensory input. These predictions are communicated to lower sensory areas by feedback connections where they are compared with incoming sensory data. Mismatches between expected and incoming data form prediction errors which are passed by feedforward connections to update or change higher-level representations, creating the most reasonable interpretation of incoming sensory input. There are different interpretations of the predictive coding framework and its neurobiological implementation in literature (e.g., Rao and Ballard, 1999; Friston, 2002, 2008; Lee and Mumford, 2003; Knill and Pouget, 2004; Yuille and Kersten, 2006; Summerfield et al., 2006; Bar, 2009; Friston and Kiebel, 2009; Rauss and Pourtois, 2013).

Prediction error minimization = minimization of uncertainty about the sensory data.

According to PEM model, brain continuously minimizes its prediction error, that is the discrepancy between predictions (expectations) about the sensory data and the actual data.

I set out from the assumption that it is neither likely, nor indeed desirable, for a prediction coding framework to ultimately form a general biological "theory of art" or aesthetic perception – a holy grail of neuroaesthetics, to which art historians and theorists remain justifiably indifferent. Even if predictive coding could assume the position of a grand unified brain theory (and some doubt it can; see, e.g., Bowers and Davis, 2012; Anderson and Chemero, 2013), there are strong grounds for being skeptical about the prospect of any comprehensive brain theory becoming also a grand biological theory of art and art perception. What such a theory could provide, however, is a framework with which to revisit some long-standing questions in image studies, the psychology of art and art history and out of which further empirical studies of art perception could evolve.

THE PREDICTION ERROR ACCOUNT OF ART – CURRENT FORMULATION AND ITS CONTRADICTION

The basic tenet of Van de Cruys and Wagemans (2011) model is that a temporary state of unpredictability (or prediction error) is important for the emergence of perceptual pleasure vis-a-vis a work of art. Understanding perception in terms of predictions means that it is possible for perceptual configuration to induce

different sequences of affect and to do so partly independently of the particular content of perception (TPEA, 1040). Accordingly, artists are supposed to intentionally create incongruities (perceived as prediction errors) that may not be possible in a natural visual environment, and viewers are able to tolerate and even enjoy the unpredictability because they expect to be surprised in their encounters with art (TPEA, 1041). By delaying prediction confirmation, artists create a positive affect: the viewer quickly runs into incongruities, which presumably generate an arousal aimed at reducing prediction errors. It is this incompatibility (or prediction error) that is the source of some of the emotionality of a work of art. In other words, artists intuitively attempt to strike the optimal balance between predictability and surprise. The mental effort required of a person in order to cope with the prediction error is a condition sine qua non for registering the perceptual pleasure of a Gestalt formation (prediction error reduction). According to Van de Cruys and Wagemans (2011): "Only by using minimal prediction errors painters can ensure that viewers will obtain their reward and not give up prematurely. Final gratification postponed as long as the artist has hidden in the painting enough micro reward the viewer can discover..." (TPEA, 1050).

This concept is appealing for a number of reasons. First, it offers a much stronger explanatory framework than the neuroscientific and psychological models of art experience, which focus exclusively on a bottom–up account of visual processing (Zeki, 1999; Shimamura, 2013) and neglect or downplay the role of the top–down, inferential activity of brain/mind. Second, although not specifically stated by the authors, the model responds to the so-called "dark room problem," the apparent paradox that, in order to minimize surprise, agents should avoid sensory stimulation altogether and should proceed directly to the least stimulating environment and stay there; they should take up a position in the nearest "dark room" and never move again. Neatly summarized: *avoid surprises and you will last longer.* Predictive coding theorists offer a simple solution to the dark room scenario: prior beliefs render dark rooms surprising. That is, agents that predict rich stimulating environments will find the "dark room" surprising and will leave at the earliest opportunity. The postulate of surprise minimization therefore by no means inhibits subjects from active, exploratory behavior and novelty-seeking, including presumably an aesthetic experience (Friston et al., 2012b; Schwartenbeck et al., 2013).

Third, the model is apparently compatible with some well-established and respected art-historical theories, notably Gombrich's (1960) theories of the prognostic character of the perception of pictures, the role of the beholder's share and the viewer's inferences in perception, and his notion of the artist working through a cycle of scheme and correction. Indeed, Gombrich's (1960) famous maxim that "[t]o read the artist's picture is to mobilize our memories and experience of the visible world and to test his image through tentative projections" leaves the door open to the Bayesian brain perspective; an interesting challenge would be to recast Gombrich's account in the explanatory terms of predictive coding, but this will not be pursued here. Furthermore, it provides scientific footing to some philosophical interpretations of aesthetic experience, most notably Gadamer's (1975/2004) hermeneutical

scenario, which highlights the activity of the perceiving subject vis-à-vis the aesthetic object. Gadamer describes the nature of this exchange as ongoing and dynamic, suggesting that understanding is an open-ended or at least an extended process that does not end the moment the representational content is identified or the information embedded in the work of art obtained, but also includes a more complex response and understanding: "all encounter with the language of art is an encounter with an unfinished event and is itself part of this event...There is no absolute progress and no final exhaustion of what lies in work of art" (Gadamer, 1975/2004, p. 85).

PREDICTION ERROR OPTIMIZATION VERSUS THE RUSH TO THE OBJECT

Any theory should be measured against empirical findings, in this case on what we know about how people actually interact with works of art. Seen in this light, the main objection to tentative prediction error immediately becomes apparent: the model describes an ideal situation, which represents a distinct minority of actual encounters with art works. This does not invalidate the theory as such, but addressing this discrepancy paves the way to pursuing some crucial aspects of predictive mind in art experience.

When observing people's reactions in front of works of art in a museum or gallery, one quickly notices that many viewers are content with performing the simple act of recognition, displacing the visual substance of the work as soon as possible with the kind of understanding that evidently does not prompt or entice further viewing. This situation is eloquently captured in an anecdote recounted by Rudolph Arnheim:

> "I remember once watching a teacher with her second-graders approaching a piece of abstract sculpture in a museum gallery. 'What is this?' asked the children. The teacher, very unsure herself, went closer and looked at the label. 'Gift of Mr. and Mrs. Oscar Verlinski,' she read. The children, satisfied, moved to the next object."
>
> (Arnheim, 1992, p. 61)

Besides providing a depressively true account of the nature of many encounters between visitors and works of art in a museum, of the way in which people harvest meaning from works of art, Arnheim's (1992) observation captures some essential aspect of the process in which viewers make sense of art images. It describes the moment when the visual object, presented to view and soliciting an understanding, gets an "answer." As a paradigmatic example of an act of translation, or displacement, it highlights the violent and terminal substitution of the visual presence of an art work with a label in the mind's eye of the viewer. The moment the exhibit becomes the "gift of Oscar Verlinski" in the eyes of the inquiring children, it ceases to be a sculpture, an object endowed with visual and aesthetic interest. In other words, the story neatly describes the psychological reality of surprise minimization in an encounter with an art work. By fixating on identification, the viewer's interrogation of an object is effectively concluded. And as much as if in the given case the object had been answered "correctly" (as e.g., "abstract sculpture," or "work by Anthony Caro"). The moment the visible content of the art object is recognized as what it depicts, the viewing is concluded, instead of

a series of exchanges between the image and the viewer opening up and ushering in reciprocal play, inviting the viewer into the rich possibilities of dialog[1]. Very often, moreover, the viewer recognizes the content of the image as its subject – that is, the culturally ingrained capacity for recognizing subjects in pictorial content is grafted onto the biologically ingrained propensity for perceptual identification – and thereby translates or displaces the pictorial meaning[2]. This mode of grasping a painting or sculpture can be seen as an extension of the evolutionary programmed operation of visual awareness, the role of which is to produce the best current interpretation of the visual scene, in the light of past experience, either our own or of our ancestors, and to make it available to the parts of the brain that plan and execute voluntary motor outputs (Crick and Koch, 1995). The biologically adaptive function of the human visual system, which ontogenetically evolved in the service of assuring physical and social survival, cannot be easily switched off when applied to works of art.

The viewing of art images is thus frequently characterized by a tension between unconscious reaction, a biologically ingrained need on the part of the viewer to understand a visual scene unambiguously, and the fact that stopping at the moment of this initial displacement, "answering" the image in the moment of identification, threatens to annihilate its most valuable asset – the possibility for deeper engagement. One could say that works of art are often victims of the biologically hard-wired operation of the human system for recognition and identification, victims of the "rush to the object" (to use Michael Baxandall's apt term; Baxandall, 2003, p. 130). Recognition and the resulting swift displacement, moreover, need not just relate to the visible content of the image or object, but, depending on the level of expertise and the viewer's ability to categorize, might rather involve the recognition of a style or an author[3].

Different works of art naturally present different types of constraints. On the one hand, some of them are perceptually incomplete or unstable depictions – such as many examples of modern art, whose very identity depends on the active inferential involvement of the viewer (Gamboni, 2002). The challenge of resolving visual ambiguity and understanding what the painting or sculpture represents, makes the viewer aware of the interpretive process. While some aspects of the labor of this filling-in might be unconscious, it often also requires a conscious effort from the viewer, in which the process of making sense of an

---

[1]Support for this view can be found in some psychological investigations that found that abstract and semi-abstract paintings were rated as more meaningful when they were accompanied by their titles as opposed to without titles. This finding is consistent with the assumption that titles carry information and this information aids the meaningful interpretation of the painting (Russell and Milne, 1997).

[2]As current experiments have demonstrated, it takes no longer to determine an object's category than simply to detect its presence (see Grill-Spector and Kanwisher, 2005).

[3]Identification, resulting in the substitution of the artists' name or particular style in place of the object itself, is likely to occur in cases of non-representational art (Hasenfus et al., 1983). Summarizing earlier findings, as well as their own empirical studies, Leder et al. (2004) argue that experts interpret art work on the basis of art-specific concepts, while non-experts or naïve viewers tend to use stereotypical responses, probably drawing upon their "everyday repertoire" of concepts, such as "personal experience, feelings and a similarity of personal surroundings" (Augustin and Leder, 2006).

image is brought fully into awareness. In many artistic styles, on the other hand, the representational content is recognized and identified swiftly (in the viewer's phenomenology instantaneously), without him being aware of any effort[4]. In such a case the beholder's share consists of the unconscious deployment of perceptual-cognitive routines that render the recognition of visual object. Perceptually unstable or otherwise challenging images generate prediction errors related to the recognition and identification that the viewer (as argued by Van de Cruys and Wagemans, 2011) may find challenging and rewarding to resolve, much like ambiguous perceptual occurrences in natural vision. The problem, as just described, is that once the viewer arrives at a certain "solution" (with or without some external source of information, such as a label), there is often no need to go any further beyond the recognition of content and subject and to engage in a prolonged experience. This has been repeatedly observed in empirical studies of museum visitors. For instance, psychologists who were commissioned to undertake one of the most comprehensive empirical studies of aesthetic experience wrote in their book: "*Most people, when confronted with a work of art, simply do not know what to do. Without a goal, a problem to solve, they remain on the outside, unable to interact with the work*" (Csikszentmihalyi and Robinson, 1990, p. 83).

In many (if not most) encounters with works of art, the beholder's share thus consists of sensory and conceptual predictions which are generated and resolved at the level of scene/subject recognition and identification, without proceeding to higher levels of cognitive and emotional predictions. Predictions, that is, *expectations* about the immediate sensory environment are transferred to the perception of works of art, without any adjustment being made to expectations about the *special kind of sensory environment* that works of art (and museum or gallery setting) demand and present for the benefit of the viewer. This leads to two alternative interim conclusions. The more radical (and skeptical) one suggests that the PEM account of real-world perception, which involves "explaining away' the driving (incoming) sensory signal by matching it with a cascade of predictions pitched at a variety of spatial and temporal scales" (Clark, 2013, p. 7), is incompatible with a deep experience of a work of art, as this kind of object resists being simply 'explained away.' Accordingly, the basic tenet of the Bayesian framework that the brain minimizes unpredictability is at odds with the ontological nature of works of art, which defy predictability. The less radical and more accommodating option suggests that the deeper experience of works of art requires that viewers be able to tune or adapt their prediction mechanisms to the specific visual environment of the museum and that they have the ability to create or temporally sustain a productive flow of predictions across hierarchical levels, from low-level sensory to abstract semantic and conceptual levels. It demands that viewers consciously or implicitly come to realize that the objects in artistic representations are not like natural objects, that they are depictions, which means that they

are mediated by the representational structure of a given artistic medium[5].

CASE STUDY

To obtain a better idea of how predictions operates in art perception, it is necessary to support a theoretical model with at least a minimal phenomenological account of actual viewing experience of specific work of art. I shall do this with a remarkable painting by contemporary American painter Vincent Desiderio, which is an analytically rewarding case of an image that is not perceptually unstable, so the viewer is able to swiftly recognize the depicted objects; however, identification does not yield to understanding in the sense of a well-understood subject or an established symbolic/iconographic theme. The viewer's (I assume a motivated viewer, willing and able to endure more than a fleeting encounter with the painting) initial response to the picture within PEM is plausibly explained by the mechanisms by which sensory predictions subserving recognition are exercised. In addition to contextual modulation, the low spatial frequency information in the image that encodes its gross properties triggers object and category information, which in turn serves as a prediction template to guide further sensory processing, i.e., the high spatial frequency perception that conveys details (Bar, 2003, 2009; Rauss et al., 2011; Panichello et al., 2013)[6]. But despite the relative ease with which most individual objects in a pictorial space can be identified, the beholder is left puzzled as to what is transpiring in the depicted scene, what the meaning of the painting is. There is no help to be obtained from applying the usual strategy of seeking external guidance, for looking at the caption and learning that the title of the work is *Spiegel im Spiegel* (**Figure 1**) offers no explanation or definite clue. The opacity of the painting in terms of its recognizable meaning will lead the motivated viewer to further attempt to minimize prediction error by engaging in an active search, and doing so by using both general strategies of PEM at the same time: by changing sensory input through action (that is, performing an active visual search in front of the painting) and by alternating the predictions through perception, that is, making the model fit the sensory input (Hohwy, 2013). Without a high-order generative model or predictive "template" (Summerfield et al., 2006; Summerfield and Egner, 2009) against which to match the observed sensory data, the viewer will try both to adjust his expectations and simultaneously to explain away those visible aspects that resist imminent understanding. While most objects in the pictorial space can indeed be easily identified as such, there are mismatches waiting to be resolved – most notably understanding the expressions and gazes of the faces of both figures,

---

[4]Observers can comprehend a variety of perceptual and semantic information within the first fixation, at a mere glimpse, occurring in less than 100 ms. When consolidated, within a few 100 ms, the conceptual gist of what is being viewed can be represented in the form of a verbal description of the scene's image, a description that includes both what was perceived and what was inferred.

[5]These alternatives could be also situated within the broader question (and ongoing discussion) whether and to what extend human observers are Bayes optimal. Critics point out that the basic tenet of Bayesian models that human behavior can be explained solely in terms of optimal probabilistic inference in any given situation is inadequate for explaining psychological phenomena (see e.g., Jones and Love, 2011; Bowers and Davis, 2012). Others, however, argue that humans are not "Bayes optimal" in an absolute sense (Clark, 2013, p. 9), but relative to the immediate context of the situation.

[6]However, Panichello et al. (2013), in their overview, add a cautionary note: "Given that the specific neural processes that give rise to conscious perception remain unclear, it is difficult to conjecture precisely how predictive feedback influences the contents of awareness."

FIGURE 1 | Vincent Desiderio, Spiegel im Spiegel (2010), oil on canvas.
© Vincent Desiderio, courtesy of Marlborough Gallery, New York.

or precisely defining the area in which they come closest to each other.

Prolonged engagement with the image unleashes a cycle of PEM, which enters conscious awareness and may be verbalized (and even socialized if the viewer interacts with a companion) as, for example: what exactly is the bandaged man clutching in his hand? What is the function of the white band around the boy's neck? In particular, what do the figures' expressions convey? What state is the boy in – is he sleeping, comatose, dying? The content of the pictorial scene at the level of individual objects will be almost completely resolved as observed information is "iteratively reconciled across multipe levels of visual processing hierarchy, resulting in a progressive reduction in prediction error as the visual system settles on a single perceptual interpretation of the sensory input" (Summerfield and Egner, 2009, p. 406). However, the identification of individual objects does not directly enable inferences about their relationships and hence about the meaning of the whole scene. It thus triggers a succession of higher-level, semantic predictions, which unfold through an ongoing exploration of the painting and concern above all the nature of the interaction between the depicted figures (is their enigmatic implied relationship indeed what the painting is "about"?). Similarly, the representational status of the scene remains uncertain: is it to be perceived as a real scene, or as a fantasy image (dream, vision) of the artist, or is to be seen as the state and content of the momentary state of consciousness of one of the depicted protagonists? Moreover, for an attentive viewer, the process of prediction error minimization does not transpire just at the level of (virtual) depicted objects and their relationships. Rather, there is a concomitant awareness of how objects arise from the painterly medium – in this case from Desiderio's rich and deeply textured brushwork, with individual marks oscillating between representational, mimetic and non-mimetic function. On that level, one of the mismatches relates to the perception of the spatial setting and its uncertain representational status. The ambiguous

spatial construction thwarts and frustrates attempts to recognize it in terms of some kind of empirical environment; the viewer's perception of the virtual pictorial space as some sort of non-descript enclosed space, delineated by the ground and the wall, easily shifts to an awareness of the picture plane and markings, which seem to deny any claim to mimetic optical veracity.

AFFECTIVE PREDICTIONS

The entire process of individual meaning-making out of this harrowing (others used even more expressive terms such as "strikingly frightening and nightmarish") painting, within the predictive coding framework, is steeped in affective significance. While individual feelings will naturally differ, a central part of a viewer's encounter with the image is his or her emotional reaction to it. Recent models in affective neuroscience insist that affective meaning is not something superadded on the perceptual act pure and simple; rather, while the brain is engaged in object recognition, it concurrently extracts affective value from the observed scene, particularly its valence (Russell, 2003; Barrett and Bar, 2012; see also Jaspers, 1913/1963, p. 255 for an early formulation). Affective (and interoceptive) predictions occur in rapid timescales and concurrently, not as a separate step, consequently conscious percepts are intrinsically infused with affective value[7]. According to Barrett and Bar's (2012) model, the brain's prediction about the meaning of a visual sensation includes some representation of affective impact (or similar sensations) from the past. Moreover, the affective state of the perceiver at the moment of the initial encounter exerts top–down effects on visual processing (Anderson et al., 2011), thereby constraining the formation of predictions from the sensory level upward.

Some authors have further argued that viewers recognize the "emotional gist" of the scene, referred to as the global emotionality, whereby the scene can be rapidly identified as positive, negative, or neutral without having to explore the individual (local) features of the scene. In the given case, the phenomenology of the viewing experience does not support the notion of instantaneous emotional gist, but rather that of an affective reaction that unfolds throughout the entire duration of seeing. Its initial stages appear to be related to the perception of affective affordances within the pictorial space, especially bodily postures and facial expressions of both depicted figures[8]. As recent eye-tracking studies have demonstrated, the eye initially tends to fixate on emotional objects rather than more salient, neutral ones, and emotional saliency can override visual saliency defined by features such as intensity, color and orientation (Humphrey et al., 2012; Niu et al., 2012). At the same time, people process the emotional implications of biologically emotional stimuli related to survival or

---

[7]Barrett and Bar (2012) note that the affective value of the object is sometimes represented as a property of the object, and other times it is represented as a person's reaction to that object.

[8]I use the concept of affective affordance in much the same sense as the notion of affective salience, defined as "the tendency of the item to stand out relative to its neighbors due to an association between the semantic meaning and emotional arousal" (Todd et al., 2012).

reproduction automatically, but engage in more elaborative processing when confronted with socially emotional stimuli (Sakaki et al., 2012).

The generally weak correlation between emotions and their predicted expressions (Fernández-Dols and Crivelli, 2013) is in this case further augmented by the fact that neither figure's expression can be quickly interpreted, as they are both partly obscured and partly indeterminate. While neither figure constitutes emotional affordance of the kind usually used in neuroimaging experiments (that is, a stimulus with imminent threat or fear value), both are imbued with a saliency that involves the basic repertoire of emotional reactions. The boy's indeterminate expression signals a loss of consciousness, illness, or perhaps even death – all of them possibilities suffused with a strong affective valence. Similarly, the man's body, completely wrapped in bandages, constitutes a powerful affordance whose negative valence is linked to instinctive fears of illness, accident, or disfiguration, that is, basic emotions related to bodily harm and/or survival. However, it is not just such biologically determined associations that determine how the affective response unfolds, as the process will likely also involve memories of culturally transmitted contents. Thus the eerie feeling the bandaged human figure elicits may not derive from individual experiences of bodily harm or medical treatment (or fears thereof), but also from affectively inscribed memories of experiencing and reacting to similar representations of uncanny objects – e.g., in horror movies. One should likewise note that while low-level visual properties generally contribute to an object's perceived valence (Lebrecht et al., 2012), this factor is even more important in the case of an art work. The affective salience of the figures is thus inherently constituted by the painterly medium, that is, by the way Desiderio's brushwork and handling of color depict them as objects in the pictorial space. Furthemore, affective affordances are contextually pretuned (Todd et al., 2012); in art perception such pre-tuning is significantly influenced by the general experiential context (e.g., one's feeling about being in a museum or gallery, the presence/absence of "museum fatigue"), as well as by the affective state engendered by cumulative past experience of viewing art works.

To sum up so far: in the motivated viewer's encounter with *Spiegel im Spiegel* there occurs a complex interplay of predictions that span every level within the hierarchical structure of the mind/brain, from sensory/recognitional to high-level semantic predictions. But whereas in a real-life situation, perception, cognition, and action are associated with the successful suppression ("explaining away") of prediction error or the reduction of surprise (Clark, 2013), the viewer's encounter with Desiderio's painting will almost inevitably result in a subjective awareness of not being able to arrive at an acceptable solution as to what the painting represents, what it "is about." At the end, the viewer has to settle with the best solution for the moment; in other words, the process of optimizing relative precision of empirical (top–down) priors and (bottom–up) sensory evidence (Friston, 2009) temporarily subsides, only for the viewer to be aware that the solution is precarious and provisional. Uncertainty about the meaning, as a few available records suggest,

is accompanied by an unspecified, but powerful feeling, which can be summed up in such words as uncanny, troubling, depressing, or nightmarish[9]. This overall affective response is related both to the perception of pictorial content and the medium (as discussed above) and to the affective component of prediction error minimization itself, as we shall presently discuss.

THE LIMITATIONS OF THE DOPAMINERGIC REWARD ACCOUNT

According to Van de Cruys and Wagemans (2011) tentative prediction error hypothesis, it is the incompatibility (prediction error) that causes a part of the emotionality that viewers encounter in works of art. Their main thesis is that the reduction of unpredictability is experienced as positive and pleasurable. Thus: "The effort of mental work one has to do to cope with the prediction error is a condition sine qua non for receiving perceptual pleasure of a Gestalt formation (prediction error reduction)" (TPE, 1046). The authors suggest that the degree of mental effort viewers make to compensate for unpredictability is related to reward. This is then further linked to dopaminergic reward modulation, whereby unexpected reward are associated with increased dopamine peaks. Artists are thus in the business of postponing the final gratification: by using minimal prediction errors painters can ensure that viewers will obtain their reward and not give up prematurely. Similar claims have recently been made by Kandel (2012), who argues "that the response of dopaminergic neurons to anticipated pleasure may be the physiological basis of the pleasure we experience when looking at art. Art may give rise to feelings of well-being because it predicts biological reward, even though further reward beyond the pleasure of viewing and vicariously experiencing may never materialize" (Kandel, 2012, pp. 428–429), and other recent opinions concur that positive emotional valence, or pleasure, is elicited in the transition from a state of high to low surprise (Joffily and Coricelli, 2013).

These views align with recent research on the neurobiology of reward. It is well-established that reward has a direct, non-volitional impact on perception, changing the salience of objects for attention (Hickey et al., 2010). Dopamine receptors were found to mediate prefrontal control of signals in the visual cortex (Noudoost and Moore, 2011). Studies based on the monetary reward prospect paradigm reveal that reward leads to the tuning of sensory neurons and modulates the neural dynamics of early visual category processing (Apitz and Bunzeck, 2012; for an overview of reward-related modification of sensory processing in the cortex, see FitzGerald et al., 2013; Overton et al., 2014). More specifically, Biederman and Vessel (2006) proposed that the interpretation of a novel and richly stimulating visual pattern leads to feelings of pleasure, because such patterns initially activate an abundant set of associations in the ventral visual pathway that manifest dense mu-opioid receptors. Such neurobiological accounts are supported by psychological research

---

[9]I derive this observation both from the few available comments on this painting and from ten interpretations of it written by participants in my graduate seminar on ekphrasis and the description of works of art.

that suggests that ambiguity in works of art may be pleasurable (Jakesch et al., 2013) and collectively these findings lend scientific support to old insights that perceptual interpretation in itself is rewarding, something that is summed up in Gombrich's (1982) observation on "pleasure of recognition." However, as I shall argue, there are grounds for a much more cautionary approach in postulating links between reward and pleasure in visual art perception.

First, the pleasure of recognition (or perceptual Gestalt formation) should be considered a kind of cognitive reward, which differs from more basic reward (Schultz, 2000); while acknowledging the biological basis of desire for or liking of art, it should probably be distinguished from incentive salience ("wanting") as a specific form of Pavlovian-related motivation for reward (for such distinctions, see Berridge, 2012). It is as yet unclear to what extent cognitive reward share neurobiological mechanisms with basic forms of reward, those which for instance activate a craving for food or a drug. Second, conceptualizing reward in relation to art experience in the terms of perceptual pleasure is too narrow, as it accounts for a distinct minority of encounters with visual art. If phenomenology is to be taken as a reliable guide, pleasure may indeed characterize the nature of the viewing process in the case of works in which perceptual recognition/identification is more complicated – as in the case of many modern art styles. Upon encountering Pablo Picasso's *Guitar and Violin*, or Lyonel Feininger's *Sailboats,* recognizing the objects in the depicted scenes – that is, perceptual prediction error minimization *tout court* – clearly is related to pleasure for many viewers, as Van de Cruys and Wagemans (2011) propose. In most instances, however, pleasure may have a rather limited – if any – role in the reward that is associated with viewing art works. A more inclusive account thus needs to reflect the recent conceptualization of reward, which sees it as separated into several components at the psychological and the neurosystemic level – pleasure, incentive motivation and learning (Dickinson and Balleine, 2010; Smith et al., 2012; Miller et al., 2014) – or alternatively insists on dissociating motivation from reward.

Finally, it should be briefly noted that the role of dopamine in mediating perceptual (or broadly aesthetic) pleasure is not very clear. The authors of one recent study confirmed the presence of dopaminergic activity during the anticipation and actual experience of peak emotion in music (Salimpoor et al., 2011), but there is little to suggest the same mechanism operates in the perception of a visual medium as well. The authors of this and other studies (Vuust and Frith, 2008; Schaefer et al., 2013) point to the significant role of temporal phenomena such as expectation, tension, surprise prediction, and anticipation in evoking emotion from music. In contrast, although perceiving painting or sculpture is likewise a temporal event, the role of expectation and anticipation is of a different order. On the other hand, cognitive wanting was found not to be directly affected by mesolimbic dopamine fluctuations (Wassum et al., 2011) and much evidence was accumulated against the long-standing view that dopamine mediates pleasure (Berridge, 2012, p. 1131).

GRASPING MEANING – THE OUTCOME OF PREDICTIVE ERROR OPTIMIZATION

As discussed above, in many (if not most) encounters with works of art, the recognition of the subject is accomplished instantaneously and does not lead to extensive engagement with the work, that is, once the minimization of prediction error at the level of object recognition concludes the viewing, there is no expectation of further reward and hence no motivation for a prolonged viewing and thinking about the work. Returning to the case study of Vincent Desiderio's painting, two distinct patterns of response can be postulated. In each case, the viewer, having more or less effortlessly accomplished the recognition of objects in the scene, is left puzzling over the meaning of the painting – the image itself does not provide sufficient clues with which to optimize prediction error. For viewer A, the semantic opacity of the painting does not constitute a challenge to be engaged with, resulting in a negatively valenced experience, which provides no incentive for further viewing (or for repeating such an experience). The well-documented aversion to modern art can be partly explained within this framework. On the other hand, viewer B, in the course of a much more extended, consciously reflected, viewing, experiences a cascade of prediction errors minimizations, which entails the simultaneous formation of new predictions, and thus again arrives at no final "solution" as to the meaning of the depicted scene. At a certain point she leaves the painting with the best interpretation available at the moment; in the terms of PEM, she explains away the image (and her own reaction to it) given her continuously updated generative model. The optimization of prediction error concludes with the best possible outcome for the moment, but that outcome remains tentative, as subjectively the painting retains its enigma, lingering in memory and even generating new associations. The experience itself, although subjectively felt as something disquieting, troubling and certainly not inherently pleasurable, may ultimately be perceived as rewarding, and as providing motivation for another encounter of this kind. This fully accords with some recent accounts of aesthetic experience as being disruptive and transformative at its core (Pelowski and Akiba, 2011)[10].

This is not the end of story, however, as our case study provides an apt opportunity to observe how the response to art work within the PEM framework is further, and perhaps decisively, affected by the viewer's access to external facts, some kind of extra-pictorial information that cannot be gathered from the visible configuration of the image itself. In the given case, the key information is the knowledge that in this (and several other paintings) Desiderio depicted his severely physically and mentally handicapped son Sam, who needs a tube to breathe and whom he has been constantly caring for. The viewer realizes that the painting is not to be understood entirely as a fantasy image or a dream, and that the visible content of the scene refers to an existing aspect of reality. Some remaining sensory mismatches are consequently minimized ("the white tube around the boy's neck is the breathing tube"), while simultaneously new (semantic) ones are generated ("If the

---

[10]Van de Cruys and Wagemans (2011) note that people may experience positive emotions when faced with an "unsolvable painting" by reappraising negative prediction error in a safe context, or, alternatively, by misattribution (TPEA, 1053).

boy is the artist's son, is it likely that the bandaged figure is the painter himself? Is the whole painting, then about the artist's (a father's) relationship/communication, or rather about the inability to communicate with his gravely disabled child?" A brief quote from comments made by an exceptionally perceptive art critic captures eloquently the range of associations and feelings that may unfold for a perceptive viewer, and the following, in fact, is a verbal transcription of the process of PEM unfolding in the critic's mind: "*Empathically identified with him, and bent over him in imploring care, Desiderio is unable to establish intimacy with his son, all the more so because he seems lost in an dream world of his own. His son will remain a child in spirit and body all his life. Damaged beyond repair, the boy will never become a man. I know no greater image of human suffering in contemporary art, no subtler image of modern alienation – the absurd nightmare that is modern life*" (Kuspit, 2011). Most viewers may not be able to form as elaborated and sophisticated an association as an experienced art critic, but their understanding of the representational content of the image will nevertheless alter the profile of their affective response, in which the unspecified negative feeling, as described above, can be replaced by empathic engagement with the figures depicted. This empathic reaction will in turn entail both affective resonance and cognitive perspective-taking based on generating imaginary scenarios, that is, it will involve both cognitive and interoceptive (Seth et al., 2012; Seth, 2013) predictions about what it would be like and how it would feel to be in the same situation. Therefore, reducing the uncertainty related to the representational content of the painting by minimizing prediction error, far from "explaining the painting away," decisively changes the generative model of the motivated viewer, thus triggering a new productive cycle of perceptual, cognitive, and affective inferences.

PREDICTIVE ERROR MINIMIZATION AND THE CULTURALITY OF VISION

In this section I shall point to some strategic implications of the predictive coding framework for art history and visual studies. The observation that predictions in visual art are dependent on the specific history of stimulation (Van de Cruys and Wagemans, 2011, pp. 1044–1045, see also Clark, 2013) corresponds to the long-standing and widely shared understanding that perception depends on one's personal and cultural background (Arnheim, 1954; Segall et al., 1966). Recent research has extended these views by examining the qualitative difference in the ability to predict between experts and novices, emphasizing that experts have more resources for generating predictions, but also that they make more elaborate and accurate predictions in a given context (Cheung and Bar, 2012; Panichello et al., 2013). But importantly, the inter-individual differences in art perception do not stem from expertise alone, but need to be conceived more broadly. Summarizing earlier insights, Gombrich (1960) argued that perceptual experience depends on "mental set," without much elaborating this notion. More precisely, as I shall argue, the inter-individual differences in art perception depend on three variables: (i) personality traits/affective style – that is, how and why individuals differ in how they respond to emotional incentives (Davidson, 2004). These have

a strong modulatory effect, especially with respect to the affective aspects of art perception, so, for example, individuals with neurotic and anxious personality traits are more sensitive to processing facial or bodily expressions in particular (Bishop, 2007; Cunningham et al., 2010; Cunningham and Brosch, 2012); (ii) culture-cognitive capital related to the experiential situation, that is, the skills and knowledge related to visual perception and viewing art works; and (iii) the momentary psychosomatic state of the observer.

Jointly, these three aspects form a mindset, which determines the generation of top–down expectations and predictions. Alternatively, mindset itself can be conceived of as the sum total – or repertoire – of predictions that pertain to the given task (Bar, 2009) and as such it is further primed by the given experiential situation. In the case of a typical visual art experience, entering the museum or art gallery (or mere prospect thereof) serves to prime the mindset, forming a global expectation about the experience – a potentially fascinating and enjoyable event for person A, or the prospect of something boring and tedious that has to be endured for person B, with many variations in between. This general expectation conditioned by all three variables thus sets the stage for specific predictions to be generated vis-a-vis the individual works of art encountered during the visit.

Importantly, experience-based individual differences in viewing art works, which are partly dependent on culture-cognitive capital, link the predictive coding account of art perception to a major issue in art history and visual studies – the problem of the culturality and sociality of vision. In these and related disciplines, the biological-social continuum of seeing is routinely conceptualized as a distinction between vision (as a biological act) and visuality (as culturally and socially determined; e.g., Nelson, 1996; Davis, 2011). As recently articulated by Davis (2011, p. 230): "When we speak of visuality, rather than simply vision or visual perception, we address the difference introduced into human seeing by traditional cultural meaning consolidated and reconfigured in images." This conceptual distinction, however, is problematic and calls for alternative framework, which would have to consist of four levels in order to capture the process of vision in its biological and social complexity with more precision (Kesner, 2009, 2014). Moving in a top–down fashion, these are the levels of:

(1) Concepts, attitudes, values, and motives (and their discursive articulation) about images, vision and representation— that is, visuality in the strict sense of the term; these develop and persist on a time-scale of years to centuries. (2) The level of cognitive factors, which is strongly shaped by the environment and culture and roughly corresponds to Baxandall's (1972) notion of the "period eye": semantic categories, patterns of inference experience and training in the range of representational conventions etc., that is, factors that operate in stretches ranging from the minutes of psychological time in individual perception to the historical time of years. (3) Perceptual strategies and processes – such as mechanisms of recognition, object identification and classification, patterns of saccadic eye movement, of selective visual attention, processes of unconscious embodied emotional and emphatic response, motor

reaction activated by perception etc. These are processes that operate on both the conscious and the unconscious level and span a time frame of 100s of milliseconds to minutes. (4) Mechanisms of detection of essential aspects of the scene, such as lines and edges, movement, color, binocular disparity, and related aspects of low-level vision – that is, biologically hard-wired unconscious events occurring on a time scale of up to ∼250 ms. Naturally, such a scheme implies neither a strict hierarchy nor hard boundaries between these stages, nor their mutual encapsulation. To the contrary, there is an ongoing, reciprocal relationship and feedback, whereby biologically embedded mechanisms interact with the higher levels of vision that can be modified by culture. Prediction errors at level 4, i.e., low-level vision, concerning mainly contrast and orientation, are sent further on in the processing stream and integrated into more complex messages concerning object identities, and then further on into semantic categories. The two topmost levels – of visuality and cognitive categories – are in a strong sense culturally relative. But importantly, there is an increasing body of evidence suggesting that also at level 3, concerning perceptual strategies and processes, both culture and individual perceptual history penetrate perception (for an overview of this evidence, see Kitayama and Uskul, 2011; Rule et al., 2013). These stages operate on different time-scales and to an extent can be mapped as occurring in different areas of the brain, where higher levels represent the context in which the lower levels unfold (see also Kiebel et al., 2008; Hohwy, 2013, p. 61). Representations thus depend on and interact with representations at other levels both within the topography of the brain and within the hierarchical conceptual scheme of vision outlined above.

The experience-mediated perceptual-cognitive routines and skills that ensue from a viewer's participation in a shared sphere of cultural habits and protocols of seeing and a sensitivity that is attuned to styles of representation ingrained in the viewer's culture are all the produce of perceptual learning and – at the neuronal level – the mechanism of synaptic plasticity. There is an extensive body of recent research providing evidence on experience-dependent plasticity in adult brains and specifically on how perceptual expertise alters visual processing, e.g., by determining nature of object representation in the visual system (Curby and Gauthier, 2009; Spolidoro et al., 2009; May, 2011; Wong et al., 2012; Folstein et al., 2013; Lövdén et al., 2013; Kok and de Lange, 2014; Vetter and Newen, 2014). It has been shown that the integration of top–down expectations and bottom–up sensory input can already be observed in the early visual cortex (Kok et al., 2013) and that past experience modulates shape assignment and perceptual grouping (Kimchi and Hadad, 2002; Trujillo et al., 2010). Cast in the terms of the PEM framework, previous experience and perceptual expertise generate distinct set of expectations (or priors) which determine interpretation of the image. Formation of a prediction error is achieved by adjusting synaptic efficacies both between and within levels of the processing hierarchy (Schultz, 2000; Friston, 2008). Post-synaptic effects either may be short-lived, directly impacting perception, or may control the updating and storage of predictions by inducing changes in synaptic growth (den Ouden et al., 2012). It is likely that it is by this mechanism that perceptual expertise becomes stabilized in the individual mind and by which it can even become collective

in the sense of characterizing the perceptual habits of a certain group of people, so that commonalities can be observed in the social world of a particular group of viewers that govern how they form predictions in the perception of works of art ("period eye" according to one influential art-historical paradigm – cf. Baxandall, 1972).

This kind of expertise is both enabling and constraining, as the following example will show. Chinese literati and literati painting form a well-defined group of expert viewers of a well-defined body of art that to be understood requires a strong and specific form of visual-cognitive skills, some aspects of which can be indirectly inferred from their extensive writings reflecting on the subject. The predictions involved in their experience of painting operated at all four levels of the hierarchy: at the level of visuality, value-based judgments and preferences suggested what was worth looking at and was deemed to be of aesthetic value for the given social group and this determined the most general parameters of the perceptual encounter (e.g., by constraining attentional allocation according to the perceived value of the painting). At the level of cognitive factors, literati culture shaped predictions related to the semantic categorization of paintings, thus enabling the viewer-expert to differentiate and conceptualize not only specific topics and subjects or various styles, but also the variety of brushwork seen in the painting (so-called *cunfa* brushstrokes). Finally, on the level of perceptual routines, distinct patterns of visual attention shaped by experience, along with saccadic eye movements or an embodied kinesthetic proprioceptive feeling of brush wielding determined the formation of predictions related to the imminent perceptual processing of these paintings.

This expertise, comprising a vast store of memory representations and semantic associations, but also internalized motor representations and body schemas, provided the expert viewer with the ability to unfold successive strings of prediction errors, from which a rich set of expectations could be formed, entertained and finally minimized. In contrast, the naive viewer, lacking such cultural equipment with which to approach this particular kind of art work has at his disposal a far more limited repertoire of resources to respond and create meaning. Note, however, that this type of culture-specific expertise attuned to a specific form of representation was also constraining, as the insider-viewer was equipped with a fixed set of routine and rigid associations, generating expectations and predictions that shut out some possible other ways in which these works could be perceived- hence the well-documented narrow focus of literati painters on the quality of the brushwork in a painting, while ignoring the representational quality and mimetic status of the image. The above example would typically have involved a continuous bidirectional flow of vision across the hierarchy of levels, from high-level predictions (visuality) down to the level of basic perceptual mechanisms and routines, and vice versa. If culture can be seen as generator of hyperpriors, essential to the hierarchical network of the Bayesian brain, which is the source of a person's perceptual fantasy of the world (Patton et al., 2013), the crucial task for understanding this is ultimately to provide a naturalistic account of how both innate and learned priors, or contextual and structural expectations (Seriès and Seitz, 2013), or informed vs. fixed priors (van

Wassenhove, 2013) arise and are mutually co-determined and modulated.

A likely candidate for the central, enabling hyperprior[11] operating in art perception is the general pictorial competence and the capacity of seeing in/as, that is, the ability of the human observer to see a certain object in a depictive configuration on a pictorial surface, to see "through" the depiction to its referent and at the same time to the nature of the relation between representation and referent. This capacity, as psychological research demonstrates, is acquire *de novo* in infants and develops ontogenetically in humans, and it can thus be considered universal (DeLoache et al., 2003; DeLoache, 2004; for a historical perspective on the role of seeing in/as in the origins of picture-making see Davis, 1986, 2011). But as such, it is decisively modulated by many culture-specific effects, which include a range of representational conventions, the nature of a particular visual medium and many other factors spanning the hierarchy of vision from visuality to perceptual routines. It would be enormously complex to formulate a fully naturalistic account of how hyperpriors, co-constituted at the intersection of biology and culture, enable and constrain the formation of predictions when viewing particular art works. But importantly, some aspects of it are open – at least in principle – to experimental study. As a brief example, consider the case of the perception of a visual symbolic form by the excessive action computer-game players. In this well-defined group of image users, research has provided empirically measurable evidence of the modification of certain perceptual routines (e.g., Boot et al., 2008; Donohue et al., 2010; Bavelier et al., 2011) and even of players' affective responses to imagery (Montag et al., 2012). One way of broadly summarizing these findings may be that the habit of on-line acting upon, motorically responding to a stream of constantly changing images resculpts the perceptual-cognitive architecture of these players on a synaptic level, ultimately giving rise to specific variations in their general pictorial competence – a hyperprior expectation that images are entities to be acted upon instantly, not something to be perused and contemplated in a prolonged fashion. The explanation for how such individual, experience-dependent expertise and plasticity characteristic of this kind of temporally delimited collective visual experience may become stabilized, more permanent and transmitted to successive generations within the same cultural milieu remains a major desideratum for further research, but this is a topic beyond the scope of present article.

CONCLUDING REMARKS AND FUTURE DIRECTIONS

I shall conclude by briefly outlining several possible directions in which the present accounts of the predictive coding framework for the visual arts can be further productively elaborated. The first direction, mentioned above, concerns the possibility of making some key theoretical accounts in art history and visual studies – such as Gombrich's (1960) classic account of the rise of naturalistic depiction, or "a general theory of visual culture"

(Davis, 2011) – compatible with the prediction error minimization framework. The second direction is to elaborate the model on the basis of further case studies of specific types of visual art objects. The theory could likely be productively applied to many pre-modern works of art, which in their original context of use served as objects endowed with specific functions, and where the "viewing" in the original setting was inextricably bound up with (or accompanied by) some sort of embodied action. Such works can be said to contain their own script for action; the experience of the original audience could thus be modeled through action-oriented predictive processing (Clark, 2013), which suggests that motor intentions, as they unfold into detailed motor actions, actively elicit continuous streams of sensory results that our brains predict. Furthermore, while Van de Cruys and Wagemans (2011) note that the predictive framework does not explain the popularity of realist art, "which depicts the world as it is, thereby confirming rather than violating prediction error," (TPEA, 1056) this genre of painting need not be discounted and the PEM framework can be elaborated for the naturalistic/realistic spectrum of artistic representations as well[12].

The third challenge is even more complex, but offers the potential for truly interdisciplinary dialog between theorists of art and neuroscientists. It has been argued that the actual virtue of predictive coding is the fact that it is typically implemented at a level of abstraction that is intermediate between that of low-level, biophysical, circuits and that of high-level, psychological, behaviors (Spratling, 2013). While this indeed seems to be the case, the greatest challenge (as noted in the opening section) is linking the accounts of predictions at the neuronal level with those on the cognitive-psychological and cultural levels (see also Clark, 2013). While current research is providing an increasingly detailed insight into neuronal mechanisms, including an account of the interactions between prediction and error signals (Kveraga et al., 2007; Summerfield and Egner, 2009; Rauss et al., 2011; Friston et al., 2012a; Koster-Hale and Saxe, 2013; Kok et al., 2013), the relationship between the various levels of prediction operating in visual-arts perception, between the neuronal-architecture and cognitive-psychological levels of prediction, has hitherto been at best tentatively and sketchily explained. In particular, future work needs to address the nature of representational formats of hierarchically different levels of predictions. In other words, the notions of predictions (priors) needs to be related to range of terms currently used both within and outside the predictive coding framework to characterize disparate contents of mental representations underlying the recognition and interpretation of sensory content from across the hierarchy of vision, including, for example, the "generative image model" (Yuille and Kersten, 2006), the "pictorial schema" (Gombrich, 1960), the "image schema" (Lakoff, 2006), or "subjective internal representation" (Smith et al., 2012). Much will also depend on whether (and how) some alignment can be made between the predictive coding framework and other current accounts of image perception, such as incremental grouping theory (Roelfsema and Houtkamp, 2011).

---

[11]Following Newsome (2013), I take hyperpriors to refer to general facts, or expectations, about "embodied interaction with the physical world that structures and constrains prior predictions and how they interact."

[12]As one of the reviewers of this paper suggested, realistic art heightens the tension between inferring the depicted object as a cause and painterly medium as a cause. Either interpretation gives a prediction error, leading to alternations in perception.

The fourth challenge relates to determining how affective and empathic inference, as conceived in the PEM framework, arises out of integrated interaction between large-scale brain networks (Bressler and Menon, 2010; Oosterwijk et al., 2012; Barrett and Satpute, 2013). As Friston (2013, p. 1331) remarks, "the most prescient challenge to formal description of the brain as inference machines is how one can accommodate emotions." Of prime importance we need to decipher the nature of the interaction that occurs between the affective feelings elicited by prediction errors formed as the content and medium of an image are processed and the affective valence elicited by the outcome of the process of PEM (or, in conformity with a recent conceptualization of this problem, we need to decipher the relationship between unsigned perceptual and cognitive prediction errors and signed reward prediction errors (den Ouden et al., 2012). Finally, future experimental work, building on the neuroimaging and eye-tracking work done with non-art images (e.g., Itti and Baldi, 2009) and, most notably, combining subjective, behavioral and neuronal data, may untangle some aspects of prediction in art perception in different hierarchies and their interaction. Analysis of cultural hyperpriors, on the other hand, will likely have to be performed within the territory of individual disciplines, rather than predictive coding framework alone thus necessitating a multi-level, multi-disciplinary methodology (Clark, 2013; Newsome, 2013). The challenge for the Bayesian account of art perception (and equally for the neuroscience of art research generally) lies in reconciling the demands of the subjective, phenomenological description of the experience of a work of art with the rhetoric and the argumentational style of discourse used in mind and brain science and with a formalized account of prediction error theory. Recognizing the inherent difficulty involved in meaningfully bridging these two modes of argumentation (a problem that undoubtedly surfaced in present article) I still wish to argue that real progress at the intersection of art and mind/brain science can only be achieved by attempting this kind of direct interfacing.

ACKNOWLEDGMENT

REFERENCES

Anderson, E., Siegel, E., and Barrett, L. (2011). What you feel influences what you see: the role of affective feelings in resolving binocular rivalry. *J. Exp. Soc. Psychol.* 47, 856–860. doi: 10.1016/j.jesp.2011.02.009

Anderson, M., and Chemero, T. (2013). The problem with brain GUTs: conflation of different senses of "prediction" threatens metaphysical disaster. *Behav. Brain Sci.* 36, 204–205. doi: 10.1017/S0140525X1200221X

Apitz, T., and Bunzeck, N. (2012). Reward modulates the neural dynamics of early visual category processing. *Neuroimage* 63, 1614–1622. doi: 10.1016/j.neuroimage.2012.08.046

Arnheim, R. (1954). *Art and Visual Perception. A Psychology of the Creative Eye. The New Version.* Los Angeles, CA: University of California Press.

Arnheim, R. (1992). "For your eyes only: seven exercises in art appreciation," in *To the Rescue of Art*, ed. R. Arnheim (Los Angeles, CA: University of California Press).

Augustin, M. D., and Leder, H. (2006). Art expertise: a study of concepts and conceptual spaces. *Psychol. Sci.* 48, 135–156.

Bar, M. (2003). A cortical mechanism for triggering top-down facilitation in visual object recognition. *J. Cogn. Neurosci.* 15, 600–609. doi: 10.1162/089892903321662976

Bar, M. (2009). The proactive brain: memory for predictions. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 364, 1235–1243. doi: 10.1098/rstb.2008.0310

Barrett, L. F., and Bar, M. (2012). See it with feeling: affective predictions during object perception. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 364, 1325–1334. doi: 10.1098/rstb.2008.0312

Barrett, L. F., and Satpute, A. B. (2013). Large-scale brain networks in affective and social neuroscience: towards an integrative functional architecture of the brain. *Curr. Opin. Neurobiol.* 23, 361–372. doi: 10.1016/j.conb.2012.12.012

Bavelier, D., Green, C. S., Han, D. H., Renshaw, P. F., Merzenich, M. M., and Gentile, D. A. (2011). Brains on video games. *Nat. Rev. Neurosci.* 12, 763–768. doi: 10.1038/nrn3135

Baxandall, M. (1972). *Painting and Experience in Fifteenth-Century Italy: A Primer in the Social History of Pictorial Style.* Oxford: Oxford University Press.

Baxandall, M. (2003). *Words for Pictures. Seven Papers on Renaissance Art and Criticism.* New Haven, CT: Yale University Press.

Berridge, K. (2012). From prediction error to incentive salience: mesolimbic computation of reward motivation. *Eur. J. Neurosci.* 35, 1124–1143. doi: 10.1111/j.1460-9568.2012.07990.x

Biederman, I., and Vessel, E. (2006). Perceptual pleasure and the brain. *Am. Sci.* 94, 249–255. doi: 10.1511/2006.59.995

Bishop, S. (2007). Neurocognitive mechanisms of anxiety: an integrative account. *Trends Cogn. Sci.* 11, 307–316. doi: 10.1016/j.tics.2007.05.008

Boot, W. R., Kramer, A. F., Simons, D. J., Fabiani, M., and Gratton, G. (2008). The effects of video game playing on attention, memory, and executive control. *Acta Psychol.* 129, 387–398. doi: 10.1016/j.actpsy.2008.09.005

Bowers, J., and Davis, C. (2012). Bayesian just-so stories in psychology and neuroscience. *Psychol. Bull.* 138, 389–414. doi: 10.1037/a0026450

Bressler, S., and Menon, V. (2010). Large-scale brain networks in cognition: emerging methods and principles. *Trends Cogn. Sci.* 14, 277–290. doi: 10.1016/j.tics.2010.04.004

Cheung, O., and Bar, M. (2012). Visual prediction and perceptual expertise. *Int. J. Psychophysiol.* 83, 156–163. doi: 10.1016/j.ijpsycho.2011.11.002

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behav. Brain Sci.* 36, 181–204. doi: 10.1017/S0140525X12000477

Crick, F., and Koch, C. (1995). Are we aware of neural activity in primary visual cortex? *Nature* 375, 121–123. doi: 10.1038/375121a0

Csikszentmihalyi, M., and Robinson, R. (1990). *The Art of Seeing. An Interpretation of the Aesthetic Encounter.* Malibu, CA: J. Paul Getty Museum/Getty Center for Education in the Arts.

Cunningham, W. A., Arbuckle, N., Jahn, A., Mowrer, S., and Abdul-jalil, A. (2010). Aspects of neuroticism and the amygdale: chronic tuning from motivational styles. *Neuropsychologia* 48, 3399–3404. doi: 10.1016/j.neuropsychologia.2010.06.026

Cunningham, W., and Brosch, T. (2012). Motivational salience: amygdala tuning from traits, needs, values, and goals. *Psychol. Sci.* 21, 54–59.

Curby, K., and Gauthier, I. (2009). To the trained eye: perceptual expertise alters visual processing. *Topics Cogn. Sci.* 2, 189–201. doi: 10.1111/j.1756-8765.2009.01058.x

Davidson, R. (2004). Well-being and affective style: neural substrates and biobehavioural correlates. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 359, 1395–1411. doi: 10.1098/rstb.2004.1510

Davis, W. (1986). Origins of image-making. *Curr. Anthropol.* 27, 193–215. doi: 10.1086/203422

Davis, W. (2011). *General Theory of Visual Culture.* Princeton, NJ: Princeton University Press.

DeLoache, J. (2004). Becoming symbol-minded. *Trends Cogn. Sci.* 8, 66–70. doi: 10.1016/j.tics.2003.12.004

DeLoache, J., Pierroursakos, S., and Uttal, D. (2003). The origins of pictorial competence. *Curr. Dir. Psychol. Sci.* 12, 114–118. doi: 10.1111/1467-8721.01244

den Ouden, H. E. M., Kok, P., and de Lange, F. P. (2012). How prediction errors shape perception, attention, and motivation. *Front. Psychol.* 3:548. doi: 10.3389/fpsyg.2012.00548

Dickinson, A., and Balleine, B. (2010). "Hedonics: the cognitive-motivational inter-face," in *Pleasures of the Brain*, eds M. L. Kringelbach and K. C. Berridge (New York: Oxford University Press), 74–84.

Donohue, S. E., Woldorff, M. G., and Mitroff, S. R. (2010). Video game players show more precise multisensory temporal processing abilities. *Atten. Percept. Psychophys.* 72, 1120–1129. doi: 10.3758/APP.72.4.1120

Fernández-Dols, J., and Crivelli, C. (2013). Emotion and expression: naturalistic studies. *Emot. Rev.* 5, 24–29. doi: 10.1177/1754073912457229

FitzGerald, T., Friston, K., and Dolan, R. (2013). Characterising reward outcome signals in sensory cortex. *Neuroimage* 83, 329–334. doi: 10.1016/j.neuroimage.2013.06.061

Folstein, J. R., Palmeri, T. J., and Gauthier, I. (2013). Category learning increases discriminability of relevant object dimensions in visual cortex. *Cereb. Cortex* 23, 814–823. doi: 10.1093/cercor/bhs067

Friston, K. (2002). Functional integration and inference in the brain. *Prog. Neurobiol.* 590, 1–31. doi:10.1016/j.neuroimage.2005.08.035

Friston, K. (2008). Hierarchical models in the brain. *PLoS Comput. Biol.* 4:e1000211. doi: 10.1371/journal.pcbi.1000211

Friston, K. (2009). The free-energy principle: a rough guide to the brain? *Trends Cogn. Sci.* 13, 293–301. doi: 10.1016/j.tics.2009.04.005

Friston, K. (2013). The fantastic organ. *Brain* 136, 1328–1332. doi: 10.1093/brain/awt038

Friston, K., and Kiebel, S. (2009). Predictive coding under the free-energy principle. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 364, 1211–1221. doi: 10.1098/rstb.2008.0300

Friston, K., Shiner, T., FitzGerald, T., Galea, J. M., Adams, R., Brown, H., et al. (2012a). Dopamine, affordance and active inference. *PLoS Comput. Biol.* 8:e1002327. doi: 10.1371/journal.pcbi.1002327

Friston, K., Thornton, C., and Clark, A. (2012b). Free-energy minimization and the dark-room problem. *Front. Psychol.* 3:130. doi: 10.3389/fpsyg.2012.00130

Gadamer, H. (1975/2004). *Truth and Method*, trans. J. Weinsheimer and D. Marshall, London-New York: Continuum.

Gamboni, D. (2002). *Potential Images. Ambiguity and Indeterminacy in Modern Art*, London: Reaktion.

Gombrich, E. (1960). *Art and Illusion*. London: Phaidon Press.

Gombrich, E. (1982). *The Image and the Eye. Further studies in the Psychology of Pictorial Representation*. Ithaca-London: Phaidon Press.

Gregory, R. (1980). Perception as hypotheses. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 290, 181–197. doi: 10.1098/rstb.1980.0090

Grill-Spector, K., and Kanwisher, N. (2005). Visual recognition. as soon as you know it is there, you know what it is. *Psychol. Sci.* 16, 152–160. doi: 10.1111/j.0956-7976.2005.00796.x

Hasenfus, N., Martindale, C., and Birnbaum, D. (1983). Psychological reality of cross-media artistic styles. *J. Exp. Psychol.* 9, 841–863.

Helmholz, H. (1860/1962). *Handbuch der Physiologischen Optik*. Leipzig: Leopold Voss.

Hickey, C., Chelazzi, L., and Theeuwes, J. (2010). Reward changes salience in human vision via the Anterior Cingulate. *J. Neurosci.* 30, 11096–11103. doi: 10.1523/JNEUROSCI.1026-10.2010

Hohwy, J. (2013). *The Predictive Mind*. Oxford: Oxford University Press. doi: 10.1093/acprof:oso/9780199682737.001.0001

Humphrey, K., Underwood, G., and Lambert, T. (2012). Salience of the lambs: a test of the saliency map hypothesis with pictures of emotive objects. *J. Vis.* 12, 1–15. doi: 10.1167/12.1.22

Itti, L., and Baldi, P. (2009). Bayesian surprise attracts human attention. *Vision Res.* 49, 1295–1306. doi: 10.1016/j.visres.2008.09.007

Jakesch, M., Leder, H., and Forster, M. (2013). Image ambiguity and fluency. *PLoS ONE* 8:e74084. doi: 10.1371/journal.pone.0074084

Jaspers, K. (1913/1963). *General Psychopathology*, trans. H. Hamilton, Manchester: Manchester University Press.

Joffily, M., and Coricelli, G. (2013). Emotional valence and the free-energy principle. *PLOS Comput. Biol.* 9:e1003094. doi: 10.1371/journal.pcbi.1003094

Jones, M., and Love, B. (2011). Bayesian fundamentalism or enlightenment? On the explanatory status and theoretical contributions of Bayesian models of cognition. *Behav. Brain Sci.* 34, 169–231. doi: 10.1017/S0140525X10003134

Kandel, E. (2012). *The Age of Insight. The Quest to Understand the Unconscious in Art, Mind, and Brain*. New York: Random House.

Kersten, D., Mamassian, P., and Yuille, A. (2004). Object percep-tion as Bayesian inference. *Annu. Rev. Psychol.* 55, 271–304. doi: 10.1146/annurev.psych.55.090902.142005

Kesner, L. (2009). Gombrich and the problem of relativity of vision. *Hum. Aff. Postdiscip. J. Hum. Soc. Sci.* 19, 266–273. doi: 10.2478/v10023-009-0040-x

Kesner, L. (2014). The Warburg/Arnheim effect: linking cultural/social and perceptual psychology of art. *J. Art Historiogr.* 11/LK1.

Kiebel, S., Daunizeau, J., and Friston, K. (2008). A hierarchy of time-scales and the brain. *PLoS Comput. Biol.* 4:e1000209. doi: 10.1371/journal.pcbi.1000209

Kimchi, R., and Hadad, B.-S. (2002). Influence of past experience on perceptual grouping. *Psychol. Sci.* 13, 41–47. doi: 10.1111/1467-9280.00407

Kitayama, S., and Uskul, A. K. (2011). Culture, mind and the brain: current evidence and future directions. *Annu. Rev. Psychol.* 62, 419–449. doi: 10.1146/annurev-psych-120709-145357

Knill, D., and Pouget, A. (2004). The Bayesian brain: the role uncertainty in neural coding and computation. *Trends Neurosci.* 27, 712–719. doi: 10.1016/j.tins.2004.10.007

Kok, P., Brouwer, G., van Gerven, M., and de Lange, F. (2013). Prior expectations bias sensory representations in visual cortex. *J. Neurosci.* 33, 16275–16284. doi: 10.1523/JNEUROSCI.0742-13.2013

Kok, P., and de Lange, F. (2014). Shape perception simultaneously up-and down-regulates neural activity in the primary visual cortex. *Curr. Biol.* 24, 1–5. doi: 10.1016/j.cub.2014.05.042

Koster-Hale, J., and Saxe, R. (2013). Theory of mind: a neural prediction problem. *Neuron* 79, 836–848. doi: 10.1016/j.neuron.2013.08.020

Kuspit, D. (2011). *Pitiless Pathology: Vincent Desiderios Paintings*. Avail-able at: http://www.artnet.com/magazineus/features/kuspit/vincent-desiderio-at-marlborough-chelsea-2011.asp

Kveraga, K., Ghuman, A., and Bar, M. (2007). Top-down predictions in the cognitive brain. *Brain Cogn.* 65, 145–168. doi: 10.1016/j.bandc.2007.06.007

Lakoff, G. (2006). "The neuroscience of form in art," in *The Artful Mind. Cognitive Science and the Riddle of Human Creativity*, ed. M. Turner (Oxford: Oxford University Press), 153–169.

Lebrecht, S., Bar, M., Barrett, L. F., and Tarr, M. J. (2012). Micro-valences: perceiving affective valence in everyday objects. *Front. Psychol.* 3:107. doi: 10.3389/fpsyg.2012.00107

Leder, H., Belke, B., Oeberst, A., and Augustin, D. (2004). A model of aes-thetic appreciation and aesthetic judgments. *Br. J. Psychol.* 95, 489–508. doi: 10.1348/0007126042369811

Lee, T., and Mumford, D. (2003). Hierarchical Bayesian inference in the visual cortex. *J. Opt. Soc. Am.* 20, 1434–1448. doi: 10.1364/JOSAA.20.001434

Lövdén, M., Wenger, E., Mårtensson, J., Lindenberger, U., and Bäckman, L. (2013). Structural brain plasticity in adult learning and development. *Neurosci. Biobehav. Rev.* 37(Pt B), 2296–2310. doi: 10.1016/j.neubiorev.2013.02.014

May, A. (2011). Experience-dependent structural plasticity in the adult human brain. *Trends Cogn. Sci.* 15, 475–482. doi: 10.1016/j.tics.2011.08.002

Miller, E. M., Shankar, M., Knutson, B., and McClure, S. M. (2014). Dissociating motivation from reward in human striatal activity. *J. Cogn. Neurosci.* 26, 1075–1084. doi: 10.1162/jocn_a_00535

Montag, C., Weber, B., Trautner, P., Newport, B., Markett, S., Walter, N. T., et al. (2012). Does excessive play of violent first-person –shooter-video-games dampen brain activity in response to emotional stimuli? *Biol. Psychol.* 89, 107–111. doi: 10.1016/j.biopsycho.2011.09.014

Nelson, R. (ed.). (1996). *Visuality Before and Beyond the Renaissance. Seeing as Others Saw*. Cambridge: Cambridge University Press.

Newsome, W. (2013). Complementing predictive coding. *Front. Psychol.* 3:554. doi: 10.3389/fpsyg.2012.00554

Niu, Y., Todd, R., and Anderson, A. K. (2012). Affective salience can reverse the effects of stimulus-driven salience on eye movements in complex scenes. *Front. Psychol.* 3:336. doi: 10.3389/fpsyg.2012.00336

Noudoost, B., and Moore, T. (2011). Control of visual cortical signals by prefrontal dopamine. *Nature* 474, 372–375. doi: 10.1038/nature09995

Oosterwijk, S., Lindquist, K. A., Anderson, E., Dautoff, R., Moriguchi, Y., and Barrett, L. F. (2012). States of mind: emotions, body feelings, and thoughts share distributed neural networks. *Neuroimage* 62, 2110–2128. doi: 10.1016/j.neuroimage.2012.05.079

Overton, P. G., Vautrelle, N., and Redgrave, P. (2014). Sensory regulation of dopaminergic cell activity: phenomenology, circuitry and function. *Neuroscience* 282C, 1–12. doi: 10.1016/j.neuroscience.2014.01.023

Panichello, M., Cheung, O., and Bar, M. (2013). Predictive feedback and conscious visual experience. *Front. Psychol.* 3:620. doi: 10.3389/fpsyg.2012.00620

Patton, B., Skewes, J., Frith, C., and Hohwy, J. (2013). Skull-bound perception and precision optimization through culture. *Behav. Brain Sci.* 36, 222–223. doi: 10.1017/S0140525X12002191

Pelowski, M., and Akiba, F. (2011). A model of art perception, evaluation and emotion in transformative aesthetic experience. *New Ideas Psychol.* 29, 80–97. doi: 10.1016/j.newideapsych.2010.04.001

Rauss, K., Schwartz, S., and Pourtois, G. (2011). Top-down effects on early visual processing in humans: a predictive coding framework. *Neurosci. Biobehav. Rev.* 35, 1237–1253. doi: 10.1016/j.neubiorev.2010.12.011

Rauss, K., and Pourtois, G. (2013). What is bottom-up and what is top-down in predictive coding? *Front. Psychol.* 4:276. doi: 10.3389/fpsyg.2013.00276

Rao, R. P. N., and Ballard, H. D. (1999). Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive –field effects. *Nat. Neurosci.* 2, 79–87. doi: 10.1038/4580

Roelfsema, P., and Houtkamp, R. (2011). Incremental grouping of image elements in vision. *Atten. Percept. Psychophys.* 73, 2542–2572. doi: 10.3758/s13414-011-0200-0

Rule, N., Freeman, J., and Ambady, N. (2013). Culture in social neuroscience: a review. *Soc. Neurosci.* 8, 3–10. doi: 10.1080/17470919.2012.695293

Russell, J. (2003). Core affect and the psychological construction of emotion. *Psychol. Rev.* 110, 145–172. doi: 10.1037/0033-295X.110.1.145

Russell, P., and Milne, S. (1997). Meaningfulness and hedonic value of paintings: effects of titles. *Empir. Stud. Arts* 15, 61–73. doi: 10.2190/EHT3-HWVM-52CB-8QHJ

Sakaki, M., Kazuhisa, N., and Mather, M. (2012). Beyond arousal and valence: the importance of the biological versus social relevance of emotional stimuli. *Cogn. Affect. Behav. Neurosci.* 12, 115–139. doi: 10.3758/s13415-011-0062-x

Salimpoor, V. N., Benovoy, M., Larcher, K., Dagher, A., and Zatorre, R. J. (2011). Anatomically distinct dopamine release during anticipation and experience of peak emotion to music. *Nat. Neurosci.* 14, 257–262. doi: 10.1038/nn.2726

Schaefer, R., Overy, K., and Nelson, P. (2013). Affect and non-uniform characteristics of predicitive processing in musical behaviour. *Behav. Brain Sci.* 36, 226–227. doi: 10.1017/S0140525X12002373

Schultz, W. (2000). Multiple reward signals in the brain. *Nat. Rev. Neurosci.* 1, 199–207. doi: 10.1038/35044563

Schwartenbeck, P., FitzGerald, T., Dolan, R. J., and Friston, K. (2013). Exploration, novelty, surprise, and free energy minimization. *Front. Psychol.* 4:710. doi: 10.3389/fpsyg.2013.00710

Segall, M. H., Campbell, D. T., and Herskovits, M. J. (1966). *The Influence of Culture on Visual Perception.* Indianapolis, IN: Bobbs-Merril.

Seriès, P., and Seitz, A. (2013). Learning what to expect in visual perception. *Front. Hum. Neurosci.* 7:668. doi: 10.3389/fnhum.2013.00668

Seth, A. (2013). Interoceptive inference, emotion, and the embodied self. *Trends Cogn. Sci.* 17, 565–573. doi: 10.1016/j.tics.2013.09.007

Seth, A., Suzuku, K., and Critchley, H. (2012). An interoceptive predictive coding model of conscious presence. *Front. Psychol.* 2:395. doi: 10.3389/fpsyg.2011.00395

Shimamura, A. (2013). *Experiencing Art. In the Brain of the Beholder.* Oxford: Oxford University Press.

Smith, M., Gosselin, F., and Schyns, P. G. (2012). Measuring internal representations from behavioral and brain data. *Curr. Biol.* 22, 191–196. doi: 10.1016/j.cub.2011.11.061

Spolidoro, M., Sale, A., Berardi, N., and Maffei, L. (2009). Plasticity in the adult brain: lessons from the visual system. *Exp. Brain Res.* 192, 335–341. doi: 10.1007/s00221-008-1509-3

Spratling, M. (2013). Distinguishing theory from implementation in predictive coding accounts of brain function. *Behav. Brain Sci.* 36, 231–232. doi: 10.1017/S0140525X12002178

Summerfield, C., Egner, M., Greene, E., Koechlin, E., Mangels, J., and Hirsch, J. (2006). Predictive codes for forthcoming perception in the frontal cortex. *Science* 314, 1311–1314. doi: 10.1126/science.1132028

Summerfield, C., and Egner, T. (2009). Expectation (and attention) in visual cognition. *Trends Cogn. Sci.* 13, 403–409. doi: 10.1016/j.tics.2009.06.003

Todd, R., Cunningham, W. A., Anderson, A. K., and Thompson, E. (2012). Affect-biased attention as emotion regulation. *Trends Cogn. Sci.* 16, 365–372. doi: 10.1016/j.tics.2012.06.003

Trujillo, L., Allen, J., Schnyer, D., and Peterson, M. (2010). Neurophysiological evidence for the influence of past experience on figure-ground perception. *J. Vis.* 10, 1–21. doi: 10.1167/10.2.5

Van de Cruys, S., and Wagemans, J. (2011). Putting reward in art: a tentative prediction error account of visual art. *Iperception* 2, 1035–1062.

van Wassenhove, V. (2013). Speech through ears and eyes: interfacing the senses with the supramodal brain. *Front. Psychol.* 4:388. doi: 10.3389/fpsyg.2013.00388

Vetter, P., and Newen, A. (2014). Varieties of cognitive penetration in visual perception. *Conscious. Cogn.* 27C, 62–75. doi: 10.1016/j.concog.2014.04.007

Vuust, P., and Frith, C. (2008). Anticipation is the key to understanding music and the effects of music on emotion. *Behav. Brain Sci.* 31, 599–600. doi: 10.1017/S0140525X08005542

Wassum, K. M., Ostlund, S. B., Balleine, B. W., and Maidment, N. T. (2011). Differential dependence of pavlovian incentive motivation and instrumental incentive learning processes on dopamine signalling. *Learn. Mem.* 18, 475–483. doi: 10.1101/lm.2229311

Wong, Y. K., Folstein, J. R., and Gauthier, I. (2012). The nature of experience determines object representations in the visual system. *J. Exp. Psychol. Gen.* 141, 682–698. doi: 10.1037/a0027822

Yuille, A., and Kersten, D. (2006). Vision as Bayesian inference: analysis by synthesis? *Trends Cogn. Sci.* 10, 301–308. doi: 10.1016/j.tics.2006.05.002

Zeki, S. (1999). *Inner Vision. An Exploration of Art and the Brain.* Oxford: Oxford University Press.