frontiers

RESEARCH TOPICS

LINKING PERCEPTION AND COGNITION

Topic Editors
Michela C. Tacca and Arnon Cahen

ABOUT FRONTIERS

Frontiers is more than just an open-access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

FRONTIERS JOURNAL SERIES

The Frontiers Journal Series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing.

All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the Frontiers Journal Series operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

DEDICATION TO QUALITY

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews.

Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view.

By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

WHAT ARE FRONTIERS RESEARCH TOPICS?

Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area!

Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: researchtopics@frontiersin.org

LINKING PERCEPTION AND COGNITION

Topic Editors:
Michela C. Tacca, Heinrich-Heine University Düsseldorf, Germany
Arnon Cahen, Ben Gurion University of the Negev, Israel

Image by Arnon Cahen.

Perception and cognition are tightly related. As our primary mode of contact with the world, perception is the informational and causal foundation of our cognitive processes; it is fundamental to our empirical thinking, believing, and action planning. Traditional accounts of the mind consider perception and cognition to be distinct, yet highly interrelated, systems. Much interdisciplinary empirical and theoretical work, from cognitive scientists and philosophers, has attempted to elucidate the complex relations holding among these systems, suggesting the existence not only of influences of perception on cognition but also vice versa.

However, many questions are left unanswered. Given that perception is a guide to our thinking about and acting upon the world appropriately, the two systems must be able to 'talk' to one another; information carried by perception must be of a form adequate to be 'taken in' by our various cognitive systems. The question arises: What kinds of structural similarities must hold between perceptual and cognitive representations for such 'communication' to be possible? And how exactly do perceptual and cognitive representations interact? Some researchers have argued that the links between perception and cognition are so tight as to instill doubt as to the significance of the distinction between the two systems in the first place. Yet others insist that perception and cognition remain distinct.

The aim of this Research Topic is to deepen our understanding of the kinds of interaction among perception and cognition and of the nature of the representational structures that would enable such interaction. It proposes to do so by bringing together theoretical and empirical contributions that discuss the relation between perception and cognition from the following perspectives:

First, the relationship between perception and cognition call upon us to ask: What are the similarities between the respective representational structures and processes of perception and cognition?

Second, what mechanisms, if any, mediate between perceptual and cognitive processes? It has recently been argued that attention might be such a mechanism. However, it is still unclear how exactly attention carries out this mediating role, and whether it is both a necessary and sufficient condition for interaction between the two systems. Other mechanisms might be implicated as well.

Third, thinking about the relations between perception and cognition calls for the investigation of such relations involving distinct perceptual modalities. In fact, it is an open question whether the different senses share the same structural properties. It may be that similarities holding among certain perceptual modalities and cognition do not obtain with others.

The Research Topic is open to both theoretical and empirical contributions from different fields (e.g., philosophy, psychology, and neuroscience) in the form of original research articles, hypothesis and theory articles, reviews, and commentaries.

Table of Contents

Linking perception and cognition

Arnon Cahen[1]* and Michela C. Tacca[2]

[1] Department of Philosophy, Haifa University, Haifa, Israel
[2] Department of Philosophy, Heinich-Heine University Düsseldorf, Düsseldorf, Germany
*Correspondence: acahen@wustl.edu

As our primary mode of contact with the world, perception is the causal and informational foundation for our higher cognitive functions—it guides our thinking about and acting upon the world. It is therefore unsurprising that so much empirical and theoretical research is devoted to the study of the complex interrelations between perception and cognition. Nor is it surprising that such research spans traditional disciplinary boundaries and attracts the interest and efforts of researchers from the full spectrum of the cognitive sciences, psychology, neuroscience, philosophy, and others. This Research Topic aims to contribute to this expansive research project—the exploration of the perception/cognition interface—while respecting its essentially interdisciplinary character.

Given that perception is *the* input to cognition, the two systems must be able to "talk" to each other; at the very least, information carried by perception must be of a form adequate to be "taken in" by our various cognitive systems. The central questions of this Research Topic surround the nature of their "communication." In particular, we ask: what kinds of structural relations must hold between perceptual and cognitive representations for such communication to be possible? To what extent, if at all, is it a "dialog?" And, what mechanisms mediate the transitions between the two systems? As the papers here included exemplify, all these questions are open to exploration from within a variety of different disciplines and perspectives.

The question on the structural relation between perception and cognition is addressed by Tacca (2011) who argues that there are important structural similarities between early vision and higher cognitive systems underlying active empirical thought—in particular, she argues that both involve systematic representations. Given that systematicity is a central feature of representations employing conceptual capacities, the systematicity of early vision implies that it might count as an early type of conceptual representation. Such a view helps explain how perceptual representations can inform our thoughts about the world. Raftopoulos (2011), on the other hand, proposes that *early* vision might not involve cognition-like structural properties, though he continues to argue that *late* vision is cognitively penetrable and not purely visual. Nonetheless, unlike representations involved in higher cognition, late vision representations do not support inferences and discursive reasoning. This view is shared by Tacca, who, for this reason, considers perceptual representations as an early type of conceptual representations.

Roth and Franconeri (2012) suggest a further property shared between vision and high-level cognition: asymmetry.

The asymmetry in question is found in spatial language: in sentences like "A is above B," A and B are assigned different roles: the first is the "figure" and the other "ground." Similarly, they argue, in vision, the perception of A being above B may be asymmetrically encoded such that the one is at the "spotlight" of attention, and has a certain, behaviorally manifest, priority over the other. They further implicate spatial attention as the mechanism underlying this asymmetric encoding.

Potter (2012) explores which mechanism might mediate between perception and cognition by reviewing evidence for conceptual short-term memory—a mental buffer in which perceptual stimuli and their related concepts come together for a brief time. This allows for the identification of meaningful patterns and structures. Potter considers perception and cognition as roughly continuous—as different processing stages operating on the same representational format.

Another possible mechanism linking perception and cognition is attention. Brown et al. (2011) expand upon the thought that attention mediates spatial visual perception and investigate its possible role in biasing proprioceptive signals in the motor system. They investigate further possible similarities between the attentional mechanisms underlying spatial perception and those operating on proprioceptive inputs.

Goldstone et al. (2011) investigate a specific perceptual mechanism that might be influenced by cognition: the extraction of distant similarities. They suggest that high-level cognition directs attention so as to modulate what they call "categorical perception." They argue that perception is plastic and that "it is not just perceptual sensitivities that are driving the categories, but rather the acquired categories are also driving perceptual sensitivities, (p. 385)." They further emphasize: "There is little, if any, gap between perception and high-level cognition because perceptual systems adapt to fit the needs of high-level cognition, (p. 385)." Perception is suffused by cognition. Goldstone et al.'s account is based on the so-called feature-based approach to concepts; namely, the idea that concepts are composed of perceptual stimuli.

Stöckle-Schobel (2012) reviews the success of the feature-based approach to concept acquisition (see, Schyns and Rodet, 1997; Goldstone and Barsalou, 1998; Barsalou, 1999; Goldstone et al., 2011). He argues that if a feature-based theory of concept learning is to be successful it must address a number of philosophical challenges originally posed by Fodor (1981, 2008) against the possibility of acquiring genuinely novel concepts.

He argues that these challenges have not been fully met by the feature-based approach, and provides recommendations for how proponents of such an approach might eventually meet them.

Tanaka et al. (2012) further explore the role of categorization in the process of perceptual recognition. They do so via an exploration of the atypicality bias—the finding that objects that are equally similar to two examples of the same category will be perceived as more similar to the more atypical of the two. This suggests that perceptual recognition of some object is influenced by the density and organization of exemplars in the object's category space.

Lebrecht et al. (2012) focus on the affective content of perceptual representations. They argue that even the very subtle valence, or micro-valence, of an object has a significant unconscious influence on how those objects are perceptually represented and recognized. They argue that such effects are so robust that affect can be understood as a straightforwardly perceptible property of objects.

Finally, Mroczko-Wąsowicz and Werning (2012) provide a sensory-motor account of synesthesia that considers the role of top–down associations in shaping the synesthetic experience. Particularly, they argue that swimming style synesthesia can be seen as a case of hyperbinding that—unlike normal binding (Barsalou, 2008)—combines sensory attributes that do not normally form a concept frame. This interpretation calls for a larger integration between visuo-motor representations and conceptual representations.

Taken together the papers in "Linking Perception and Cognition" offer a wide perspective of the theoretical and empirical research on the nexus between how we perceive the world and how we think about it. These papers provide useful guides toward better understanding the connections that make our mental life so rich, indeed, that make it possible at all.

REFERENCES

Barsalou, L. W. (1999). Perceptual symbol system. *Behav. Brain Sci.* 22, 566–660.

Barsalou, L. W. (2008). Grounded cognition. *Annu. Rev. Psychol.* 59, 617–645.

Brown, H., Friston, K., and Bestmann, S. (2011). Active inference, attention, and motor preparation. *Front. Psychol.* 2:218. doi: 10.3389/fpsyg.2011.00218

Fodor, J. A. (1981). "The present status of the innateness controversy," in *Representations*, ed J. Fodor (Cambridge: MIT Press), 257–316.

Fodor, J. A. (2008). *LOT 2: The Language of Thought Revisited.* Oxford: Oxford University Press.

Goldstone, R. L., and Barsalou, L. W. (1998). Reuniting perception and conception. *Cognition* 65, 231–262.

Goldstone, R. L., Landy, D., and Brunel, L. C. (2011). Improving perception to make distant connections closer. *Front. Psychol.* 2:385. doi: 10.3389/fpsyg.2011.00385

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

Mroczko-Wąsowicz, A., and Werning, M. (2012). Synesthesia, sensory-motor contingency, and semantic emulation: how swimming style-color synesthesia challenges the traditional view of synesthesia. *Front. Psychol.* 3:279. doi: 10.3389/fpsyg.2012.00279

Potter, M. C. (2012). Conceptual short term memory in perception and thought. *Front. Psychol.* 3:113. doi: 10.3389/fpsyg.2012.00113

Raftopoulos, A. (2011). Late vision: processes and epistemic status. *Front. Psychol.* 2:382. doi: 10.3389/fpsyg.2011.00382

Roth, J. C., and Franconeri, S. L. (2012). Asymmetric coding of categorical spatial relations in both language and vision. *Front. Psychol.* 3:464. doi: 10.3389/fpsyg.2012.00464

Schyns, P. G., and Rodet, L. (1997). Categorization creates functional features. *J. Exp. Psychol. Learn. Mem. Cogn.* 23, 681–696.

Stöckle-Schobel, R. (2012). Perceptual learning and feature-based approaches to concepts – a critical discussion. *Front. Psychol.* 3:93. doi: 10.3389/fpsyg.2012.00093

Tacca, M. C. (2011). Commonalities between perception and cognition. *Front. Psychol.* 2:358. doi: 10.3389/fpsyg.2011.00358

Tanaka, J. W., Kantner, J., and Bartlett, M. (2012). How category structure influences the perception of object similarity: the atypicality bias. *Front. Psychol.* 3:147. doi: 10.3389/fpsyg.2012.00147

Commonalities between perception and cognition

Michela C. Tacca*

Department of Philosophy, Heinrich-Heine University Düsseldorf, Düsseldorf, Germany

Perception and cognition are highly interrelated. Given the influence that these systems exert on one another, it is important to explain how perceptual representations and cognitive representations interact. In this paper, I analyze the similarities between visual perceptual representations and cognitive representations in terms of their structural properties and content. Specifically, I argue that the spatial structure underlying visual object representation displays systematicity – a property that is considered to be characteristic of propositional cognitive representations. To this end, I propose a logical characterization of visual feature binding as described by Treisman's Feature Integration Theory and argue that systematicity is not only a property of language-like representations, but also of spatially organized visual representations. Furthermore, I argue that if systematicity is taken to be a criterion to distinguish between conceptual and non-conceptual representations, then visual representations, that display systematicity, might count as an early type of conceptual representations. Showing these analogies between visual perception and cognition is an important step toward understanding the interface between the two systems. The ideas here presented might also set the stage for new empirical studies that directly compare binding (and other relational operations) in visual perception and higher cognition.

Keywords: systematicity, generality constraint, conceptual content, non-conceptual content, attention

INTRODUCTION

Perception and cognition are tightly related. Perceptual information guides our decisions and actions, and shapes our beliefs. At the same time our knowledge influences the way we perceive the world (Brewer and Lambert, 2001). To the extent that perception and cognition seem to share information, it seems there is no sharp division between the realm of cognitive abilities and that of perceptual abilities. An example is visual perception. Visual processing is composed of different stages (Marr, 1982): early, intermediate, and late vision. Roughly, at early stages of the visual system, processes like segregation of figure from background, border detection, and the detection of basic features (e.g., color, orientation, motion components) occur. This information reaches intermediate stages, where it is combined into a temporary representation of an object. At later stages, the temporary object representation is matched with previous object shapes stored in long-term visual memory to achieve visual object identification and recognition. While early visual processes are largely automatic and independent of cognitive factors, late visual stages are more influenced by our knowledge (Raftopoulos, this issue). Examples of cognitive influence on how we perceive the world – that modulates late vision – are visual search and attention (Treisman, 1993). Knowing the color or shape of an object helps a person to quickly identify that particular object in a cluttered visual scene (Wolfe and Horowitz, 2004). Phenomena like visual search highlight the fact that visual perception at later stages depends on both sensory and cognitive factors. Late vision is at what philosophers call the personal level: we have conscious access to information represented at this stage and we can exploit it for action planning and thinking (Lamme, 2003; Block, 2005). This is apparently not the case for early visual stages, which

occur at a subpersonal level, without a person being aware of the information being processed at that stage. Intermediate stages, on the other hand, are probably accessible at a personal level. The degree of representational awareness occurring at this stage is commonly identified with phenomenal consciousness (Lamme, 2003; Raftopoulos and Mueller, 2006): we get a gist of the perceived scene, but it is not possible to retrieve detailed information of the objects' features. It is a matter of debate to what extent intermediate stages of visual processing are influenced by our knowledge (i.e., are cognitively penetrable). Some authors argue that those stages are purely visual (Raftopoulos and Mueller, 2006) and that the transition between pure perception to cognition occurs only at later visual stages, when temporary object representations are matched for recognition and identification. In this paper, I will not propose an argument for whether early and intermediate stages of visual perception are cognitively penetrable. However, I would like to stress that some of the common properties between visual perception and cognition that I will consider already occur at intermediate stages, thus, casting doubt on the claim that mid-level vision is purely perceptual.

Cognitive information influences perceptual processes, but, at the same time, cognitive processes depend on perceptual information (Goldstone and Barsalou, 1998). Recent work in philosophy brought new vigor to the hypothesis originally proposed by British Empiricists that cognition is inherently perceptual (Prinz, 2002): cognitive/conceptual tasks have their roots in perception and they rely on perceptual mechanisms for their processing. Such theoretical proposals are supported by empirical findings from psychology. Work on concept acquisition shows that functions (e.g., categorization, inference) that are associated with cognition have their

basis in perceptual systems (Barsalou, 1999) and that perceptual and conceptual processes share common mechanisms (Pecher et al., 2004). The basic hypothesis is that a concept is represented by means of a simulation at the sensory level of an experience of that to which the concept truly applies. For example, to represent the concept APPLE[1], perceptual systems for vision, action, and touch partially produce the experience of a particular apple. Taken together, work on the influence of knowledge on the character of one's perceptual experience and on perceptual information shaping one's conceptual abilities provides evidence for perception and cognition being related systems.

Though it seems to be common ground that cognitive and perceptual representations influence each other, they are not taken to be the same kind of representations. Neurophysiological studies distinguish different functional areas for sensory and cognitive systems. Those areas process specific inputs and specialize in different kinds of information processes (Zeki, 1978; Felleman and Van Essen, 1991). And distinct sensory areas can be treated as separate modules (Barrett, 2005) that deal with their specific representational primitives.

From a philosophical point of view, visual perception and cognition process information by means of representations that differ in both their structure and content (Heck, 2007; Fodor, 2008). One of the main characteristics of cognitive states, paradigmatically of thoughts, is that they have a propositional combinatorial structure that satisfies the requirement of the Generality Constraint (Evans, 1982). The Generality Constraint describes the pervasive ability of humans to entertain certain thoughts that they have never had before on the basis of having entertained the components of these new thoughts in other preceding situations. For example, from the fact that a person can think that the sky is blue and the car is gray, she can also think that the sky is gray and the car is blue, even if she has never had this thought before. The new thought depends on her conceptual ability to combine already acquired concepts in different ways. This regularity of human thinking is explained by appealing to the fact that thoughts are mental representations with a sentential combinatorial structure (Fodor, 1975). Thoughts are built up by combining primitive constituents according to propositional rules. The thought 'the car is gray' depends on the tokening and combination of the concepts CAR and GRAY and the rule of composition for the verb 'to be.' Recombination of concepts in cognitive processes displays a constituent structure. The constituent structure of thought is such that whenever a complex representation is tokened its constituents are simultaneously tokened. Failure to represent car or grayness leads to failure to represent that the car is gray. The appeal to the constituent structure of cognitive representations allows us to explain a further property of these representations: their systematicity (Fodor and Pylyshyn, 1988). Systematicity, similar to the Generality Constraint, describes the human ability to entertain semantically related thoughts. For example, the ability to entertain a certain thought about cars is connected to the ability to entertain certain other thoughts about cars: thoughts like 'the car is

gray' and 'the car is blue' share the same constituent 'car.' That is, the semantic systematicity of thought is explained by postulating a system of representations with a combinatorial syntax.

Systematic recombinations are necessary to satisfy the Generality Constraint but not sufficient. In fact, systematicity is a weaker requirement than the Generality Constraint since it lacks the "generality" part. According to the Generality Constraint, once a thinker can entertain a thought, elements of this thought could be in principle indefinitely recombined with every other appropriate concept that a person possesses. This requirement is not part of systematicity, since it leaves open whether it is in principle possible that a finite type of systematicity exists (Fodor and Pylyshyn, 1988). For what concerns the analysis of the structure of visual representations, I will mostly focus on whether those representations implement a systematic structure of constituents. I will then discuss the "generality" requirement in the analysis of the content of visual representations.

Acceptance of the Generality Constraint, or the weaker systematicity requirement, also affects how we characterize the content of cognitive and perceptual representations. Philosophers distinguish between two types of content: conceptual and non-conceptual content (Evans, 1982; Bermúdez, 2007). Typical cases of mental states with conceptual content are cognitive mental states, like thought, belief, desire, and so on: their content – what is thought, believed, and desired – is a function of the concepts a person possesses, where concepts are taken to be the constituents of thoughts and other cognitive states. Mental states with non-conceptual content, on the other hand, are states the having of which does not depend on the subject's possessing any of the concepts required to specify the content of that state. Perception, both personal and subpersonal, is considered a paradigmatic example of states with non-conceptual content. In other words, to have the thought that an apple is red, one has to possess the concepts involved in that thought, but to have a perceptual experience characteristic of seeing a red apple one does not need to possess the concepts involved in the specification.

One way of distinguishing conceptual and non-conceptual content appeals to a mental representation's satisfaction of the systematicity requirement (Toribio, 2008; Camp, 2009). It has been argued that perceptual representations, specifically visual representations, do not satisfy the requirement of systematicity, and, hence, unlike cognitive representations, do not have conceptual content (Heck, 2007). The argument is based on the idea that visual representations have a pictorial nature. Pictorial theories equate visual representations to images or maps. Like images or maps, visual representations are spatially characterized: at each point in an image or map a specific trait (color, shape, etc.) occurs. Furthermore, like images or maps, visual representations have a holistic character. Unlike cognitive representations, there is no unique structured propositional representation that determines the content of a visual representation. There are many distinct possible decompositions of the same image, such that it is impossible to both identify which are its constituent parts and disentangle the role of these parts in the building up of the pictorial representation. Thus, visual representations, like maps, seemingly lack the syntactic structure of constituents typical of cognitive representations. The lack of a constituent structure entails that visual

---

[1]I follow the common practice in philosophy to capitalize terms that refer to concepts.

representations are not systematic. Satisfying systematicity is a necessary condition on satisfying the Generality Constraint. For the reasons above, visual representations do not seem to satisfy systematicity, and hence the Generality Constraint. Therefore, they have a content of a different kind than the content of cognitive representations: they have non-conceptual content.

If visual perception and cognition do indeed have different structural properties and content, then it becomes difficult to understand how perceptual representations are "translated" into cognitive representations. This is both an empirical and theoretical question. From the philosophical point of view, finding out the relationship between perception and cognition will be of benefit to explain phenomena as different as concept formation and acquisition, belief justification, and demonstrative thinking, each of which partly depends on perceptual information.

In this paper, I will focus on commonalities between visual perception and cognition that might help explain the communication between those systems. In the first part, I will show that the spatial recombination underlying visual object recognition satisfies the requirement of systematicity. The analysis will take into account the so-called Feature Integration Theory (Treisman and Gelade, 1980); a model that explains visual object representation by considering the spatial nature of visual representations. Although Feature Integration Theory characterizes visual representations as spatially organized, it differs from pictorial theories of visual representations, since it does not commit to the view that visual representations are holistic. In fact, visual representations can be seen as states of the visual system that can be neuronally specified, such that each part of an object representation can be spelled out by considering the different neuronal activations (Treisman and Gelade, 1980; Goldstone and Barsalou, 1998). Each neuronal activation roughly corresponds to a part, or primitive constituent, of the representation. Thus, one can decompose an object representation into its primitive constituents and analyze whether a systematic structure of constituents is displayed by visual spatial recombinations (Tacca, 2010). In the second part, I will argue against the claim that visual representations have non-conceptual content. Based on the analysis in the first part of the paper, I will propose that, if one takes systematicity to be a necessary requirement for having conceptual content, visual representations might be an *early* type of conceptual representations. I conclude that understanding the link between perception and cognition requires considering whether they satisfy common requirements in terms of structure and content. These similarities might be at the basis of the translation of perceptual representations into cognitive representation and elucidate the mechanism of their interaction.

PRIMITIVE VISUAL FEATURES AND THE BINDING PROBLEM

Recombination in cognitive processes depends on operations on primitive constituents. A primitive constituent is an entity that corresponds to the smallest meaningful representation carrying relevant information for the processing of more complex representations. Different theories posit different types of primitive constituents (Smolensky, 1990; Fodor, 1998). However, there is agreement that the primitive mental representations involved in thought and other cognitive processes, like belief and desire, are concepts. According to an atomistic perspective, concepts cannot

be further decomposed into more primitive elements and as such they are the building blocks of thoughts (Fodor, 1975). However, others have argued that concepts can be further decomposed into their perceptual components (e.g., Barsalou, 1999). For example, the concept APPLE can be decomposed into its constituent concepts: COLOR, TEXTURE, SHAPE, etc. At the same time, each part can be further decomposed into more elementary constituents like GREEN, BROWN, SMOOTH, and ROUND. Those elementary constituents are taken to be symbolic perceptual representations stored at late perceptual stages that become part of cognitive recombinations. Therefore, they share with cognitive representations systematicity, compositionality, and productivity (Barsalou, 1999). In the following, I will show that intermediate visual representations that contribute to object perception but are not yet stored at late visual stages also display systematicity.

The hypothesis that concepts have a structure of constituents that involves perceptual representations is based on anatomical, physiological, and psychophysical evidence for the existence of distinct representations for primitive visual features. Neurobiological (Zeki, 1978; Livingstone and Hubel, 1988; Felleman and Van Essen, 1991) and psychophysical studies (Treisman and Gelade, 1980) report the existence in visual areas of so-called feature maps. Feature maps code for specific object features, like color, motion, and orientation. They are also topographically organized; namely, they represent a specific feature and the specific location in which the feature occurs in the visual field. Thus, any visual object we perceive is first decomposed into its primitive components and only later those components are recombined into a coherent object representation. But what makes color, motion, and orientation count as primitive features not further decomposable? To provide an answer to this question is important, since if we can show that there is an empirically reasonable standard for primitive recombinable features, then we can challenge one of the central motivation for thinking that visual perception does not display systematicity and that the content of visual representations is non-conceptual; namely, the claim of pictorial theories for which there is no unique decompositions of visual representations into a proper structure of constituents.

The definition of a primitive visual feature not further decomposable depends on experimental consideration (Wolfe, 1998). First, a primitive feature allows for efficient visual search when embedded in a cluttered scene of unlike distracters. The efficiency of visual search is indicated by the so-called "pop-out" of the target that is independent of how many items are present in the visual field. Second, a primitive feature supports effortless texture segregation. For example, a region of vertical lines in a field of horizontal lines will be immediately segregated from the background and perceived as a figure. Color, orientation, and motion justify the criteria of efficient search and effortless segmentation, and are, thus, primitive features. Furthermore, these features are represented by different visual cortical areas, each of which is retinotopically organized. Taken together, neurophysiological and psychophysical findings uncover the fact that visual features are the primitive constituents of visual object representations.

Once primitive visual features have been individuated, the subsequent main question is how those features are combined. In light of the complexity of natural visual scenes, it is striking that features

are almost never miscombined in our perception. In fact, this is even remarkable for the simplest possible scenes, such as one with a red-horizontal bar and a green-vertical bar and another one with a green-horizontal bar and a red-vertical bar. These scenes contain identical features that are combined in different ways. The challenge consists in individuating objects by their unique combination of features, so as to distinguish, for example, the red-vertical bar from the green-horizontal bar. Jackson (1977) described the problem of feature recombination as the Many-Property problem. Research in vision science has approached this problem under the label of "binding problem" (Roskies, 1999). An example of what the binding problem involves comes from studies of visual conjunction search (Treisman and Gelade, 1980). A typical case of feature integration is to show a subject a scene in which red-vertical bars, red-horizontal bars, green-horizontal bars, and one green-vertical bar are presented together. The subject is asked to identify the green-vertical bar. In order to detect the right target, something like a comparison between the right orientation and the right color has to occur. It has been shown that in the case of identification of objects that share different features (orientation and color in the example case) selective attention is at play (Treisman, 1996). Further evidence for the binding problem being solved by an attentional mechanism comes from studies of illusory conjunctions in healthy subjects (Treisman and Schmidt, 1982) and patients suffering from Balint's syndrome (Robertson, 2003). Healthy subjects are asked in a laboratory setting to report properties of visually presented stimuli under high attentional load. Results show that they report a high number of illusory conjunctions. For example, when shown a screen with blue squares and red triangles, they report wrong recombinations of presented features, e.g., a blue triangle. A high rate of illusory conjunctions occurs if similar experiments are performed with Balint's syndrome patients. These patients suffer, among other things, from an attentional disruption, providing more evidence for the role of attention in successful binding.

The reported findings support the so-called Feature Integration Theory (Treisman and Gelade, 1980). Feature Integration Theory is one of the most influential models of visual feature binding that considers the role of attention and the spatial layout of feature maps as the basic ingredients for successful feature binding. Other influential models have been proposed for explaining the binding process, such as the hypothesis of binding by synchrony that considers synchronized neuronal mechanism as the basic binding mechanism (Engel et al., 1991). Furthermore, besides the spatial-attentional mechanism posited by Feature Integration Theory, also object-based attention might be necessary to integrate features (Blaser et al., 2000). The hypotheses of binding by spatial attention, synchrony, and of the role of object-based attention are not mutually exclusive (Tacca, 2010). It might be that all these factors are at play during the binding process. Indeed, empirical studies show the relation between spatial attention and synchrony (Fries et al., 2001) and between object-based and spatial attention (Scholl, 2009) in building up an object representation. Here, I will only focus on the role of spatial attention to bind features, in order to show that spatial representations display systematicity in a way similar to cognitive-sentential representations.

According to Feature Integration Theory, selective attention acts as the active binding mechanism. Whenever a person focuses her attention on a specific object location in the visual field, the features at that location are represented in the corresponding location in the feature maps. By selecting all the features occupying a specific location, attention integrates these into a coherent object representation. More specifically, the focus of attention selects an object location within a topographically organized master map of location (Treisman, 1993) or saliency-map (Koch and Ullman, 1985). This saliency-map represents the saliency of objects at each location of the visual field, because it combines the information about all features' saliency from all the specific feature maps, which it receives via topographically organized connections from the feature maps. Within each feature map, the saliency at a given location is determined by two classes of factors (Wolfe, 1998): (i) bottom-up saliency, that is, the local feature gradient (Koch and Ullman, 1985); and (ii) top-down factors, like the match between a stimulus feature and the features of the object that a person is currently searching for (Wolfe, 1998).

Independently of whether the saliency of individual locations is governed by bottom-up or top-down factors, the saliency representation in the saliency-map is always generated by combining the outputs from feature maps in a fashion that preserves topography. That is, the saliency-map receives information about the different object locations – suppose that locations are indexed with i, l, m, n, etc. – and their conspicuity values from distinct feature maps. If location$_i$, signaled by the feature map$_\alpha$, is the same as location$_l$ ($i = l$), signaled by the feature map$_\beta$, they will activate the same portion of the saliency-map. The saliency value of this location will then depend on the conspicuity of both location$_i$ and location$_l$. The saliency-map only codes for saliency at a given location. Thus, the saliency-map represents the locations of objects but has no information about which features occur at those locations. In order to recover which features determine the object's shape and surface, information within the topographic feature maps has to be selected for binding and further processing of object identity. A "winner take-all" mechanism selects the location in the saliency-map that is the most salient at any given moment (Koch and Ullman, 1985). This determines where the focus of attention will next move. Via topographically ordered feedback connections from the saliency maps to the corresponding locations in the feature maps, the features at that location (e.g., features occurring at both location$_i$ in the feature map$_\alpha$ and location$_l$ in the feature map$_\beta$, since $i = l$) are jointly selected for further processing, and, in this way, bound. These integrated features are stored as temporary representations – called by some authors an object-file (Kahneman et al., 1992) – in which their constituting information of location is indexed. Hence, in models based on Feature Integration Theory, the representation of objects' locations is fundamental for integrating their features.

In this framework, the difference between saliency being governed by bottom-up or top-down factors amounts to the distinction between exogenous and endogenous attention. Exogenous attention is governed by stimulus property: it is attracted by the conspicuity of an object in the perceived scene. If you are attending a seminar and a fly suddenly enters into the room, you will immediately spot and follow it. No matter how much you are interested in the seminar. Endogenous attention is governed by a subject's tasks and plans. You want to wear your favorite

pullover and you go through the content of your messy closet to find it. You will drive your attention to the location where you thought the pullover should be, if you are lucky your search is over, but, as often happens, you will have to scan through different locations before you can find it among all the other similar cloths.

Note that, in the sequence of processes postulated by Feature Integration Theory, the binding process is separate from the representation of location saliency. In principle, binding can be disrupted without a disruption of the saliency representation. Thus, in this framework, attention and binding can come apart. To illustrate a scenario in which such dissociation occurs, let us assume that we selectively interrupt the feedback connections from the saliency-map to the feature maps, leaving everything else intact. Then, there will still be a most salient location selected in the saliency-map and only the final process in the above sequence will be disrupted. Suppose that the perceived scene is one with a green-vertical bar and a red-horizontal bar. Object features are represented in feature maps according to their location: $green_i$, $vertical_l$, red_m, and $horizontal_n$. Information about feature locations is sent to the saliency-map, which computes the most salient location. In the saliency-map, $location_i$ and $location_l$ activate the same area ($location_i = location_l$), since they bring information about the same object location, and $location_m$ and $location_n$ activate the same area ($location_m = location_n$) that is different from the location of the object signaled by $location_i$ and $location_l$. Suppose that the location of i and l is the most salient, then attention will be directed to this location and a signal to select features "indexed" i and l will be sent to the feature maps. Since the feedback connections from the saliency-map to the feature maps are disrupted, features in the feature maps belonging to the same location cannot be selected. The feature maps will encode for features and their locations, but there is no selective feedback signal that routes only those features from the selected location to the next step of object processing that binds them. This might result in perceptual misbinding because features from many locations are spuriously sent on to higher-level object processing. In fact, one possibility is that psychophysical manipulations leading to illusory conjunctions (Treisman and Gelade, 1980) work by interrupting the feedback from the saliency-map to the feature maps, just as in this thought experiment. For proper binding, information about features occupying the same identical location has to be routed from the feature maps to higher processing stages.

Also note that, even with disturbed saliency representation, and thus disturbed attention, some feature binding (even if erroneous) occurs. An empirical example for this can be found in Balint's syndrome patients. Spatial attention in these patients is disrupted, yet they still report a (wrong) recombination of features. Thus, even without spatial attention, some erroneous form of binding can occur. The fact that attentional selection and feature binding are tightly related, yet distinct processes, is of importance for the analysis of the binding process in the logical terms that are proposed in the next section.

Briefly, the main ingredients of Feature Integration Theory are the representation of primitive features, their spatial location, and attention. The interaction of these elements gives rise to the perception of objects in a scene in which features are correctly conjoint. This might solve the Many-Property or binding problem at least in the case of visual object representation.

SYSTEMATIC RECOMBINATION OF SPATIALLY ORGANIZED REPRESENTATIONS

Models based on Feature Integration Theory describe visual object representations as the outcomes of recombinations of primitive visual constituents. This contrasts with pictorial theories of visual perception in philosophy (e.g., Heck, 2007; Fodor, 2008) that argue that visual representations have a holistic nature. Visual representations, like images or maps, can be decomposed in many different ways: to each visual representation might correspond a different decomposition of constituents. That means that any kind of decomposition of a visual representation into its constituents makes the same contribution to the final object representation. The decomposition of, for example, a visual representation of a flower into (petals, stem, leaves) is as good as the decomposition (part of petal₁, roots, sepal, stalk). Therefore, visual representations are, unlike cognitive representations, not canonically decomposable (Fodor, 2007): while the decomposition of a thought representation allows only a unique decomposition – e.g., 'John loves Mary' decomposes into JOHN, LOVES, and MARY – iconic representations have infinitely many decompositions, none of which is canonical. Having a structure of primitive constituents depends on the individuation of the unique parts of a canonical decomposition. Since visual representations seemingly fail to canonically decompose, they lack a structure of primitive constituents. To implement a structure of constituents is a pre-requisite for explaining the systematic behavior of cognitive processes. The relation of constituency is defined as a mereological relation; namely, as a relation of parts to whole (Fodor and Mclaughlin, 1990): every time the expression E is tokened, its constituents $<e_1, \ldots, e_n>$ are tokened, too. In a classical account of thought processes, systematicity results from processes that are sensitive to the structure of constituents: the ability to entertain related thoughts depends on the fact that different combinations of constituents have the same syntactic structure. As an example, the thoughts 'John loves Mary' and 'Mary loves John' share the same structure, even if the constituents are differently arranged. According to the pictorialists, because of the holistic character of visual representations, those representations fail to implement such a structure of constituents, and, as a consequence, they do not display systematicity.

Empirical evidence casts doubt on the main assumption of pictorial theories: that perceptual representations have a holistic character, and therefore lack systematicity. Evidence from vision science shows that visual object representations depend on the recombination of neuronally specified primitive features. These features can be uniquely determined in terms of neuronal activations, and they are represented in distinct feature maps. Experimental considerations make clear that features represented in the feature maps are primitive and not further decomposable. Object representations then depend on the spatial recombination of those features. It seems plausible that such recombinations display systematicity; namely that visual scenes that are structurally related (e.g., to see a red-vertical bar to the left of a green-horizontal

bar and *vice versa*) share the same primitive visual features (i.e., 'green,' 'horizontal,' 'red,' and 'vertical'). In order to show that this is indeed the case, one has to first argue that visual representations implement a mereological structure of constituents, such that every time an object representation is tokened its primitive features are tokened, too; and, second, that the visual system implements a systematic structure of constituents; namely, that visual features make the same contribution in structurally related visual scenes.

The analysis of the type of structure implemented in the process of binding by attention, as described by Feature Integration Theory, can be given in logical terms (Clark, 2004a; Tacca, 2010). Binding involves predication and identity: features are considered to be the predicates of the same sensory individual that, in the case of Feature Integration Theory, is the object location. The reason for introducing identity is that a pure conjunction of terms might lead to different representations of the same scene, each of which would be valid. Consider, for example, the simple visual scene with a red-vertical bar and a green-horizontal bar. Its decomposition only by means of conjunction would be: (red and vertical and green and horizontal). The recombination of those features could lead to two distinct visual scenes: one in which there are a red-vertical bar and a green-horizontal bar, and one in which there are a red-horizontal bar and a green-vertical bar. This kind of ambiguity does not occur in object perception. The binding process normally produces a unique representation of the objects in the environment. This unique representation is partly achieved when features are processed as occurring at the same location. Ideally, the process within the visual system can be seen as doing something like scanning a location and applying a specific tag to the features occurring at that location (maybe by keeping track of that location within object files). For example, all the features occurring at the location i are indexed or tagged with i, and all features occurring at a distinct location m are indexed with m. If the location m and i do not overlap; namely, features in i and m do not occur at the same location, then features are bound into two separate object representations. In real-world perception of cluttered visual scenes, attention serially selects one location after the other, binding the features at each of them. To this extent, the role of attention is to secure identification: it determines when features have a common subject matter and allows for the identification of, and discrimination between, different objects (Clark, 2004a). Object location is, thus, the key element that secures a successful binding of features. This process can be logically characterized as follows:

<red-vertical bar>:
(at loc_i is R; at loc_l is V; $loc_i = loc_l$ ∴ at loc_i is R and V)
<green-horizontal bar>:
(at loc_m is G; at loc_n is H; $loc_m = loc_n$ ∴ at loc_m is G and H)

The logical characterization of visual feature integration has the advantage of outlining the structure of the binding operations. This characterization is an important tool to compare the spatial structure of visual representation with the propositional structure of thought. I argue that the structure of visual representation resembles the structure of constituents of thought.

In fact, the schema above indicates that the representation of an object depends on its constituents being explicitly represented. If not, the derived object representation is only partial. To determine whether vision has a systematic structure of constituents, it is necessary to investigate whether structurally related visual scenes – i.e., scenes that involve different recombinations of objects or features – share the same constituents, and whether visual constituents contribute in the same way, during the binding processes operating on structurally related scenes, to determine the objects of which they are parts. If visual binding mechanisms meet those requirements, then the binding process has a systematic structure of constituents. A systematic recombination of the example visual scene – a green-horizontal bar to the left of a red-vertical bar – requires that at least one of the features belonging to one of the objects in the scene is shifted, so that, as a result, this feature will change its position. Consider a visual scene with a red-horizontal bar to the left of a green-vertical bar. The representation of the example visual scene and the structurally related scene just described can be schematized as follows:

∗<green-horizontal bar to the left of a red-vertical bar>:
(at loc_i is R; at loc_l is V; $loc_i = loc_l$ ∴ at loc_i is R and V)
(at loc_m is G; at loc_n is H; $loc_m = loc_n$ ∴ at loc_m is G and H)

∗∗<red-horizontal bar to the left of a green-vertical bar>:
(at loc_j is R; at loc_k is H; $loc_j = loc_k$ ∴ at loc_j is R and H)
(at loc_b is G; at loc_c is V; $loc_b = loc_c$ ∴ at loc_b is G and V)

The above configurations show how visual features can be recombined in a systematic fashion by means of combining predicates (features) in a formal language. However, according to Feature Integration Theory, vision does not combine its constituents by means of propositional rules but according to the features' spatial locations. Therefore, it is necessary to provide an argument to explain how visual processes implement the structure just described by means of spatial recombinations.

When two instantiations of the same feature occur at different locations in the world, the feature map coding for that feature will be active. Particularly, it will signal that this specific feature occurs at two distinct locations, corresponding to its locations in the world. In the case of (∗) and (∗∗), the same color maps for green and red, and the same orientation maps for horizontal and vertical are active. But the colors are swapped in the two scenes, leading to different object configurations. The difference between the two configurations is encoded in the change of the activated locations in the color maps. The color map signaling green will be active, to simplify, in its "left side" when representing the location of the green feature in scene (∗), while it will be active in its "right side" when representing green in scene (∗∗). The converse applies for the feature map coding for red. Thus, whenever two visual scenes are structurally related (as in this case), attentional scanning through the scenes will select object locations, thereby leading to a diverse binding of the features in the structurally related scenes. This results in different object representations in the case of (∗) and (∗∗). The binding process is such that primitive constituents are simultaneously tokened with the complex representation. In other words, lacking one of the constituents will result in failure of the binding process.

Thus, feature binding turns out to be more than an associative process that merely links inputs to outputs. In fact, visual binding by spatial attention displays a systematic competency: first, the visual system implements a mereological structure of constituents, rather than processing arbitrarily correlated inputs. Second, the proposed model of visual feature binding displays a systematic structure of constituents. As outlined above, structurally related visual scenes share the same, but differently arranged, primitive features.

Systematicity is a property displayed by both sentential-cognitive representations and spatial representations. This conclusion is in contrast with previous works in philosophy (Clark, 2004b; Fodor, 2008), according to which only representations with a language-like format combine constituents in a way such that a small set of primitive representations can be recombined to form different types of complex representations. In particular, Clark (2004b, p. 571) suggests that sensory states "have something like a subject–predicate structure, though they are not sentential and do not manifest most of the hallmarks of compositionality." In a classical account, a systematic structure of constituents is a distinctive feature of, and tightly related to, compositionality (Fodor, 1998). The requirement of systematicity is explained in terms of the syntactic structure of constituent recombination in thought, whilst compositionality concerns the content of propositional representations. The main idea is that the content of a thought depends on the content of its constituents and the way they are syntactically combined. The reason Clark argues that visual representations do not have traits that satisfy compositionality is because those representations, arising from the binding of primitive features, provide the basis for the conceptual identification of particulars but do not themselves involve conceptual identification; namely, visual primitive representations do not contribute their content to the content of the final object representation.

I argue, instead, that if a system has a structure of contentful constituents, then this system displays at least one of the hallmarks of compositionality: systematicity. It can also be shown that visual representations satisfy a deflationary notion of compositionality – a weaker form of compositionality than the one mentioned here (Tacca, 2010). A deflationary account only requires that (i) vision has a systematic structure, and that (ii) visual primitive constituents have a specific content. But it remains neutral on which types of semantic properties compose, as required by a classical account of compositionality (Fodor and Lepore, 2001). This is a consequence of the spatial, rather than sentential, character of visual representations. In fact, as Clark notices, visual representations are indeed not sentential. This seems to be the case for both primitive features that are bound at intermediate visual stages and for more complex representations that occur at late visual stages.

The spatial nature of visual representations also makes the systematicity of visual representations different from the systematicity of cognitive representations. The explanation of the systematicity of thought involves two parts (Cummins et al., 2001): (i) it entails that having a thought requires having mental representations that express that thought. This also applies to visual representations, since to represent a visual object, the primitive representations that code for its characteristics have to be tokened; and (ii) it entails that mental representations have a language-like combinatorial syntax (and semantic). This is not the case for visual representations. Spatial recombinations underlying visual object representation lack the operational repertoire of language-like recombinations. Visual feature binding requires the integration (and spatial grouping) of local, primitive features. To this extent, operations like conjunction and identity are required. But it is not possible to characterize any of the processes involved in binding in terms of other logical operations. No "visual negation" or "visual disjunction" take place. There is no feature integration that is the negation of any of the integrations that occur within the visual system, and, in contrast with feature conjunction, an explicit feature disjunction does not exist in vision: either features are conjoint or they are not combined at all. In sum, vision does not possess the rich propositional structure that higher-cognitive processes seem to have.

The fact that visual representations do not have a propositional nature highlights the difference in combinatorial processes between the visual and perceptual systems but it does not rule out the possibility that systems with different combinatorial structures can implement the same combinatorial requirement, even if in different ways. This is the case for visual representations that, even if they do not allow for propositional recombinations, display systematicity. Thus, the requirement of systematicity can be considered as a general property that does not depend on the type of operations performed on the primitive constituents.

THE CONTENT OF INTERMEDIATE VISUAL REPRESENTATIONS

Another difference between visual perception and cognition concerns the content of their representations. While cognitive representations have conceptual content, the content of perceptual experience is better described as non-conceptual content. Non-conceptual content is often defined in the following way (Bermúdez and Cahen, 2011): a mental state has non-conceptual content if and only if the subject of that state does not need to possess the relevant concepts required to specify its content.

How to define then the non-conceptual content of perceptual states? Heck (2007) argues that what kind of content perceptual and cognitive states have is a question about what kinds of representations those states involve. Heck's analysis starts from the premise that the conceptual content of beliefs is structured in a way that fulfills the requirement of the Generality Constraint. The debate over non-conceptual content then turns out to be about whether the cognitive abilities one exercises when one thinks that tomatoes are red are also exercised when one veridically perceive a ripe tomato, and whether it would be impossible for one to perceive the tomato as one does were one not able to think as one can. Thus, the question of what kind of content one should take perceptual experience to have has to be answered by investigating the structural characteristics of perceptual representations. The content of perception will be conceptual only if the Generality Constraint is satisfied (Heck, 2007). But, according to Heck, this is not the case,

since visual representations, as described by pictorial theories, have a spatial structure that violates even the weaker requirement of systematicity. Satisfying systematicity is a necessary condition on satisfying the Generality Constraint. Therefore, since visual representations do not display systematicity, their content is non-conceptual.

The analysis proposed in this paper of how visual representations spatially combine leads, instead, to a different conclusion: the appeal to the spatial structure of vision seems to count in favor of the conceptualist thesis, rather than providing a strong argument for the existence of representations with a non-conceptual content. This is because visual representations satisfy the requirement of systematicity – i.e., a necessary condition to satisfy the Generality Constraint. Systematicity is a weak-syntactic reading of the Generality Constraint that states that there is a certain kind of pattern in our cognitive capacities. In this form, the requirement of systematicity describes representational composites as depending on syntactic recombinations involving the same constituents. Recombinations of cognitive representations entail that a person has conceptual abilities (Mclaughlin, 2009). In the case of visual perception, systematic recombinations of primitive features involve the ability of a subject to identify particular features. This ability might correspond to an *early* type of conceptual ability, since visual representations, like cognitive representations, are constituted by primitive constituents that make the same contribution in structurally related representations. Particularly, the representation of features within feature maps is such that whenever a feature is tokened in the feature map (e.g., "red"), this feature will contribute in the same way to the final object representations in which the color red is involved (e.g., a red-vertical bar, a red-horizontal bar). While the contribution of the feature representation is the same in different object representations, those representations will differ from each other as a function of the spatial configuration of their features, since, for different object representations, feature locations are different. This is similar to what occurs in propositional representations, for which, although the same constituent (e.g., the concept RED) contributes in the same way to thoughts regarding red things, the final complex representations depend on the syntactic configurations of the primitive constituents.

However, unlike propositional representations, the possession of systematic perceptual skills is not sufficient to satisfy the Generality Constraint in its strong form, and, thus, not enough to establish both necessary and sufficient conditions for the conceptuality of perceptual representations. The idea behind the Generality Constraint is that conceptual representations involve not only a systematic recombination of primitive constituents but also an abstract grasp on the way things are. Thought representations, and propositional representations in general, are not constrained to any mode of access (Peacocke, 2001). We can, in principle, entertain an indefinite number of thoughts. This is based on the idea that human thoughts have an unbound competence that is not limited by our performance (Fodor and Pylyshyn, 1988; Tacca, 2010). Instead, our perceptual representation of the world is bound to the limit of the perceptual system in use. We cannot perceive an indefinite number of visual scenes, since what we can perceive depends on the physical constitution of our visual system. There is no such thing as an abstract visual competence.

Nevertheless, it can be argued that failure to satisfy the Generality Constraint in its fullest version – that is, by showing both systematic combinability and abstract competence – does not exclude intermediate visual representations from being a specific type of conceptual representations. Perceptual representations might count as an early type of conceptual representations that will become more abstract only with full possession of conceptual resources. These early types of conceptual representations display systematic recombinability – a necessary even if not sufficient condition for a person to possess conceptual abilities. Moreover, the definition of visual representations as early types of conceptual representations will also provide a link between human and animal cognition. Some of the criteria analyzed here, particularly systematicity, have been reported as basic criteria for showing concept possession in animals, too (Newen and Bartels, 2007). Thus, the distinction of the content of perception and cognition based on satisfaction of systematicity does not show that the content of conscious perceptual experience is nonconceptual. At best, one can argue that satisfaction of the requirement of systematicity shows that intermediate stage visual representations, the ones involved in the binding process, might be an early type of conceptual representations. The abstract grasp on ways of representing the world, required by the full satisfaction of the Generality Constraint, is then a criterion to distinguish fully conceptual–cognitive representations from early types of conceptual–perceptual representations; rather than to distinguish conceptual from nonconceptual representations. However, while visual representations at intermediate stages have properties that characterize their content as conceptual, it is still possible that visual representations at early visual stages (e.g., feature segregation, boundary representation) have non-conceptual content. At this stage, there is hardly any influence from cognitive processes, and recombination of primitive constituents that satisfy the requirement of systematicity does not seem to occur. Thus, it might be that the transition between representations with nonconceptual and conceptual content occurs already between early and intermediate visual stages.

CONCLUDING REMARKS

To claim that perception and cognition are tightly related makes sense only if one can explain how those systems, which are individuated in different brain areas and process different types of information, communicate. In this paper, I argue that visual representations share a structural property with cognitive representations; namely, that spatial recombination of visual representations into an object representation displays systematicity. This conclusion contrasts the traditional view in philosophy, according to which only sentential-cognitive representations implement a systematic structure of constituents, and it is in line with findings in physiology and psychology of how the visual system creates object representations.

The fact that visual perceptual representation, even if not sentential, displays systematicity poses a further problem for philosophical theories that see systematicity as a hallmark of representations with conceptual content. I argue that if one takes the satisfaction of this requirement as a necessary condition for having conceptual content then the content of visual representations

amounts to an early type of conceptual content that does not allow for the same kind of abstraction that is typical of human cognitive abilities. This type of early conceptual and perceptual content might be a characteristic that humans have in common with animals.

Moreover, showing that visual representations display systematicity makes it easier to see how visual perception and cognition might relate and share representational information. In fact, one of the problems of claiming that visual perception and cognition have different structure and content is that it becomes unclear how they can share information. It might be that implementing a systematic structure is a basic way of recombination that is shared by different brain areas. This might be a general code of assembling information that makes more efficient its processing in different modalities.

To conclude: my analysis adds to the debate on how perception and cognition are related. It shows that visual representations and cognitive representations display the same structural properties and might have an analogous type of content. This conclusion, based on theoretical grounds, can be tested empirically in future experiments that apply analogous manipulations to relational operations in visual perception and higher-order processes (e.g., Reverberi et al., 2011). Moreover, my ideas might lay a theoretical foundation for novel exchanges between the fields of perceptual and cognitive psychology.

REFERENCES

Barrett, H. C. (2005). Enzymatic computation and cognitive modularity. *Mind Lang.* 259–287.

Barsalou, L. W. (1999). Perceptual symbol systems. *Behav. Brain Sci.* 577–660.

Bermúdez, J. L. (2007). What is at stake in the debate on nonconceptual content? *Philos. Perspect.* 21, 55–72.

Bermúdez, J. L., and Cahen, A. (2011). "Nonconceptual mental content," in *Stanford Encyclopedia of Philosophy*, Summer 2011 Edn, ed. E. N. Zalta. Available at: http://plato.stanford.edu/entries/content-nonconceptual/

Blaser, E., Pylyshyn, Z. W., and Holcombe, A. O. (2000). Tracking an object through feature space. *Nature* 408, 196–199.

Block, N. (2005). Two neural correlates of consciousness. *Trends Cogn. Sci. (Regul. Ed.)* 9, 46–52.

Brewer, W. F., and Lambert, B. L. (2001). The theory-ladenness of observation and the theory-ladenness of the rest of the scientific process. *Philos. Sci.* 3, 176–186.

Camp, E. (2009). Putting thoughts to work: concepts, systematicity, and stimulus-independence. *Philos. Phenomenol. Res.* 78, 275–311.

Clark, A. (2004a). Feature-placing and proto-objects. *Philos. Psychol.* 17, 443–469.

Clark, A. (2004b). Sensing, objects, and awareness: reply to commentators. *Philos. Psychol.* 17, 553–579.

Cummins, R., Blackmon, J., Byrd, D., Poirier, P., Roth, M., and Schwarz, G. (2001). Systematicity and the cognition of structured domains. *J. Philos.* 98, 167–185.

Engel, A. K., Kreiter, A. K., Konig, P., and Singer, W. (1991). Synchronization of oscillatory neuronal responses between striate and extrastriate visual cortical areas of the cat. *Proc. Natl. Acad. Sci. U.S.A.* 88, 6048–6052.

Evans, G. (1982). *The Varieties of Reference.* Oxford: Oxford University Press.

Felleman, D. J., and Van Essen, D. C. (1991). Distributed hierarchical processing in the primate cerebral cortex. *Cereb. Cortex* 1, 1–47.

Fodor, J. A. (1975). *The Language of Thought.* Cambridge, MA: Harvard University Press.

Fodor, J. A. (1998). *Concepts: Where Cognitive Science Went Wrong.* Oxford: Clarendon Press.

Fodor, J. A. (2007). "The revenge of the given," in *Contemporary Debates in Philosophy of Mind*, eds B. P. Mclaughlin and J. Cohen (New York: Blackwell Publishing), 105–116.

Fodor, J. A. (2008). *LOT 2: The Language of Thought Revisited.* Oxford: Oxford University Press.

Fodor, J. A., and Lepore, E. (2001). Why compositionality won't go away: reflections on Horwich's 'deflationary' theory. *Ratio* 14, 350–368.

Fodor, J. A., and Mclaughlin, B. P. (1990). Connectionism and the problem of systematicity: why Smolensky's solution doesn't work. *Cognition* 35, 183–204.

Fodor, J. A., and Pylyshyn, Z. W. (1988). Connectionism and cognitive architecture: a critical analysis. *Cognition* 28, 3–71.

Fries, P., Reynolds, J. H., Rorie, A. E., and Desimone, R. (2001). Modulation of oscillatory neuronal synchronization by selective visual attention. *Science* 291, 1560–1563.

Goldstone, R., and Barsalou, L. W. (1998). Reuniting perception and conception. *Cognition* 65, 231–262.

Heck, R. G. (2007). "Are there different kinds of content?" in *Contemporary Debates in Philosophy of Mind*, eds B. P. Mclaughlin and J. Cohen (New York: Blackwell Publishing), 117–138.

Jackson, F. (1977). *Perception: A Representative Theory.* Cambridge: Cambridge University Press.

Kahneman, D., Treisman, A., and Gibbs, B. J. (1992). The reviewing of object files: object-specific integration of information. *Cogn. Psychol.* 24, 175–219.

Koch, C., and Ullman, S. (1985). Shifts in selective visual attention: towards the underlying neural circuitry. *Hum. Neurobiol.* 4, 219–227.

Lamme, V. A. F. (2003). Why visual attention and awareness are different. *Trends Cogn. Sci. (Regul. Ed.)* 7, 12–18.

Livingstone, M., and Hubel, D. (1988). Segregation of form, color, movement, and depth: anatomy, physiology, and perception. *Science* 240, 740–749.

Marr, D. (1982). *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information.* New York: Freeman.

Mclaughlin, B. P. (2009). Systematicity redux. *Synthese* 170, 251–274.

Newen, A., and Bartels, A. (2007). Animal minds and the possession of concepts. *Philos. Psychol.* 20, 283–308.

Peacocke, C. (2001). Does perception have a nonconceptual content? *J. Philos.* 98, 239–264.

Pecher, D., Zeelenberg, R., and Barsalou, L. W. (2004). Sensorimotor simulations underlie conceptual representations: modality-specific effect of prior act. *Psychon. Bull. Rev.* 11, 164–167.

Prinz, J. (2002). *Furnishing the Mind: Concepts and Their Perceptual Basis.* Cambridge, MA: MIT Press.

Raftopoulos, A., and Mueller, V. C. (2006). The phenomenal content of experience. *Mind Lang.* 21, 187–219.

Raftopoulos, A. (this issue). Late vision: processes and epistemic status. *Front. Psychol.*

Reverberi, C., Gorgen, K., and Haynes, J. D. (2011). Compositionality of rule representations in human prefrontal cortex. *Cereb. Cortex.* doi: 10.1093/cercor/bhr200. [Epub ahead of print].

Robertson, L. C. (2003). Binding, spatial attention, and perceptual awareness. *Nat. Rev. Neurosci.* 4, 93–102.

Roskies, A. L. (1999). The binding problem. *Neuron* 24, 111–125.

Scholl, B. J. (2009). "What have we learned about attention from multiple-object tracking (and vice versa)?" in *Computation, Cognition, and Pylyshyn*, eds D. Dedrick and L. Trick (Cambridge, MA: MIT Press), 49–78.

Smolensky, P. (1990). Tensor product variable binding and the representation of symbolic structures in connectionist networks. *Artif. Intell.* 46, 59–216.

Tacca, M. C. (2010). *Seeing Objects: The Structure of Visual Representations.* Paderborn: Mentis.

Toribio, J. (2008). State versus content: the unfair trial of perceptual nonconceptualism. *Erkenntnis* 69, 351–361.

Treisman, A. M. (1993). "The perception of features and objects," in *Attention: Selection, Awareness, and Control*, eds A. Baddeley and L. Weiskrantz (Oxford: Clarendon Press), 1–36.

Treisman, A. M. (1996). The binding problem. *Curr. Opin. Neurobiol.* 6, 171–178.

Treisman, A. M., and Gelade, G. (1980). A feature-integration theory of attention. *Cogn. Psychol.* 12, 97–136.

Treisman, A. M., and Schmidt, H. (1982). Illusory conjunctions in the perception of objects. *Cogn. Psychol.* 14, 107–141.

Wolfe, J. M. (1998). "Visual search," in *Attention*, ed. H. Pashler (London: University College London Press), 13–73.

Wolfe, J. M., and Horowitz, T. S. (2004). What attributes guide the deployment of visual attention and how do they do it? *Nat. Rev. Neurosci.* 5, 495–501.

Zeki, S. (1978). Functional specialization in the visual cortex of the rhesus monkey. *Nature* 274, 423–428.

Conflict of Interest Statement: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Late vision: processes and epistemic status

Athanassios Raftopoulos *

Department of Psychology, University of Cyprus, Nicosia, Cyprus

In this paper, I examine the processes that occur in late vision and address the problem of whether late vision should be construed as a properly speaking perceptual stage, or as a thought-like discursive stage. Specifically, I argue that late vision, its (partly) conceptual nature notwithstanding, neither is constituted by nor does it implicate what I call pure thoughts, that is, propositional structures that are formed in the cognitive areas of the brain through, and participate in, discursive reasoning and inferences. At the same time, the output of late vision, namely an explicit belief concerning the identity and category membership of an object (that is, a recognitional belief) or its features, eventually enters into discursive reasoning. Using Jackendoff's distinction between visual awareness, which characterizes perception, and visual understanding, which characterizes pure thought, I claim that the contents of late vision belong to visual awareness and not to visual understanding and that although late vision implicates beliefs, either implicit or explicit, these beliefs are hybrid visual/conceptual constructs and not pure thoughts. Distinguishing between these hybrid representations and pure thoughts and delineating the nature of the representations of late vision lays the ground for examining, among other things, the process of conceptualization that occurs in visual processing and the way concepts modulate perceptual content affecting either its representational or phenomenal character. I also do not discuss the epistemological relations between the representations of late vision and the perceptual judgments they "support" or "guide" or "render possible" or "evidence" or "entitle." However, the specification of the epistemology of late vision lays the ground for attacking that problem as well.

Keywords: late vision, visual awareness, visual understanding, conceptualization, perceptual beliefs, essential indexicals

INTRODUCTION

In earlier work (Raftopoulos, 2009), I analyzed early vision, which I claimed is a pre-attentional visual stage unaffected by top-down conceptual/cognitive modulation. (In what follows when I refer to top-down processes I mean cognitively driven processes, although there is top-down flow of signals within the visual areas. In addition, where I refer to attention I mean cognitively driven attention, unless I state otherwise.) Thus, early vision is a cognitively impenetrable stage of visual processing. I have related the content of the states of early vision with the non-conceptual content (NCC) of perception by arguing that the cognitive impenetrability of some states and contents is a necessary and sufficient condition for these states and contents to be non-conceptual. I also underlined Pylyshyn's (2003) distinction between early vision and late vision. The latter is cognitively penetrated and involves the modulation of processing by either spatial or object/feature centered attention.

In this paper, I examine the processes that occur in late vision and discuss whether late vision should be construed as a perceptual stage or as a thought-like discursive stage. I argue that late vision, its (partly) conceptual nature notwithstanding, does not consist in pure thoughts, that is, propositional structures that are formed in the cognitive areas of the brain and participate in discursive reasoning and inferences. The content of the output of late vision, that

is, an explicit belief concerning the identity of an object (recognitional belief), enters into discursive reasoning. Using Jackendoff's (1989) distinction between visual awareness, which characterizes perception, and visual understanding, which characterizes pure thought, I claim that the contents of late vision belong to visual awareness and not to visual understanding. Although late vision implicates beliefs, either implicit or explicit, these beliefs are hybrid visual/conceptual constructs and not pure thoughts. Distinguishing between these hybrid representations and pure thoughts lays the ground for examining the conceptualization of perceptual content and the way concepts modulate it affecting either its representational or its phenomenal character. I do not discuss these problems here, as I do not discuss the epistemological relations between the representations of late vision and the perceptual judgments they "support" or "evidence" or "entitle." However, the specification of the epistemic status of late vision lays the ground for attacking this problem as well.

In the first section, I sketch early vision. Then, I discuss late vision with an emphasis on its role in object recognition. The purpose is to examine some of the contents and processes of late vision and their timing. In the third section, I argue that late vision should be considered as a perceptual rather than as a discursive stage involving understanding, that is, a stage of thought processing involving pure thoughts and inferences from

propositionally structured premises to the identity of objects. My argument is based on considerations regarding the sorts of contents and processes formed in early and late vision.

EARLY VISION

Early vision includes a feed forward sweep (FFS) in which signals are transmitted bottom-up. In visual areas (from LGN to IT) FFS lasts for about 100 ms. It also includes a stage at which lateral and recurrent processes that are restricted within the visual areas and do not involve signals from cognitive centers occur. Recurrent processing starts at about 80–100 ms and culminates at about 120–150 ms. Lamme (2003) calls it local recurrent processing (LRP). The unconscious FFS extracts high-level information that could lead to categorization, and results in some initial feature detection. LRP produces further binding and segregation. The representations formed at this stage are restricted to information about spatio-temporal and surface properties, color, texture, orientation, motion, and affordances of objects, in addition to the representations of objects as bounded, solid entities that persist in space and time.

By not involving signals from the cognitive areas of the brain, FFS and LRP are cognitively impenetrable/conceptually encapsulated, since the transmitting of signals within the visual system is not affected by top-down signals produced in cognitive areas. Early vision processing is not affected directly by top-down signals from cognitive states through attention – that is, attention does not affect the early visual processes although it may affect pre-perceptual and post-perceptual stages of vision. I have argued that this leads to the thesis that early vision has NCC, provided that concepts do not figure inherently in the perceptual system, a possibility that I have rejected (Raftopoulos, 2009). The processes during early vision that result in states with personal-level NCC correspond to Dretske's (1995) phenomenal seeing[1].

LATE VISION

The conceptually[2] modulated stage of visual processing is called late vision. Starting at 150–200 ms, signals from higher executive centers including mnemonic circuits intervene and modulate perceptual processing in the visual cortex and this signals the onset of global recurrent processing (GRP). In 50 ms low spatial frequency (LSF) information reaches the IT and in 100 ms high spatial frequency (HSF) information reaches the same area (Kihara and Takeda, 2010). (LSF signals precede LSF signals. LSF information is transmitted through fast magnocellular pathways, while HSF information is transmitted through slower parvocellular pathways.) Within 130 ms post-stimulus, parietal areas in the dorsal system but also areas in the ventral pathway (IT cortex) semantically process the LSF information and determine the gist of the scene based on stored knowledge that generates predictions about

the most likely interpretation of the input, even in the absence of focal attention.

This information reenters the extrastriate visual areas and modulates (at about 150 ms) perceptual processing facilitating the analysis of HSF, for example, by specifying certain cues in the image that might facilitate target identification (Barr, 2009; Kihara and Takeda, 2010; Peyrin et al., 2010). Determining the gist may speed up the FFS of HSF by allowing faster processing of the pertinent cues, using top-down connections to preset neurons coding these cues at various levels of the visual pathway (Delmore et al., 2004). Thus, at about 150 ms, specific hypotheses regarding the identity of the object(s) in the scene are formed using HSF information in the visual brain and information from visual working memory (WM). The hypothesis is tested against the detailed iconic information stored in early visual circuits including V1. ERP's waveforms that distinguish scenes and objects in object recognition tasks are registered at about 150 ms in extrastriate areas and are thought to be early indices of P3[3] (Fabre-Thorpe et al., 2001; Johnson and Olshausen, 2005). This testing requires that top-down signals reenter the early visual areas of the brain, and mainly V1. Indeed, evidence shows that V1 is reentered by signals from higher cognitive centered mediated by the effects of object/feature centered attention at 235 ms post-stimulus (Chelazzi et al., 1993; Roelfsema et al., 1998). This leads to the recognition of the object(s) in the visual scene. This occurs, as signaled by the P3 ERP waveform, at about 300 ms in the IT cortex, whose neurons contribute to the integration of LSF and HSF information.

A detailed analysis of the form that the hypothesis testing might take is provided by Kosslyn (1994). Note that one need not subscribe to some of the assumptions presupposed by Kosslyn's account (see Raftopoulos, 2010 for criticism), but these disagreements do not undermine the framework. Suppose that one sees an object. A retinotopic image is formed in the visual buffer, which is a set of visual areas in the occipital lobe that is organized retinotopically. An attentional window selects the input from a contiguous set of points for detailed processing. This is allowed by the spatial organization of the visual buffer. The information included in the attention window is sent to the dorsal and ventral system where different features of the image are processed. The ventral system retrieves the features of the object, whereas the dorsal system retrieves information about the location, orientation, and size of the object. Eventually, the shape, the color, and the texture of the object are registered in anterior portions of the ventral pathway. This information is transmitted to the pattern activation subsystems in the IT cortex where the image is matched against representations stored there, and the compressed image representation of the object is thereby activated. This representation (which is an hypothesis regarding the identity of an object) provides imagery feedback to the visual buffer where it is matched against the input image to test the hypothesis against the fine pictorial details registered in the retinotopical areas of the visual buffer. If the match is satisfactory, the category pattern activation subsystem sends the relevant pattern code to associative

---

[1] In Raftopoulos (2009) I argue that a state with NCC does not have a propositional content, and that two states cannot have the same content and one have NCC and the other conceptual content. In this paper, I assume both theses. I also assume that part of the NCC is content at the personal level and that one has *phenomenal awareness* of that content.

[2] Concepts are constant, context independent, and freely repeatable elements that figure constitutively in propositional contents; they correspond to lexical items.

[3] The P3 waveform is elicited about 250–600 ms and is generated in many areas in the brain and is associated with cognitive processing and the subjects' reports. P3 may signify the consolidation of the representation of the object(s) in working memory.

or WM, where the object is tentatively identified with the help of information arriving at the WM through the dorsal system (information about, size, location, and orientation). Occasionally the match in the pattern activation subsystems is enough to select the appropriate representation in WM. On other occasions, the input to the ventral system does not match well a visual memory in the pattern activation subsystems. Then, a hypothesis is formed in WM. This hypothesis is tested with the help of other subsystems (including cognitive ones) that access representations of such objects and highlight their more distinctive feature. The information gathered shifts attention to a location in the image where an informative characteristic can be found. The attention window zooms on object's distinctive feature, and the pattern code for it is sent to the pattern activation subsystem and to the visual buffer where a second cycle of matching commences.

ERP experiments registering the time onset of various waveforms related to specific processes in the brain largely confirm this analysis. The N2 ERP component that signifies cognitively driven spatial–attentional effects on the extrastriate cortex is registered at about 170–200 ms. Thus, by 170 ms spatial attention directly modulates visual processing. However, cognitive top-down modulation of the extrastriate cortex, mainly V4, from the IT and parietal cortex is found as early as 150 ms, which, as we saw, is the first sign of the process of object identification.

Eventually there is considerable competition since only few items can enter in interactions with the higher hierarchically processing levels. Further selection becomes necessary when several stimuli reach the brain but only one response is possible. Attentional selection intervenes to resolve this competition. The selection results from the combination of bottom-up information processing with WM and long-term memory (LTM) that recover the meaning of input and relate it to the subject's current goals. In the biased competition account of attention (Desimone and Duncan, 1995), attention is the competition between neuronal populations that encode environmental stimuli. All the stimuli in a visual scene are initially processed in parallel and activate neuronal assemblies that represent them. These assemblies eventually engage in competitive interactions for several reasons (when, for example, some behaviorally relevant feature or object must be selected among all present stimuli).

Recurrent interactions with areas outside the visual stream make storage in visual WM possible and give rise to GRP. In GRP, standing knowledge, that is, information stored in the synaptic weights of the neurons is activated (becoming part of WM) and modulates visual processing, which up to that point was conceptually encapsulated. Consequently, during GRP the conceptualization of perceptual content starts and the states formed during this stage have (perhaps partly) conceptual and eventually propositional contents[4]. This is the stage where the 3D sketch is formed,

since the recovery of the 3D sketch, that is, the representation of an object independently of the viewer's perspective, cannot be the output of early vision. This recovery cannot be purely data-driven, since what is regarded as an object depends on the subsequent usage of the information, and thus depends on the knowledge about objects. It follows that the formation of the 3D sketch requires constitutively the application of concepts[5]. Seeing 3D sketches of objects is an instance of amodal perception, i.e., the representation of object parts or features that are not visible from the viewer's standpoint. Thus, late vision involves a synergy of perceptual bottom-up processing and top-down processing, where knowledge from past experiences guides the formation of hypotheses about the identity of objects present in the visual scene. Late vision is also responsible for the experience of the 3D sketch.

There are two sorts of completion. In modal completion the viewer has a distinct visual impression of a hidden contour or other hidden features even though these features are not occurrent sensory features. The perceptual system fills in the missing features, which thus become as phenomenally occurrent as the occurrent sensory features of the object. In amodal completion, one does not have a perceptual (imagination is not perception) impression of the object's hidden features since the perceptual system does not fill in the missing features as it happens in modal perception; the hidden features are not perceptually occurrent.

There are cases of amodal perception that are purely perceptual, that is, bottom-up. In these cases, although no direct signals from the hidden features impinge on the retina (there is no local information available), the perceptual system can extract information regarding them from the global information contained in the visual scene without any cognitive involvement, as the resistance of the ensuing percepts to beliefs indicates. However, in such cases, the hidden features are not perceived. One simply has the visual impression of a single concrete object that is partially occluded and not the visual impression of various disparate image regions. Therefore, in these perceptually driven amodal completions there is no mental imagery involved, since no top-down signals from cognitive areas are required for the completion, and since the hidden features are not phenomenologically present.

There are also cases of amodal completion that are cognitively driven (Briscoe, 2011 calls them C-completions[6]), such as the formation of the 3D sketch of an object, in which the hidden features of the object are represented through the top-down activation of the visual cortex from the cognitive centers of the brain. In some of these cases, top-down processes activate the early visual areas and fill in the missing features that become phenomenologically present. In other cases of C-completion, the viewer simply forms a pure thought concerning the hidden structure in the absence of any activation of the visual areas and, thus, in the absence of

---

[4]This means that some conceptual content in late vision may not be propositionally structured, although recognitional beliefs have propositional structure. It is also possible that some states in late vision have both NCC and conceptual content. I will not elaborate on these issues here. Note that if some of the states of late vision can have conceptual contents that are not propositionally structured, my thesis that late vision does not involve inferences is strengthened because inferences relate propositional structures.

---

[5]The view that the formation of the viewer independent representation of an object relies on object knowledge is common in theories of the formation of the 3D viewer independent representation. Biederman (1987) thinks that object recognition is based on part decomposition, which is the first stage in forming a structural description of an object. This decomposition cannot be determined by general principles reflecting the structure of the world alone, since the decomposition appears to depend upon knowledge of specific objects.

[6]Briscoe's paper analyzes Nanay's (2010) account of the role of imagination in amodal completion.

mental imagery. As the latter possibility may threaten my thesis that C-completion takes place in late vision, I discuss it in the next section.

Before I proceed, allow me to delve on "mental imagery," since the way it is used may cause some confusion concerning the top-down processes in late vision. Imagery is central in Kosslyn's (1994) account of object recognition. As we saw, Kosslyn thinks that visual imagery is involved in all cases of perception and covers all the top-down flaw of information either from the associative areas of the brain or the pattern activation subsystems in the IT cortex. Strawson (1974) also holds that object recognition involves visual imagery. Discussions on amodal completion emphasize the role of imagery in completing the hidden features by representing them and occasionally making them phenomenologically present even though they are perceptually absent (Nanay, 2010)[7]. In discussing late vision, I emphasized the role of top-down processes that are necessary for object recognition. Now, it is well known that many of the neural systems engaged in mental imagery are also actively involved in the formation of the percept, most notably the early visual areas. Since mental imagery is usually related to top-down processes, imagery could be assimilated to late vision, which involves top-down processes too. As mental imagery involves top-down activation of the visual areas, it is tempting to claim that the top-down processes in late vision are instances of visual imagery, especially so in the case of C-completion in which the object or feature that is represented through mental imagery is absent from the visual scene.

To decide the issue one should define mental imagery. Usually mental imagery is related to the mental construction of the image of an object or feature in its absence. The image formed from actual (perceptual) experience is called a percept to distinguish this image from an imagined or mental image. When a subject is asked to recall a visual object, the image formed in memory is called a mental image. The mental image is constructed via top-down processes (when, for example, subjects are presented with a lower case letter and are asked to form a mental image of the upper case letter, a task that is cognitively driven since it requires knowledge of the upper case letter), while the percept is constructed through a synergy of top-down and bottom-up processes. Thus, mental imagery is usually construed as (i) involving only top-down cognitively driven processes, and (ii) taking place in the absence of the imagined object or feature. This is how I use the term.

Kosslyn (1994) and Strawson (1974), in contrast, uses the term to designate the top-down processes in object recognition. Kosslyn talks about imagery feedback to the visual buffer both from the associative concept involving areas of the brain, and the pattern activation subsystems that Kosslyn thinks store non-conceptualized information. Therefore, mental imagery can be either cognitively driven or data-driven, which goes against the usual construal of mental imagery. Moreover, mental imagery is engaged in perceptual tasks of object recognition, which means that Kosslyn foregoes the second trait of mental imagery as

---

[7]The phenomenal/non-phenomenal distinction is orthogonal to the discussion on mental imagery since mental imagery, exactly like perception, can either be accompanied by consciousness, or it can be implicit (as in implicit perception). I wish to thank a reviewer for suggesting this.

well. Nanay (2010, pp. 244–246, 250) uses visual imagery to account for cognitively driven amodal completion, and specifically, to designate the top-down knowledge-driven effects on visual processing.

Mental imagery is perceptually and not propositionally coded, even though it may start with the activation of concepts in associative memory (Kosslyn, 1994). However, the activation of the visual areas in a top-down manner in mental imagery is not the same as the activation of these same areas by sensory signal. For example, the top-down induced activation in the absence of retinal input is weaker and, thus, the modal "mode" associated with mental imagery is not as strong or lively as in perception. Although it is true that when an object is imagined as opposed to merely thought about a number of properties must be added to the description, these properties fall far short of all those that would be present in perception. Not only some features may be omitted, but also precise iconic and metric information is lost in mental imagery. Since the concepts that activate the visual cortex represent abstract categorical information, such as bright, red, and not the determinate color say red21 (which is why one cannot recall the determinate color of an object but only its category membership), not all visual details of the actual visual scene can be the contents of a state of visual imagery (Raftopoulos, 2010). In late vision, on the other hand, the presence of the visual object allows conceptual demonstratives to rely on the presence of the sample and overcome any conceptual limitations.

Since late vision constitutively involves a synergy of bottom-up and top-down processing, whereas mental imagery, as I construe it, involves only top-down flow of information to early visual areas in the absence of sensory stimulation, I prefer (pace Kosslyn and Nanay) not to use "imagery" to designate the top-down activation of the visual cortex in late vision, even in those cases in which top-down processing completes hidden features of objects. Mental imagery differs from seeing in that it uses only the late processing components of the perceptual system when the early processing sensory-driven processes are unavailable (as when there is no sensory stimulation). Visual imagery activates the (inactive) visual processing areas to recreate to a certain extent a visual scene. As such, mental imagery, unlike late vision, involves only top-down processes. Although in both cases the early visual areas are reentered from signals emanating from cognitive centers, in late vision the cognitive centers are activated through bottom-up signals from the visual cortex, while in visual imagery the cognitive centers are activated in the absence of any sensory stimulation on the retina. Thus, I think that the top-down processes in late vision should be distinguished from mental imagery in that the former are essentially engaged by the existence of sensory stimuli on the retina, whereas in the latter there are no sensory stimuli.

IS LATE VISION A VISUAL STAGE OR A DISCURSIVE THOUGHT-LIKE STAGE?

THE PROBLEM

Jackendoff (1989) distinguishes visual awareness from visual understanding. There is a qualitative difference between the experience of a 3D sketch and the experience of a 21/2D sketch. One is aware of the 3D sketch or of category based representations,

however, this is not visual awareness but some other kind of awareness. Visual awareness is awareness of Marr's 21/2D sketch, which is the viewer-centered representation of the visible surfaces of objects, while the awareness of the 3D sketch is visual understanding. Thus, the 3D sketch, which includes the unseen surfaces that are not represented in the 21/2D sketch, is a result of an inference; amodal completion is an inference. Jackendoff's views belong to the so-called belief-based account of amodal completion: the 3D sketch is the result of beliefs inferred from the object's visible features and other background information from past experiences.

The problem is whether object identification and C-completion that occur in late vision and are both dependent on concepts should be thought of as cases of vision or as cases of discursive understanding involving inferences. If late vision involves conceptual contents and if the role of concepts and stored knowledge consists, among other things, in providing some initial interpretation of the visual scene and in forming hypotheses about the identity of objects that are tested against perceptual information, one is tempted to say that this stage relies on inferences (this is what hypothesis testing amounts to) and, thus, differs in essence from the purely perceptual processes of early vision. Perhaps it would be better to construe late vision as a discursive stage involving thoughts, in the way of Jackson's (1977) epistemic seeing, where "seeing" is used in a metaphorical non-perceptual sense, as where one says of his friend whom she visited "I see he has left," based on perceptual evidence. It is, also possible that Dretske (1993, 1995) thinks that seeing in the doxastic sense is not a visual but, rather, a discursive stage.

One might object that abandoning this usage of "to see" violates ordinary usage. A fundamental ingredient of visual experience consists in meaningful 3D solid objects. Adopting this proposal would mean that one should resist talking of seeing tigers and start talking about seeing viewer-centered visible surfaces. "By this criterion, much of the information we normally take to be visually conscious would not be, including the 3D shape of objects as well as their categorical identity" (Palmer, 1999, p. 649).

The arguments to common language notwithstanding, I think that one should not assume either that late vision is an inferential discursive stage that constitutively involves thoughts in the capacity of premises in inferences whose conclusion is the content of the states of late vision (although implicit hypotheses play a role), or that late vision consists in discursively entertaining thoughts. The reason is twofold. First, I think that seeing an object is not the result of an inference, that is, a movement in thought from some premises to a conclusion and, thus, a discursive process, even though it involves concepts. Second, late vision is a stage in which conceptual modulation and perceptual processes form an inextricable link that differentiates late vision from discursive stages and renders it a different sort of a set of processes than understanding, even though late vision involves implicit beliefs regarding objects that guide the formation of hypotheses concerning object identity, and an explicit belief of the form "that O is F" eventually arises in the final stages of late vision. Late vision has an irreducible visual ingredient, which makes it different from discursive understanding. Before I discuss all these, let me clarify some terminological issues.

Beliefs

Traditionally judgments are occurrent states, whereas beliefs are dispositional states. To judge that O is F is to predicate Fness to O, while endorsing the predication (McDowell, 1994). To believe that O is F is to be disposed to judge, under the right circumstances, that O is F. This is the first sense in which beliefs are dispositional items. Now, as the reader recalls, I have distinguished between standing knowledge – information stored in LTM – and information that is activated in WM. The belief that O is F may be a standing information in LTM, a memory, because, say, one has seen O to be F in the past, even though presently one does not have an occurrent thought about O. Beliefs need not be consciously or unconsciously recalled or apprehended in order to be possessed by a subject, which means that beliefs are dispositional rather than occurrent items; this is a second sense in which beliefs are dispositional. When this information is activated, the occurrent thought that O is F emerges in WM. In the literature one finds the distinction between "thought" and "standing knowledge" (Prinz, 2002, p. 148). Accordingly, all thoughts are occurrent states by being activated in WM. Thus, I use "occurrent thought" and "thought" as synonymous.

It follows that a belief *qua* dispositional state may be either a piece of standing knowledge, in which case it is dispositional in the sense that when activated it becomes a thought, or a thought that awaits endorsement to become a judgment, in which case the belief is dispositional in the sense that it has the capacity to become a judgment. In the first case, if beliefs are stored in LTM as standing knowledge and if thoughts are occurrent states, beliefs are not the same as thoughts although a belief when activated becomes a thought. In the second case, a belief is a thought held in WM, albeit one that has not been yet endorsed. There are interesting epistemological implications but they are irrelevant here. In what follows, I assume that beliefs are either thoughts or pieces of standing information, which have not been endorsed and, thus, are not judgments. One might wonder how is it possible to understand a belief as an occurrent thought that is not endorsed? An explanation has to wait until I have explained why late vision does not involve inferences.

State consciousness

It is important for the discussion that follows to clarify another problem, namely, under which conditions are beliefs conscious or not. An intuitive answer is that, as a matter of course, one may entertain beliefs or judgments and use them for various purposes (for example to draw conclusions in inferences or guide actions) even though one is not conscious that one entertains these beliefs or judgments (as in the case of using implicit premises in an argument); these beliefs are implicit. Underneath this intuitive view one discerns the assumption that a state is conscious if the person who has it is conscious that she is in that state. Either that person has a second order thought that she is entertaining such a belief – that is, she has fact-awareness that she is entertaining that state – or she has a second order experience or inner sense that she is in such a state – that is, she has thing-awareness of the state – where "thing-awareness" and "fact-awareness" are used in the way Dretske (1993) defines them. If one subscribes to this view, what makes a mental state of a person conscious is the person's

awareness of the state. However, Dretske (1993) argues that what renders a person's state conscious is not some sort of second order awareness that one is in such and such state, or that she is having that state. A state is conscious

> being a certain sort of representation, it makes one aware of the properties (of x) and objects (x itself) of which it is a sensory representation...[A] certain belief is conscious, not because the believer is conscious of it (or conscious of having it), but because it is a representation that makes one conscious of the fact (that P) that it is a belief about...beliefs are conscious, not because you are conscious of them, but because, so to speak, you are conscious *with* them (Dretske, 1993, pp. 437–438).

Beliefs that are thought of as implicit but play a cognitive role in making a person aware of some facts or things are conscious (a first-order consciousness). Dretske does not claim that everything that happens to one when one becomes conscious of some object or event is conscious. However, a perceptual experience or a belief has to be conscious in order for a person to be made aware of things and events. I do not assess Dretske's thesis, which is only among many views on consciousness (some of which are higher-order theories that Dretske resists), and I remain neutral as to how conscious state should be construed. By "implicit belief" I mean the belief held by a person who is not aware that she is having that belief.

Inference

My claim is that the processes in late vision are not inferential processes where "inference" is understood as discursive, that is, as a process that involves drawing propositions–conclusions from other propositions acting as premises by applying (explicitly or implicitly) inferential rules that are also represented. These inferences are distinguished from "inferences" as understood by vision scientists according to whom any transformation of signals carrying information according to some rule is a form of inference. "Every system that makes an estimate about unobserved variables based on observed variables performs inference... We refer to such inference problems that involve choosing between distinct and mutually exclusive causal structures as causal inference" (Shams and Beierholm, 2010).

LATE VISION, HYPOTHESIS TESTING, AND INFERENCE

I think that the states of late vision are not inferences from premises that include the contents of early vision states, even though it is usual to find claims that one infers that a tiger, for example, is present from the perceptual information retrieved from a visual scene. An inference relates some propositions in the form of premises with some other proposition, the conclusion. However, the objects and properties as they are represented in early vision do not constitute contents in the form of propositions, since they are part of the non-propositional NCC of perception. In late vision, the perceptual content is conceptualized but the conceptualization is not a kind of inference but rather the application of stored concepts to some input that enters the cognitive centers of the brain and activates concepts by matching their content. Thus, even though the states in late vision are

formed through the synergy of bottom-up visual information and top-down conceptual influences, they are not inferences from perceptual content.

Late vision involves hypotheses regarding the identity of objects and their testing against the sensory information stored in iconic memory. One might think that inferences are involved since testing hypotheses is an inferential process even though it is not an inference from perceptual content to a recognitional thought. It is, rather, an argument of the form if A and B then (conclusion) C, where A and B are background assumptions and the hypothesis regarding the identity of an object respectively, and C is the set of visual features that the object is likely to have. A consists of implicit beliefs about the features of the hypothesized visual object. If C is what obtains in the visual areas, that is, if the predicted visual features match those that are stored in iconic memory then the hypothesis about the identity of the object is likely correct. However, the test basis or evidence against which these hypotheses are tested for a match, that is, the iconic information stored in the sensory visual areas, is not a set of propositions but patterns of neuronal activations whose content is non-propositional.

There is nothing inference-like in this matching. It is just a comparison between the activations of neuronal assemblies that encode the visual features in the scene and the activations of the neuronal assemblies that are activated top-down from the hypotheses. If the same assemblies are activated then there is a match. If they are not, the hypothesis fails to pass the test. This can be done through purely associational processes of the sort employed, say, in connectionist networks that process information according to rules and, thus, can be thought of as instantiating processing rules, without either representing these rules or operating on language-like symbolic representations. Since inferences are carried out through rules that are represented in the system, and operate on symbolic structures, the processing in a connectionist network does not involve inferences, although it can be described in terms of inference making. Thus, even though seeing an object in late vision involves the application of concepts that unify the appearances of the object and of its features under some category, it is not an inferential process. The processes in late vision despite their reliance on background beliefs do not entail by themselves a recognitional belief.

Spelke (1988, p. 458)[8] argues that "perceiving objects may be more akin to thinking about the physical world than to sensing the immediate environment." The reason is that the perceptual system, to solve the underdetermination problem of both the distal object from the retinal image and of the percept from the retinal image, employs a set of object principles and that reflect the geometry and the physics of our environment. Since the contents of these principles consist of concepts, and thus, the principles can be thought of as some form of knowledge about the world, perception engages in discursive, inferential processes. Against this, I have argued (Raftopoulos, 2009) that for a variety of reasons the processes that constrain the operations of the visual system should not be construed as inferences. Rather, they constitute the

---

[8]Spelke echoes Rock's (1983) views that the perceptual system combines inferentially information to form the percept. For example, from visual angle and distance information, one infers and perceives size.

modus operandi of the perceptual system, they are hardwired in the perceptual circuits, and they are not represented anywhere.

Being hardwired is another reason why perceptual processes should not be assimilated to inference making. Inferences presuppose that the subject applies explicitly or implicitly inferential rules that are represented in the subject. But the operations by means of which signals are transformed from one into the other in the visual system are not represented at all; they are just hardwired in the perceptual system. For this reason, perceptual operations should not be construed as inference rules, although they are describable in terms of inference rules.

LATE VISION AND DISCURSIVE UNDERSTANDING

Even if I am right that seeing in late vision is not the result of a discursive inference, it is still arguable that late vision should be better construed as a stage of discursive understanding rather than as a visual stage. If object recognition involves forming a belief about class-membership, even if the belief is not the result of an inference, why not say that recognizing an object is an experience-based belief that is a case of understanding rather than vision.

Late vision is more than object recognition

A first problem with this view is that late vision involves more than a recognitional belief. Suppose that S sees an animal and recognizes it as a tiger. In the parallel preattentive early vision, the proto-object that corresponds to the tiger is being represented amongst the other objects in the scene. The relevant activations enter the parietal and temporal lobes, and the prefrontal cortex, where the neuronal assemblies encoding the information about tigers are activated and this activation spreads through top-down signals to the visual areas of the brain where visual sensory memory stores the proto-objects extracted from the visual scene. The cells encoding the proto-object corresponding to the animal and its properties have their activations strengthened and win the competition against the assemblies encoding the proto-objects corresponding to the other objects in the scene. After a proto-object has been selected, the object recognition system forms hypotheses regarding the identity of the object. However, for the subject's confidence to reach the threshold that will allow her to form beliefs about the identity of the object and report it, these hypotheses must be tested (Treisman, 2006).

To test these hypotheses the visual system allocates resources to features and regions that would confirm or disconfirm the hypotheses. Conceptual information about a tiger affects visual processing and after some hypothesis testing the animal is recognized as a tiger through the synergy of visual circuits and WM. At this point the explicit belief "O is F" is formed. This occurs after 300 ms, when the viewer consolidates the object in WM and identifies it with enough confidence to report it, which means that beliefs are formed at the final phases of late vision. However, semantic modulation of visual processing and the process of conceptualization that eventually leads to object recognition starts at about 130–200 ms. There is, thus, a time gap, between the onset of conceptualization and the recognition of an object, which is a prerequisite for the formation of an explicit recognitional belief.

As Treisman and Kanwisher (1998) observe, although the formation of hypotheses regarding the categorization of objects can occur within 130–200 ms after stimulus onset (the time depends on the saliency of the object), it takes another 100 ms for subsequent processes to bring this information into awareness so that the perceiver could be aware of the presence of an object and be able to report it. To form the recognitional belief that "O is F," one must be aware of the presence of an object token and construct first a coherent representation. This requires the enhancement through attentional modulation of the visual responses in early visual circuits that encode rich sensory information in order to integrate them into a coherent representation, which is why beliefs are delayed in time compared with the onset of conceptualization. It follows that not all of late vision involves explicit beliefs.

Late vision and consciousness

The beliefs involved in late vision in the form of hypotheses are not in the stream of the perceiver's consciousness; they are not explicit. The processing in late vision is done automatically and is outside both of the cognitive control of the viewer and of her awareness. Matching an input to a stored template is not under anyone's cognitive control and is not a process of which one is aware; neither is the determination of the gist of a visual scene. The conceptualization of the content of perception is not under anyone's control. Furthermore, for a thought to be conscious the person who has it must have access awareness to the contents of the thought; the perceiver reports, as it were, the content of her thoughts to herself. Thus, she must have some kind of a higher-order thought about the contents of her thought. Such a higher-order thought is not required in order to be able to recognize objects. Report awareness occurs in 500 ms, when the object has been categorized. This marks a difference between late vision and thought. Most of the contents and their transformations that occur during late vision cannot be in the realm of awareness, although the outcome of late vision is. Propositional inferences, by contrast, can be available to awareness.

Late vision as a synergy of bottom-up and top-down information processing

A third reason why the beliefs formed in late vision are partly visual constructs and not pure thoughts is that the late stage of late vision in which explicit beliefs concerning object identity are formed constitutively involves visual circuits (that is, brain areas from LGN to IT in the ventral system). Pure thought involves primarily an amodal form of representation formed in higher centers of the brain, even though these amodal representations can trigger in a top-down manner the formation of mental images and can be triggered by sensory stimulation. The point is that amodal representations can be activated without a concomitant activation of the visual cortex (see Prinz, 2002 notion of default concepts that are amodal representations). Perceptual representations in late vision, in contrast, are modal since they constitutively involve visual areas. Thus, what distinguishes late vision beliefs and pure thoughts is not so much their modal or amodal character (pure thoughts can also be accompanied by some sort of phenomenology), as the fact that the beliefs in late vision are formed through a synergy of bottom-up and top-down activation and their maintenance requires the active participation of the visual circuits. Pure

thoughts can be activated and maintained in the absence of any activation in visual circuits.

The constitutive reliance of late vision on the visual circuits suggests that late vision relies on the presence of the object of perception; it cannot cease to function as a perceptual demonstrative that refers to the object of perception, as this has been individuated though the processes of early vision (Raftopoulos and Muller, 2006; Burge, 2010, p. 542). As such, late vision is constitutively context dependent since the demonstration of the perceptual particular is always context dependent. Thought, on the other hand, by its use of context independent symbols, is free of the particular perceptual context. Even though both recognitional beliefs in late vision and pure perceptual beliefs involve concepts (pure attributive elements Burge, 2010), the concepts function differently in the two contexts. As Burge (2010, p. 545) claims "perceptual belief makes use of the singular and attributive elements in perception. In perceptual belief, pure attribution is separated from, and supplements, attributive guidance of contextually purported reference to particulars... Correct conceptualization of a perceptual attributive involves taking over the perceptual attributive's range of applicability and making use of its (perceptual) mode of presentation."

Note that the attributive and singular elements in perception correspond to the perceived objects and their properties and not to concepts concerning these objects and properties. The attributive elements (properties in perception) guide the contextual reference to particulars (the objects of perception) since the referent in a demonstrative perceptual reference is fixed through the properties of the referent as these properties are presented in perception – what I have called the non-conceptual mode of presentation of the object in perception (Raftopoulos and Muller, 2006). As such, they belong to the NCC content of perception (Burge, 2010, p. 538) Concepts enter the game in their capacity as pure attributions that make use of the perceptual mode of presentation. Burge's claim that in perceptual beliefs pure attributions supplement attributions that are used for contextual reference to particulars may be read to mean that perceptual beliefs are hybrid states involving both visual elements (the contextual attributions used for determining reference to objects and their properties) and conceptualizations of these perceptual attributives in the form of pure attributions. In this case, the role of perceptual attributives is ineliminable and, thus, Burge's perceptual beliefs map onto my recognitional beliefs of late vision. In late vision, unlike in pure beliefs, there can be no case of pure attribution, that is, of attribution of features in the absence of perceptually relevant particulars since the attributions are used to single out these particulars.

The concepts that figure in perceptual beliefs in late vision need not correspond to perceptual attributives, that is, they need not be restricted to concepts that late vision employs when it takes over the mode of presentation of the perceptual content. Visual systems have perceptual attributives for features such as shape, size, spatial relations, color, motion, orientation, texture, and affordances (Pylyshyn, 2003; Raftopoulos, 2009; Burge, 2010, p. 546), which are matched (partly, because one does not have concepts for all perceptual attributives) by the salient concepts. However, they do not have perceptual attributives for tigers, yet one does have perceptually based beliefs about tigers. They are perceptual in that

even though they do not conceptualize perceptual content and do not take over the mode of presentation of perceptions (category membership does not have a perceptual a mode of presentation), they depend for their empirical applications on perceptual attributives (the concept "tiger" depends for its application on perceptual attributives such as size, shape, and color).

I said that visual systems do not have perceptual attributives for category membership, which means that these higher-order properties cannot be visually represented; one does not perceive, say, tigerness, as Bayne (2009) and Siegel (2006) argue. Let me explain this. The fact that late vision outputs recognitional beliefs that are not pure beliefs does not entail that one has visual awareness of the high-level properties that figure in the recognitional beliefs. The IT cortex (which is the highest visual area) may represent objects in 3D, their 2D projections, viewer-centered representations, viewer independent representations, whole objects, and parts of objects, but not category membership. One has cognitive access awareness (CAA) of higher-level properties. (CAA is about perceptual content that is accessed by cognition becoming available to introspection and refers to episodes of thinking about the contents of one's perceptual experience.) These beliefs are inextricably linked to a perceptual context but this does not entail that there is a visual phenomenology of category membership. It means, however, that the belief modulates top-down the processing in the visual areas of the brain and enhances the activation of the visible features that knowledge of the category membership highlights. Thus, having recognized an object affects the perception of some of its visible features by changing their representation and phenomenology, but one does not have visual awareness of high-level features of objects.

The inextricable link between thought and perception in late vision explains the essentially contextual, in Perry's (2001) and Stalnaker's (2008, pp. 78–82) sense, character of beliefs in late vision. The proposition expressed by the belief cannot be detached from the perceptual context in which it is believed, and cannot be reduced to some other belief in which some third person or objective content is substituted for the indexicals that figure in the thought (in the way one can substitute via Kaplan's characters the indexical terms with their referents and get the "objective" truth-evaluable content of the belief). The reason is that the belief is tied to a idiosyncratic viewpoint by making use of the viewer's physical presence and occupation of a certain location in space and time; the context in which an essentially indexical thought is believed is essential to the information conveyed. There are not, to paraphrase Stalnaker (2008, pp. 86–87), some relevant objective facts that the person (S1) who entertains the objective thought that purports to express the essentially indexical content has to learn in order to entertain the same content as S2 who uses the essentially contextual thought. This means that the way the world is thought by S2 is different form the way the world is thought by S1 not because there are some different facts the two thoughts are about, but because S1's and S2's perspectives on the same facts are different.

Perception individuates objects in a visual scene by assigning object-files based primarily on spatio-temporal information. The perception itself has the demonstrative reference force of "that object" and, thus, perceptual objects are determined relationally (Burge, 2010). For an object to be an object of a perceptual state

it must stand in a certain kind of relation to that state. Being acquainted perceptually with an object means that one is in direct contact with the object itself and retrieves information from it and not through a description (Burge, 1977). Perception puts one in a *de re* relationship with the object (as opposed to a *descriptivist* relationship). The content of an object-file is the *de re* mode of presentation of the object in perception (Raftopoulos and Muller, 2006).

Since recognitional beliefs rely on the presence of the object (reference to the object is fixed through a demonstrative as in "That x is F"), they are *de re* beliefs. Pure perceptual beliefs, on the other hand, have their referents fixed through a description of the object in memory. The *de re* relationship to a visual object eventually results in the formation of a *de re* belief about it. The outcome of late vision is a *de re* belief tied into a perceptual context. In contradistinction, pure thoughts and the pure attributions they render possible can be used outside any perceptual context and they descriptivist beliefs[9].

It is sometimes argued that the main difference between thoughts and perceptions is that perceptual experiences, unlike thoughts, have a sensory quality to them (Dretske, 1993, p. 436). Although the amodal character of cognitive states as opposed to the modality-specific character of perceptions is a good place to start, this should be qualified because thoughts are not in a sense necessarily purely amodal since they may be accompanied by experiences that have a phenomenal character. The thought "the orange is round and yellow" has a modality-specific content, in that when one holds this thought, visual areas of one's brain encoding color and shape are also be activated (Prinz, 2002). However, things are complicated. First, this activation does not entail that there is visual awareness of these features. Second there is a large literature on this issue with conflicting results. I am raising this issue to urge the reader not to take in views like Dretske's uncritically.

BELIEFS: TAKE TWO

If the recognitional beliefs formed in late vision are not endorsed to become judgments, they are in some sense hypotheses. Suppose that upon viewing a scene containing an object O, S comes to believe that O is F. Since things may not be as they seem, S refrains from judging that O is F; S does not endorse the content of her perceptual belief. How is this recognitional belief different from the hypotheses or implicit beliefs that are constructed during the earlier stages of late vision in order to establish the identity of the object beyond the fact that the one is explicit, while the other is implicit?

In my view, the main difference consists in that the early hypotheses are tested against the iconic information stored in visual areas. This is an unconscious process that is outside the control of the viewer who is usually aware only of the content

of the winner, that is, the content of the explicit recognitional belief. However, the recognitional belief of late vision must be tested against a different sort of evidence in order to become a judgment. It must be tested against other sorts of propositional structures, that is, pure beliefs in which the predicate terms function as pure attributions. The aim of the testing is to put aside various possible defeaters of the belief. For example, the viewer has to decide whether she is the victim of some hallucination, etc. The processes involved in this testing may be available to the viewer's consciousness, they are usually under her control, and they have the form of inferences from propositional contents to propositional contents, unlike the processes in late vision. The viewer tries to determine whether she should take the content of her late vision at face value. This is why testing the recognitional belief against other pure beliefs is a discursive process that is within the space of reasons, whereas testing the implicit hypotheses to come up with a recognitional belief belongs to late vision. In this sense, the recognitional beliefs formed in late vision are at the interface between the space of reasons and the perceptual space and, thus, have a pivotal role to play in accounts of justification of perceptual judgments.

I can explain now my claim that a belief is a dispositional state as opposed to a judgment that is an occurrent state. I tried to express the thought that perception gives us a *prima facie* inclination to believe that O is F but other evidence may override this and preclude us from forming the judgment that O is F. For example, some illusions give us a *prima facie* reason to believe that O is F but we do not endorse this because we do not believe that O is F. Undoubtedly, when O appears F in one's experience, one is inclined to form (this is what I mean by "*prima facie*") the recognitional belief that O is F. However, one need not endorse that thought. That O appears F in one's experience should not be equated with one endorsing that O is F. To do that, one has to consider other relevant beliefs. Thus, to transform the belief to a judgment, one has to integrate it in the nexus of other beliefs, putting it, thus, within the space of reasons. This is possible because the recognitional belief already has a propositional structure.

There are two notions of belief here. The one is related to the expression of the content of a conceptual perceptual state, the recognitional belief, and the other is constitutively related to the notion of judgment. The relation of the belief in the first sense to late vision contents is not inferential. The relation of the same recognitional belief with the nexus of other beliefs is an inferential relation; if endorsed, the belief becomes a judgment. The belief is, thus, a disposition to make judgments (McDowell, 1994, p. 60), which do not introduce some new content but simply endorse the content of the recognitional belief.

Johnston (2006, p. 282) argues that the judgments that perception outputs are not inferentially based on perceptual content. "My judgment does not go beyond its truthmaker, which sensory experience has made manifest. Its truth is thus guaranteed by its origins. This is how immediate perceptual judgments often have the status of knowledge. There is no evidence from which they are inferred; instead they are reliable formed out of awareness of their truth maker, often in the absence of any evidence to the contrary." Johnston talks about immediate perceptual judgments, whereas I talk about recognitional beliefs that may or may not

---

[9]In a *de re* belief, one retrieves information from the object itself and not through a description. In late vision where information in WM guides the formation of hypotheses about object identity, these hypotheses are based on descriptions in addition to visual information, since the knowledge stored in memory is a description of the object. Thus, the ensuing recognitional belief is based on a combination of information deriving from the object and from a description of it in memory. It is not a pure *de re* belief.

become judgments. Johnston's view that perceptual judgments are not inferred from perceptual evidence is correct. Our difference stems from considerations pertaining to the sentence "often in the absence of evidence to the contrary." I have claimed that to examine possible evidence against a recognitional belief, the belief must be inferentially tested against other pure beliefs (perceptual or otherwise). Only when it passes the test it becomes a judgment. Thus, I qualify Johnston's view that perceptual judgments are not inferred from any evidence, by distinguishing between perceptual beliefs and perceptual judgments and by adding that the former are not inferred from any evidence as outputs of late vision, but to become judgments they have to enter into inferential relations with possible defeaters.

I do not claim that recognitional beliefs are always tested this way to become judgments. Under normal conditions they are not tested at all. One might argue, however, that the absence of testing means that the viewer thinks that there is no reason to doubt the recognitional belief, which in itself is a sort of implicit inference. Or, one might think that in these normal cases, the recognitional belief becomes automatically a judgment without any inferential involvement. Still, the distinction holds because on certain occasions the recognitional belief is inferentially tested against other beliefs in order to become a judgment and, thus, recognitional beliefs and perceptual judgments belong to different categories, the first being a state that has the potential to become a judgment, even if the potentiality is actualized on certain occasions automatically.

LATE VISION AND AMODAL COMPLETION

Nanay (2010) thinks that mental imagery is necessary to account for amodal completion. He also (Nanay, 2010, p. 252) thinks that amodal completion in some cases is accompanied by some sort of phenomenology subserved by the activation of the early visual areas. In this sense, the hidden parts and features of an object are not merely believed in but are present in the object of perception as actualities by being imagined. Moreover, even in cases of amodal completion that are not accompanied by some sort of phenomenology, the hidden parts or features are perceptually represented. This is good point to delineate further the distinction between visual awareness and visual understanding and why late vision is a case of visual awareness. Briscoe (2011, pp. 165–167), argues that although imagery is sufficient for amodal completion, it is not necessary since one could either C-complete a visual scene by forming beliefs about the hidden parts of an object based on its visible features without projecting a mental image (the belief-based account of C-completion), or one could amodally complete a scene in bottom-up perceptual ways, in the way explained in the third section[10].

Briscoe (2011) remarks that there are cases of C-completion, for example, the 3D sketch of an object whose backsides are hidden from view, which are cognitively driven in that to complete the hidden parts the viewer must draw from object knowledge. This may produce activation of the visual cortex, such that one has a mental image of the hidden parts, or it may produce simply

a thought that there are some parts hidden from view without any mental images, or it may produce both (Briscoe, 2011, p. 158). If the visual cortex is involved in C-completing the picture there is a synergy of bottom-up and top-down processes. 3D completion occurs in late vision where certain visual processing areas are activated.

If C-completion involves a pure perceptual thought about the hidden parts that results from an inference based on past experience and the current visual evidence, this is a case of visual understanding and not of visual awareness. I do not think that this possibility undermines my thesis that seeing the 3D sketch takes place in late vision. First, it is not clear whether there is empirical evidence for C-completion through pure thought and in the absence of any activation in visual areas. Second, if there are such cases, this only shows that sometimes C-completion does not occur in late vision but in discursive reasoning. Third, Briscoe's example from which he argues that C-completion may involve a pure thought involves a picture of the backside of what looks like a horse. In this case C-completion takes the form of a pure thought that this is a horse without any visual awareness. This is clearly a case of an inference involving visual understanding that occurs in the space of reasons and not in late vision. My claim is, on the other hand, that seeing the 3D sketch is a case of C-completion that takes place in late vision and involves visual awareness. Thus, even if there are cases of C-completions through pure thoughts, there are sorts of C-completions, such as seeing the 3D sketch, that take place in late vision and are cases of visual awareness.

Consider the white surface of a wall seen in a shadow and perceived as gray. Even though the viewer knows that the gray shade is caused by the shadow cast on a white wall, the phenomenal character of her experience is that of gray. The phenomenal character of her experience of the situation dependent color property (Schellenberg, 2008) or the phenomenal property (Shoemaker, 2006) is gray not white. Of course, being aware of the shadow she could infer the intrinsic (Schellenberg, 2008) or objective (Shoemaker, 2006) color of the wall but this is an inference based on the visible grayness, knowledge of the effects of shadows on surfaces, etc. In this case, one does not perceive the whiteness in any sense of "seeing" and, thus, the output of late vision is not the belief that the color of the wall is white. That the wall is experienced in late vision as gray is a case of visual awareness, where the concomitant belief takes over the mode of presentation of the object of experience. One may form the judgment that the wall is white even though it looks gray, but this representation is in the realm of pure thought. It is a case of visual understanding, a process in which one draws a conclusion based on the evidence of one's senses and other relevant information.

Suppose now that one sees one's hand moving back and forth. One sees the hand having the same size, a case of size constancy. If the constancy is due to cues that are available in the retinal image, the viewer is phenomenally aware of the same size despite differences in the viewing conditions. If size constancy is not effectuated through visual information and cognitive sources are needed, it is achieved in late vision; the viewer believes that the size is constant and has the phenomenal experience of a constant size. Should visual information be insufficient for perceptual constancies and should the non-visual information that ensures constancy be not

---

[10]Note that Nanay (2010, p. 244) seems to talk about a perceptually driven amodal completion that is insensitive to other beliefs.

available (as where attention is diverted elsewhere), the viewer would be aware of changes in size. This is what Epstein and Broota (1986) show by demonstrating that when attention is directed elsewhere, the size constancy operations fail. Thus, the experience of a stable size is the product of late vision, created by the knowledge of the size and stability of our hand in synergy with visual information coming from the hand. There is a large amount of literature supporting the view that many a perceptual constancy rely on object knowledge (Granrud, 2004; Cohen, 2008; Hatfield, 2009). Despite the role of thoughts in late vision, these cases should be better construed as visual awareness and not as visual understanding because, first, the states of late vision do not consist in pure thoughts but in hybrid states and, second, because the processes that lead to perceptual constancies are not discursive inferences.

To recapitulate, in pure thought the beliefs formed result from discursive processes (which may include perceptual information cast in a propositional form) and their attributives are context free, while in late vision there are no discursive processes but only conceptually modulated visual processing and the relevant attributives are context bound. These differences result from the constitutive involvement in late vision of visual circuits, an involvement that is absent in pure thought. This view entails that in amodal completion, which is one of the processes that take place in late vision, the missing or occluded features are nor represented by pure perceptual beliefs, a view that is also supported by (partially) independent considerations offered by Nanay (2010, pp. 243–246).

CONCLUDING COMMENTS

Some philosophers consider that there is a sharp distinction between vision and thought and attempt to explain various phenomena (such as modal and amodal completion, or cognitive effects on perception) either (exclusive "either…or") as perceptual or thought-based. McPherson (in press) considers evidence for the effects of knowledge of the colors of objects on the perception of these colors and after having rejected a thought-based explanation of these effects goes on to argue that knowledge affects perception itself through the processes of mental imagery and, consequently that perception is cognitively penetrable. The main reason that at a last analysis drives McPherson to conclude that color perception is cognitively penetrable is that cognition affects the phenomenology of the way colors look and this cannot be explained by a belief-based account but only by admitting that it is the perceptual stage itself that is cognitively effected. However, if one allows for the possibility of a stage of visual processing in which visual processing and cognitive effects coexist and, consequently, allows for a stage of visual processing that is cognitively penetrated and has its own phenomenology, one can explain the cognitive effects on visual phenomenology without drawing the conclusion that all visual processes are cognitively penetrable, since early vision may still be cognitively impenetrable. There is a hybrid stage of vision/thought in which perception and cognition are intermingled. This is the cognitively penetrated stage of late vision. Since late vision does not involve pure thoughts, the belief-based accounts are wrong but that does not entail that early vision is cognitively penetrable.

REFERENCES

Barr, M. (2009). The proactive brain: memory for predictions. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 364, 1235–1243.

Bayne, T. (2009). Perception and the reach of phenomenal content. *Philos. Q.* 39, 385–405.

Biederman, I. (1987). Recognition by components: a theory of human image understanding. *Psychol. Rev.* 94, 115–147.

Briscoe, R. E. (2011). Mental imagery and the varieties of amodal perception. *Pac. Philos. Q.* 92, 153–173.

Burge, T. (1977). Belief de re. *J. Philos.* 74, 338–362.

Burge, T. (2010). *Origins of Objectivity*. Oxford: Clarendon Press.

Chelazzi, L., Miller, E., Duncan, J., and Desimone, R. (1993). A neural basis for visual search in inferior temporal cortex. *Nature* 363, 345–347.

Cohen, J. (2008). Colour constancy as counterfactual. *Australas. J. Philos.* 86, 61–92.

Delmore, A., Rousselet, G. A., Mace, M. J.-M., and Fabre-Thorpe, M. (2004). Interaction of top-down and bottom up processing in the fast visual analysis of natural scenes. *Brain Res. Cogn. Brain Res.* 19, 103–113.

Desimone, R., and Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annu. Rev. Neurosci.* 18, 193–222.

Dretske, F. (1993). Conscious experience. *Mind* 102, 263–283. [Reprinted in Noë and Thompson (2002). *Vision and Mind.* Cambridge, MA: MIT Press].

Dretske, F. (1995). *Naturalizing the Mind.* Cambridge, MA: MIT Press.

Epstein, W., and Broota, K. D. (1986). Automatic and attentional components in perception of size-at-a-distance. *Percept. Psychophys.* 40, 256–262.

Fabre-Thorpe, M., Delorme, A., Marlot, C., and Thorpe, S. (2001). A limit to the speed of processing in ultra-rapid visual categorization of novel natural scenes. *J. Cogn. Neurosci.* 13, 171–180.

Granrud, C. E. (2004). "Visual metacognition and the development of size constancy," in *Thinking and Seeing*, ed. D. T. Levin (Cambridge: MIT Press), 75–95.

Hatfield, G. (2009). *Perception and Cognition.* Oxford: Clarendon Press.

Jackendoff, R. (1989). *Consciousness and the Computational Mind.* Cambridge, MA: MIT Press.

Jackson, F. (1977). *Perception: A Representative Theory.* Cambridge: Cambridge University Press.

Johnson, J. S., and Olshausen, B. A. (2005). The earliest EEG signatures of object recognition in a cued-target task are postsensory. *J. Vis.* 5, 299–312.

Johnston, M. (2006). "Better than mere knowledge: the function of sensory awareness," in *Perceptual Experience*, eds T. S. Gendler and J. Hawthorne (Oxford: Clarendon Press), 260–291.

Kihara, K., and Takeda, Y. (2010). Time course of the integration of spatial frequency-based information in natural scenes. *Vision Res.* 50, 2158–2162.

Kosslyn, S. M. (1994). *Image and Brain.* Cambridge, MA: MIT Press.

Lamme, V. A. F. (2003). Why visual attention and awareness are different. *Trends Cogn. Sci. (Regul. Ed.)* 7, 12–18.

McDowell, J. (1994). *Mind and World.* Cambridge, MA: Harvard University Press.

McPherson, F. (in press). Cognitive penetration of colour experience: rethinking the issue in light of an indirect mechanism. *Philos. Phenomenol. Res.*

Nanay, B. (2010). Perception and imagination: amodal perception as mental imagery. *Philos. Stud.* 150, 239–254.

Palmer, S. (1999). *Vision Science.* Cambridge, MA: MIT Press.

Perry, J. (2001). *Knowledge, Possibility, and Consciousness.* Cambridge, MA: MIT Press.

Peyrin, C., Michel, C. M., Schwartz, S., Thut, G., Seghier, M., Landis, T., Marendaz, C., and Vuilleumier, P. (2010). The neural processes and timing of top-down processes during coarse-to-fine categorization of visual scenes: a combined fMRI and ERP study. *J. Cogn. Neurosci.* 22, 2678–2780.

Prinz, J. J. (2002). *Furnishing the Mind.* Cambridge, MA: MIT Press.

Pylyshyn, Z. (2003). *Seeing and Visualizing: It's not what you Think.* Cambridge, MA: MIT press.

Raftopoulos, A. (2009). *Cognition and Perception: How do Psychology and the Neural Science Inform Philosophy?* Cambridge, MA: MIT Press.

Raftopoulos, A. (2010). Can nonconceptual content be stored in visual memory? *Philos. Psychol.* 23, 639–668.

Raftopoulos, A., and Muller, V. (2006). Nonconceptual demonstrative reference. *Philos. Phenomenol. Res.* 72, 251–285.

Rock, I. (1983). *The Logic of Perception.* Cambridge, MA: MIT Press.

Roelfsema, P. R., Lamme, V. A. F., and Spekreijse, H. (1998). Object-based attention in the primary visual cortex of the macaque monkey. *Nature* 395, 376–381.

Schellenberg, S. (2008). The situation dependency of perception. *J. Philos.* 105, 55–84.

Shams, L., and Beierholm, U. R. (2010). Causal inference in perception. *Trends Cogn. Sci. (Regul. Ed.)* 14, 425–432.

Shoemaker, S. (2006). "On the way things appear," in *Perceptual Experience*, eds T. S. Gendler and J. Hawthorne (Oxford: Clarendon Press), 461–481.

Siegel, S. (2006). "Which properties are represented in perception?" in *Perceptual Experience*, eds T. S. Gendler and J. Hawthorne (Oxford: Clarendon Press), 481–504.

Spelke, E. S. (1988). "Object perception," in *Readings in Philosophy and Cognitive Science*, ed. A. I. Goldman (Cambridge, MA: MIT Press), 447–461.

Stalnaker, R. C. (2008). *Our Knowledge of the Internal World.* Oxford: Clarendon Press.

Strawson, P. (1974). "Imagination and perception," in *Freedom and Resentment*, ed. P. Strawson (London: Methuen), 45–65.

Treisman, A. (2006). How the deployment of attention determines what we see. *Vis. cogn.* 14, 411–443.

Treisman, A., and Kanwisher, N. G. (1998). Perceiving visually presented objects: recognition, awareness, and modularity. *Curr. Opin. Neurobiol.* 8, 218–226.

Conflict of Interest Statement: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Perceptual learning and feature-based approaches to concepts – a critical discussion

Richard Stöckle-Schobel *

School of Philosophy, Psychology and Language Sciences, University of Edinburgh, Edinburgh, UK

A central challenge for any theory of concept learning comes from Fodor's argument against the learning of concepts, which lies at the basis of contemporary computationalist accounts of the mind. Robert Goldstone and his colleagues propose a theory of perceptual learning that attempts to overcome Fodor's challenge. Its main component is the addition of a cognitive device at the interface of perception and conception, which slowly builds "cognitive symbols" out of perceptual stimuli. Two main mechanisms of concept creation are unitization and differentiation. In this paper, I will present and examine their theory, and will show that two problems hinder this reply to Fodor's challenge from being a successful answer to the challenge. To amend the theory, I will argue that one would need to say more about the input systems to unitization and differentiation, and be clearer on the representational format that they are able to operate upon. Until these issues have been addressed, the proposal does not deploy its full potential to threaten a Fodorian position.

Keywords: concept learning, perceptual learning, computationalism

Cognitive psychology has recently seen the development of several new models positing a perceptual basis for conceptual systems. The panoply of views ranges from proposals to eliminate the distinction between concepts and percepts altogether (Barsalou, 1999) over more modest appraisals of the relations between the two (Goldstone and Barsalou, 1998) to proposals for the creation of cognitive processes through experience with perceptual stimuli (Schyns and Rodet, 1997). As one important contribution to this line of research, Robert Goldstone's perceptual learning approach stands out and shall be at the center of our present investigation in the links between the perceptual and the conceptual.

Among the specific questions related to perception and learning, Goldstone and his colleagues and collaborators discuss the possibility and mechanisms of perceptual learning (Goldstone, 1998), the influence of perception on categorization (Landy and Goldstone, 2005), the role of features of objects in categorization (Schyns et al., 1998), and learning in early ontogeny (Goldstone et al., 2011). Their target is the more conservative fixed-feature approach – a form of computationalism, which holds that new concepts are constructed by using pre-existent, cognitively fixed features. One avid defender of this view is Fodor (1975, 1981, 2008), whose classic argument against concept learning especially affects perceptually based "empiricist" theories. The upshot of this argument is that learning concepts needs to be based on a vocabulary in which hypotheses about these concepts are formulated. But that vocabulary itself already needs to contain the concept that is just being "learned." Fodor infers from this that all concepts must be innate, or at least not learned. In what follows, I will call this "Fodor's Challenge for theories of concept learning." Taking its conclusion as an undesirable outcome for any theorist who wants to maintain a notion of genuine learning, one might ask the following question.

Assuming that cognition is at least a partly computational process, is there any way of having new symbols from perceptual origins entering the internal symbol system? As Goldstone answers this to the affirmative, I will discuss his proposal and point out two problems with it that need more consideration.

FODOR'S CHALLENGE FOR THEORIES OF CONCEPT LEARNING

In order to see where Goldstone and colleagues aim when they criticize the fixed feature position, I will first briefly set out this position and the challenge it poses to research in perceptual learning. Fodor goes through the following steps to reach the conclusion that concepts, and by that also features, cannot be learned.

First, Fodor construes learning mechanisms as "rational–causal processes" (Fodor, 1981, p. 273) – Being a rational process, learning is mediated by psychological states, such as beliefs. Also, regarding the possible constituents of thought, Fodor argues that what is not learned is innate or acquired in some other non-rational way.

A further premise of his argument concerns possession conditions for concepts: "A sufficient condition for having the concept C is: being able to think about something *as (a)* C" (Fodor, 2008, p. 138, original emphasis.). This means that the main act of concept use is using the concept in forming beliefs (or other types of thoughts) – as contrasted with using the concept to categorize new sensory experiences, or to act upon a thing in the world. For Fodor, thinking is prior to perceiving and acting in the order of concept use.

Now, from this position one needs to have a model of how a concept can enter this realm of thought. Fodor argues that the only available, empirically tested model for learning is the following: learning the concept C consists in forming hypotheses about C and testing them against the available evidence. Thus, learning

is a process of inductive inference. Forming a hypothesis about the concept C requires bringing the property expressed by C before one's mind. One needs to think about a piece of evidence "x" as (a) C to (dis-)confirm the hypothesis about C. To learn which things are green, one must judge something to be (or not be) green. This act of judging is a mental going-on for which one needs to be able to think about green things, or about a thing as a green thing. Now, what is already used for hypothesis formation is not learned in the application (confirmation or disconfirmation) of the hypothesis. C was already available to form the hypothesis, thus C was not learned.

Fodor's conclusion is that all concepts are either innate or non-rationally (brute-causally, see Fodor, 1981) acquired. This conclusion is supposed to affect theorizing about learning concepts in all areas of the cognitive sciences, from developmental psychology to artificial intelligence research, since it affects the theory choices one has to explain the phenomena of these disciplines. Consider artificial intelligence: an important research aim of AI is to develop systems with human-like intelligence – computer programs that play chess like grandmasters, robots that move like biological organisms, and the like. In order to arrive at a theoretical foundation for such systems, philosophers explore the possibility and extent of the computational theory of mind (CTM) hypothesis – roughly, that the human mind can be best described as a system that works like a computer that operates on symbolic representations. For a Fodor-type computationalist, the number of symbols would be predetermined by the system, and so the symbols would be innate. Given the additional constraint that each symbol of such a computational mind equals one concept, one has arrived at the point where Fodor's challenge and the computationalist program tie in. Landy and Goldstone (2005) describe such conceptions of CTM as being essentially linked to the idea that a fixed store of primitive, basic symbols is sufficient for successful cognition, and continue by saying that this classical version of CTM "entails a fixed set of primitives, or at least demands that any alterations to the primitive set are not cognitively interesting acts" (Landy and Goldstone, 2005, p. 346). Thus, we have characterized one stance toward Fodor's challenge, the fixed-features approach: it accepts the conclusion of Fodor's challenge, embraces the radical Concept Nativism that it entails, and denies any transformational effect on the cognitive system that would count as learning a new primitive concept.

By challenging the sufficiency of a fixed set of symbols for explaining human cognition, and by denying that changes to primitive symbols are not cognitively interesting, Landy and Goldstone set out their alternative to the Fodorian position and by that give us the second theory that we will presently take into account as a reply to Fodor's challenge. They want to argue for the creation of cognitive symbols from perceptual materials, and they want to argue for the possibility of manipulating "systems of high-level categories" (Landy and Goldstone, 2005, p. 346) to better fit the demands of the cogniser. The question motivating the present investigation thus is: Can Goldstone's theory of perceptual learning, and especially Landy and Goldstone's stance against fixed-feature languages, stand against Fodor's challenge and the acceptance of the fixed-feature approach, and can it give a mechanistically and computationally credible account of human concept learning?

PERCEPTUAL LEARNING AS A REPLY TO FODOR'S CHALLENGE

The question before us is whether it is possible to enrich a symbol system through the manipulation or introduction of perceptual information, or perceptual symbols. Learning features, like other forms of concept learning, can in an important sense be seen to hinge on the possibility of arriving at thoughts one was not able to hold or express before, and thus on having an alternative to Fodor's innateness conclusion by ways of providing an alternative empirical model for concept learning (rejecting the premise that hypothesis formation and testing is the only empirically available model for concept learning). The Fodorian CTM perspective, on Landy and Goldstone's (2005) account, denies this possibility, whereas several recent contributors to the debate have tried to develop models that support an affirmative answer. One main inspiration for this project comes from Gibson's (1963) theory of perceptual learning. A second major theoretical development was initiated by the work of Philippe Schyns on feature creation through experience with stimuli (Schyns and Murphy, 1991, 1994; Schyns and Rodet, 1997), leading up to the unified account of Schyns et al. (1998). The idea that the learning of a novel vocabulary of features yields new categorizations, which will be introduced below as a part of Landy and Goldstone's (2005) account, is rooted in the groundbreaking work of Schyns and his colleagues.

With their proposal, Landy and Goldstone mainly challenge Fodor's assumption that the primary use of concepts is in forming thoughts, as opposed to using concepts in dealing with the world via reacting on (sensory) inputs and acting in it/producing (behavioral) outputs. Grounding concept use and concept learning in perception does however not preclude the use of new perceptual concepts in higher cognitive activities – this is an important point made by Landy and Goldstone, e.g., in their discussion of changes in scientific reasoning through perceptual changes[1]. It is worth dwelling on this aspect of Goldstone's theory before turning to the core of Landy and Goldstone's (2005) proposal. Goldstone sees the possibility of what he calls perceptual learning, following Gibson (1963):

> Any relatively permanent and consistent change in the perception of a stimulus array, following practice or experience with this array, will be considered as perceptual learning. (Gibson, 1963, p. 29)

On this definition, perceptual learning is a sensory as well as a cognitive process: the changes in focus, or attentional center, to give two examples, in seeing something are at the same time changes in the categories pertaining to the perceived object. The repeated sensory contact with a certain class of objects will bring about a change in the way one thinks about these objects, which will in turn influence its perception, i.e., the sensory processes.

Goldstone explicitly wants to trace the ties between these perceptual changes and the possible conceptual changes that accompany them. He holds that one traditionally neglected aspect of the relation between perception and conception is the influence that

---

[1]Landy and Goldstone (2005) discuss changes in ontology, cognitive properties of groups of scientists, and changes in scientific practice through new perceptual capacities as cases in point.

the conceptual system has on perception. In categorical perception, the learned categories influence the performance in perceptual tasks. Especially in the sciences, there are multiple examples for this. Mathematicians can name several properties of a function just by looking at its graph. Similarly, after studying the geological categories and training to differentiate various stone samples, geologists have a sharper grasp of the differences between stone types and are able to name them much faster than any layperson could (Goldstone, 1994; Goldstone and Hendrickson, 2010; Goldstone et al., 2012). This is also the second point Goldstone and Barsalou (1998) stress:

> (...) perception's usefulness in grounding concepts comes from several sources. First, perception provides a wealth of information to guide conceptualization. Second, perceptual processes themselves can change as a result of concept development and use. Third, many of the constraints manifested by our perceptual systems are also found in our conceptual systems. (Goldstone and Barsalou, 1998, p. 232)

The first statement of this quote, that our perception can be a source of information for our conceptual system, does not sound very controversial since it is not very informative and specific in itself. In what way does perception inform conception? Even on Fodor's account, perception informs conception in so far as perceiving an object x can cause the triggering of the accompanying concept X. For Goldstone, and especially for Barsalou (1999), there needs to be a more detailed description of the way in which perceptual information touches upon our concepts; a description which probably even does away with the distinction between perception and conception. The second point has bearing on the present question in so far as it is the converse of the claim that Landy and Goldstone (2005) put forward to challenge Fodor: if both of these directions of influence were part of the actual workings of the human mind, then the strongly computationalist position would either lose a lot of its plausibility, or would have to be reformulated to accommodate these interrelations. Such an accommodation would however run against the self-proclaimed Rationalist position that Fodor adopts. Finally, the third point is especially important for Barsalou's (1999) project, but beyond the scope of the current investigation.

With these preliminaries set out, let's see how they form a frame for Landy and Goldstone's (2005) answer regarding Fodor's challenge. A short characterization of the main aims of argumentation, with references to extended presentations and discussions of these points, can be put as follows:

1. In learning about things we do not already understand, our cognitive system constructs specialized variable-feature languages that deal with these novel things (cf. Schyns and Murphy, 1991, 1994; Schyns and Rodet, 1997; Quinn et al., 2006).
2. The vocabulary of these languages consists of stimuli that we perceptually pick up and group as belonging to features, or feature dimensions (Schyns et al., 1998).
3. New features can be learned by applying the grouping mechanisms of unitization and differentiation, as the main players among other perceptual mechanisms (Goldstone, 1998, 2003; Goldstone and Landy, 2010).

It is generally assumed that concepts are the tools for, or the components of, thought. Thus, they are rather highly developed parts of our mental lives – conceptual thought is at the upper end of the scale of cognitive activity. Many things that we think about are very specific to a problem domain, like choosing a move in a chess game, while others are central to many modern human activities, like deciding which way to go to reach the nearest restaurant. Keeping with the computational tradition in the study of cognition, one can speak of different "vocabularies" or symbol stores for different tasks, with some being used for a more diverse range of activities than others.

Landy and Goldstone (2005) frame the debate as pertaining to languages of cognitive systems, which is not an uncommon level of discussion, given that Computationalism treats cognition as symbol-manipulation, and a number of symbols, combined with operations over these symbols, can with some right be called a "language." In the context of this paper, I propose to call such a language a computational language (Language$_C$), to highlight that the sense of "language" is somewhat restricted as compared to a spoken language. Computationalists like Newell and Simon (1976) or Fodor and Pylyshyn (1981, 1988) favor a fixed Language$_C$, whose symbols are inherent in the cognitive system, and sufficient for any kind of cognitive activity within that system – there is no need to import new symbols, since the given stock is supposed to express any proposition that the system would need to process. Biederman's (1987) geon model is another example for such a fixed Language$_C$, with the added twist that he attempts to posit perceptual representations – representations of basic geometrical forms – as a part of the innate stock of symbols.

To counter this model, Landy and Goldstone present what they call a "variable-feature language" (Landy and Goldstone, 2005, p. 347): a Language$_C$ that can be enriched with new primitive symbols, if new perceptual tasks require this. In Landy and Goldstone (2005), they characterize these enrichments as additions to particular sets of symbols, constrained by the category, or task, they are used for. In this, they follow Schyns and Murphy's (1991, 1994) major contribution to feature-based approaches to concept learning[2]. This leads to changes in highly specialized vocabularies, and need not necessarily affect the foundations of the Language$_C$. Landy and Goldstone talk of special-purpose Languages$_C$ and general-purpose Languages$_C$ in the cognitive system. While the latter are not excluded by Landy and Goldstone to be innate, given that they are ubiquitous in the most basic cognitive functions, the former need to be learned on their account. This is because the tasks that they are needed for are highly specialized in one way or another: examples that their paper discusses are fine perceptual discriminations such as discriminating brightness and saturation, and scientific theorizing and theory construction. Landy and Goldstone (2005, p. 348) compare the cognitive symbol system to LEGO blocks: some objects can only be constructed in a very cumbersome manner if only using the standard blocks (think of sails for a pirate's ship), so adding LEGO sails to their

---

[2] I want to thank an anonymous reviewer for pointing this out to me. The functionality principle, that functional demands shape the perceptual processes of categorizing new stimuli and forming new featural discriminations, from Schyns and Murphy (1994), has been a cornerstone of recent work in this area.

repertoire facilitates that specific kind of building process. The disadvantage of these special parts, however, is that they cannot serve for much else except their originally intended function. This, again, echoes the constraints on special-purpose Languages$_C$: the concept "color saturation" only has a very limited set of tasks for which it is needed, whereas the concept "not" has a scope that's equivalent to the generic LEGO blocks.

Now, in terms of the mechanisms of learning, Landy and Goldstone's theory's main component is the addition of a cognitive device at the interface of perception and conception, which slowly builds "cognitive symbols" out of perceptual stimuli. By adding these new symbols to the symbolic building blocks of thought, this device is the agent of concept learning and conceptual change. The main operations in this system are unitization and differentiation, two mechanisms which either unite previously separated conceptual elements, or split a vaguely bounded element class into finer groupings. In my present investigation, I will focus on these two mechanisms, since they are central to the argument by Landy and Goldstone (2005). When linking their theories with other, related work in the field, like in Goldstone et al. (2011), or Goldstone and Landy (2010), they also discuss other ways of learning. These include processes that Fodor would classify as brute-causal acquisition rather than genuine learning, which raises some questions I can only hint at in this investigation. Nevertheless, I will introduce these mechanisms later on.

Unitization can be described as a process of grouping several previously independent categories under one heading: "When elements co-vary together and their co-occurrence predicts an important categorization, the elements tend to be unitized" (Landy and Goldstone, 2005, p. 350). Here is an example of a process of unitization learning: suppose you learn what a cup is by seeing various different cups and not-quite-cuplike objects. Something qualifies as a cup if it consists of a cylindrical container and a handle to the side of the container. The contrast class of cuplike objects consists of other configurations of containers and handles, like a handle spanning the top or the bottom of the container (the former looking a bit like a bucket), or with a handle only connecting with the container at one point (looking like a horn attached to the cylinder), or even just unconnected handles and containers. The rules of unitization would incline you to unite the two featural elements (cylinder and handle) into a token of the concept if and only if they are in the right spatial configuration (handle on the side, both parts properly connected). Unitization allows you to conceive of the two parts as one object, and with that also to keep unfitting combinations, which do not satisfy the perceptual constraints, out of the class of cups.

It may be necessary to distinguish two kinds of uniting learning cases: associative chunking and perceptual unitization [3]. Associative chunking is the process through which two elements that co-occur regularly become associated: if one is accustomed to getting a glass of water with an ordered cup of coffee, then being served just a cup of coffee will create the expectation of a glass of water that has to follow: "drinking coffee" as an activity-concept has these two elements.

Let me consider a case in which two feature dimensions are reliably correlated so that the occurrence of either one is a reliable sign of the occurrence of the higher-order phenomenon. Would such a case be more aptly described as unitzation or as associative chunking?

Suppose that fire fighters always take big red vehicles that sound off a siren alternating between the first and the fourth tone of a scale (say C and F), and that all other emergency sirens use a different interval, say the prime and the fifth (C and G). Upon learning about the visual properties of fire engines, one might form a concept "fire engine" that is related to big red vehicles. Having also learned that the peculiar siren sound of the prime and the fourth is the fire fighter siren (having formed a concept "fire fighter siren"), one has formed the basis for putting together those two stimuli as the two most reliable signs for the presence of a fire engine. Thus, either of the stimuli can be used to trigger the concept "fire engine," despite the lack of the other. While more elaborate than the coffee-and-water case, from Goldstone's perspective this would still count as chunking, since the co-occurrence is not based on spatial, but on causal and temporal contiguity, which supports the formation of two separate feature elements that later get a common "heading" "fire engine."

The difference between unitization and chunking can be made clearer by another example, this time adapted from Landy and Goldstone (2005, p. 352): in the same way that a photograph of a group of people is a combination of pictures of the individuals, a unitized concept is a combination of features that stand in certain cognitively interesting, complex relations – possibly with spatial configuration as the main combinatorial criterion (again, as in the photograph case). In light of the perceptual constraints on unitization, which are the main point of difference to associative chunking, it would be more prudent to not expect unitization over different sense modalities such as visual and sound perception (as in the fire engine example) – at least until robust experimental data supports this idea.

The second mechanism for concept learning is dimension differentiation, "by which dimensions that are originally psychologically fused together become separated and isolated" (Goldstone, 2003, p. 249). Especially in differentiating dimensions, perceptual constraints influence the process: while it is easy for adult perceivers to separate the properties "size" and "brightness," it is much more difficult for the non-specialist to separate other fused dimensions such as brightness and hue. Differentiation might also be at work in separating non-dimension features, as in the fire engine example above. Suppose one never has paid much attention to siren sounds, and so has never noticed the difference between the fire fighters' siren and all other sirens – one has one single concept "emergency siren." Upon learning about the tonal difference between the two intervals, probably in a music class, one might start noticing the difference, and thus differentiate one's concept into "fourth interval siren" and "fifth interval siren," and then even relate these to the appropriate kinds of emergencies ("there's been a robbery next door, I'm quite sure that I'll soon hear the fifth interval police siren").

To make the differences between the perceptual learning perspective and classical computationalism clearer, here is another rephrasing with an example. For the fixed feature approach, new

---

[3]The distinction was pointed out to me by Robert Goldstone.

mental representations are new combinations of previously available primitive elements. Associative chunking, as in "glass of water" and "cup of coffee" as components of the concept "things that I drink when having coffee," requires the availability of the components that are combined. In the variable-feature approach, new representations need not necessarily be primitive elements, or psychologically pre-available elements. Rather, they can be stimulus elements with "no parsing in terms of psychological primitives" (Landy and Goldstone, 2005, p. 350) – so, what Landy and Goldstone want to argue for is the import of perceptual tokens into the cognitive system. As with LEGO blocks, constructing the concept "cup" from perception is like designing a LEGO cup, with the restrictions that this brings to the use of the concept (you can mainly use the cup to use it as a cup, and not to, e.g., build a LEGO house from LEGO cups).

One already alluded to example comes from Burns and Shepp's (1988) study on color vision. Their main idea is that the three defining features of any given color – its brightness (value), saturation (chroma), and hue – are difficult to selectively attend to for an untrained observer since color perception is the perception of quite holistic stimuli. If this were the case, then one would expect that test subjects would have difficulty separating these dimensions when comparing a range of samples. This is just what their experiments demonstrated. In their study, Burns and Shepp also found that differentiating brightness and saturation is easier for trained individuals, such as artists. Landy and Goldstone take this as evidence for the creation of new feature detectors: if there was only one detector for "color" before the training, and the subjects were able to differentiate the brightness and saturation of a range of colors after their training, then the perceptual task must have been the cause for feature learning, and for the creation of new perceptual, discriminatory capacities. And surely, if a person did not know the difference between the brightness and the saturation of a color before, and could make a discriminatory judgment after the study, then a new concept has been learned.

A final, important aspect of Goldstone's proposal is that his and his colleagues' studies do not rely on predetermined, fixed stimuli sets, but on totally novel ones that often cannot readily be parsed into already known structural elements. An example can be found in Schyns and Murphy's (1994) "Martian rocks" studies. The study employed various black blobs with several kinds of round or prolate appendices, without any indication of possible fragments, or parts of the whole object.

In their argument for using such alternative materials, Schyns et al. (1998) express that they want to exclude the possibility of using known categories in their experimental tasks, and that they want to get a better understanding of the ways in which totally new categories are learned. If a shape is almost certain to not represent any possibly innate, fundamental shape primitive, then it should be very likely that learning to pick out that shape is a case of learning something new.

With alternative materials, many different interpretations are possible, there are multiple features that could be encoded, and the analog format (as opposed to the digital signs one also finds in fixed feature experiments) make it possible to study something akin to real-life concept learning, where the interesting or learn-worthy features also are not as plainly recognizable.

With this picture of Landy and Goldstone's (2005) reply to Fodor's challenge in mind, let me now confront the question if their proposal can stand up to the challenge.

CONCEPT LEARNING OR CONCEPTUAL CHANGE?

Given that the perceptual learning approach can do the things described above, does it actually answer Fodor's Challenge for concept learning? I want to argue that it doesn't, because of two problems. Goldstone and colleagues have thus far left decisive questions pertaining to the central elements of their account, namely the details of the integration of perceptual symbols into the representational system, and the role of features and stimuli in that process, unanswered.

First and foremost, it is not clear whether the model provides prospects for concept learning at all. One might agree that the phenomena of unitization and differentiation are a form of learning, since they are mechanisms of restructuring previously available categories, and thereby they are means of grouping information in new ways that might lead to new beliefs. Consider somebody who finds out that two animals which she knew to be dogs belonged to two different breeds, say Labrador Retrievers and Dalmatians. This could clearly count as learning something one had not previously understood. But does it really count as introducing two new "psychological primitives"? An alternative view along fixed-feature lines would be to grant that "Labrador Retriever" and "Dalmatian" are indeed new, but only as names for two objects that had already been processed in thought in a different way, say as p and q (the letters standing for the symbols standing for the individual dog tokens). So, what has been added were not new symbols, but rather new labels for old symbols, or new beliefs about these symbols, as in "p is a token of 'Labrador Retriever'."

It has been suggested to me to look at a more perceptually taxing kind of differentiation process, since this might support Landy and Goldstone's position. Suppose that, in a psychological experiment, a subject is rewarded for identifying tokens of pacman shapes with a 92° "mouth" angle, and not rewarded if she chooses pacman shapes with a 90° "mouth" angle. Wouldn't one want to say that learning to appropriately keep those two shapes apart counts as learning a new concept? While the example is intriguing, and representative for a class of psychological experiments on categorization, I would argue that it does not count as a case of learning a new concept of, say, "92° pacman." Since the goal of the subject lies in getting the reward, it seems more appropriate to speak of the concept "choice that gets me a reward" and the perceptual input that is related to a token of the concept – a choice of a given answer, say "A" or "B" (if "A" and "B" stand for the answers related to the respective 92°/90° pacman). If one carries this thought further, the discriminatory input does not become involved in the conceptual content of the concept that is applied in the task – a 92° pacman is not a token of "choice that gets me the reward," but it is a prompt to apply the concept by acting in a certain way. This is not to say that fine perceptual discriminations can never become conceptually relevant, or the topic of conceptual development: examples like Smith and Kemler's (1978) study of changes in the integrality of dimensions such as color and shape surely count as evidence to the contrary. The theoretical status of these developmental changes, however, is exactly the topic of this paper,

even though I presently cannot go into a detailed treatment of the developmental literature for lack of space.

Returning to the dog example, the tricky question for the fixed-feature computationalist at this point, therefore, is not whether the new dog breeds were learned, but rather how the symbol, the name, and the object that the symbol and the name denote are causally related. This is a thorny question for philosophers like Fodor (1998), as they have to defend a very specific type of metaphysical theory of causation to make their analysis stick (see, e.g., Cowie, 1998 for discussion). Perceptual learning would evade this metaphysical question ("How can an innate symbol refer to anything that was encountered perceptually?"), but at the price of creating a psychological one ("How is it possible to import new cognitive symbols from perceptual origins into a Language$_C$?"), which will create the second worry that is identified further below.

At this point, one might be tempted to postulate that there are several kinds of learning: in one kind of learning experience, some genuinely new primitive psychological token (be it a new feature, or symbol, or whatever kind of enrichment one might be interested in) is incorporated into the cognitive system. A case in point would be turning a perceptual stimulus into a cognitively usable symbol that can be used for category judgments, forming thoughts, or other conceptual tasks. Here, something that has not been part of the Language$_C$ would be transferred into that same language. Fodor's challenge is concerned with this kind of learning.

In another kind of learning experience, the available pieces of information get re-ordered, linked to other bits of information, or get categorized in a finer raster. Strictly speaking, nothing new enters the store of cognitive symbols, but the differentiation between different kinds of already available symbols will be finer, or coarser, depending on the type of change. The question remains if Landy and Goldstone would be happy with "only" providing a model for the second type of learning, since the aim of their article, in their own words, clearly was to give a model for the first type:

> our alternative to fixed-primitive languages involves not giving up computationalism, but enriching it with mechanisms which allow the construction of new psychological primitives that are not just combinations of other known categories. (Landy and Goldstone, 2005, p. 347)

On one interpretation of the perceptual learning approach, the main processes of unitization and differentiation seem to fail to introduce new concepts, since they only operate on existing concepts, which are modified to be either more general or more specific regarding certain features of a given category. As Landy and Goldstone (2005) openly state, "feature creation simply involves alterations to the organization of stimulus elements into features" (Landy and Goldstone, 2005, p. 349). But a more strict computationalist, or a Nativist, could easily argue that this process does not strictly speaking add any new information to the cognitive system, as, e.g., Fodor (2008) does. Rearranging old concepts, on this view, cannot be counted as learning, since there is no new information added, but only a regrouping of old concepts. Like in the dog case above, there would only be the addition of new labels for objects that have previously been parts of the Language$_C$. If, however, one wants to object to this analysis, and maintain that unitization and

differentiation mainly work on percepts, then the worries raised in the next part of this paper will apply.

A variant of Fodor's hypothesis-testing paradox can be formulated that transfers the point into the feature-based learning Goldstone endorses: in order to categorize a stimulus as being evidence for/being a token of a certain psychological feature, one needs to know what feature that is – in order to perceive a sound as a fire fighter siren, one needs to know what a fire fighter siren sounds like (fourth interval). And to know that, one surely needs a feature category that is available before having a stimulus to categorize accordingly. If there is not more to what perceptual learning can do for our understanding of what concept learning should be, then it does not have an explanatory advantage to Fodor's Nativism in this regard, and need not even be incompatible with his large and fixed basic vocabulary of the mind – the primitive symbols being given, while more elaborate concepts might well be composed by mechanisms like unitization and differentiation.

Gauker (1998) addresses a similar worry concerning Schyns et al. (1998). Gauker in fact poses a double dilemma to any "concepts as (composed out of) features" approach. Suppose that concepts are composed of features, and that learning a concept involves learning a certain amount of the properties that are associated with the concept. Learning the concept "bird" might be linked to associating the concepts "has wings," "has a beak," "flying animal," or any other combination of attributes, to the concept "bird." Yet, this would require these attributes to be developmentally more basic than the concept "bird." How could that be? Gauker sees only two possibilities, which pose a dilemma for Goldstone. Either one accepts that there is a developmental hierarchy of features and concepts. This option is in principle open to anybody, but has special advantages for fixed-feature theorists. One could postulate that there are certain primitive features at the basis of the more elaborated conceptual constructions that are learned in cases like learning the concept "bird." These features are already parts of the cognitive system that did not need to be learned, and so would form an (in some way) innate basis for our more superordinate concepts. If these features were innate, or pre-specified, they would be fixed. As laid out above, Goldstone wants to have some room for flexible features, so only relying on this option does not seem to be a viable option, especially since fixed-feature theorists might just postulate a big enough or flexible enough primitive basis of concepts that could really ground any supposedly perceptually learnable concept. The other option is to deny that feature concepts need to be more primitive than the superordinate concepts – Gauker associates Schyns et al. with this view. The problem for this view is that it requires an explanation of exactly how "truly new features are created" (Gauker, 1998, p. 27) – features that do not have a previous history of, e.g., having been fused with other features, forming a less differentiated category.

Landy and Goldstone (2005) attempt to answer these kinds of criticism by pointing out the changes that they have observed in their studies and in similar studies. They cite evidence that "early perceptual devices can be systematically and physically altered by the environment to change their representational capacities" (p. 351) to support the claim that new features can be created. For example, in simulations by Rumelhart and Zipser (1985), connectionist systems were able to create new detectors for different kinds

of stimuli in a competitive learning task. But while the evidence might be supporting this claim, it certainly need not support the connected claim that such a change in representational capacities causes changes in the Language$_C$, and by this causes the learning of new features. A fixed feature theorist might be very happy with the first claim, linking it to the activation or triggering of a certain store of symbols that affects early perceptual devices: environmental influences would first cause changes in the (already present) symbolic system, which would in turn result in changes in perception. The change in representational capacities thus might just be a change in frequencies of triggering certain symbols. One might call this a form of learning, since there would be changes in the perceptual domain, but the corresponding changes in the conceptual domain – starting to use previously available symbols for hitherto unperformed perceptual tasks – would not be substantial enough to warrant the label "concept learning."

Adding to this point, one can enlist another example of a learning system that Landy and Goldstone (2005) briefly discuss it in their, and which they revisit in Goldstone and Landy (2010): the Pask device[4]. A Pask device is "an array of electrodes partially immersed in an aqueous solution of metallic salts" (Landy and Goldstone, 2005, p. 351) that will physiologically change when electric currents are applied. Now, changes in electrical configuration in the device come with changes in functionality – the device will start reacting discriminatively to two kinds of sound frequencies: a "new ear" for the circuit has been trained while it got constructed. From Fodor's perspective, it would however be a mistake to call this learning. These changes have all the characteristics that brute-causal acquisition of concepts in humans also has, so by definition, they do not amount to concept learning. This issue is independent of the question whether the Pask device actually is a representational system – if a certain reaction to frequency A counts as representing that frequency. Following Prinz and Barsalou (2000), I am inclined to regard the Pask device as a representational system, and in that sense a fitting analogy to a cognitive system. Focusing on the question whether the Pask device's development of an electronic ear is more like knowing the difference between smoke and steam after being hit on the head or more like learning from observation about the difference, I submit that it is decidedly more like the former, and thus not a case of learning in Fodor's sense[5].

Based on this reasoning, it seems right to focus on unitization and differentiation as the main players in feature learning when seen as concept learning, while acknowledging that Fodor's kind of brute-causal acquisition – as demonstrated by human analogs to the Pask device's "learning" – plays a transformative role in human cognition that might be seen as enabling concept learning.

Up to this point, the perceptual learning approach has not succeeded in answering Fodor's challenge, since the alternative fixed-feature theory has been shown to give equally powerful explanations of phenomena like changes in representational capacity, while not having the problem of having to explain how new cognitive symbols could be created from perceptual materials. Also, the perceptual learning approach has not given a full model for the latter task, and thereby only stands on a partial base of providing good explanations for the influence of the conceptual on the perceptual. There is however another set of conceptual problems that call for a resolution before the perceptual learning approach can get off the ground and before we can assess whether an inference to the best explanation would support the fixed-feature approach or the perceptual learning approach.

THE NOTIONS "FEATURE" AND "STIMULUS"

A second worry that directly follows from the first one has to do with the notion of features and stimuli in concept learning. In Goldstone's theory, concepts are (created out of) features; features are created from stimuli. Stimuli are in a format that is supposedly compatible to the symbolic vocabulary of cognition. Compatibility is a decisive criterion here, and can be defined as follows:

> Compatibility $=$ (Def.) A set of symbols S is compatible to a cognitive mechanism M iff inputting S into M yields a (symbolic) output S_O which can be used by further cognitive mechanisms. A set of symbols S is compatible to a second set of symbols S_2 iff S and S_2 can both serve as input for M individually, and iff symbols from S and S_2 in combination also yield a symbolic output S_O which can be used by further cognitive mechanisms.

Without a fulfilled compatibility criterion, it would not be possible to incorporate the perceptually based symbols into the previously available and exercised cognitive activities. Speaking in terms of Language$_C$, all symbols need to be combinable to form correct sentences from them[6]. Or again in Landy and Goldstone's (2005) terms:

> the claim that novel perceptual features can be learned sounds murky, or even mystical, without the clarification that the novel features are always drawn from a larger, more expressive, more primitive language embodying the physical and pre-conceptual constraints on what can be incorporated into features in the first place. (Landy and Goldstone, 2005, p. 348)

What is a stimulus then? In any dictionary of behavioral and cognitive science, one will find descriptions of perception in terms of proximal and distal stimuli. The distal stimulus of a visual experience might be the tree one sees, whereas the proximal stimulus would be the light reflection arriving at the eyes. At which level do, e.g., Landy and Goldstone (2005) individuate stimuli? This question is pertinent to our present investigation since Landy and Goldstone should address it to make clear where exactly they would see the origins of perceptual experiences, and with that the origins

---

[4]The example of the Pask device has been introduced into the feature creation literature in the section "Authors' response" of Schyns et al. (1998), where it is used to link the ideas of perceptual learning and of emergent properties in learning.

[5]This investigation invariably leads to the question whether perceptual learning in humans should count as Fodor-type learning or as brute–causal acquisition. Discussing this point is beyond the scope of the current paper, but can be developed into a different argument against the methodological set-up of Fodor's challenge – the distinction between brute–causal and rational acquisition might not be as helpful as Fodor would like it to be, and cases like the Pask device might work in favor of giving up the distinction altogether.

[6]This point is related, though not identical, to the issue of the compositionality of thought, raised, e.g., by Fodor and Pylyshyn (1988).

of perceptually based concepts: are they in the world (i.e., distal stimuli) or are they in sensory activations (i.e., proximal stimuli)?

Just looking at their (Landy and Goldstone, 2005) paper, they discuss examples of roughly half-moon shaped figures combined of five segments and call these objects stimuli. In another example, they talk about "pieces of physical information [that are related to, or] packaged together in the same psychological feature" (p. 349) and use this as a synonym for "feature." These are exemplary – or metaphorical, respectively – descriptions of what features or stimuli should be, and yet, these are the most concrete mentions of those terms. A look at earlier renderings of the theory might help. Schyns et al. (1998) commit themselves to the following characterization of the meaning of the term "feature":

> The term "*feature* will refer to any elementary property of a distal stimulus that is an element of cognition, an atom of psychological processing. This does not imply that people are consciously aware of these properties. Instead, features are identified by their functional role in cognition; for example, they allow new categorizations and perceptions to occur." (Schyns et al., 1998, p. 1, original emphasis)

Here, first, features are described as "elementary properties" of distal stimuli. They are also implied to be "elements of cognition," i.e., Schyns et al. (1998) postulate a transition of perceptual properties into cognitive functions. The second point concerns a property's role in cognition: it is supposed to be functional. By this description, Schyns and colleagues want to counter the objection that some feature of an object might not enter a perceiver's conscious awareness and thus should not count as an element of cognition. While the wording in the quote above suggests a definition in terms of distal stimuli, it also alludes to the psychological role of features, which is even more obvious in Goldstone (2003):

> A psychological feature [. . .] is a set of stimulus elements that are responded to together, as an integrated unit. That is, a feature is a package of stimulus elements that [. . .] reflects the subjective organization of the whole stimulus into components. (Goldstone, 2003, p. 242)

So, I suggest we should understand Landy and Goldstone (2005) as taking proximal stimuli as the larger background language. Still, there is the unanswered question how these perceptual signals are transferred into language-like symbols that can be used in the same cognitive operations as either innate or previously acquired symbols. How are these vocabularies matched to each other? Let me push the language analogy a little further with an example: suppose the Language$_C$ is like English – approximately every English word corresponds to a mental symbol. Now suppose that the cognitive vocabulary gets enriched with a number of specialized features developed from a perceptual task, like learning chess moves. Perceptually learning a chess move, as opposed to learning it from a written description, might work as follows in the case of the rook's permissible movements on the board: straight lines along the horizontal and the vertical axis, but not on diagonals. The correct movements can be observed by watching rooks in a large sample of chess moves, or video clips from chess matches, also with a variety of token rooks (made from different materials, or shaped in a variety of ways), and one might even learn how to tell

whether a chess piece is a rook or a queen. The interesting question then is: would a thought about a situation in a chess match be multimodal – would it involve the perceptually learned symbol for the rook as well as the previously available non-domain-specific vocabulary? Let's take "[]" as a replacement for the perceptual symbol related to the rook moving one square to the left, just for this example, and phrase a thought like "If the rook moves one square to the left, the players will stop playing" multimodally: "If [], the players will stop playing" (given, e.g., that the result is a checkmate). Is it possible to infer the consequent of this conditional from being presented with a representation of []? While this last question might be for further empirical studies to decide, it already hints at the more general worry about perceptual tokens of some sort and their role in cognitive operations: given that originally, a certain cognitive function is performed by a mechanism using symbols of a (possibly innate) Language$_C$, how can the mechanism adapt to new symbols being introduced into it and filling that cognitive function? That is, how can a perceptual symbol store and an innately fixed symbol store become compatible, as defined above? Landy and Goldstone do not offer a model for this, and so I conclude that, as it stands, the construction of variable-feature language has not been sufficiently based on a model of transferring perceptual symbols into conceptual systems.

Sticking to the notion of features as primarily relevant to building mental representations, one could bring the theory of feature detectors into play, as in Barlow (2001). This, specifically in visual perception, would be an (obvious) way out of the problem, yet with the (obvious) problem that the feature detectors would have to be tuned to some specific inputs, and then the question arises again: How did the feature detectors come into being, and how did they get the tuning they exhibit? Feature detectors are an instance of (possibly innate) processes that one might use to explain concept learning, under the assumption that concepts need not be built into a system as long as there are built in processes that will be able to import concepts into the system. The problem, however, remains the same: a nativist can always argue that built in processes need to be tuned to their inputs in some way. In the perceptual learning literature (broadly defined), one can find several references to Gestalt laws (Schyns and Murphy, 1994; Quinn et al., 2006; Bhatt and Quinn, 2011), and even proposals to explain that acquiring Gestalt laws is possible through neural network simulations (Gerganov et al., 2007). Yet, how should adherence to a given Gestalt law, e.g., good continuity, be possible for a cognitive system without having concepts that are able to express the law and that would help classifying perceptions according to the principle?

Appealing to maturation, or another form of innateness, would yield no new variable-feature language in the sense of Landy and Goldstone (2005), as its large explanatory trump is its independence of the specific type of stimuli the system is confronted with. A classic example in point of this could be imprinting in newborn ducks, as discussed by Fodor (1981): any moving object will trigger the "concept" "mother," having the duck following the moving object. The important thing is that the duck's sensory system is predetermined to follow the closest moving object.

If on the other hand the specific stimuli play a role, or matter in some sense for the creation of psychological primitives, how do they influence the creation of a feature detector?

We seem to have gone the circle back to the original question, and the appeal to a larger background language of stimuli has not advanced us very far. Maybe, looking at the problem from a different perspective, under the heading of "physical information" will clear things up. After all, Landy and Goldstone (2005), also refer to the materials from which to get new features as "physical information" (p. 349).

Now, the task is to disambiguate the notion "physical information" and to link it to the question of the proximity of stimuli. Either, physical information, or a bundle of features, is supposed to be informationally structured before or in being perceived. This would require a form of direct realism, or of direct perception: for example, a given object affords to be perceived as a tree, so we as perceivers pick up the right kind of information in order to treat it as a tree. Or, on the other hand, physical information is the cognitive content that has been extracted from the experience; this could be something like a representation of a tree. The retina has registered a certain image, sent it to the visual cortex in one way or another and there, a representation of the tree is formed, or accessed, or activated. In line with the above decision to talk of stimuli in terms of proximal stimuli, it seems sensible to choose the interpretation of physical information as cognitive content. This interpretation, however, just skips the interesting question, which is: "How does a representation of a new object come to be included in a cognitive system?", and leaves the field open for any kind of Nativist reply to the effect that the representation was triggered by some process, or that the perceptual stimulus got paired with an arbitrary symbol from the wealth of symbols in the representational mind. To avoid this, any proponent of perceptual learning has to go the long way and show just how perceptual content enters cognition and by which means new symbols, or new bits of a feature-"language," are added to the system. Landy and Goldstone do not offer a model for this form of feature learning, and so I have to conclude that – while it is not conceptually excluded that such a model is possible – their proposal still has some way to go before it can pose a fully developed alternative to their fixed-feature opponents.

CONCLUSION: PERSPECTIVES FOR AN AMENDMENT OF THE PERCEPTUAL LEARNING APPROACH

To amend the theory, one would need to say more about the input systems to unitization and differentiation, and be clearer on the representational format that they are able to operate upon. Specifically, the following questions are still unanswered.

How can a cognitive mechanism that was presumably first stocked with innate computational symbols grow to work with learned perceptual features as input to and vocabulary for its activity? And is it possible to mix symbols of different origins and formats (amodal/modal) – to have "multi-lingually" integrated cognition?

Until the issues raised in this article have been addressed, the proposal does not deploy its full potential to threaten a fixed-feature approach à la Fodor: even if both approaches can be construed as having similar levels of explanatory power, one of them satisfies Fodor's challenge while the other one does not yet overturn its empirical premise. The disadvantages that stem from the problems identified in this investigation weaken the perceptual learning approach's appeal and thereby put its opponent in the stronger argumentative position for now.

After this discussion, one might also be tempted to conclude that the notion of a feature language, flexible, or variable, is misguided as it invariably brings the issue of translation into the debate. How to translate a stimulus (and which stimulus) into a mental symbol? Also, it suggests an ordered, or "grammatical" structure in the non-mental/physical world that is the object of perception. This is dangerous, because the world as appearing to us might not actually be best carved into the perceived (natural?) kinds, but into theoretical kinds that we only perceive through mediation. One does not see the chemical structure of object x without being somewhat of a trained chemist, if at all, or at least it is not clear in the Perceptual Learning approach whether the interaction between perception and cognition leads to such depths of theory-ladenness of perception (as in *seeing* chemical structure) as opposed to a quick-and-dirty inferential connection between perceiving certain visual properties of a chemical sample – e.g., observing a deep green flame when burning a sample of a chemical powder, and identifying it as copper(II)-sulfate (maybe a perceptual–cognitive process), as well as the sample's being a salt (an inference from that observation). In making the distinction between perceptual–cognitive and inferential, the adherence of a given process to perceptual constraints might indicate that the process is of the former type, whereas observations following theoretical, or "conceptual," rules can be properly classified as the latter[7]. Still, the distinction is not always a clearly cut one.

This does not just touch upon Landy and Goldstone's (2005) proposal, but more generally on their still dominating opponents: if the language metaphor does not work for features, as one might conclude from the problems raised in the previous section, then why should one be inclined to see the strong analogy between computers – symbol crunchers – and human minds with brains and nervous systems underlying them (in one way or another) as necessary? Perhaps the mind only becomes symbolic by starting to use symbols, but does not reflect this symbol-mindedness in the elements of cognition. Dissociating materials, or vehicles of thought, on the one hand, and thought contents on the other hand, might be a prudent move until a clearer picture of the connections between vehicles and contents is available.

---

[7]It was Robert Goldstone who pointed this out to me.

REFERENCES

Barlow, H. (2001). "Feature detectors," in *The MIT Encyclopedia of the Cognitive Sciences*, eds R. A. Wilson and F. C. Keil (Cambridge: MIT Press), 311–314.

Barsalou, L. W. (1999). Perceptual symbol systems. *Behav. Brain Sci.* 22, 577–660.

Bhatt, R. S., and Quinn, P. C. (2011). How does learning impact development in infancy? The case of perceptual organization. *Infancy* 16, 2–38.

Biederman, I. (1987). Recognition-by-components: a theory of human image understanding. *Psychol. Rev.* 94, 115–117.

Burns, B., and Shepp, B. E. (1988). Dimensional interactions and the structure of psychological space: the representation of hue, saturation, and brightness. *Percept. Psychophys.* 43, 494–507.

Cowie, F. (1998). Mad dog nativism. *Br. J. Philos. Sci.* 49, 227–252.

Fodor, J. A. (1975). *The Language of Thought.* New York: Thomas Crowell Publishing.

Fodor, J. A. (1981). "The present status of the innateness controversy," in *Representations*, ed. J. Fodor (Cambridge: MIT Press), 257–316.

Fodor, J. A. (1998). *Concepts: Where Cognitive Science Went Wrong.* Oxford: Oxford University Press.

Fodor, J. A. (2008). *LOT 2: The Language of Thought Revisited.* Oxford: Oxford University Press.

Fodor, J. A., and Pylyshyn, Z. W. (1981). How direct is visual perception? Some reflections on Gibson's "ecological approach." *Cognition* 9, 139–196.

Fodor, J. A., and Pylyshyn, Z. W. (1988). Connectionism and cognitive architecture: a critical analysis. *Cognition* 28, 3–71.

Gauker, C. (1998). Building block dilemmas. *Behav. Brain Sci.* 21, 26–27.

Gerganov, A., Grinberg, M., Quinn, P. C., and Goldstone, R. L. (2007). "Simulating conceptually-guided perceptual learning," in *Proceedings of the Twenty-Ninth Annual Conference of the Cognitive Science Society* (Nashville: Cognitive Science Society), 287–292.

Gibson, E. J. (1963). Perceptual learning. *Annu. Rev. Psychol.* 14, 29–56.

Goldstone, R. L. (1994). Influences of categorization on perceptual discrimination. *J. Exp. Psychol. Gen.* 123, 178–200.

Goldstone, R. L. (1998). Perceptual learning. *Annu. Rev. Psychol.* 49, 585–612.

Goldstone, R. L. (2003). "Learning to perceive while perceiving to learn," in *Perceptual Organization in Vision: Behavioral and Neural Perspectives*, eds R. Kimchi, M. Behrmann, and C. Olson (Mahwah: Lawrence Erlbaum Associates), 233–278.

Goldstone, R. L., and Barsalou, L. W. (1998). Reuniting perception and conception. *Cognition* 65, 231–262.

Goldstone, R. L., Braithwaite, D. W., and Byrge, L. A. (2012). "Perceptual learning," in *Encyclopedia of the Sciences of Learning*, ed. N. M. Seel (Springer), 2580–2583.

Goldstone, R. L., and Hendrickson, A. T. (2010). Categorical perception. *Wiley Interdiscip. Rev. Cogn. Sci.* 1, 69–78.

Goldstone, R. L., and Landy, D. (2010). Domain-creating constraints. *Cogn. Sci.* 34, 1357–1377.

Goldstone, R. L., Son, J. Y., and Byrge, L. (2011). Early perceptual learning. *Infancy* 16, 45–51.

Landy, D., and Goldstone, R. L. (2005). How we learn about things we don't already understand. *J. Exp. Theor. Artif. Intell.* 17, 343–369.

Newell, A., and Simon, H. A. (1976). Computer science as empirical inquiry: symbols and search. *Commun. ACM* 19, 113–126.

Prinz, J., and Barsalou, L. W. (2000). "Steering a course for embodied representation," in *Cognitive Dynamics: Conceptual Change in Humans and Machines*, eds E. Dietrich and A. Markman (Cambridge: MIT Press), 51–77.

Quinn, P. C., Schyns, P. G., and Goldstone, R. L. (2006). The interplay between perceptual organization and categorization in the representation of complex visual patterns by young infants. *J. Exp. Child. Psychol.* 95, 117–127.

Rumelhart, D. E., and Zipser, D. (1985). Feature discovery by competitive learning. *Cogn. Sci.* 9, 75–112.

Schyns, P. G., Goldstone, R. L., and Thibaut, J.-P. (1998). The development of features in object concepts. *Behav. Brain Sci.* 21, 1–17.

Schyns, P. G., and Murphy, G. L. (1991). "The ontogeny of units in object categories," in *Proceedings of the XIII Meeting of the Cognitive Science Society* (Hillsdale: Erlbaum), 197–202.

Schyns, P. G., and Murphy, G. L. (1994). "The ontogeny of part representation in object concepts," in *The Psychology of Learning and Motivation*, ed. D. L. Medin (Waltham: Academic Press), 305–349.

Schyns, P. G., and Rodet, L. (1997). Categorization creates functional features. *J. Exp. Psychol. Learn. Mem. Cogn.* 23, 681–696.

Smith, L. B., and Kemler, D. G. (1978). Levels of experienced dimensionality in children and adults. *Cogn. Psychol.* 10, 502–532.

Conceptual short term memory in perception and thought

Mary C. Potter *

Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, MA, USA

Edited by:
Arnon Cahen, Ben Gurion University in the Negev, Israel
Michela C. Tacca, Heinrich-Heine University Düsseldorf, Germany

Reviewed by:
Mark Nieuwenstein, University of Groningen, Netherlands
Candice Coker Morey, Rijksuniversiteit Groningen, Netherlands

*Correspondence:
Mary C. Potter, Department of Brain and Cognitive Sciences, 46-4125, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA 02139, USA.
e-mail: mpotter@mit.edu

Conceptual short term memory (CSTM) is a theoretical construct that provides one answer to the question of how perceptual and conceptual processes are related. CSTM is a mental buffer and processor in which current perceptual stimuli and their associated concepts from long term memory (LTM) are represented briefly, allowing meaningful patterns or structures to be identified (Potter, 1993, 1999, 2009). CSTM is different from and complementary to other proposed forms of working memory: it is engaged extremely rapidly, has a large but ill-defined capacity, is largely unconscious, and is the basis for the unreflective understanding that is characteristic of everyday experience. The key idea behind CSTM is that most cognitive processing occurs without review or rehearsal of material in standard working memory and with little or no conscious reasoning. When one perceives a meaningful stimulus such as a word, picture, or object, it is rapidly identified at a conceptual level and in turn activates associated information from LTM. New links among concurrently active concepts are formed in CSTM, shaped by parsing mechanisms of language or grouping principles in scene perception and by higher-level knowledge and current goals. The resulting structure represents the gist of a picture or the meaning of a sentence, and it is this structure that we are conscious of and that can be maintained in standard working memory and consolidated into LTM. Momentarily activated information that is not incorporated into such structures either never becomes conscious or is rapidly forgotten. This whole cycle – identification of perceptual stimuli, memory recruitment, structuring, consolidation in LTM, and forgetting of non-structured material – may occur in less than 1 s when viewing a pictured scene or reading a sentence. The evidence for such a process is reviewed and its implications for the relation of perception and cognition are discussed.

Keywords: CSTM, RSVP, scene perception, reading, short term memory, attentional blink

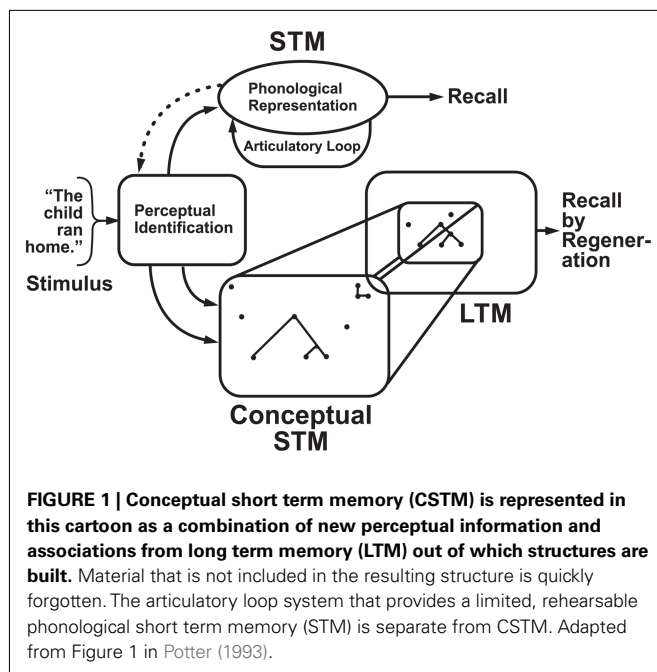## INTRODUCTION: CONCEPTUAL SHORT TERM MEMORY

Conceptual short term memory (CSTM) is a construct based on the observation that most cognitive processing occurs without review or rehearsal of material in standard working memory and with little or no conscious reasoning. CSTM proposes that when one perceives a meaningful stimulus such as a word, picture, or object, it is rapidly identified and in turn activates associated information from long term memory (LTM). New links among concurrently active concepts are formed in CSTM, shaped by parsing mechanisms of language or grouping principles in scene perception, and by higher-level knowledge and current goals. The resulting structure is conscious and represents one's understanding of the gist of a picture or the meaning of a sentence. This structured representation is consolidated into LTM if time permits. Momentarily activated information that is not incorporated into such structures either never becomes conscious or is rapidly forgotten. **Figure 1** shows a cartoon of CSTM in relation to LTM and one component of conventional STM.

CSTM IN RELATION TO OTHER MEMORY SYSTEMS AND OTHER MODELS

Conceptual short term memory is a processing and memory system that differs from other forms of short term memory. In vision, *iconic memory* (Sperling, 1960) maintains a detailed visual

representation for up to about 300 ms, but it is eliminated by new visual stimulation. Meaning plays little or no role. *Visual short term memory* (VSTM) holds a limited amount of visual information (about four items' worth) and is somewhat resistant to interference from new stimulation as long as the information is attended to (Coltheart, 1983; Phillips, 1983; Luck and Vogel, 1997; Potter and Jiang, 2009). Although VSTM is more abstract than perception in that the viewer does not mistake it for concurrent perception, it maintains information about many characteristics of visual perception, including spatial layout, shape, color, and size. In audition, the *phonological loop* (Baddeley, 1986) holds a limited amount (about 2 s worth) of recently heard or internally generated auditory information, and this sequence can be maintained as long as the items are rehearsed (see **Figure 1**).

Conceptual short term memory differs from these other memory systems in one or more ways: in CSTM, new stimuli are rapidly categorized at a meaningful level, associated material in LTM is quickly activated, this information is rapidly structured, and information that is not structured or otherwise consolidated is quickly forgotten (or never reaches awareness). In contrast, standard *working memory*, such as Baddeley's articulatory/phonological loop and visuospatial sketchpad together with a central executive (Baddeley, 1986, 2007), focuses on memory systems that support cognitive processes that take place over several seconds or minutes. A

FIGURE 1 | Conceptual short term memory (CSTM) is represented in this cartoon as a combination of new perceptual information and associations from long term memory (LTM) out of which structures are built. Material that is not included in the resulting structure is quickly forgotten. The articulatory loop system that provides a limited, rehearsable phonological short term memory (STM) is separate from CSTM. Adapted from Figure 1 in Potter (1993).

memory system such as the phonological loop is unsuited for conceptual processing that takes place within a second of the onset of a stream of stimuli: it takes too long to be set up and it does not represent semantic and conceptual information directly. Instead, Baddeley's working memory directly represents articulatory and phonological information or visuospatial properties: these representations must be reinterpreted conceptually before further meaning-based processing can occur.

More recently, Baddeley (2000) proposed an additional system, the *episodic buffer*, that represents conceptual information and may be used in language processing. The episodic buffer is "a temporary store of limited capacity. . . capable of combining a range of different storage dimensions, allowing it to collate information from perception, from the visuo-spatial and verbal subsystems and LTM. . . representing them as multidimensional chunks or episodes. . ." (Baddeley and Hitch, 2010). Baddeley notes that this idea is similar to CSTM as it was described in 1993 (Potter, 1993).

Although Baddeley's multi-system model of working memory has become the dominant model of short term memory, it neglects the evidence that stimuli in almost any cognitive task rapidly activate a large amount of potentially pertinent information, followed by rapid selection and then decay or deactivation of the rest. That can happen an order of magnitude faster than the setting up of a standard, rehearsable STM representation, permitting the seemingly effortless processing of experience that is typical of cognition. Of course, not all cognitive processing is effortless: our ability to engage in slower, more effortful reasoning, recollection, and planning may well draw on conventional short term memory representations.

Relation to other cognitive models
Many models of cognition include some form of processing that relies on persistent activation or memory buffers other than standard working memory, tailored to the particular task

being modeled. CSTM may be regarded as a generalized capacity for rapid abstraction, pattern recognition, and inference that is embodied in a more specific form in models such as ACT-R (e.g., Budiu and Anderson, 2004), the construction–integration model of discourse comprehension (Kintsch, 1988), the theory of long term working memory (Ericsson and Kintsch, 1995); and models of reading comprehension (e.g., Just and Carpenter, 1992; see Potter et al., 1980; Verhoeven and Perfetti, 2008).

EVIDENCE FOR CSTM
RAPID SERIAL VISUAL PROCESSING AS A METHOD FOR STUDYING CSTM
The working of CSTM is most readily revealed when processing time is limited. A simple way to limit processing time for visual stimuli is to use a single visual mask, immediately after a brief presentation. The idea is to replace or interfere with continued processing of the stimulus by presenting a new image that will occupy the same visual processors. Backward masks, if they share many of the same features (contours, colors, and the like) as the target stimulus, do produce interference and may prevent perception or continued processing of the target. Much has been learned about how we perceive, using backward masking. However, for complex stimuli such as pictures or written words, conceptual processing may continue despite the mask; to interfere with understanding or memory for the target, the mask itself must engage conceptual processing that will interfere with that of the target. An effective way to create such a limitation is to present a rapid sequence of visual stimuli, termed rapid serial visual processing (RSVP) by Forster (1970). By using RSVP in which all the stimuli (pictures or words) are meaningful and need to be attended, one can obtain a better measure of the actual processing time required for an individual stimulus or for the sequence as a whole (Potter, 1976; Intraub, 1984; Loftus and Ginn, 1984; Loftus et al., 1988). This method was used in many of the studies cited in the present review.

SUMMARY OF THE EVIDENCE
The evidence is summarized here before presenting some of it in more detail. Three interrelated phenomena give evidence for CSTM:
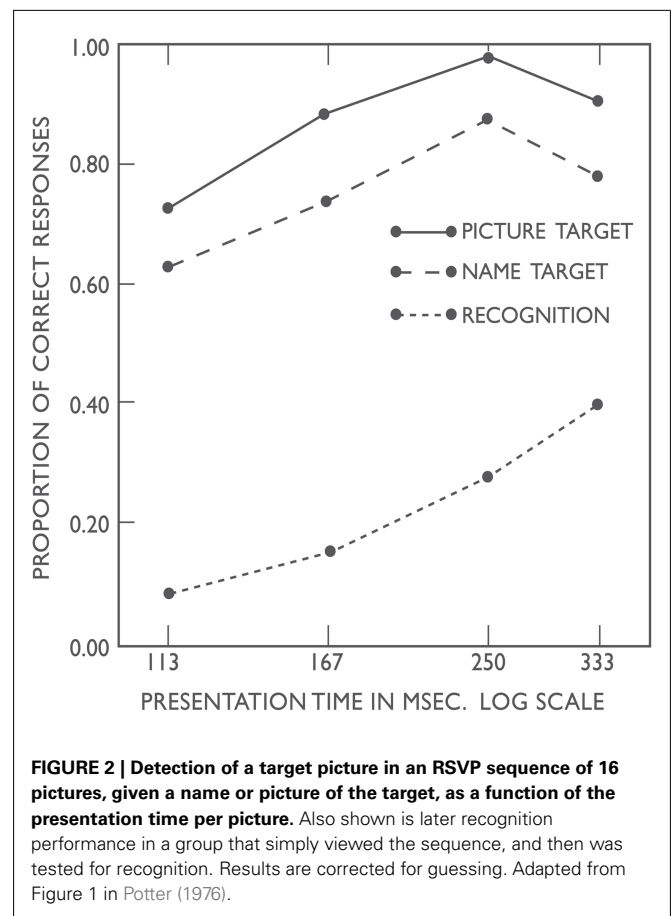
(1) There is rapid access to conceptual (semantic) information about a stimulus and its associations. Conceptual information about a word or a picture is available within 100–300 ms, as shown by experiments using semantic priming (Neely, 1991), including masked priming (Forster and Davis, 1984) and so-called fast priming (Sereno and Rayner, 1992); eye tracking when reading (Rayner, 1983, 1992) or looking at pictures (Loftus, 1983); measurement of event-related potentials during reading (Kutas and Hillyard, 1980; Luck et al., 1996); and target detection in RSVP with letters and digits (Sperling et al., 1971; Chun and Potter, 1995), with pictures (Potter, 1976; Intraub, 1981; Potter et al., 2010), or with words (Lawrence, 1971b; Potter et al., 2002; Davenport and Potter, 2005). To detect a target such as an animal name in a stream of words, the target must first be identified (e.g., as the word *tiger*) and then matched to the target category, an animal name (e.g., Meng and Potter, 2011). Conceptually defined targets can

be detected in a stream of non-targets presented at rates of 8–10 items/s or higher, showing that categorical information about a written word or picture is activated and then selected extremely rapidly. These and other experimental procedures show that semantic or conceptual characteristics of a stimulus have an effect on performance as early as 100 ms after its onset. This time course is too rapid to allow participation by slower cognitive processes, such as intentional encoding, deliberation, or serial comparison in working memory.

(2) New structures can be discovered or built out of the activated conceptual information, influenced by the observer's task or goal. Activated conceptual information can be used to discover or build a structured representation of the information, or (in a search task) to select certain stimuli at the expense of others. A major source of evidence for these claims comes from studies using RSVP sentences, compared with scrambled sentences or lists of unrelated words. Studies by Forster (1970), Potter (1984, 1993), Potter et al. (1980), and Potter et al. (1986) show that it is possible to process the structure in a sentence and hence to recall it subsequently, when reading at a rate such as 12 words/s. In contrast, when short lists of unrelated words are presented at that rate, only two or three words can be recalled (see also Lawrence, 1971a). For sentences, not only the syntactic structure, but also the meaning and plausibility of the sentence is recovered as the sentence is processed (Potter et al., 1986). Because almost all sentences one normally encounters (and all the sentences in these experiments) include new combinations of ideas, structure-building is not simply a matter of locating a previously encountered pattern in LTM: it involves the instantiation of a new relationship among existing concepts. The same is true when viewing a new pictured scene: not only must critical objects and the setting be identified, but also the relations among them: the gist of the picture. Structure-building presumably takes advantage of as much old structure as possible, using any preexisting associations and chunks of information to bind elements (such as individual words in a list) together.

(3) There is rapid forgetting of information that is not structured or that is not selected for further processing. Conceptual information is activated rapidly, but the initial activation is highly unstable and will be deactivated and forgotten within a few hundred milliseconds if it is not incorporated into a structure. As a structure is built – for example, as a sentence is being parsed and interpreted – the resulting interpretation can be held in memory and ultimately stabilized or consolidated in working or LTM as a unit, whereas only a small part of an unstructured sequence such as a string of unrelated words can be consolidated in the same time period.
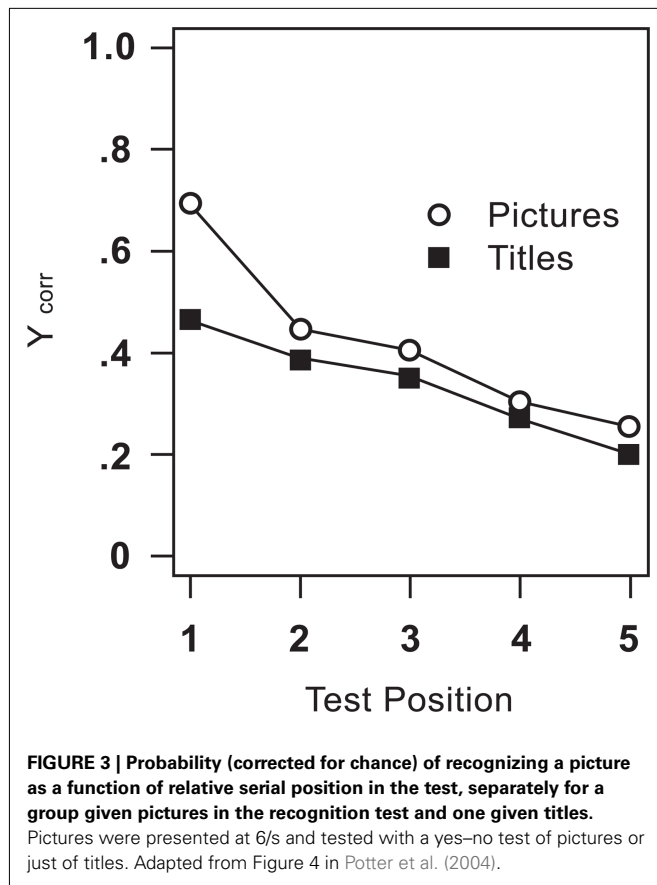
UNDERSTANDING PICTURES AND SCENES

In studies in which unrelated photographs are presented in RSVP, viewers can readily detect a picture when given a brief descriptive title such as *wedding* or *two men talking,* at rates of presentation up to about 10 pictures/s, even though they have never seen that picture before and an infinite number of different pictures could fit the description (Potter, 1976; Intraub, 1981; see **Figure 2** and Potter (2009) Demo 1 for a demonstration). As **Figure 2** shows,

FIGURE 2 | Detection of a target picture in an RSVP sequence of 16 pictures, given a name or picture of the target, as a function of the presentation time per picture. Also shown is later recognition performance in a group that simply viewed the sequence, and then was tested for recognition. Results are corrected for guessing. Adapted from Figure 1 in Potter (1976).

detection is almost as accurate when given only a name, as when shown the target picture itself.

More recent research shows that viewers can detect named pictures in RSVP sequences at above-chance levels at still higher rates, even for durations as short as 13 ms (Potter, Wyble, and McCourt, in preparation). Evidently viewers can extract the conceptual gist of a picture rapidly, retrieving relevant conceptual information about objects and their background from LTM. Having spotted the target picture, viewers can continue to attend to it and consolidate it into working memory – for example, after the sequence they can describe the picnic scene they were looking for, or recognize it in a forced choice task. Yet, when they are not looking for a particular target, viewers forget most pictures presented at that rate almost immediately, as shown in **Figure 2** (Potter and Levy, 1969; see also Intraub, 1980). The rate must be slowed to about 2 pictures/s for viewers to recognize as many as half the pictures as familiar when tested minutes after the sequence.

If a test picture is presented immediately after the sequence, however, viewers are usually able to recognize it, even if the pictures have been presented at a rate such as 6/s (Potter et al., 2004). That is, for a second they will remember most of the presented pictures, but memory drops off rapidly over the first few seconds thereafter, as memory is tested (**Figure 3**). Picture memory includes understanding of the gist of a picture, not just specific visual features, as shown by the ability of viewers to call a picture to mind when given a descriptive title as a recognition cue. Just as a viewer can

FIGURE 3 | Probability (corrected for chance) of recognizing a picture as a function of relative serial position in the test, separately for a group given pictures in the recognition test and one given titles. Pictures were presented at 6/s and tested with a yes–no test of pictures or just of titles. Adapted from Figure 4 in Potter et al. (2004).

FIGURE 4 | When participants has viewed the picture on the left and judged whether they had seen the picture on the right, they made more false yeses than when the new picture was unrelated in meaning to any pictures they had viewed. Adapted from Figure 7 in Potter et al. (2004).

detect a picture that matches a target description given in advance, such as "baby reaching for a butterfly," so can viewers recognize that they saw a picture that matches a name if they are given the name shortly *after* viewing the sequence. Recognition memory is not quite as good when tested by a title instead of showing the picture itself, however, and in both cases performance falls off rapidly (**Figure 3**). Thus, gist can be extracted rapidly, but is quickly forgotten if the presentation was brief and was followed by other stimuli.

Further evidence for the extraction of gist when viewing a rapid sequence of pictures is that viewers are more likely to falsely recognize a new picture with the same gist than a new picture with an unrelated gist. For example, if they saw a picture of a camel, they are more likely to falsely say yes to a very different picture of a camel than to a totally new picture (**Figure 4**), indicating that at some level they knew they had seen a camel (Potter et al., 2004). Other studies with single pictures presented briefly and masked by a following stimulus have shown that objects in the foreground are more easily recognized if they are consistent with the background (and the background is more readily recognized if it is consistent with the foreground), showing that relationships within a single picture are computed during initial recognition (Davenport and Potter, 2004; Davenport, 2007).

Conclusions: Pictures
Consistent with the CSTM hypothesis, the evidence shows that meaningful pictures can be understood extremely rapidly, permitting the detection of targets specified by name (i.e., by meaning)

and at least momentary understanding of the gist of a picture, although there is rapid forgetting if the picture is not selected for further processing.

UNDERSTANDING RSVP SENTENCES
Our ability to read rapidly and continuously, with comprehension and substantial memory for the meaning of what we have read, is a strong indication that we are able to retrieve rapidly a lot of information not only about word meanings but also general knowledge about the world and specific episodic memories. Good readers can read about 300 words/min, or 5 words/s. Eye tracking studies have shown that the length of time that the eyes rest on a given word varies with the frequency of that word, its predictability at that point in the text, and other factors that indicate that the word's meaning and its fit to the context is retrieved fast enough to affect whether the eyes linger on that word or move on (for a review, see Staub and Rayner, 2007).

Instead of measuring eye movements when reading, one can use RSVP to control the time available for processing each word in a sentence. The first to try this was Forster, who presented short sentences at 16 words/s, three times faster than a typical good reader would read spontaneously. He varied the syntactic complexity of the sentences and showed that recall was less accurate for more complex sentences, implying that sentence syntax was processed as sentences were read (Forster, 1970; Holmes and Forster, 1972).

As shown in studies reviewed above, searching for a specific target in an RSVP sequence can be easy even at high rates of presentation, whereas simply remembering all the items in the presentation can be difficult, if they are unrelated. A claim which is central to the CSTM hypothesis, however, is that associative and other structural relations among items can be computed rapidly, assisting in their retention. This section reviews some of the evidence for this claim.

Differences between lists and sentences
A person's memory span is defined as the number of items, such as unrelated words or random digits, that one can repeat back accurately, after hearing or seeing them presented at a rate of about 1/s. For unrelated words a typical memory span is five or six words. The memory span drops, however, as the rate of presentation increases. In one experiment (Potter, 1982) lists of 2, 3, 4, 5, or 6 nouns were presented at rates between 1 and 12 word/s, for immediate recall. The results are shown in **Figure 5**.

FIGURE 5 | Immediate recall of RSVP lists of 2, 3, 4, 5, or 6 nouns presented at rates between 1 and 12 words/s. Adapted from Figure 2 in Potter (1993).
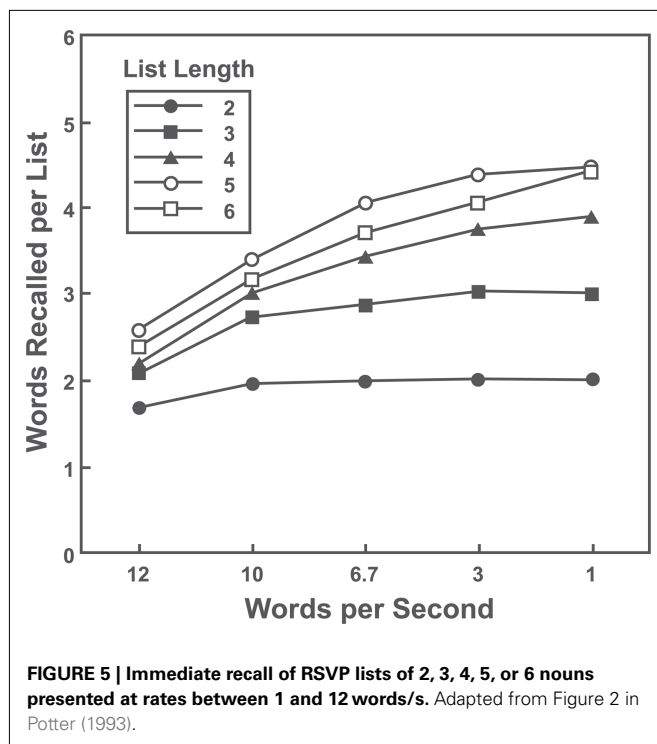
With five words, recall accuracy declined from a mean of 4.5 at the 1 s rate to 2.6 words at the rate of 12 words/s This was evidently not because participants could not recognize each of the words at that rate, because a list of just two words (followed by a mask) was recalled almost perfectly at 12/s: instead, some additional process is necessary to stabilize the words in short term memory. Note that this finding is similar to that for memory of rapidly presented pictures, in that one can detect a picture at a rate that is much higher than the rate required for later memory. In another study I found that the presentation of two related words on a five-word RSVP list (separated by another word) resulted in improved recall for both words, suggesting that the words were both activated to a level at which an association could be retrieved. This hinted at the sort of process that might stabilize or structure information in CSTM.

In contrast to lists, 14-word sentences presented at rates up to at least 12 words/s can be recalled quite accurately (see Potter, 1984; Potter et al., 1986). The findings with sentences versus lists or scrambled sentences strongly support the CSTM assumption that each word can be identified and understood with an 83- to 100-ms exposure, even when it is part of a continuing series of words. (See Potter, 2009, Demo 2, for a demonstration.) The results also support the second assumption that representations of the words remain activated long enough to allow them to be bound into whatever syntactic and conceptual structures are being built on the fly. When, as with a list of unrelated words, there is no ready structure to hand, all but two or three of the words are lost.

How are RSVP sentences recalled? The regeneration hypothesis

Before addressing the question of how rapidly presented sentences are retained, one should address the prior question of why

sentences heard or read at *normal* rates are easy to repeat immediately, even when they are two or three times as long as one's memory span (the length of list that can be repeated accurately). The difference in capacity between lists and sentences is thought to be due to some form of chunking, although it has also been assumed that sentences can be stored in some verbatim form temporarily (see the review by Von Eckardt and Potter, 1985). Before continuing, try reading the sentence below once, cover it, and then read the five words on the next line, look away, and write down the sentence from memory. (We will come back to this exercise shortly.)

The knight rode around the palace looking for a place to enter.

Anchor forest castle oven stocking [look away and write down the sentence].

Instead of assuming that people remember sentences well because they hold them in some verbatim form, we (Potter and Lombardi, 1990) proposed a different hypothesis: immediate recall of a sentence (like longer-term recall) is based on a conceptual or propositional representation of the sentence. The recaller regenerates the sentence using normal speech-production processes to express the propositional structure (what the sentence means). That is, having understood the conceptual proposition in a sentence, one can simply express that idea in words, as one might express a new thought. We proposed that recently activated words were likely to be selected to express the structure. In consequence, the sentence is normally recalled verbatim.

To test this hypothesis we (Potter and Lombardi, 1990) presented distractor words (like the five words below the sentence you read above) in a secondary task immediately before or after the to-be-recalled sentence, and on some trials one of these distractor words was a good substitute for a word in the sentence (such as "castle" for "palace"). As we predicted, that word was frequently intruded in recall, as long as the sentence meaning as a whole was consistent with the substitution. (Did you substitute "castle" for "palace"?) In the experiments, half the participants had lure words like "castle" on the word list, and half did not, allowing us to show that people are more likely to make the substitution when that word has appeared recently. Thus, recall is guided by a conceptual representation, not by a special verbatim representation such as a phonological representation.
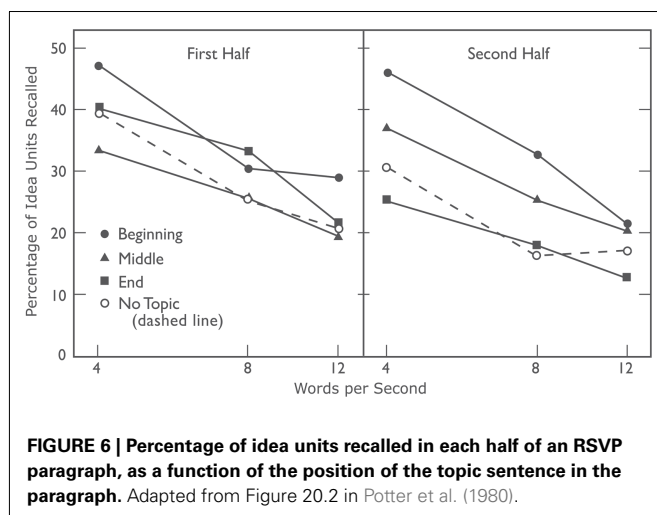
Further studies (Lombardi and Potter, 1992; Potter and Lombardi, 1998) indicated that syntactic priming from having processed the sentence plays a role in the syntactic accuracy of sentence recall. Syntactic priming (e.g., Bock, 1986) is a temporary facilitation in the production of a recently processed syntactic structure, as distinguished from direct memory for the syntactic structure of the prime sentence.

The Potter–Lombardi hypothesis that sentences are comprehended and then regenerated rather than "recalled verbatim" is consistent with the CSTM claim that propositional structures are built rapidly, as a sentence is read or heard. In one of the Potter and Lombardi (1990) experiments the sentences were presented at a rate of 10 words/s, rather than the moderate 5 words/s of our other experiments: the intrusion results were essentially the same, showing that the relevant conceptual processing had occurred at the higher rate, also.

Reading RSVP paragraphs: More evidence for immediate use of structure

A single RSVP sentence is apparently easy to comprehend and recall when presented as fast as 12 words/s, so that recall of a single sentence at that rate is close to ceiling. Does that mean that longer-term retention of the sentence will be as good as if it had been presented more slowly? To answer that question Potter et al. (1980) presented RSVP paragraphs of about 100 words at three rates, 4, 8, and 12 words/s, with the equivalent of a two-word pause between sentences (the net rate averaged 3.3, 6.7, or 10 words/s). (See Demo 3 in Potter, 2009, for a demonstration.) Participants wrote down the paragraph as accurately as possible, immediately after presentation. To evaluate both single-word perception and use of discourse-level information, paragraphs were used that appeared to be ambiguous and poorly integrated unless the reader knew the topic (see Dooling and Lachman, 1971; Bransford and Johnson, 1972). Only one sentence mentioned the one-word topic (e.g., "laundry"), and this sentence appeared either at the beginning, the middle, or the end of the paragraph, or was omitted.

Consistent with the prediction that knowing the topic would enhance recall of the other sentences (and that the topic word itself would be recalled because of its relevance to the paragraph), recall was improved for the part of the paragraph after the topic was presented (but not the part before), at all three rates of presentation. Even at the highest rate the discourse topic could be used to structure the paragraph (**Figure 6**). This suggests that the discourse topic, once it became evident, remained active as a source of structure as the rest of the paragraph was read. (The topic word was perceived and recalled by more than 80% of the subjects regardless of rate or condition.) At the same time, there was a marked main effect of rate: recall declined as rate of presentation increased, from 37% of the idea units at 4 words/s to 26% at 8 words/s to 20% at 12 words/s, averaging over all topic conditions. Clearly, even though there was internal evidence that discourse-level structuring was occurring at all rates of presentation, some process of consolidation was beginning to fail as rate increased.

FIGURE 6 | Percentage of idea units recalled in each half of an RSVP paragraph, as a function of the position of the topic sentence in the paragraph. Adapted from Figure 20.2 in Potter et al. (1980).

Conclusions: words, sentences, and paragraphs.

Putting the paragraph results together with those for word lists and single sentences, we see that structuring can occur rapidly, and that more structure results in better memory (comparing lists to sentences, or comparing a string of seemingly unrelated sentences to sentences structured by having a topic). Nonetheless, rapid conceptual processing is not sufficient for accurate retention if there is no additional time for consolidation: the gist may survive, but details will be lost in immediate recall, just as they are in longer-term memory.

MECHANISMS OF STRUCTURING IN RSVP SENTENCE PROCESSING

Although I have repeatedly invoked the idea that there is rapid structuring of information that is represented in CSTM, I have had little to say about just how this structuring occurs. In the case of sentences, it is evident that parsing and conceptual interpretation must occur virtually word by word, because any substantial delay would outrun the persistence of unstructured material in CSTM. Here I will describe briefly three studies that have investigated the process of selecting an appropriate interpretation of a given word in an RSVP sentence, a key process in comprehension given the extent of lexical ambiguity in English and in most other languages.

The influence of sentence context on word and non-word perception

One study (Potter et al., 1993) took advantage of the propensity of readers to misperceive a briefly presented non-word as an orthographically similar word. Non-words such as *dack* that are one letter away from two other words (*deck*, *duck*) were presented in RSVP sentences biased-toward one or the other of these words or neutral between them, as in the following examples. Note that when we presented a real word in the biased sentences, it was always the mismatching word. The sentence was presented at 10 words/s. Participants recalled the sentence; they were told to report misspelled words or non-words if they saw them.

Neutral: "The visitors noticed the *deck/duck/dack* by the house"
Biased: "The child fed the *deck/dack* at the pond"
"The sailor washed the *duck/dack* of that vessel."

Our main interest is what participants reported when shown a non-word. Non-words were reported as the biased-toward word (here, *duck*) on 40% of the trials, compared with only 12% with the neutral sentence and 3% with the biased-against sentence (the non-words were reported correctly as non-words on 23% of the neutral trials). Similar although smaller effects of context were shown when the biased-against word (rather than a non-word) appeared in the sentence. Thus, context can bias word and non-word perception even when reading at 10 words/s. More surprisingly, even selective context that did not appear until as much as three words (300 ms) after the critical word or non-word influenced perceptual report, suggesting that multiple word candidates (and their meanings) are activated as the non-word or word is perceived, and may remain available for selection for at least 300 ms after the word or non-word has been masked by succeeding words. This supposition that multiple possible words and their meanings are briefly activated during word perception accords with the Swinney (1979) hypothesis that multiple meanings of ambiguous

words are briefly activated: both results are consistent with the CSTM view. In the present study and in the case of ambiguous words the process of activation and selection appears to occur unconsciously for the most part, an issue considered in a later section.

Double-word selection

In another study (Potter et al., 1998) two orthographically distinct words were presented simultaneously (one above and one below a row of asterisks) in the course of an RSVP sentence, as illustrated below. Participants were instructed to select the one that fit into the sentence and include it when immediately recalling the sentence. We regarded this as an overt analog of lexical ambiguity resolution. The sentence was presented for 133 ms/word and the two-word ("double word") display for 83 ms.

```
                                pencil
Maggie carried the kitten in a ****** to her house
                                basket
```

Sentence context had a massive influence on selection: the appropriate word was included in recall in 70% of the sentences and the non-matching word in only 13%. This ability to pick the right word was evident both when the relevant context arrived before the double words and when it arrived later (up to 1 s later, in one experiment), showing that readers could activate and maintain two distinct lexical possibilities. Subjects were asked to report the "other" word (the mismatching word) after they recalled the sentence, but most of the time they were unable to do so, showing that the unselected word was usually forgotten rapidly. Again, this illustrates the existence of fast and powerful processors that can build syntactically and pragmatically appropriate structures from briefly activated material, leaving unselected material to be rapidly forgotten.

Lexical disambiguation

Miyake et al. (1994) carried out two experiments on self-paced reading of sentences with ambiguous words that were not disambiguated until two to seven words after the ambiguous word. They found that readers with low or middling reading spans were slowed down when the disambiguation was toward the subordinate meaning, especially with a delay of seven words. (High-span readers had no problems in any of the conditions.) In an unpublished experiment we presented subjects with a similar set of sentences that included an ambiguous word, using RSVP at 10 words/s; the task was to decide whether or not the sentence was plausible, after which we gave a recognition test of a subset of the words, including the ambiguous word. The sentence was implausible with one of the meanings, plausible with the other. Our hypothesis was that sentences that eventually turned out to require the subordinate meaning of an ambiguous word would sometimes be judged to be implausible, implying that only the dominant reading had been retrieved. Unambiguously implausible and plausible sentences were intermixed with the ambiguous sentences.

Subjects were more likely to judge a plausible sentence to have been implausible (1) when a subordinate meaning of the ambiguous word was required (27 versus 11% errors), (2) when the disambiguating information appeared after a greater delay (23

versus 16% errors), and especially (3) when there was both a subordinate meaning and late disambiguation (32% errors, versus 9% for the dominant/early condition). A mistaken judgment that the sentence was implausible suggests that on those trials only one meaning, the wrong one, was still available at the point of disambiguation. Interestingly, the ambiguous word itself was almost always correctly recognized on a recognition test of a subset of words from the sentence, even when the sentence had mistakenly been judged implausible. The results suggest that although multiple meanings of a word are indeed briefly activated (in CSTM), the less frequent meaning will sometimes fall below threshold within a second, when sentences are presented rapidly.

Conclusion: lexical interpretation and disambiguation

The results of all three studies show that context is used immediately to bias the perception or interpretation of a word, consistent with the CSTM claim that processing to the level of meaning occurs very rapidly in reading.

SELECTIVE SEARCH AND THE ATTENTIONAL BLINK
The attentional blink

In brief, the attentional blink (AB) is a phenomenon that occurs in RSVP search tasks in which two targets are presented among distractors. When the rate of presentation is high but still compatible with accurate report of a single target (e.g., a presentation rate of 10/s, when the task is to detect letters among digit distractors), two targets are also likely to be reported accurately except when the second target appears within 200–500 ms of the onset of the first target. This interval during which second target detection drops dramatically was termed an AB by Raymond et al. (1992).

The AB is relevant to CSTM because it provides evidence for rapid access to categorical information about rapidly presented items and at the same time shows that selective processing of specified targets has a cost. When the task is to pick out targets from among distractors, the experimental findings suggest that there is a difference in time course between two stages of processing, a first stage that results in identification of a stimulus (CSTM) and a second stage required to select and consolidate that information in a reportable form (Chun and Potter, 1995).

Consider a task in which targets are any letter of the alphabet, presented in an RSVP sequence of digit distractors. Presumably a target letter must be identified as a specific letter in order to be classified as a target (see Sperling et al., 1971). At rates as high as 11 items/s the first letter target (T1) is detected quite accurately, consistent with evidence that a letter can be identified in less than 100 ms. This initial identification is termed Stage 1 processing, which constitutes activation of a conceptual but short-lasting representation, i.e., a CSTM representation.

But a second target letter (T2) that arrives soon after the first one is likely to be missed, suggesting that a selected target (T1) requires additional processing beyond identification: Stage 2 processing. Stage 2 processing is necessary to consolidate a selected item into some form of short term memory that is more stable than CSTM. However, Stage 2 processing is serial and limited in capacity. The items following the first target (T1) continue to be processed successfully in Stage 1 and remain for a short time in CSTM; the problem is that as long as Stage 2 is tied up with T1,

a second target may be identified but must wait, and thus may be lost from CSTM before Stage 2 is available. When this happens, T2 is missed, producing an AB. Although the duration of the AB varies, it is strong at 200–300 ms after the onset of the first target, diminishes thereafter, and is usually gone by 500 ms.

Consistent with the CSTM hypothesis, there is both behavioral and ERP evidence that stimuli that are not reported because of an AB are nonetheless momentarily comprehended, because they activate an ERP mismatch marker when they are inconsistent with prior context (Luck et al., 1996). Similarly, word targets that are related in meaning are more accurately detected even when the second word occurs within the time period that produces an AB (e.g., Chun et al., 1994; Maki et al., 1997; Vogel et al., 1998; Potter et al., 2005).

Attention only blinks for selection, not perception or memory

As shown in earlier sections, meaningful items in a continuous stream, such as the words of a sentence, are easy to see and remember, which makes the difficulty of reporting a second target (the AB) surprising. When there is an uninterrupted sequence of targets, as happens when a sentence is presented and recalled as a whole, there is no AB, whereas if the task is to report just the two words in a sentence that are marked by color or by case, as in the following example, there is an AB (Potter et al., 2008).

Our tabby cat CHASED a MOUSE all around the backyard

In a more recent study (Potter et al., 2011), participants did both tasks simultaneously: they reported the sentence and then marked the red (or uppercase) words. In another block they only reported the two target words. An AB for marking or reporting the second target word was observed in both blocks. Surprisingly, the target words were highly likely to be reported as part of the sentence even when the participant could not mark them correctly. What seemed to happen was that the feature (color or case) that defined the target was detected, but in the AB interval that feature was displaced to a different word: the AB interfered with the binding of the target feature to the correct word.

In subsequent experiments the targets (Arabic digits or digit words) were inserted between words of the sentence as additional items, and again participants either reported the digits and then the surrounding sentence, or just reported the digits. The following is an example:

Our 6666 tabby cat 2222 chased a mouse all around the backyard

When participants reported just the digits, they were overall more accurate, but showed an AB for the second digit string or number word. When they reported the digits and then the whole sentence, they did not show an AB for the digits. Evidently the continuous attention associated with processing the sentence included the inserted digits, allowing them to be selected afterward rather than during initial processing. We concluded that on-line, immediate selection generates an AB, whereas continuous processing with delayed selection does not (cf. Wyble et al., 2011).

Summary: CSTM and the attentional blink

Studies of the visual AB demonstrate a dissociation between an early stage of processing sufficient to identify letters or words presented at a rate of about 10/s, and a subsequent stage of variable duration (up to about 400 ms) required to stabilize a selected item in reportable STM. The AB thus provides evidence for the central claims of CSTM.

FURTHER QUESTIONS ABOUT CSTM
HOW DOES STRUCTURING OCCUR IN CSTM?

Structuring in CSTM is not different in principle from individual steps in the slower processes of comprehension that happen as one gradually understands a difficult text or an initially confusing picture, or solves a chess problem over a period of seconds and minutes. But CSTM structuring occurs with a relative absence of awareness that alternatives have been weighed and that several possibilities have been considered and rejected, at least implicitly. As in slower and more conscious pattern recognition and problem solving, a viewer's task set or goal makes a major difference in what happens in CSTM, because one's intentions activate processing routines such as sentence parsing, target specifications in search tasks, and the like. Thus the goal partially determines what enters CSTM and how structuring takes place.

Conceptual short term memory is built on experience and learned skills. Seemingly immediate understanding is more likely for material that is familiar, as becomes evident when one learns a new language or a new skill such as chess. Our ability to understand a new pictured scene in a fraction of a second also depends on our lifetime of visual experience.

Compound cuing and latent semantic analysis

The presence of many activated items at any moment, in CSTM, allows for compound cuing – the convergence of two or more weak associations on an item. The visual system is built on converging information, with learning playing a major role, at least at higher levels in the visual system (e.g., DiCarlo et al., 2012), enabling familiar combinations of features to converge on an interpretation in a single forward pass. The power of converging cues, familiar to any crossword puzzle fan, is likely to be central to structure-building in CSTM. A radical proposal for the acquisition and representation of knowledge, latent semantic analysis (LSA; Landauer and Dumais, 1997), provides a suggestive model for how structure may be extracted from loosely related material. However, there is no syntactic parser in LSA and it is clear from RSVP research that we do parse rapidly presented sentences as we read; thus, the LSA approach is at best a partial model of processing in CSTM.

IS CSTM CONSCIOUS?

The question is difficult to answer, because we have no clear independent criterion for consciousness other than availability for report. And, by hypothesis, report requires some form of consolidation; therefore, only what persists in a structured form will be reportable. Thus, while the evidence we have reviewed demonstrates that there is conceptual processing of material that is subsequently forgotten, it does not tell us whether we were briefly conscious of that material, or whether the activation and selection occurred unconsciously.

It seems unlikely that multiple competing concepts (such as the multiple meanings of a word) that become active simultaneously could all be conscious in the ordinary sense, although

preliminary structures or interpretations that are quickly discarded might be conscious. For example, people do sometimes become aware of having momentarily considered an interpretation of a spoken word that turns out to be mistaken. As noted earlier, most words are ambiguous, yet we are only rarely conscious of multiple meanings (except when someone makes a pun). In viewing rapid pictures, people have a sense of recognizing all the pictures but forgetting most of them. But such experiences in which we are aware of momentary thoughts that were immediately lost seem to be the exception, rather than the rule. Thus, much of CSTM activation, selection, and structuring happens before one becomes aware. It is the structured result, typically, of which one is aware, which is why perception and cognition seem so effortless and accurate.

THE LIMITS OF CSTM

Conceptual short term memory is required to explain the human ability to understand and act rapidly, accurately, and seemingly effortlessly in response to the presentation of richly structured sensory input, drawing on appropriate knowledge from LTM. Working memory as it is generally understood (e.g., Baddeley and Hitch, 2010) comes into play when a first pass in CSTM does not meet one's goal. Then, more conscious thought is required, drawing on working memory together with continued CSTM processing. Systematic reasoning, problem solving, recollection, and planning are slower and more effortful, however. They typically involve a series of steps, each of which sets up the context for the next step. CSTM may be involved in each step.

SUMMARY: RAPID CONCEPTUAL PROCESSING FOLLOWED BY RAPID FORGETTING

In each of the experimental domains discussed – comprehension and retention of RSVP word lists, sentences, and paragraphs; studies of word perception and selection; experiments on picture perception and memory; and the AB – there is evidence for activation of conceptual information about a stimulus early in processing (possibly before conscious awareness), followed by rapid forgetting unless conditions are favorable for retention. The two kinds of favorable conditions examined in these studies were selection for attention (e.g., the first target in the AB procedure, or selection of a target picture from among rapidly presented pictures) and the availability of associations or meaningful relations between momentarily active items (as in sentence and paragraph comprehension and in perception of the gist of a picture). The power of these two factors – selective attention that is defined by conceptual properties of the target, and the presence of potential conceptual structure – is felt early in processing, before conventional STM or working memory for the stimuli has been established, justifying the claim that CSTM is separate from those forms of working memory.

Outside of the laboratory, we usually have control over the rate of presentation: we normally read at a rate sufficient not only for momentary comprehension, but also for memory of at least the gist of what we are reading. Although we cannot control the rate of input from radio, TV, movies, or computer games, producers are adept at adjusting the rate to fit the conceptual and memory requirements of their audience. Similarly, in a conversation

the speaker adjusts his or her rate of speech to accommodate the listener's signals of comprehension. Rapid structuring can only occur if the material permits it and if the skills for discovering latent structure are highly practiced: object and word recognition, lexical retrieval, sentence parsing, causal inference, search for a target, and the like. These cognitive skills, built up over a lifetime, make comprehension seem trivially easy most of the time.

Conceptual short term memory is the working memory that supports these processes, lasting just long enough to allow multiple interpretation to be considered before one is selected and the remaining elements evaporate, in most cases without entering awareness. The labored thoughts and decisions we are aware of pondering are a tiny fraction of those we make effortlessly. Even these worked-over thoughts may advance by recycling the data through CSTM until the next step comes to mind. We may be aware of slowly shaping an idea or solving a problem, but not of precisely how each step occurs. More work will be needed to gain a full understanding of this largely preconscious stage of cognitive processing.

IMPLICATIONS FOR THE RELATION BETWEEN PERCEPTION AND COGNITION

In the present account, perception is continuous with cognition. Information passes from the sense organs to the brain, undergoing transformations at every stage, combining with information from other senses, evoking memories, leading to conscious experience and actions shaped by one's goals. These events are continually renewed or replaced by new experiences, thoughts, and actions. From the moment the pattern of a falling cup appears on the retina (to take one example) to the moment when one reaches to catch it, a series of events has passed through the visual system and to the motor system by way of the conceptual and goal-directed systems, at every stage influenced by prior experience already represented in these systems. It is to some extent a matter of convention that we break up such an event into perceptual, cognitive, goal-directed, and motor parts, when in reality these parts are not only continuous but also interact. In this example, perception of the tipping cup combines with conceptual knowledge to elicit the goal and action almost simultaneously, and the action anticipates the subsequent perceptual sequence. There is no clear separation between mental/brain activity originating from outside the observer ("perception") and that from the observer's internally generated thoughts and memories ("cognition"); any experience is likely to be a blend of these sources. Thinking with the eyes closed in a dark, silent room might seem to come close to a "pure" cognitive experience, but our propensity to augment our thoughts with visual, auditory, and other sense images suggests that the perceptual system is ubiquitous in cognition. Whether such self-generated perceptual images are the heart of cognition (as embodiment theories suggest) or only play a supporting role, is in dispute, however. I keep Helen Keller in mind as an apparent counterexample to the claim that cognition is entirely embodied in perception and action: Keller could neither see nor hear, but those who knew her as an adult had no doubt about her cognitive abilities and her command not only of language but also of world knowledge. Was Keller a kinesthetic/proprioceptive

zombie who simply simulated human understanding, or did she have an intact capacity for cognitive abstraction, once an access route for language was established? The latter seems more likely.

Authorization for the use of experimental animals or human subjects: The experimental studies described here were approved by the Internal Review Board of Massachusetts Institute of Technology, and all participants signed consent forms.

REFERENCES

Baddeley, A. (1986). *Working Memory.* Oxford: Clarendon.

Baddeley, A. (2000). The episodic buffer: a new component of working memory? *Trends Cogn. Sci. (Regul. Ed.)* 4, 417–423.

Baddeley, A. (2007). *Working Memory, Thought, and Action.* USA: Oxford University Press.

Baddeley, A., and Hitch, G. J. (2010). Working memory. *Scholarpedia* 5, 3015.

Bock, J. K. (1986). Syntactic persistence in language production. *Cogn. Psychol.* 18, 355–387.

Bransford, J. D., and Johnson, M. K. (1972). Contextual prerequisites for understanding: some investigations of comprehension and recall. *J. Verbal Learn. Verbal Behav.* 11, 717–726.

Budiu, R., and Anderson, J. R. (2004). Interpretation-based processing: a unified theory of semantic sentence processing. *Cogn. Sci.* 28, 1–44.

Chun, M. M., Bromberg, H. S., and Potter, M. C. (1994). "Conceptual similarity between targets and distractors in the attentional blink," *Poster presented at the 35th Annual Meeting of the Psychonomic Society*, St. Louis, MO.

Chun, M. M., and Potter, M. C. (1995). A two-stage model for multiple target detection in rapid serial visual presentation. *J. Exp. Psychol. Hum. Percept. Perform.* 21, 109–127.

Coltheart, M. (1983). Iconic memory. *Philos. Trans. R. Soc. Lond. B* 302, 283–294.

Davenport, J. L. (2007). Contextual effects in object and scene processing. *Mem. Cognit.* 35, 393–401.

Davenport, J. L., and Potter, M. C. (2004). Scene consistency in object and background perception. *Psychol. Sci.* 15, 559–564.

Davenport, J. L., and Potter, M. C. (2005). The locus of semantic priming in RSVP target search. *Mem. Cognit.* 33, 241–248.

DiCarlo, J. J., Zoccolan, D., and Rust, N. C. (2012). How does the brain solve visual object recognition? *Neuron* 73, 415–434.

Dooling, D. J., and Lachman, R. (1971). Effects of comprehension on retention of prose. *J. Exp. Psychol.* 88, 216–222.

Ericsson, K. A., and Kintsch, W. (1995). Long-term working memory. *Psychol. Rev.* 102, 211–245.

Forster, K. I. (1970). Visual perception of rapidly presented word sequences of varying complexity. *Percept. Psychophys.* 8, 215–221.

Forster, K. I., and Davis, C. (1984). Repetition priming and frequency attenuation in lexical access. *J. Exp. Psychol. Learn. Mem. Cogn.* 10, 680–698.

Holmes, V. M., and Forster, K. I. (1972). Perceptual complexity and underlying sentence structure. *J. Verbal Learn. Verbal Behav.* 11. 148–156.

Intraub, H. (1980). Presentation rate and the representation of briefly glimpsed pictures in memory. *J. Exp. Psychol. Hum. Learn. Mem.* 6, 1–12.

Intraub, H. (1981). Rapid conceptual identification of sequentially presented pictures. *J. Exp. Psychol. Hum. Percept. Perform.* 7, 604–610.

Intraub, H. (1984). Conceptual masking: the effects of subsequent visual events on memory for pictures. *J. Exp. Psychol. Learn. Mem. Cogn.* 10, 115–125.

Just, M. A., and Carpenter, P. A. (1992). A capacity theory of comprehension: individual differences in working memory. *Psychol. Rev.* 99, 122–149.

Kintsch, W. (1988). The role of knowledge in discourse comprehension: a construction-integration model. *Psychol. Rev.* 95, 163–183.

Kutas, M., and Hillyard, S. A. (1980). Reading senseless sentences: brain potentials reflect semantic incongruity. *Science* 207, 203–205.

Landauer, T. K., and Dumais, S. T. (1997). A solution to Plato's problem: the latent semantic analysis theory of acquisition, induction and representation of knowledge. *Psychol. Rev.* 104, 211–240.

Lawrence, D. H. (1971a). Temporal numerosity estimates for word lists. *Percept. Psychophys.* 10, 75–78.

Lawrence, D. H. (1971b). Two studies of visual search for word targets with controlled rates of presentation. *Percept. Psychophys.* 10, 85–89.

Loftus, G. R. (1983). "Eye fixations on text and scenes," in *Eye Movements in Reading: Perceptual and Language Processes*, ed. K. Rayner (New York: Academic Press), 359–376.

Loftus, G. R., and Ginn, M. (1984). Perceptual and conceptual masking of pictures. *J. Exp. Psychol. Learn. Mem. Cogn.* 10, 435–441.

Loftus, G. R., Hanna, A. M., and Lester, L. (1988). Conceptual masking: how one picture captures attention from another picture. *Cogn. Psychol.* 20, 237–282.

Lombardi, L., and Potter, M. C. (1992). The regeneration of syntax in short term memory. *J. Mem. Lang.* 31, 713–733.

Luck, S. J., and Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature* 390, 279–281.

Luck, S. J., Vogel, E. K., and Shapiro, K. L. (1996). Word meanings can be accessed but not reported during the attentional blink. *Nature* 383, 616–618.

Maki, W. S., Frigen, K., and Paulson, K. (1997). Associative priming by targets and distractors during rapid serial visual presentation: does word meaning survive the attentional blink? *J. Exp. Psychol. Hum. Percept. Perform.* 23, 1014–1034.

Meng, M., and Potter, M. C. (2011). An attentional blink for nontargets? *Atten. Percept. Psychophys.* 73, 440–446.

Miyake, A., Just, M. A., and Carpenter, P. A. (1994). Working memory constraints on the resolution of lexical ambiguity: maintaining multiple interpretations in neutral contexts. *J. Mem. Lang.* 33, 175–202.

Neely, J. H. (1991). "Semantic priming effects in visual word recognition: a selective review of current findings and theories," in *Basic Processes in Reading: Visual Word Recognition*, eds D. Besner and G. W. Humphreys (Hillsdale, NJ: Earbaum), 264–336.

Phillips, W. A. (1983). Short-term visual memory. *Philos. Trans. R. Soc. Lon. B* 302, 295–309.

Potter, M. C. (1976). Short-term conceptual memory for pictures. *J. Exp. Psychol. Hum. Learn. Mem.* 2, 509–522.

Potter, M. C. (1982). Very short-term memory: in one eye and out the other. *Paper presented at the 23rd Annual Meeting of the Psychonomic Society*, Minneapolis.

Potter, M. C. (1984). "Rapid serial visual presentation (RSVP): a method for studying language processing," in *New Methods in Reading Comprehension Research*, eds D. Kieras and M. Just (Hillsdale, NJ: Erlbaum), 91–118.

Potter, M. C. (1993). Very short-term conceptual memory. *Mem. Cogn.* 21, 156–161.

Potter, M. C. (1999). "Understanding sentences and scenes: the role of conceptual short term memory," in *Fleeting Memories: Cognition of Brief Visual Stimuli*, ed. V. Coltheart (Cambridge, MA: MIT Press), 13–46.

Potter, M. C. (2009). Conceptual short term memory. *Scholarpedia* 5, 3334.

Potter, M. C., Dell'Acqua, R., Pesciarelli, F., Job, R., Peressotti, F., and O'Connor, D. H. (2005). Bidirectional semantic priming in the attentional blink. *Psychon. Bull. Rev.* 12, 460–465.

Potter, M. C., and Jiang, Y. V. (2009). "Visual short-term memory," in *Oxford Companion to Consciousness*, eds T. Bayne, A. Cleeremans, and P. Wilken (Oxford: Oxford University Press), 436–438.

Potter, M. C., Kroll, J. F., and Harris, C. (1980). "Comprehension and memory in rapid sequential reading," in *Attention and Performance VIII*, ed. R. Nickerson (Hillsdale, NJ: Erlbaum), 395–418.

Potter, M. C., Kroll, J. F., Yachzel, B., Carpenter, E., and Sherman, J. (1986). Pictures in sentences: understanding without words. *J. Exp. Psychol. Gen.* 115, 281–294.

Potter, M. C., and Levy, E. I. (1969). Recognition memory for a rapid sequence of pictures. *J. Exp. Psychol.* 81, 10–15.

Potter, M. C., and Lombardi, L. (1990). Regeneration in the short-term recall of sentences. *J. Mem. Lang.* 29, 633–654.

Potter, M. C., and Lombardi, L. (1998). Syntactic priming in immediate recall of sentences. *J. Mem. Lang.* 38, 265–282.

Potter, M. C., Moryadas, A., Abrams, I., and Noel, A. (1993). Word perception and misperception in context. *J. Exp. Psychol. Learn. Mem. Cogn.* 19, 3–22.

Potter, M. C., Nieuwenstein, M. R., and Strohminger, N. (2008). Whole report versus partial report in RSVP sentences. *J. Mem. Lang.* 58, 907–915.

Potter, M. C., Staub, A., and O'Connor, D. H. (2002). The time course of competition for attention: attention is initially labile. *J. Exp. Psychol. Hum. Percept. Perform.* 28, 1149–1162.

Potter, M. C., Staub, A., and O'Connor, D. H. (2004). Pictorial and conceptual representation of glimpsed pictures. *J. Exp. Psychol. Hum. Percept. Perform.* 30, 478–489.

Potter, M. C., Stiefbold, D., and Moryadas, A. (1998). Word selection in reading sentences: preceding versus following contexts. *J. Exp. Psychol. Learn. Mem. Cogn.* 24, 68–100.

Potter, M. C., Wyble, B., and Olejarczyk, J. (2011). Attention blinks for selection, not perception or memory: reading sentences and reporting targets. *J. Exp. Psychol. Hum. Percept. Perform.* 37, 1915–1923.

Potter, M. C., Wyble, B., Pandav, R., and Olejarczyk, J. (2010). Picture detection in RSVP: features or identity? *J. Exp. Psychol. Hum. Percept. Perform.* 36, 1486–1494.

Raymond, J. E., Shapiro, K. L., and Arnell, K. M. (1992). Temporary suppression of visual processing in an RSVP task: an attentional blink? *J. Exp. Psychol. Hum. Percept. Perform.* 18, 849–860.

Rayner, K. (ed.). (1983). *Eye Movements in Reading: Perceptual, and Language Processes.* New York: Academic Press.

Rayner, K. (ed.). (1992). *Eye Movements and Visual Cognition: Scene Perception and Reading.* New York: Springer-Verlag.

Sereno, S. C., and Rayner, K. (1992). Fast priming during eye fixations in reading. *J. Exp. Psychol. Hum. Percept. Perform.* 18, 173–184.

Sperling, G. (1960). The information available in brief visual presentations. *Psychol. Monogr. Gen. Appl.* 74. 1–29.

Sperling, G., Budiansky, J., Spivak, J. G., and Johnson, M. C. (1971). Extremely rapid visual search: the maximum rate of scanning letters for the presence of a numeral. *Science* 174, 307–311.

Staub, A., and Rayner, K. (2007). "Eye movements and on-line comprehension processes," in *The Oxford Handbook of Psycholinguistics*, ed. G. Gaskell (Oxford: Oxford University Press), 327–342.

Swinney, D. A. (1979). Lexical access during sentence comprehension: (re)consideration of context effects. *J. Verbal Learn. Verbal Behav.* 18, 645–659.

Verhoeven, L., and Perfetti, C. (2008). Advances in text comprehension: model, process and development. *Appl. Cogn. Psychol.* 22, 293–301.

Vogel, E. K., Luck, S. J., and Shapiro, K. L. (1998). Electrophysiological evidence for a post-perceptual locus of suppression during the attentional blink. *J. Exp. Psychol. Hum. Percept. Perform.* 24, 1656–1674.

Von Eckardt, B., and Potter, M. C. (1985). Clauses and the semantic representation of words. *Mem. Cognit.* 13, 371–376.

Wyble, B., Potter, M., Bowman, H., and Nieuwenstein, M. (2011). Attentional episodes in visual perception, *J. Exp. Psychol. Gen.* 140, 488–505.

Active inference, attention, and motor preparation

Harriet Brown[1]*, Karl Friston[1] and Sven Bestmann[2]

[1] The Wellcome Trust Centre for Neuroimaging, University College London, Queen Square, London, UK
[2] Sobell Department of Motor Neuroscience and Movement Disorders, University College London Institute of Neurology, Queen Square, London, UK

Perception is the foundation of cognition and is fundamental to our beliefs and consequent action planning. The Editorial (this issue) asks: "what mechanisms, if any, mediate between perceptual and cognitive processes?" It has recently been argued that attention might furnish such a mechanism. In this paper, we pursue the idea that action planning (motor preparation) is an attentional phenomenon directed toward kinesthetic signals. This rests on a view of motor control as active inference, where predictions of proprioceptive signals are fulfilled by peripheral motor reflexes. If valid, active inference suggests that attention should not be limited to the optimal biasing of perceptual signals in the exteroceptive (e.g., visual) domain but should also bias proprioceptive signals during movement. Here, we investigate this idea using a classical attention (Posner) paradigm cast in a motor setting. Specially, we looked for decreases in reaction times when movements were preceded by valid relative to invalid cues. Furthermore, we addressed the hierarchical level at which putative attentional effects were expressed by independently cueing the nature of the movement and the hand used to execute it. We found a significant interaction between the validity of movement and effector cues on reaction times. This suggests that attentional bias might be mediated at a low level in the motor hierarchy, in an intrinsic frame of reference. This finding is consistent with attentional enabling of top-down predictions of proprioceptive input and may rely upon the same synaptic mechanisms that mediate directed spatial attention in the visual system.

Keywords: priming, motor preparation, action selection, attention, precision, free energy, active inference

INTRODUCTION

During the preparation and execution of goal-directed movements, processing is biased toward the perceptual attributes of the goal (e.g., Baldauf and Deubel, 2010; Gherri and Eimer, 2010; Humphreys et al., 2010; Perfetti et al., 2010) and preparation or execution of an action improves perceptual processing in relevant sensory domains (Fagioli et al., 2007). This suggests motor planning and attention are inherently linked, such that "perceptual codes and action plans share a common representational medium, which presumably involves the human premotor cortex" (Fagioli et al., 2007). This relates to the concept of *motor attention* that is specific to the effectors employed (Rushworth et al., 2001) and decision making through attentional selection among motor plans (Goldberg and Segraves, 1987). Moreover, the *premotor theory of visual attention* (Rizzolatti et al., 1994) proposes that distinct maps are tuned to different effector representations and become active when a movement is prepared. In short, attention has a fundamental role in the selection and control of action; see Allport (1987) for a review.

The link between action and attention and was first proposed by James (1890) and Woodworth (1899): however, the cognitive and neural mechanisms responsible for this association remain largely unknown (Dalrymple and Kingstone, 2010). Greenwald (1970) provided evidence that attention to a particular sensory modality speeded movements that are detected in that modality: In the oculomotor system, visual discrimination performance is enhanced at the target location of a prepared saccade (Deubel and Schneider, 1996). Furthermore, stimulation of the superior colliculus can produce both eye movements (Robinson, 1972) and shifts of attention (Müller et al., 2005). Conversely, Craighero et al. (1999) showed that reaction times to visually presented objects are reduced when subjects grasp the objects being presented, illustrating the motor facilitation of sensory processing.

In this paper, we entertain the idea that motor attention uses exactly the same synaptic mechanisms as visual attention. This may sound strange because motor commands are usually considered to be outputs, whereas the visual channels selected by attention are inputs. However, recent theoretical treatments of motor control (*active inference*) regard action as being driven by proprioceptive prediction errors in exactly the same way that perception is driven by exteroceptive prediction errors (Friston et al., 2010). If true, this means that attentional modulation may operate at low levels in the motor system in the same way that it operates in the early visual system. We sought evidence for this by reproducing a classical visual attention paradigm (Posner et al., 1978; Posner, 1980) in the motor domain. Furthermore, by cueing attention to different attributes of movements we tried to locate the putative attentional modulation within the motor hierarchy. We hoped to show that attentional effects were expressed in low levels (in an intrinsic frame of reference) in much the same way that directed spatial attention operates in the early visual pathways. This paper comprises four sections. The first rehearses the theoretical background that motivated a

reaction time study described in the second section. The third section presents our results, which are discussed in relation to theoretical considerations in the final section.

ACTIVE INFERENCE AND MOTOR ATTENTION

In this section, we consider motor preparation as attention that is directed toward predicted *proprioceptive* sensations (Galazky et al., 2009), as opposed to the predicted *exteroceptive* consequences of action. This idea is motivated by the success of a recent computational model of attention in explaining reaction times benefits in visual detection tasks (Feldman and Friston, 2010). In this model, the effects of orienting cues on reaction times were explained by the Bayes-optimal encoding of precision in a hierarchical message-passing scheme (predictive coding). In this context, precision is the inverse variance or uncertainty associated with particular sensory channels, such that attention can be understood as weighting sensory signals in proportion to their precision (Feldman and Friston, 2010; Friston, 2010). In these predictive coding schemes, precision is encoded by the gain of units reporting bottom-up sensory information that has yet to be explained by top-down predictions. This sensory information is called *prediction error* and is generally associated with the activity of superficial pyramidal cells: these cells are the source of forward or bottom-up projections in the brain (Rockland and Pandya, 1979; Mumford, 1992; Friston, 2010). In these schemes, attention therefore reduces to the optimization of the postsynaptic gain of superficial pyramidal cells, of the sort associated with gamma-synchronization (e.g., Womelsdorf et al., 2006) and monoaminergic or cholinergic modulation (e.g., Herrero et al., 2008); both of which have been implicated in attention. Here, we pursue the notion that attention is the optimum weighting of prediction error in the context of action preparation (Mars et al., 2007; Bestmann et al., 2008). In short, we consider attention to boost the gain of proprioceptive channels during motor preparation, in the same way that attention selects particular visual channels when subjects prepare for a visual target. In what follows, we will briefly review predictive coding and active inference, with a special focus on the role of attention.

PREDICTIVE CODING AND ACTIVE INFERENCE

Predictive coding is based on the assumption that the brain makes inferences about the causes of its sensations. These inferences are driven by bottom-up or forward sensory information that is passed to higher brain areas in the form of prediction errors (Rao and Ballard, 1999; Friston and Kiebel, 2009). Top-down or backward connections convey predictions that try to suppress prediction errors until predictions are optimized and prediction error is minimized. This suppression rests on opposing excitatory and inhibitory effects of top-down predictions and bottom-up inputs on prediction-error units (usually considered to be superficial pyramidal cells: Mumford, 1992). Active inference (Friston et al., 2010) generalizes this scheme and proposes that exactly the same recursive message-passing operates in the motor system. The only difference is that prediction errors at the lowest level (in the cranial nerve nuclei and spinal cord) are also suppressed by movement, through classical reflex arcs. In this view, descending (cortico-spinal) signals are not motor commands *per se* but predictions of proprioceptive signals that the peripheral motor system fulfills

(see Friston et al., 2010, 2011 for details). As illustrated in **Figure 1**, a cued movement is not regarded as a simple stimulus–response mapping but is generated by a high-level (sensorimotor) percept that predicts a particular pattern of proprioceptive and exteroceptive sensory signals. This percept arises to explain prediction errors caused by a cue in the exteroceptive domain, while motor reflexes suppress the ensuing prediction errors in the proprioceptive domain. This framework has been used to explain several features of the motor system and a series of behaviors, from visual tracking (Friston et al., 2010) to action observation (Friston et al., 2011). Active inference formalizes much of what is proposed by the ideomotor theory of action (Lotze, 1852; James, 1890). The ideomotor account of motor control posits that moving causes a bidirectional association to be formed between a movement and its perceptual consequences. Learning this association allows the perceptual consequences of a movement to be predicted, and anticipating the sensory consequences of a movement can be used to select an action. At the level of the stretch receptors, the similarity is clear: signaling the predicted sensory consequences of an action (under active inference) causes the action to occur. At higher hierarchical levels, movements can still be initiated in order to change the sensory input in another sensory system; indeed the free-energy principle demands the sampling of predicted information to minimize free energy or, more simply, surprise. See **Figure 1** for a schematic illustration.

ATTENTION AND ACTIVE INFERENCE

Attention enters this picture through context or state-dependent optimization of the precision of prediction errors. This sort of prediction is about the second-order statistics of sensory signals (i.e., their variability or reliability). In predictive coding, top-down first-order predictions drive (or inhibit) neurons reporting prediction errors; while contextual, second-order predictions optimize their postsynaptic gain. It is this sort of top-down effect that is associated with attention. Neurobiologically, the distinction between first and second-order predictions can be related to the distinction between the driving and modulatory effects mediated by AMPA and NMDA receptors. Optimizing postsynaptic gain ensures that sensory information (prediction error) is weighted in proportion to its precision. This may sound complicated but is exactly the same procedure we use every day in statistics, when weighting a difference in means (prediction error under the null hypothesis) by SE (inverse precision) to form a t-statistic. Precision can thus be regarded as representing the reliability, ambiguity, or uncertainty about sensory signals. In summary, top-down predictions can have a direct (first-order) or a modulatory (second-order) effect on the responses of prediction-error units that make the ensuing predictions as efficient as possible. Reaction time (Goodman and Kelso, 1980), cortico-spinal excitability (Mars et al., 2007; Bestmann et al., 2008), and EEG data (Osman et al., 1995; Mars et al., 2008) all confirm that the motor system is highly sensitive to such second-order effects.

If ascending sensory signals are prediction errors and descending motor commands are predictions, then optimal predictions (and the resulting movements) should depend on optimizing precision in exactly the same way as in sensory processing. This suggests that, in the motor domain, cueing has a similar effect
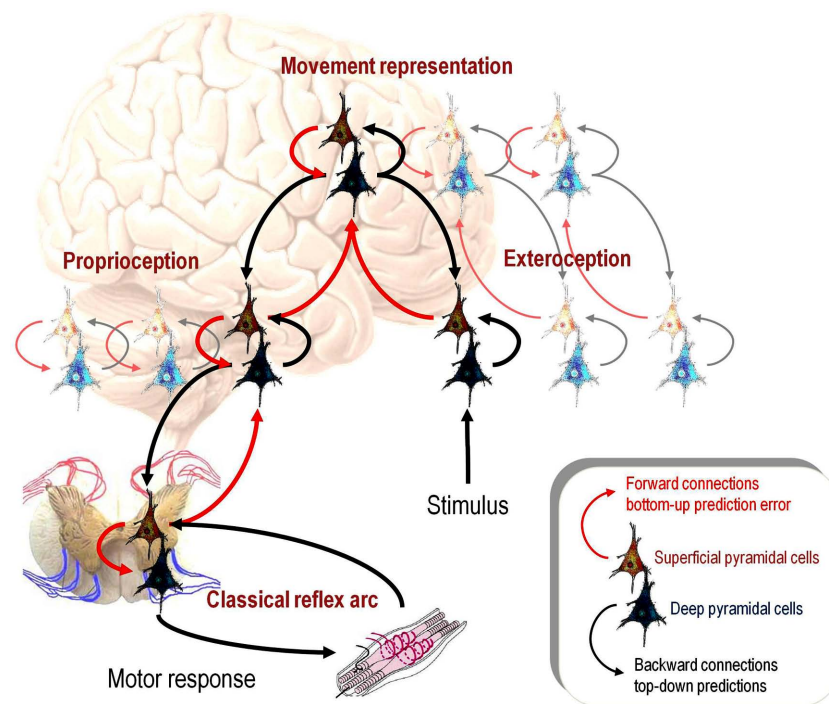
FIGURE 1 | Active inference and predictive coding: Active inference is a generalization of predictive coding that covers motor behaviors and itself is a special instance of the principle of free-energy minimization. Free energy is a statistical quantity that bounds the surprise (self-information) associated with sensory signals. This surprise is quantified in relation to a generative model of how those signals were caused. Predictive coding uses prediction error as a proxy for free energy (cf, surprise) and rests on a hierarchical model, in which prediction errors are passed up the hierarchy (red arrows) to optimize high-level representations that provide top-down predictions (black arrows). In this schematic, prediction-error units are portrayed in red and units encoding the conditional expectations of the hidden causes of sensory input are shown in blue. During perception, the best explanation for sensory input emerges when the top-down predictions can explain as much of the prediction error (at each hierarchical level) as possible. Active inference takes this one step further and notes that certain sensory modalities can use prediction errors to drive motoneurons to eliminate prediction error directly (through classical motor reflex arcs). This is shown schematically on the lower left, using units in the dorsal and ventral horns of the spinal cord. Under active inference, a movement just fulfills the predictions afforded by percepts that predict both exteroceptive (e.g., visual) and interoceptive (e.g., stretch receptor) consequences. This high-level (sensorimotor) percept is activated by an exteroceptive (sensory) cue and the ensuing top-down predictions propagate to both sensory cortex (to suppress exteroceptive prediction error) and the motor system. However, in the motor system, the predictions engender a proprioceptive prediction error that is eliminated by movement. In this schematic, we have assumed that prediction errors are reported by superficial pyramidal cells (Mumford, 1992), while conditional representations are encoded by (top-down) projecting deep pyramidal cells. Darker units highlight those activated by the presentation of a target-stimulus.

to that observed in the sensory domain: Rosenbaum (1980) first demonstrated an effect of movement cueing on reaction time in a way that is analogous to the accelerated detection of visual targets when they are preceded by valid cues in the Posner paradigm (Posner, 1980). However the movements cued in Rosenbaum (1980) were button presses, which required either visual or somatosensory attention to guide movement to the target. Thus, these non-proprioceptive aspects of button presses conflate attentional effects in visual, somatosensory, and proprioceptive domains. In other words, in previous work movements were planned in relation to an object in extra-personal space. Here, we used a simpler paradigm in which movements (wrist flexion and extension) could be performed using only proprioceptive information. This ensured that any attentional effects could be attributed to proprioception. Our motor analog of the Posner paradigm therefore allowed us to interpret our results in relation to visual attention as modeled in Feldman and Friston (2010); and to illustrate how active inference provides a framework in which to address

questions about the functional anatomy of action preparation and attention.

CUEING IN AN EXTRINSIC OR INTRINSIC FRAME OF REFERENCE?

A key question in the functional anatomy of motor attention is where biasing effects are located in the cortical hierarchy: see Grafton and Hamilton (2007) for a review of motor hierarchies. In the sensory domain, attention is usually considered to operate at the lower levels of sensory hierarchies to select among competing sensory processing channels. This is seen in both psychological (e.g., the distinction between object and spatial visual attention: Treisman, 1998; Macaluso et al., 2003) and electrophysiological treatments (e.g., biased competition models: Desimone and Duncan, 1995). If the functional anatomy of the motor hierarchy recapitulates that of sensory hierarchies, then one might expect to see attentional modulation in lower levels, which we will associate with representations in an intrinsic frame of reference.

Electrophysiological evidence demonstrates that between the ventral premotor cortex and M1 neurons change their response patterns from signaling movements in a visual (*extrinsic*) coordinate system that is independent of starting posture to a motor (*intrinsic*) coordinate system that depends on starting posture (Kakei et al., 1999, 2001, 2003). Thus in ventral premotor cortex, actions are largely encoded allocentrically, while in M1 they are predominantly encoded in terms of the joint angles and proprioceptive input required to reach the target (Soechting and Flanders, 1992). Shipp (2005) suggests that neurons representing movements in an intrinsic frame of reference send descending cortico-spinal predictions from M1. Kakei et al. (2003) provide a detailed discussion of movement representations in terms of the coordinate transformations that begin with an "extrinsic coordinate frame representing the spatial location of a target and end with an intrinsic coordinate frame describing muscle activation patterns." It should be noted however, that the segregation of intrinsic and extrinsic representations between motor and premotor cortex may not be complete or unique (Wu and Hatsopoulos, 2007).

These observations suggest two possible levels of the motor hierarchy at which attentional cueing effects could operate. Consider movements with two dimensions or attributes that are cued in an extrinsic frame of reference; for example, moving the left or right hand (*where*) inward or outward (*what*). If attention operates at high levels of the motor hierarchy, then one might expect cues to move the hand inward will facilitate inward movements, irrespective of which hand is used. This is because the representation of the movement can be primed in extrinsic coordinates, prior to transformation to intrinsic coordinates. Conversely, if attention operates at lower levels, encoding the muscle groups involved in inward movements of the left hand, then attentional priming will only be expressed when the left hand is moved inward. In short, if attention operates on prediction errors in an intrinsic frame of reference, the effect of the *what* cue will depend upon the *where* cue.

In summary, if sensorimotor constructs mediate attentional biases in an extrinsic frame of reference, we would expect to see cueing effects on both dimensions independently. Conversely, if these representations instantiate top-down biases at a lower (intrinsic) level of the motor system, then only a particular movement (in an intrinsic frame of reference) will be cued. **Figure 2** tries to make the different predictions clear in terms of top-down enabling of postsynaptic gain (indicated with blue arrows). Crucially, the profile of speeded responses (under valid and invalid cueing) is different for extrinsic and intrinsic levels of attentional gain. In the intrinsic (motor cortex) model, there should be an interaction between the validity effects of cues over both movement dimensions. Conversely, under the extrinsic (premotor cortex) model, there should be no interaction but two main effects due to the validity of both *what* and *where* aspects of the cue. It was this difference in the profile of validity effects on reaction times our experiment was designed to reveal.

Based on the results of Jentzsch et al. (2004) and the retinotopic frame of reference of attentional effects in the Posner paradigm (Woldorff et al., 1997), we hypothesize that attentional cueing

operates in an intrinsic frame of reference. We therefore expected to see an interaction between the validity effects of cueing, with speeded responses when, and only when, both *what* and *where* dimensions were valid.
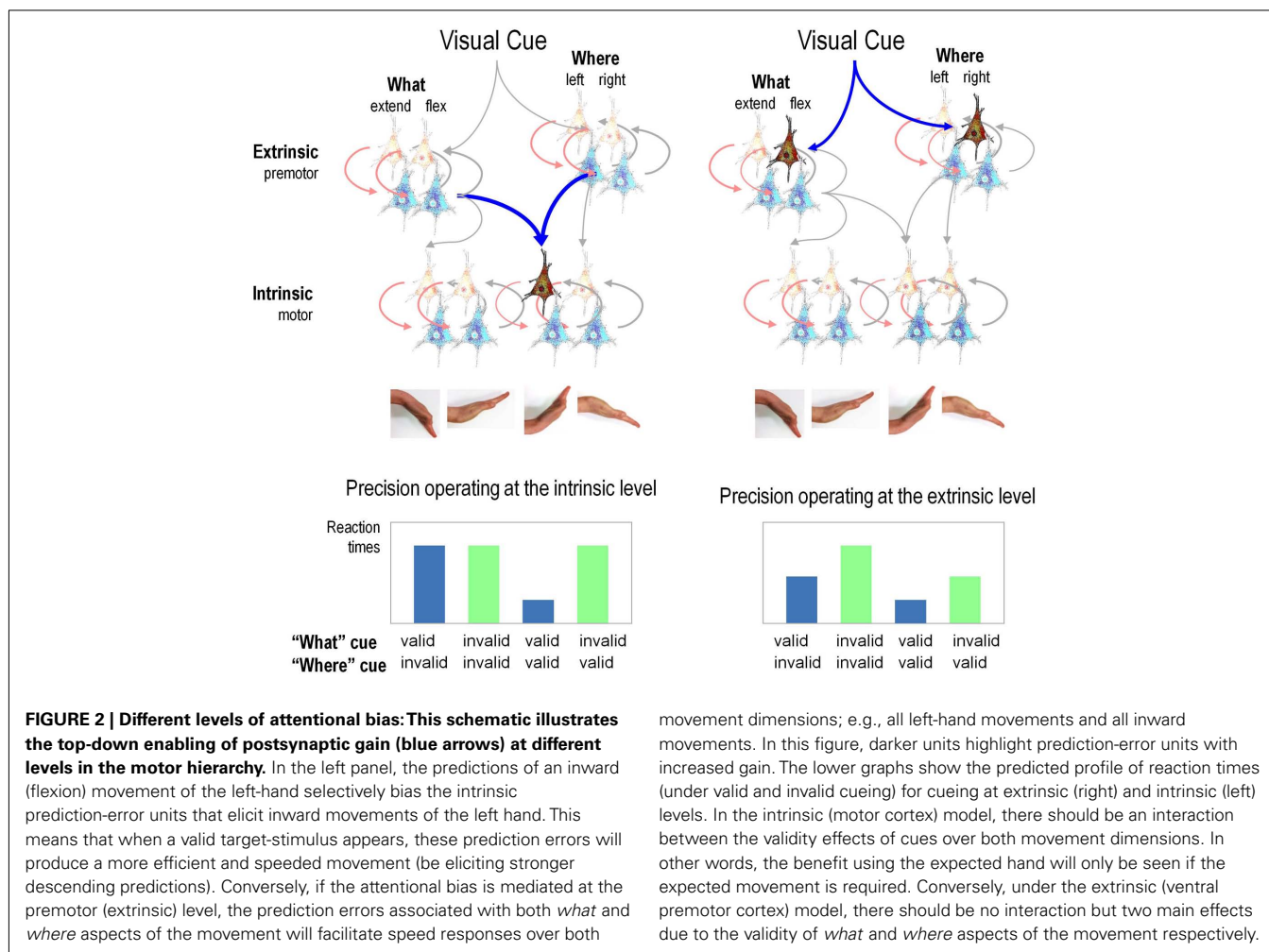
MATERIALS AND METHODS
SUBJECTS
Eight healthy right-handed volunteers (two female), aged 19–42, participated in this experiment. All subjects provided written and informed consent and the experiments were conducted in compliance with the standards established by the local ethical committee.

EXPERIMENTAL PROCEDURE AND EMG RECORDINGS
Subjects were seated in a comfortable reclining chair. Their wrists were in a semi-supine position with the palms facing each other and supported by a splint that restricted wrist and hand movement to pure flexion and extension. The hand-splints were mounted on vertical spindles, which allowed rotation in the transverse plane. The hands were positioned such that the wrist joints sat directly above the axes of rotation. Additional support of the forearms further ensured that movements were constrained to the wrists, and reduced fatigue. Stimuli were viewed on a screen placed at eye level. Each trial started with a (150 ms) cue stimulus, followed by a blank screen (see **Figure 3**). Seven hundred millisecond after the appearance of the cue, a target-stimulus appeared for 400 ms. A 50-ms white-noise mask was presented after the cue and target stimuli to prevent the appearance of visual after-effects. Participants were given 1000 ms after the appearance of the target-stimulus to make a response. No feedback was given. At the appearance of the target-stimulus, participants were required to respond as quickly as possible with the movement indicated. Four movements were possible – flexion or extension at the left or right wrist. The cue and target stimuli had two dimensions – color (blue, red) and spatial frequency (high, low). For four of the participants, the color of the stimulus cued the hand (e.g., blue = left, red = right) and the spatial frequency indicated the movement (e.g., high frequency = flexion, low frequency = extension). For the remaining four, the stimulus–response mapping was reversed, so that color indicated the movement to be made and spatial frequency the hand to be used. The stimuli subtended approximately 35° of visual angle. High-frequency stimuli were 2.5 c/deg, low frequency were 0.25 c/deg. The colors had RGB values ([128 0 0] [255 100 100]) and ([0 0 128] [100 100 255]).

Participants were required to relax their hands and lower arms until the appearance of the target-stimulus. Our paradigm independently cued which motor and (right or left) would implement one of two movements (wrist flexion or extension). Each cue contained two dimensions – one signaling the hand to be moved and one the movement. For each dimension (color, spatial frequency), cue stimuli could be valid (80%) or invalid with regards to the target-stimulus (20%). Since the validity of the cue in each dimension was independent, this gave 64% (0.8×0.8) of trials with a completely valid cue, 32% ($0.8 \times 0.2 \times 2$) of trials where either the hand or the movement required was invalidly cued and 4% (0.2×0.2) of trials where both the hand and movement were cued invalidly. The experiment comprised one training session

FIGURE 2 | Different levels of attentional bias: This schematic illustrates the top-down enabling of postsynaptic gain (blue arrows) at different levels in the motor hierarchy. In the left panel, the predictions of an inward (flexion) movement of the left-hand selectively bias the intrinsic prediction-error units that elicit inward movements of the left hand. This means that when a valid target-stimulus appears, these prediction errors will produce a more efficient and speeded movement (be eliciting stronger descending predictions). Conversely, if the attentional bias is mediated at the premotor (extrinsic) level, the prediction errors associated with both *what* and *where* aspects of the movement will facilitate speed responses over both movement dimensions; e.g., all left-hand movements and all inward movements. In this figure, darker units highlight prediction-error units with increased gain. The lower graphs show the predicted profile of reaction times (under valid and invalid cueing) for cueing at extrinsic (right) and intrinsic (left) levels. In the intrinsic (motor cortex) model, there should be an interaction between the validity effects of cues over both movement dimensions. In other words, the benefit using the expected hand will only be seen if the expected movement is required. Conversely, under the extrinsic (ventral premotor cortex) model, there should be no interaction but two main effects due to the validity of *what* and *where* aspects of the movement respectively.

and 25 experimental sessions. Each session contained 100 trials, which were balanced for the four types of cue and four movements. The large number of trials was needed to acquire sufficient data from trials with invalid cues in both dimensions. The sessions were conducted over three separate days.

Reaction times were evaluated using surface EMG. Ag/AgCl electrodes were placed on the left and right brachioradialis/extensor carpi radialis longus and flexor carpi ulnaris muscles. Muscle activity was monitored throughout the experiment to ensure the effector muscles were relaxed before the appearance of the target-stimulus. Signals were recorded via a CED 1401 laboratory interface (Cambridge Electronic Design Ltd., Cambridge, UK) and stored on a personal computer (for later analysis) at a sample rate of 5 kHz (Signal 2.0, Cambridge Electronic Design Ltd.). Data were bandpass-filtered between 3 Hz and 2.5 kHz.

STATISTICAL ANALYSIS

EMG data were smoothed with a Butterworth low-pass filter with a cutoff frequency of 600 Hz to increase signal-to-noise. After full-wave rectification the data were log-transformed to provide normally distributed time series for further analysis. The mean of 100 consecutive data points was compared with the mean of the preceding 5000 data points, using two-sample t-tests and a sliding window. Reaction times were defined operationally as the first time at which the absolute value of the t-statistic exceeded 50. This *ad hoc* threshold identified the highest number of correctly performed trials. Incorrect trials, where a muscle other than the agonist for the correct movement showed the shortest reaction time, were excluded.

A standard summary statistic method was used for statistical inference, using the log of the mean reaction times (to correct for positive skew) over each of the four conditions, for each subject. Univariate five-way ANOVA was performed in SPSS, with factors HAND CUE VALIDITY (valid vs. invalid), MOVEMENT CUE VALIDITY (valid vs. invalid), HAND (left vs. right), MOVEMENT (flexion vs. extension). Factors SUBJECT and STIMULUS–RESPONSE MAPPING were nested and were implemented in two separate ANOVA models.

RESULTS

Thirteen percentage of trials (range over subjects 8–22%) were discarded. Of these trials, in 2% no movement was made or no movement could be identified. In the remaining 11%, an incorrect movement was made (error trials). Error trial frequency varied significantly by cue type ($p < 0.001$, $\chi^2 > 400$, 1 d.f.),
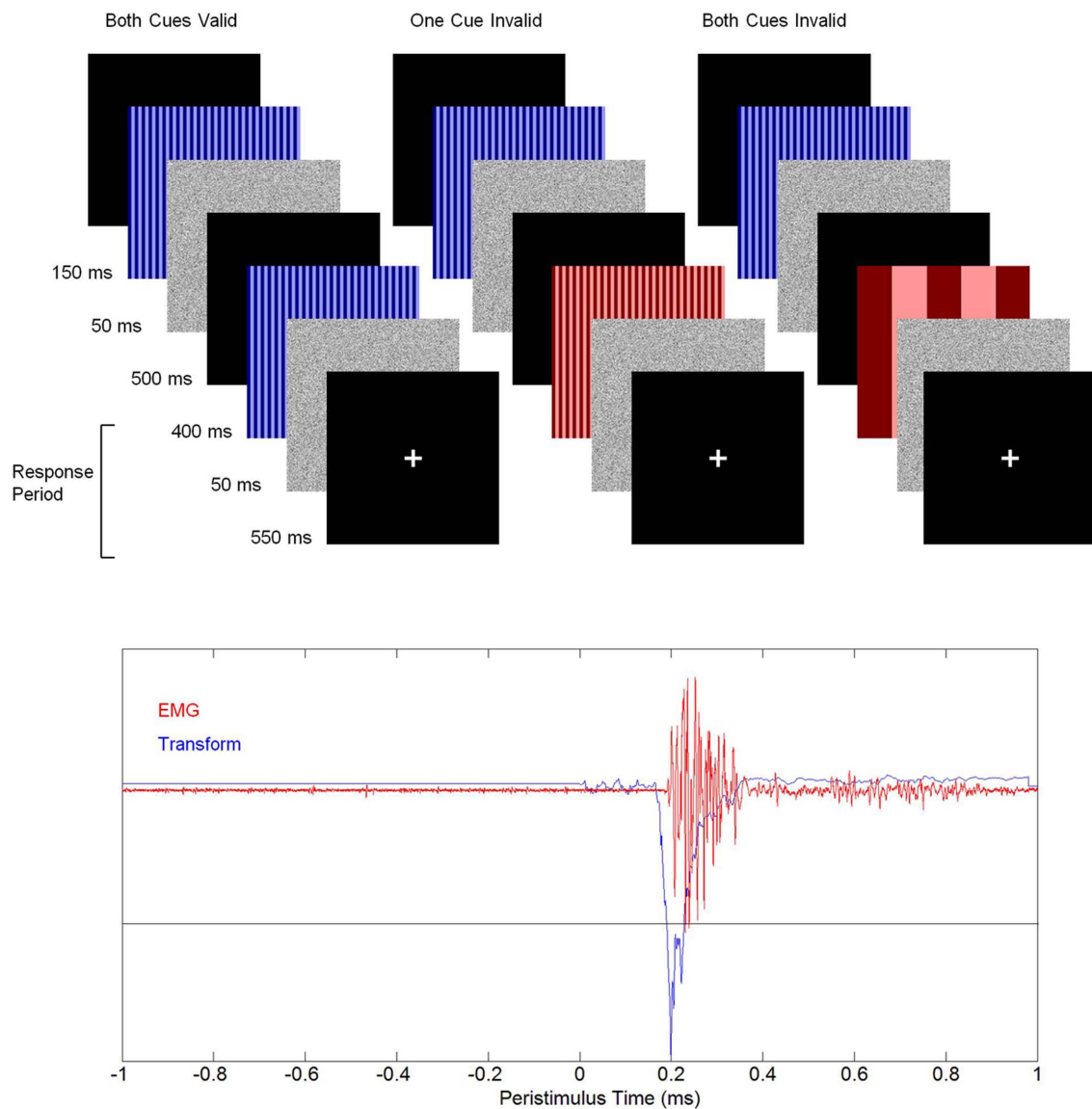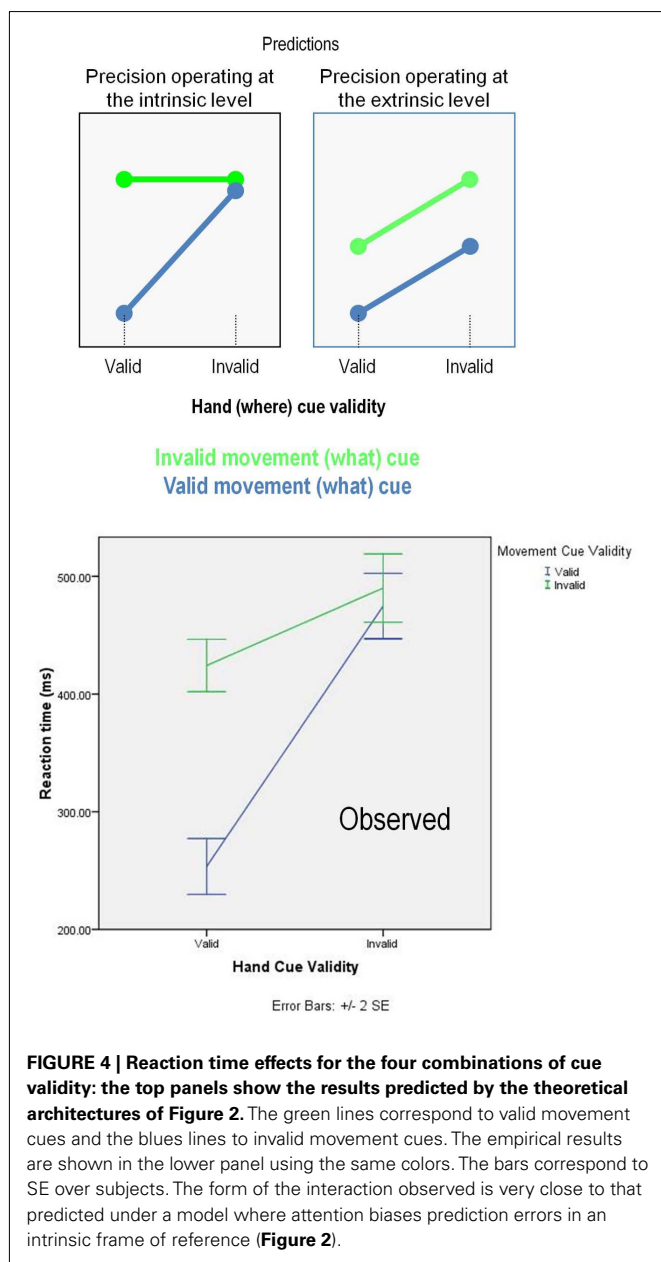
FIGURE 3 | Experimental Design: Top panel: Schematic showing the time-line of three experimental trials, which comprised cue stimuli that could be congruent (valid) or incongruent (invalid) over each of their two dimensions (what:extension vs. flexion; where:left vs. right hand). Bottom panel: Example EMG trace acquired from a single muscle, plotted with the transform used for identifying movement onset. The line shows the *ad hoc* threshold used to derive reaction times automatically.

with errors less likely on validly cued trials. The most common error (64% of errors) was making the incorrect movement with the correct hand. The least common error (10% of errors) was making the correct movement with the wrong hand. Among invalidly cued trials, performing the movement specified by the cue stimulus rather than the target-stimulus occurred significantly more often ($p < 0.05$, $\chi^2 > 6.01$, 1 d.f.). Since the EMG measured the onset of movement rather than the endpoint, changing the response before the movement was completed resulted in an error trial. This may explain the comparatively high error rate seen here, compared with more traditional button-press paradigms.

The grand average reaction time was 334 ms. There was no significant main effect of HAND, MOVEMENT, or STIMULUS–RESPONSE MAPPING, so the ANOVA model including SUBJECT as a factor was used for further analysis. There were significant main effects of HAND CUE VALIDITY [$F_{(1,7)} = 90.54$, $p < 0.001$, partial $\eta^2 = 0.928$], MOVEMENT CUE VALIDITY [$F_{(1,7)} = 171.12$, $p < 0.001$, partial $\eta^2 = 0.961$, $\eta^2 = 0.155$], and SUBJECT [$F_{(1,7)} = 9.29$, $p < 0.003$, partial $\eta^2 = 0.797$]. There were two significant two-way interactions – MOVEMENT × MOVEMENT CUE VALIDITY [$F_{(1,7)} = 4.98$, $p = 0.048$, partial $\eta^2 = 0.449$], and, as anticipated, MOVEMENT CUE VALIDITY × HAND CUE VALIDITY [**Figure 4**; $F_{(1,7)} = 233.34$,

FIGURE 4 | Reaction time effects for the four combinations of cue validity: the top panels show the results predicted by the theoretical architectures of Figure 2. The green lines correspond to valid movement cues and the blues lines to invalid movement cues. The empirical results are shown in the lower panel using the same colors. The bars correspond to SE over subjects. The form of the interaction observed is very close to that predicted under a model where attention biases prediction errors in an intrinsic frame of reference (**Figure 2**).

$p < 0.001$, partial $\eta^2 = 0.971$]. As expected, the fastest mean reaction time was seen when both cues were valid (see **Table 1**). **Figure 4** highlights the nature of this interaction with reference to the profiles predicted by high (extrinsic) and low (intrinsic) levels of facilitation in the motor hierarchy. It is clear that this profile is consistent with attentional bias at the (motor cortex) level of representation, in an intrinsic frame of reference. Quantitatively, these results suggests that the validity effect is expressed primarily when both cue dimensions were jointly valid.

Paired t-tests among the four validity categories confirmed that only one pair failed to show a significant difference (after Bonferroni correction): movement cue valid, hand cue invalid, and movement cue invalid, hand cue invalid ($p > 0.2$). All other pairwise differences were highly significant ($p < 0.001$).

Table 1 | Mean reaction time in milliseconds for each cue validity condition

		Hand cue	
		Valid	Invalid
Movement cue	Valid	253.5 (SE: 11.7)	474.7 (SE: 13.7)
	Invalid	424.2 (SE: 10.9)	490.0 (SE: 14.3)

DISCUSSION

We have pursued the idea that attention is an integral part of motor control and expresses itself through biasing the precision afforded to the proprioceptive and somatosensory consequences of an anticipated action (Galazky et al., 2009). This places previous proposals that link motor preparation and attention (cf, Allport, 1987; Goldberg and Segraves, 1987; Rizzolatti et al., 1994; Rushworth et al., 2001; Humphreys et al., 2010; see Tipper, 2004 for an overview) in the general framework of active inference and predictive coding. The important perspective provided by active inference is that movements fulfill predictions furnished by percepts with both exteroceptive (e.g., visual) and proprioceptive (e.g., stretch receptor) components.

We have previously demonstrated that the reaction time benefits of cueing can be understood as statistically optimal responses, where the associated optimization of precision can account for both psychophysical and electrophysiological phenomena fairly accurately (Feldman and Friston, 2010). In this paper, we asked whether similar reaction time benefits can be seen empirically in the motor domain. To this end, we adapted the paradigm developed by Rosenbaum (1980), in which two different visual dimensions (color and spatial frequency) cued the impending movement. As in Rosenbaum and Kornblum (1982), we predicted and confirmed that cueing effects would occur only when both cue dimensions were valid. Our predictions were based on the possible outcomes of attentional bias at different levels in the cortical hierarchy; which we associate with representations in extrinsic (higher) and intrinsic (lower) frames of reference: In an extrinsic model, one would predict that cueing effects enact their influence independently and to a comparable degree. As outlined above, the interaction between the two validity factors argues for an intrinsic model, in which hand and movement are selectively enabled in a way that cannot be separated. In the present case, the observed interaction can be accounted for by a model where precision is increased in proprioceptive channels that represent the confluence of top-down predictions about the nature of a movement and where it will be implemented (see **Figure 2**).

In addition to the interaction above, there was a small reaction time benefit from a valid hand cue, even if the movement cue was invalid. The magnitude of this effect was much smaller compared to the reaction time benefit seen for two valid cues (66 vs. 237 ms). This, and the lack of any benefit for a valid movement cue if the hand cue is invalid, means that a model in which precision operates at the intrinsic level is still the most likely. The small validity effect of a valid hand cue might be explained in the framework of active inference; because the movements performed in this experiment

were self-limited, the same muscles were recruited for both flexion and extension movements, to either initiate or terminate the movement. Thus, if the precision of the stretch receptor channels in one forearm were boosted after cuing that side, a small benefit might accrue for the opposite movement.

Rushworth et al. (1997) also demonstrated a benefit for valid cuing using a similar paradigm. Spatial cues were used, and the motor preparation time was calculated from the difference between two conditions: a simple cuing task in one movement dimension, and a control task where the movement made did not depend on the validity of the cue. A small reaction time benefit was seen for valid cues.

In Rosenbaum (1980), some aspects of the movement were left unspecified until the appearance of the target-stimulus. Unlike our study, Rosenbaum saw separable effects of cuing just the arm, the direction and the extent of the upcoming movement. However, there is a key difference between our paradigm and that of Rosenbaum (1980) that may account for the difference. The button-press responses used in Rosenbaum (1980) entail visuomotor and somatosensory–motor integration. This means that attentional cueing effects in the visual or somatosensory domains cannot be disambiguated from purely proprioceptive attention. Our paradigm avoided conflating multiple attention processes by cueing movements that could be performed using only proprioceptive channels (simple, self-terminated flexion, and extension movements). This means that one can attribute the cue validity effects to attentional modulation of proprioceptive signals, in accordance with active inference. Furthermore, Rosenbaum's cues were semantic (letters), whereas ours used low-level visual features which were arbitrarily mapped onto flexion and extension movements. The complexity of the semantic cues meant that most of the reaction time advantages seen in Rosenbaum (1980) could be accounted for by validity effects on processing visual targets and their semantic content and not on the movements *per se.* In short, the simplicity of our movements and cues suggests a motoric rather than sensory locus for attentional cueing.

A further study (Rosenbaum and Kornblum, 1982), which resembled ours except that only two of four possible movements were possible in each trial, did not find that correctly cuing one response attribute benefited reaction time. They found the opposite – violating the hand and movement cues increased reaction times relative to violating just the movement cue. Their explanation for this was that both movements were simultaneously prepared, but choosing between two movements on the same hand takes longer because the movements are more "similar." The larger number of possible movements in our experiment meant that simultaneously preparing all responses was unlikely (our flexion and extension movements used the same motor plant, while index and middle finger movements were used in Rosenbaum and Kornblum, 1982). By contrast, Miller (1982) found a contradictory effect – advance information of which hand to use gave a reaction time advantage, whereas advance information of which finger (on either hand) did not.

How can these discrepancies be resolved? Cui and Deecke (1999) found anatomically congruent movements were performed faster than spatially congruent movements, suggesting

that anatomically congruent movements are prepared together in the motor hierarchy, or, alternatively that the mapping from extrinsic to intrinsic coordinates is more efficient. Despite the anatomical distance between [pre]motor cortex in each hemisphere, activity in these areas may be influenced at an early stage during motor preparation. If left and right effectors are competing alternatives for subsequent actions (cf. Cisek and Kalaska, 2010), several (bilateral) representations can in principle occur in an intrinsic frame of reference at the same time. Our results suggest that predictions about impending movements are integrated to boost processing in effector-based (intrinsic) coordinates.

Goodman and Kelso (1980) suggested that stimulus–response mapping time is shorter for cued movements. If this were the case, we would expect cues correct in one response that mention to provide some reaction time benefit for the other. The locus of such an effect would likely be before the motor stage; i.e., early in the stimulus–response interval. However, evidence from EEG studies suggests that the effects of cueing occur relatively late, again suggesting an effect in intrinsic coordinates: for example, the lateralized readiness potential (LRP), an EEG potential evoked when one hand is cued, has been suggested to be the halfway point between premotor and motor processing (Osman et al., 1995). This is supported by the finding that it occurs nearer to the movement during trials with informative cues than those without, although the stimulus–LRP latency does not change (Jentzsch et al., 2004). Finally, we note that a locus of the motor attentional effect in intrinsic coordinates provides an interesting parallel with results from the Posner paradigm. The reaction time benefit associated with cues in most visual paradigms seems to occur in retinotopic (intrinsic) rather than allocentric (extrinsic) frames of reference (Posner and Cohen, 1984; Golomb et al., 2008).

CONCLUSION

We have explored the idea that motor preparation is an attentional phenomenon that is directed toward proprioceptive sensations (i.e., predicted sensory feedback of the anticipated motor response). This perspective suggests that attention should not be limited to perceptual processing in the exteroceptive (e.g., visual) domain but should also bias interoceptive inference during movement. We verified this prediction by adapting a classical attention (Posner) paradigm for a motor setting. Furthermore, we tried to establish the hierarchical level this attentional bias operates by cueing the movement and effector independently. Our behavioral results demonstrate an interaction between the validity of movement and effector cues. This suggests that the bias for the selected action is mediated at a low level in the motor hierarchy, in an intrinsic frame of reference. More generally, the ideas outlined above provide a heuristic framework in which to address questions about the link between motor preparation and attention, and their mechanistic underpinnings.

REFERENCES

Allport, D. A. (1987). "Selection for action: some behavioral and neurophysiological considerations of attention and action," in *Perspectives on Perception and Action,* eds H. Heuer and A. F. Sanders (Hillsdale, NJ: Erlbaum), 395–419.

Baldauf, D., and Deubel, H. (2010). Attentional landscapes in reaching and grasping. *Vision Res.* 50, 999–1013.

Bestmann, S., Harrison, L. M., Blankenburg, F., Mars, R. B., Haggard, P., Friston, K. J., and Rothwell, J. C. (2008). Influence of uncertainty and surprise on human corticospinal excitability during preparation for action. *Curr. Biol.* 18, 775–780.

Cisek, P., and Kalaska, J. F. (2010). Neural mechanisms for interacting with a world full of action choices. *Annu. Rev. Neurosci.* 33, 269–298.

Craighero, L., Fadiga, L., Rizzolatti, G., and Umilta, C. (1999). Action for perception: a motor-visual attentional effect. *J. Exp. Psychol. Hum. Percept. Perform.* 25, 1673–1692.

Cui, R.-Q., and Deecke, L. (1999). High resolution DC EEG of the Bereitschaftspotential preceding anatomically congruent versus spatially congruent bimanual finger movements. *Brain Topogr.* 12, 117–127.

Dalrymple, K. A., and Kingstone, A. (2010). Time to act and attend to the real mechanisms of action and attention. *Br. J. Psychol.* 101, 213–216.

Desimone, R., and Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annu. Rev. Neurosci.* 18, 193–222.

Deubel, H., and Schneider, W. X. (1996). Saccade target selection and object recognition: evidence for a common attentional mechanism. *Vision Res.* 36, 1827–1837.

Fagioli, S., Hommel, B., and Schubotz, R. I. (2007). Intentional control of attention: action planning primes action-related stimulus dimensions. *Psychol. Res.* 71, 22–29.

Feldman, H., and Friston, K. J. (2010). Attention, uncertainty, and free-energy. *Front. Hum. Neurosci.* 4:215. doi: 10.3389/fnhum.2010.00215

Friston, K. (2010). The free-energy principle: a unified brain theory? *Nat. Rev. Neurosci.* 11, 127–138.

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

Friston, K., Mattout, J., and Kilner, J. (2011). Action understanding and active inference. *Biol. Cybern.* 104, 137–160.

Friston, K. J., Daunizeau, J., Kilner, J., and Kiebel, S. J. (2010). Action and behavior: a free-energy formulation. *Biol. Cybern.* 102, 227–260.

Galazky, I., Schütze, H., Noesselt, T., Hopf, J. M., Heinze, H. J., and Schoenfeld, M. A. (2009). Attention to somatosensory events is directly linked to the preparation for action. *J. Neurol. Sci.* 279, 93–98.

Gherri, E., and Eimer, M. (2010). Manual response preparation disrupts spatial attention: an electrophysiological investigation of links between action and attention. *Neuropsychologia* 48, 961–969.

Goldberg, M. E., and Segraves, M. A. (1987). Visuospatial and motor attention in the monkey. *Neuropsychologia* 25, 107–118.

Golomb, J. D., Chun, M. M., and Mazer, J. A. (2008). The native coordinate system of spatial attention is retinotopic. *J. Neurosci.* 28, 10654–10662.

Goodman, D., and Kelso, J. A. S. (1980). Are movements prepared in parts? Not under compatible (naturalized) conditions. *J. Exp. Psychol. Gen.* 109, 475–495.

Grafton, S. T., and Hamilton, A. F. (2007). Evidence for a distributed hierarchy of action representation in the brain. *Hum. Mov. Sci.* 26, 590–616.

Greenwald, A. (1970). A double stimulation test of ideomotor theory with implications for selective attention. *J. Exp. Psychol.* 84, 352–398.

Herrero, J. L., Roberts, M. J., Delicato, L. S., Gieselmann, M. A., Dayan, P., and Thiele, A. (2008). Acetylcholine contributes through muscarinic receptors to attentional modulation in V1. *Nature* 454, 1110–1114.

Humphreys, G. W., Yoon, E. Y., Kumar, S., Lestou, V., Kitadono, K., Roberts, K. L., and Riddoch, J. M. (2010). Attention and its coupling to action. *Br. J. Psychol.* 101, 217–219.

James, W. (1890). *Principles of Psychology,* Vol. 2. New York: Holt.

Jentzsch, I., Leuthold, H., and Ridderinkhof, K. R. (2004). Beneficial effects of ambiguous precues: parallel motor preparation or reduced premotoric processing time? *Psychophysiology* 41, 231–244.

Kakei, S., Hoffman, D. S., and Strick, P. L. (1999). Muscle and movement representations in the primary motor cortex. *Science* 285, 2136–2139.

Kakei, S., Hoffman, D. S., and Strick, P. L. (2001). Direction of action is represented in the ventral premotor cortex. *Nat. Neurosci.* 4, 1020–1025.

Kakei, S., Hoffman, D. S., and Strick, P. L. (2003). Sensorimotor transformations in cortical motor areas. *Neurosci. Res.* 46, 1–10.

Lotze, R. H. (1852). *Medizinische Psychologie oder die Physiologie der Seele [Medical psychology or physiology of the soul].* Leipzig: Weidmann'sche Buchhandlung.

Macaluso, E., Eimer, M., Frith, C. D., and Driver, J. (2003). Preparatory states in crossmodal spatial attention: spatial specificity and possible control mechanisms. *Exp. Brain Res.* 149, 62–74.

Mars, R. B., Bestmann, S., Rothwell, J. C., and Haggard, P. (2007). Effects of motor preparation and spatial attention on corticospinal excitability in a delayed-response paradigm. *Exp. Brain Res.* 182, 125–129.

Mars, R. B., Coles, M. G., Hulstijn, W., and Toni, I. (2008). Delay-related cerebral activity and motor preparation. *Cortex* 44, 507–520.

Miller, J. (1982). Discrete versus continuous stage models of human information processing: in search of partial output. *J. Exp. Psychol. Hum. Percept. Perform.* 8, 273–296.

Müller, J. R., Philiastides, M. G., and Newsome, W. T. (2005). Microstimulation of the superior colliculus focuses attention without moving the eyes. *Proc. Natl. Acad. Sci. U.S.A.* 102, 524–529.

Mumford, D. (1992). On the computational architecture of the neocortex. II. The role of cortico-cortical loops. *Biol. Cybern.* 66, 241–251.

Osman, A., Moore, C., and Ulrich, R. (1995). Bisecting RT with lateralized readiness potentials: precue effects after LRP onset. *Acta Psychol. (Amst)* 90, 111–127.

Perfetti, B., Moisello, C., Lanzafame, S., Varanese, S., Landsness, E. C., Onofrj, M., Di Rocco, A., Tononi, G., and Ghilardi, M. F. (2010). Attention modulation regulates both motor and non-motor performance: a high-density EEG study in Parkinson's disease. *Arch. Ital. Biol.* 148, 279–288.

Posner, M. I. (1980). Orienting of attention. *Q. J. Exp. Psychol.* 32, 3–25.

Posner, M. I., and Cohen, Y. (1984). "Components of visual orienting," in *Attention and Performance*, Vol. X, eds H. Bouma and D. Bouwhuis (Hillsdale, NJ: Erlbaum), 531–556.

Posner, M. I., Nissen, M. J., Ogden, W. C. (1978). "Attended and unattended processing modes: the role of set for spatial location," in *Modes of Perceiving and Processing Information,* eds H. L. Pick and M. J. Saltzman (Hillsdale, NJ: Lawrence Erlbaum Associates), 137–157.

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

Rizzolatti, G., Riggio, L., and Sheliga, B. M. (1994). "Space and selective attention," in *Attention and Performance XV,* eds C. Umilt and M. Moscovitch (Cambridge, MA: MIT Press), 231–265.

Robinson, D. A. (1972). Eye movements evoked by collicular stimulation in the alert monkey. *Vision Res.* 12, 1795–1808.

Rockland, K. S., and Pandya, D. N. (1979). Laminar origins and terminations of cortical connections of the occipital lobe in the rhesus monkey. *Brain Res.* 179, 3–20.

Rosenbaum, D. A. (1980). Human movement initiation: specification of arm, direction, and extent. *J. Exp. Psychol. Gen.* 109, 444–474.

Rosenbaum, D. A., and Kornblum, S. (1982). A priming method for investigating the selection of motor responses. *Acta Psychol. (Amst)* 223–243.

Rushworth, M. F., Ellison, A., and Walsh, V. (2001). Complementary localization and lateralizatio of orienting and motor attention. *Nat. Neurosci.* 4, 656–661.

Rushworth, M. F., Nixon, P. D., Renowden, S., Wade, D. T., and Passingham, R. E. (1997). The left parietal and motor attention. *Neuropsychologia* 35, 1261–1273.

Soechting, J. F., and Flanders, M. (1992). Moving in three-dimensional space—frames of reference, vectors, and coordinate systems. *Annu. Rev. Neurosci.* 15, 167–191.

Shipp, S. (2005). The importance of being agranular: a comparative account of visual and motor cortex.

Philos. Trans. R. Soc. Lond. B Biol. Sci. 360, 797–814.

Tipper, S. P. (2004). "Attention and action," in *The Cognitive Neurosciences,* 3rd Edn, ed. M. S. Gazzaniga (Cambridge, MA: MIT Press), 619–630.

Treisman, A. (1998). Feature binding, attention and object perception. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 353, 1295–1306.

Woldorff, M., Fox, P., Matzke, M., Lancaster, J., Veeraswamy, J., Zamarripa, F., Seabolt, M., Glass, T., Gao, J., Martin, C., and Jerabek, P. (1997).

Retinotopic organization of the early visual spatial attention effects as revealed by PET and ERPs. *Hum. Brain Mapp.* 5, 280–286.

Womelsdorf, T., Fries, P., Mitra, P. P., and Desimone, R. (2006). Gamma-band synchronization in visual cortex predicts speed of change detection. *Nature* 439, 733–736.

Woodworth, R. S. (1899). The accuracy of voluntary movement. *Psychol. Rev.* 3, 1–119.

Wu, W., and Hatsopoulos, N. G. (2007). Coordinate system representations of movement direction in the

premotor cortex. *Exp. Brain Res.* 176, 652–657.

Improving perception to make distant connections closer

Robert L. Goldstone[1]*, David Landy[2] and Lionel C. Brunel[3]

[1] Department of Psychological and Brain Sciences, Indiana University, Bloomington, IN, USA
[2] Department of Psychology, University of Richmond, Richmond, VA, USA
[3] Department of Psychology, Université Paul-Valéry Montpellier III, Montpellier, France

One of the challenges for perceptually grounded accounts of high-level cognition is to explain how people make connections and draw inferences between situations that superficially have little in common. Evidence suggests that people draw these connections even without having explicit, verbalizable knowledge of their bases. Instead, the connections are based on sub-symbolic representations that are grounded in perception, action, and space. One reason why people are able to spontaneously see relations between situations that initially appear to be unrelated is that their eventual perceptions are not restricted to initial appearances. Training and strategic deployment allow our perceptual processes to deliver outputs that would have otherwise required abstract or formal reasoning. Even without people having any privileged access to the internal operations of perceptual modules, these modules can be systematically altered so as to better serve our high-level reasoning needs. Moreover, perceptually based processes can be altered in a number of ways to closely approximate formally sanctioned computations. To be concrete about mechanisms of perceptual change, we present 21 illustrations of ways in which we alter, adjust, and augment our perceptual systems with the intention of having them better satisfy our needs.

Keywords: perceptual learning, cognitive penetrability, transfer, priming, mathematical reasoning, science education

IMPROVING PERCEPTION TO MAKE DISTANT CONNECTIONS CLOSER

One of the prime indicators of sophisticated cognition is that it does not rely on superficial resemblances to make connections between situations. Whereas a novice physicist may group scenarios based on surface properties such as whether springs or inclined planes are involved, the expert instead groups problems on the basis of the deep law of physics required for solution, such as Newton's second law or conservation of energy (Chi et al., 1981). Whereas a child typically connects clouds to sponges via surface features such as "round and fluffy," a more experienced adult may refer to more sophisticated relations such as "stores, and then releases water" (Gentner, 1988), allowing the adult to see connections among clouds, sponges, cisterns, and reservoirs. Scientists armed with the notion of a negative feedback system can see a resemblance between toilets, heat regulation, and predator–prey dynamics – namely, that each has two variables that are related such that increases to x cause increases to y which, in turn, cause decreases to x (Goldstone and Wilensky, 2008). Even though these scenarios have little in common with one another at first sight, sophisticated cognitive processes unite these situations because they share deep properties that crucially govern their behavior.

One moral that could be drawn from these examples is that perceptual resemblances must be cast aside if one is to procure the sophisticated categories and inferences of a scientist, mathematician, or domain expert. This is precisely the moral drawn by Quine (1977) when he wrote, "I shall suggest that it is a mark of

maturity of a branch of science that the notion of similarity or kind finally dissolves, so far as it is relevant to that branch of science. That is, it ultimately submits to analysis in the special terms of that branch of science and logic" (p. 160). The sort of example that Quine has in mind is a natural kind such as gold. Prior to the discovery of atomic elements, observers presumably noticed that several geological samples resembled each other, and used the term "gold" to refer to the collection of similar objects. However, once the elemental composition of gold was identified, surface features like "yellow," "malleable," and "shiny" were no longer necessary for identifying an object as gold. Advantages of supplanting these surface features with the chemical feature "atomic number 79" are that the chemical feature offers the promise of a scientific causal account for why gold has the surface features that it does, and it provides a way of excluding objects like pyrite ("fool's gold") from the category of gold despite its possession of some of gold's surface features. Perceptual resemblances can be misleading, and a sophisticated cognizer learns when to disregard these resemblances.

Another possibility is that perceptual resemblances are not fixed, and that we may adapt our perceptions so as to better support the requirements of categories and inferences that are important for us. Another way, then, of becoming a sophisticated cognizer is to modify one's perceptual processes to generate categories and inferences that are consonant with those that are formally sanctioned. In what follows, we first describe empirical evidence that people can and do change their perceptual processes in this way. We then describe mechanisms for this perceptual plasticity, with a

particular eye toward exploring the cognitive penetrability of these perceptual adaptations.
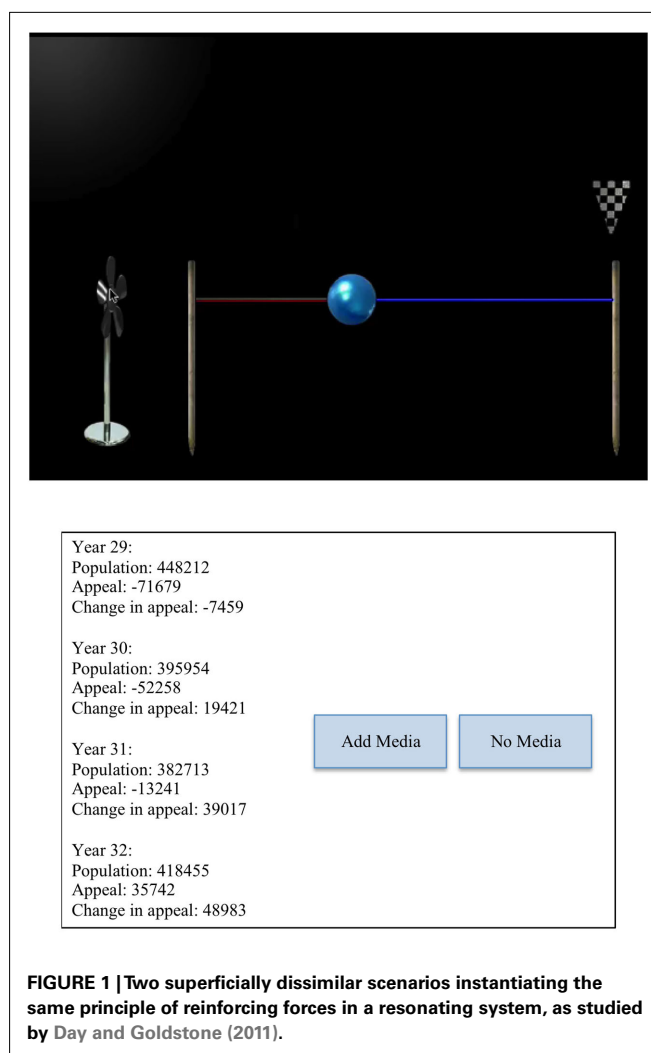
MAKING DISTANT CONNECTIONS

The examples of high-level cognition described above have a commonality – they all involve making connections between apparently distant scenarios, and/or splitting apart apparently similar scenarios. For example, an informed chemist connects a gold nugget to liquid gold dissolved in an alkaline solution, and differentiates it from pyrite. One way to draw inferentially productive yet distant connections is to equip oneself with an appropriate theory. This is the approach pursued by Quine (1977; see also, Goodman, 1972). A recently growing body of psychological evidence indicates a second way that is grounded in perception and action. Researchers in language, transfer, analogy, and cognition have found cases of people drawing connections between situations that do not seem to be superficially related. Much of this research has been associated with embodied and grounded cognition, an approach that argues that cognition is grounded in perception and action processes, rather than being associated with purely formal, amodal processing (Barsalou, 2008). This is an intriguing connection because of the *prime facie* tension between grounded accounts of cognition and connections being drawn that are not supported by perception. If cognition is inherently grounded in perception, then how are these superficially distant connections being made?

IMPLICIT ANALOGICAL TRANSFER VIA PERCEPTUAL PRIMING

One possible answer is provided by an experiment on transfer of learning by Day and Goldstone (2011). Their participants interacted with two systems that were superficially dissimilar, but both required participants to apply forces that either reinforced or opposed the system's natural resonance. The first scenario (see **Figure 1**), featured an oscillating ball suspended between two vertical poles by a rubber bands. If the ball is displaced to the right of center, then the red rubber band on the left will pull the ball back to the left. If the ball is displaced to the left, then the blue rubber band on the right will pull the ball back to the right. Given the absence of friction in the system, any perturbation of the ball's horizontal position leads to an undampened oscillation. The participants are able to apply a rightward force to the ball via a fan positioned on the left side of apparatus and facing to the right. By timing when the fan is turned on, the participants' task is either to stabilize the ball at the apparatus' midpoint without movement, or to get the ball to reach the extreme right side of the apparatus, as indicated by the checkered triangle in **Figure 1**. A Flash implementation of the simulation can be accessed at http://cognitrn. psych.indiana.edu/complexsims/Oscillatingball.html. To solve the stabilize task, the participants should turn on the fan whenever the ball is moving to the left, regardless of the ball's horizontal position. To solve the extremitize task, the participants should turn on the fan whenever the ball is moving to the right, so as to reinforce the ball's own movement.

After exploring this first simulation for several minutes, participants are given a second task without any indication of its relation to the first task. In this task, participants assume the role of mayor of a city. Whenever the population of the city is higher than 500,000, there is an intrinsic tendency for the population

FIGURE 1 | Two superficially dissimilar scenarios instantiating the same principle of reinforcing forces in a resonating system, as studied by Day and Goldstone (2011).

to decrease because of overcrowding, traffic jams, and expensive housing. Whenever the population is less than 500,000, there is a tendency for the population to increase because of living ease and inexpensive housing. Participants are given one of two goals as mayor: to stabilize the population at 500,000 citizens without fluctuation, or to make the population reach 1,000,000. To achieve these goals, participants can strategically deploy "media campaigns." At the beginning of each discrete year of the simulation, participants decide whether they will initiate a media campaign that adds a positive constant to the natural annual change (velocity) of the population.

The two tasks are isomorphic systems, governed by the same equation: $velocity_{t+1} = velocity_t + C \times (midpoint - position) + F$, where C is a constant, and F is the force that the participant strategically adds. There is a rigorous analogy in which the ball's position corresponds to the size of the population, the velocity of the ball corresponds to the year-to-year change in population, and turning on the fan corresponds to initiating a media campaign. Participants demonstrated sensitivity to these correspondences because they solved population problems more quickly when they were preceded by a congruent version of the ball task. That is, when both

tasks involved stabilize goals, or when both tasks involved extremitize goals, solutions were found more quickly than when one task required stabilization while the other required extremitization.

Interestingly, positive transfer between congruent simulations was found even when participants did not see any connection between the simulations, and could not correctly draw the correspondences listed in the previous paragraph. Similarly, when the correspondences were explicitly pointed out to participants, this did not increase the difference between congruent and incongruent conditions when performing the transfer task. In fact, the advantage of congruent over incongruent simulations was equally large when participants did versus did not demonstrate an understanding of the valid correspondences between scenarios. The observed transfer seems to be mediated by implicit priming, rather than strategic application of explicit schemas. The transfer also appears to be perceptually grounded because swapping the side of the fan from the left side (facing right) to the right side (facing left) eliminated transfer. Our interpretation of this effect is that people naturally understand population as a variable that goes from small values on the left to large values on the right, recruiting space to understand the numeric variable of population. Transfer is found only when the spatial relations in the ball scenario naturally align with the spatial interpretation of population.

The observed successes and failures of transfer across the ball and population scenarios point to both the power and fragility of perceptually grounded representations. These representations have the power to bridge across scenarios from different domains and with different interfaces, graphical elements, and timings. However, they are also fragile in that they depend upon the preservation of spatial relations that are not intrinsic to the underlying formal equations. The answer provided by these experiments to the question "If cognition is inherently grounded in perception, then how can connections be made between superficially dissimilar domains?" is that people naturally and automatically translate scenarios that are not directly spatial into spatial representations, and perceptual priming can occur between these transformed representations. In fact, perceptual priming can provide a vehicle for transfer even when more explicit, strategic avenues to transfer, such as abstract schemas (Gick and Holyoak, 1983; Detterman, 1993) or mathematical formulae (Ross, 1987) fail. Perceptual priming is effective for linking superficially dissimilar situations because people are habitually reinterpreting situations and translating them into (recently) familiar, frequently spatial, representations.

While the observed transfer apparently derives from spatial and dynamic representations, transfer is not always maximized by presenting a situation with its most intuitive embodiment. In fact, Byrge and Goldstone (2011) provide evidence that transfer from the ball to population situation is fostered by decoupling one's manual interaction with the ball simulation from its underlying resonance dynamic. The relatively unintuitive act of moving a switch to the left to make the fan blow rightwards results in better transfer to the population than when one's manual direction of motion is congruent with the fan's direction of force. The problem with incorporating highly intuitive perceptions and actions into a simulation is that people's knowledge of the simulations may become too closely tied to these groundings. If the subsequent situation does not share these groundings, then an opportunity for transfer may be missed. This result is consistent with earlier results showing that idealized, but still spatial, representations can produce particularly transferable knowledge by loosening the dependency between one's understanding of the principle and one's appreciation of the particular training domain (Goldstone and Sakamoto, 2003; Son et al., 2008; Son and Goldstone, 2009). Together with results suggesting that some action congruity effects are mediated by subjective construals rather than low-level bodily actions (Markman and Brendl, 2005), these results speak against naïvely assuming that more intuitive embodied representations will always yield superior transfer.

OTHER CASES OF GROUNDED BUT SUPERFICIALLY DISTANT CONNECTIONS BEING MADE

The above case study of cross-situational transfer that is grounded but nonetheless distant is not altogether unique. Other researchers have found examples of implicit transfer between structurally related situations despite a lack of conscious appreciation of the connection between the situations. People can solve a problem involving an "inhibition" strategy more quickly when another superficially dissimilar problem requiring inhibition was seen the previous day, even when they do not report noticing the relationship between the tasks (Schunn and Dunbar, 1996). Likewise, Gross and Greene (2007) have reported that the global structural relationships within a set of items (e.g., transitive or transverse relationships) may be transferred to a new set without participants' awareness. As a final example, structural relations involving relative clauses and scoping have been shown to transfer from mathematics equations to written sentences (Scheepers et al., 2011). Transfer across these kinds of situations have been modeled by relational priming using automatic spreading activation in neural networks (Kokinov and Petrov, 2001; Leech et al., 2008). Some results suggest that relational priming is not always automatic, but rather requires that people engage in cognitive processing that is sensitive to relations (Spellman et al., 2001). In any case, these situations provide examples of transfer across apparently dissimilar entities that reveal natural ways for people to construe their world. As with the earlier ball–population example, a cross-situation connection is forged because it does not require the cognizer to explicitly put the connection into words or equations, but rather only requires the same, grounded system to be recruited in different situations.

Another example of this generalization-by-conservation-of-systems mechanism is Hills et al. (2008, 2010) study of exploration and exploitation actions. They hypothesized that many situations fundamentally feature a decision about how much to explore new options versus exploit the options previously explored, and that there could be transfer across tasks that involve similar choice points along this tradeoff. To test this, they gave participants an initial task requiring them to forage for spatially distributed resources that were either clumped in discrete clusters or scattered. In a second task, participants came up with as many words as possible by rearranging sets of letters, exchanging old sets for new when they believed that they had effectively exhausted the potential words from their current set. Participants who foraged for distributed resources tended to exchange letter sets more often than participants who foraged for clustered resources, consistent

with the idea that training in a task that promotes exploration leads people to more exploratory behavior in a second task. Hills et al. speculate that this cross-task transfer may be mediated by dopamine. When clustered resources are present in the foraging task, then dopamine may be released as regions of highly concentrated resources are found. Dopamine is associated with an increased tendency to exploit currently known options. If this is the case, then the observed cross-task priming may be due to increased levels of dopamine in the clustered resources condition that simply remain active during the word formation task, leading to greater perseveration with a given letter set (exploiting known options). By this account, even without participants consciously appreciating that both tasks involve decisions to explore or exploit (post-experimental interviews indicated that participants did not explicitly make this connection), neural underpinnings are sensitive to the amount of exploration and exploitation required for a task, and transfer is simply a form of priming via shared task requirements.

Regardless of whether the dopaminergic hypothesis is correct, this form of explanation provides a general template for how grounded and embodied accounts of cognition can nonetheless produce surprisingly far transfer. Transfer can seem far to us because we do not have privileged access to the primitive components and parameters underlying our cognitive processes. Our conscious reflection prominently features words and justifications. However, the actual mechanisms that allow us to solve problems presented in computer simulations, recognize that a problem can be solved be inhibition, and decide whether to gamble on a new set of letters may feature other cognitive components. In particular, these components may be more perceptual, spatial, embodied, and diffuse than our reflections suggest. In these cases, transfer only seems far because we are biased to measure distance in terms of verbally expressible schemas. Perception and action provide us with unexpected connections that seem to depend on complex rationales, but this is only because our expectations are based on our consciously available justifications rather than our actual cognitive mechanisms.

ATTENTION TO THE VISUAL OBJECTS OF MATHEMATICS

Mathematical reasoning is a good place to look for connections between perception and high-level cognition. Mathematics is perhaps the pinnacle of cognitive abstraction. Mathematicians, even more so than physicists and computer scientists, strive to develop theories for increasingly general domains, covering more superordinate categories, and for more universal cases. Any particular mathematical tool, say combinatorics, can be applied to countless domains, ranging from bathroom tiling to lotteries. Much of this generality comes from the application of symbol systems, such as variables, equations, set theory, and predicate logic. These symbol systems confer on their user an ability to transcend the details of a particular domain. Given the critical role that symbol systems play in granting a cognizer distance from a domain, it is understandable that researchers have contrasted symbolic cognition from embodied cognition (Lakoff and Nuñez, 2000).

Yet, it is also worth remembering that symbol systems are physical themselves (Newell and Simon, 1976). This is especially true for external symbol systems such as mathematical notation. Rather than pitting symbolic processing versus perceptually grounded processes, we have found it productive to understand symbolic processing *via* perceptually grounded processes. Mathematical notation has changed over the millennia to be easily processed by humans (Cajori, 1928), but in addition, people change over the course of their lifetimes to more effectively manipulate and process mathematical notations. This latter, human, adaptation provides an excellent example of bridging perception and cognition by adapting perception to fit the needs of cognition that is engaged in symbolic processing.

In one line of experiments, we have studied how attentional processes are trained to facilitate algebraic reasoning (Goldstone et al., 2010). In particular, in algebra, there is an established convention of order of operations such that $3 + 4 \times 5$ equals 23 $[3 + (4 \times 5)]$ rather than 35 $[(3 + 4) \times 5]$. The mnemonic PEDMAS provides some of this order, with parenthesis – exponentiation – division and multiplication – addition and subtraction operations ordered from highest to lowest precedence. This formal system of operation precedence can be memorized and explicitly invoked when doing mathematics. However, applying explicit rules like this makes strong demands on memory and executive control. A cognitively less strenuous alternative is simply to train our visual attention in a manner that honors order of precedence without explicitly following a rule that specifies the order. In fact, people train their visual attention processes to give higher priority to notational operators that have higher precedence. The operator for multiplication, "×," attracts attention more so than does the notational symbol for the lower precedence addition operator, "+." People who know algebra show earlier and longer eye fixations to "×"s than "+"s in the context of math problems (Landy et al., 2008). Even when participants do not have to solve mathematical problems, their attention is automatically drawn toward the "×"s. When simply asked to determine what the center operator is for expressions like "$4 \times 3 + 5 \times 2$," participants' attention is diverted to the peripheral "×,"s as indicated by their inaccurate responses compared to "$4 + 3 + 5 + 2$" trials (Goldstone et al., 2010). The distracting influence of the peripheral operators is asymmetric because responding "×" to "$4 + 3 \times 5 + 2$" is significantly easier than responding "+" in "$4 \times 3 + 5 \times 2$." That is, the operator for multiplication wins over the operator for addition in the competition for attention. This is not simply due to specific perceptual properties of "×" and "+" because similar asymmetries are found when participants are trained with novel operators with orders of precedence that are counterbalanced. The results suggest that a person's attention becomes automatically deployed to where it should be deployed to get them to act in accordance with the formal order of precedence in mathematics.

BLIND AND MYOPIC FLAILING

Thus far, our argument has been that cognitive processes grounded in perception and action can still lead to surprisingly distant connections being made, because our sense of surprise is disproportionately based on our explicit rationales. Furthermore, we train our perceptual processes so that they better serve the needs of high-level cognition. The ability of our perceptual system to support far or "smart" transfer is further enhanced because of this training.

At this point, we must dispel a certain tension between the two planks of this argument. On the one hand, we are arguing that we do not have privileged access to the perceptual and grounded processes that underlie our own cognition, and so we do not realize that seemingly dissimilar ball and population simulations intrinsically involve similar force- and space-based representations, or that foraging for spatial resources and finding words involve similar processes that mediate the explore–exploit tradeoff. On the other hand, we are also arguing that we train our perceptual processes to achieve apparently more sophisticated outcomes. A critic might well press us to say, "Which way is it? Do you think we have access to the perceptual processes that underlie our cognition? If not, how can we adjust them?"

In defending our simultaneous assertion of both claims, we begin by distinguishing two senses of "*so that*" in our argument that "we selectively improve our perceptual abilities *so that* the tasks that we need to perform are performed better." By one interpretation, "so that" means "with the intention that," implying that we strategically alter our perceptual abilities. By the second interpretation, "so that" means "with the end result that," implying that our perceptual abilities are altered naturally through an automatic, non-conscious process. Our primary claim is meant in the spirit of this latter interpretation, although we shall later return to the first interpretation.

BLIND FLAILING

There is strong evidence from the field of perceptual learning that points to the importance of a learner's goal on perceptual adaptation. Granted, goals are not everything. Even perceptual information that is irrelevant to a task can become sensitized (Watanabe et al., 2001), even if this information is in the visual periphery and below the threshold for conscious detection (Seitz and Watanabe, 2005). However, there is at least as strong evidence that what is learned and how efficiently it is learned depends on the observer's task and goal. Even when sensitivity to a line orientation appears to have a relatively early locus of change, in that it does not transfer strongly across eyes or visual regions, it nonetheless depends on the observer's goals (Shiu and Pashler, 1992). Perceptual sensitization to the orientation of a line is much more robust when it is relevant for the task than when it is irrelevant. When observers are given the same stimuli in two conditions, but are required to make fine, subordinate-level categorizations in one condition and coarser, basic-level categorizations in the other, then greater selectivity of cortical regions implicated in object processing is found in the former condition (Gillebert et al., 2008). As a final example, perceptual discriminations are easier to make at boundaries between important categories for an observer, such as between a/p/and/b/phoneme that would be important for distinguishing "pats" from "bats." Evidence for this "categorical perception" effect from training studies and cross-linguistic comparisons indicates that it is not just perceptual sensitivities that are driving the categories, but rather the acquired categories are also driving perceptual sensitivities (Goldstone and Hendrickson, 2010). All of these studies show that we get better at making exactly the perceptual discriminations that help us do what we want to do.

A conservative interpretation of these results is that perceptions are changing *with the end result that* performance improves.

Strategic changes need not be implicated to account for the improvements. A mechanism that involves only random variation plus selection suffices. The effective strengths of neuronal connections are constantly varying. If a random change causes important discriminations to be made with increasing efficiency, then the changes can be preserved and extended. If not, the changes will not be made permanent. There may be other more goal-directed processes of neuronal change, but simple random variation with reinforcement that may be internally generated is all that is needed to systematically improve perceptual systems. Although apparently inefficient and "stupid," the "blind flailing" of random variation plus selection is surprisingly powerful. It features prominently in the theory of neuronal selection (Edelman, 1987), and the development of perception and action systems. In a literal application of flailing, infants often flail their arms around while learning to control them (Smith and Thelen, 1993). The flails that are relatively effective in moving the arms where desired are reinforced, allowing an infant to gradually fine-tune their motor control.

The blind flailing observed during perceptual learning can be fruitfully compared to the role of randomness in genetic algorithms. Genetic algorithms mimic some aspects of natural evolution to solve high-dimensional and difficult search problems by employing random variation and selection (Holland, 1975; Mitchell, 1996). A pool of random candidate solutions is initialized by encoding solutions in the chromosome of individuals. The fitness of each of individuals' solution is assessed and then, a new generation of solutions is formed by recombining, and adding random mutations to, the previous generations' solutions. Over several generations, genetic algorithms are often able to produce close-to-optimal solutions to difficult search problems. We are not arguing that genetic algorithms provide, in general, an accurate account of human cognition. Rather, we introduce genetic algorithms as a strong example of what blind flailing can achieve by way of macroscopically systematic progress.

For our current purposes, the important feature of genetic algorithms is that manipulations at one level, the chromosome of an individual solution, are then tested at a higher level that is effectively blind to the specific chromosomal changes that have been made. Selections of individuals are made on the basis of the results of these higher level tests. For example, a genetic algorithm might be applied to solving the traveling salesperson problem (TSP) for a given arrangement of destinations, such that the destinations are visited in a sequence that minimizes the total distance of the journey. Solutions could be encoded in an individual's chromosome as the sequence of destinations, such as "1 2 3 4 5 6" or "6 4 5 2 3 1." Mutations could involve swapping pairs of destinations, transforming "6 4 5 2 3 1" into "4 6 5 2 3 1." Each solution can be assigned a fitness defined as the total path distance, assuming that the distances between every pair of destinations is known.

Importantly, a system like this evolves fitter low-level chromosomal representations based on some other system's (e.g., the evaluator of fitness) feedback, without ever having an explicit mapping of how those low-level representations produce a good high-level result. For the TSP, it is easy to point to exactly such a mapping – namely the function that takes a sequence of destinations and

produces a total distance. However, in this case, the mapping is possessed by only one system – the high-level evaluation. More generally, the mapping between low-level chromosomes and high-level evaluations may be opaque or non-existent. When a male peacock is selected for its ability to attract females, then the mapping between the chromosome's coding of the male's body and its environmental fitness is not possessed by any single system, and is highly non-linear if it exists at all. Despite the potentially unknown/unknowable status of the mapping, it is nonetheless possible to evolve increasingly fit peacocks and TSP solutions. Similarly, blind flailing in the form of random changes to perceptual systems, combined with feedback on the changes' outcomes that is used to shape selection, can lead to systematic improvement to our perceptual processes. Both natural and artificial evolution give us strong precedents for the idea that short-term, blind flailing can lead to systematic improvement over a longer time course. Even if we completely lacked the ability to strategically refine our perception, our perceptual systems could still reliably adapt to become more congruent with the needs of high-level cognition.

MYOPIC FLAILING

A conspicuous disanalogy between perceptual learning and evolutionary algorithms is that perceptual learning occurs within sentient agents. There is no strategic agent that looks down upon evolutionary processes with the aim of increasing their efficiency by directing evolution in particular directions[1]. However, in the case of people, we may be interested in tweaking our perceptual system so that the tasks we need to perform are performed better. Now, "so that" is being used in its other sense of "with the intention that." Merely desiring that our perceptual systems perform better provides no guarantee that they *will* do so. In fact, there are arguments suggesting that they will not cooperate with our intentions. Within cognitive science it is common to argue for the cognitive impenetrability of perception – the notion that what we perceive is not influenced by our beliefs, goals, or experiences (Pylyshyn, 2006). A classic example is that the two lines of Müeller-Lyer illusion continue to *look* unequal even after we have just measured them, and hence we *know* that they are the same length. The claim for the cognitive impenetrability of perception is consistent with the notion of perceptual modules – that perceptual processes are generally structured such that we have access to their outputs, but no ability to adjust their internal workings (Fodor, 1983).

However, humans are impressively resourceful, and we have found a number of ways of affecting our perception at many different levels of processing. People purposefully "hack" their perceptual systems in order to facilitate performance. Often times, these hacks are still flailing, but they are not completely blind, merely myopic. To better appreciate the resourceful with which people manage to change their perceptual systems in order to accomplish tasks that they would have difficulty accomplishing otherwise, we outline a variety of pertinent cases. These cases illustrate mechanisms by which we alter, adjust, or adjust our perceptual abilities due to our intentional actions.

---

[1] However, in the case of genetic algorithms, there is current research interest in systems that guide evolution by creating new heuristics that will then constrain future fitness evaluation (Burke et al., 2009).

(A) Changing our perceptual equipment
   (1) Cupping one's hands behind one's ear to allow us to hear better in a particular direction.
   (2) Pushing the skin around one's eye's to deform the eye's shape to make an image sharper.
   (3) Clamping one's jaws tight to make one's ears less sensitive to noise.
   (4) Arranging our fingers so as to create a small aperture in front of our eye with the intention of creating a sharper image of an object.

(B) Strategically employing perceptual equipment
   (5) When wine tasting, sloshing the wine around one's mouth so that it covers more taste buds, also sucking in a bit of air to make more molecules airborne, thus intensifying olfactory response.
   (6) In a Stroop interference task, purposefully squinting one's eyes to facilitate ignoring the word that the colored ink forms.
   (7) Explicitly remind oneself to assess the characteristics of clarity, cut, caret, and color when judging the quality of diamonds.
   (8) Looking at a dim star not directly but in the periphery of one's eyes, where the concentration of rods is greater, and hence one's ability to detect faint light is greater.
   (9) When trying to see a pass-through rather than bounce event in the ambiguous apparent motion sequence shown in **Figure 2**, track with one's eyes a ball moving persistently from left to right.

(C) Long-term efforts to accelerate perceptual learning
   (10) Self-exposure to important stimuli. For instance, the communal collection, publication, and distribution of sets of "interesting" and "non-interesting" results from
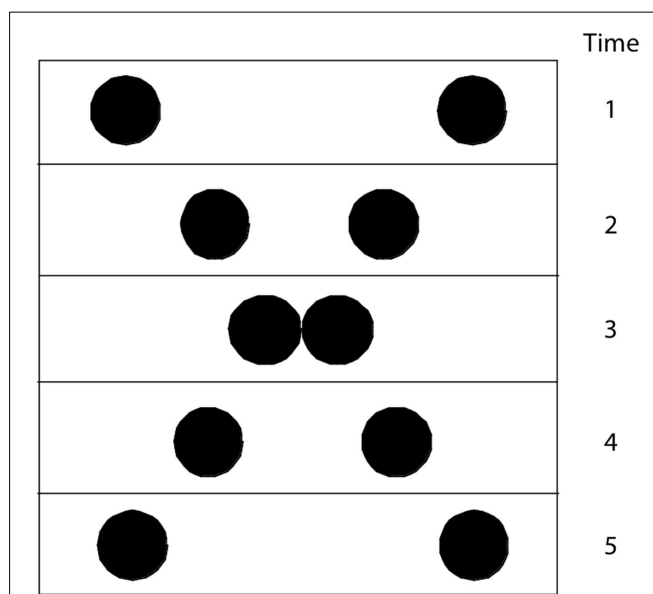

FIGURE 2 | Five frames of an ambiguous apparent motion sequence.
Two balls can be either seen passing through each other or as bouncing off one another.

cloud chamber experiments, in order to train new observers (Galison, 1997).

(11) When trying to learn the distinction between monarch and viceroy butterflies, explicitly juxtaposing pairs of the butterflies to exploit the benefit of simultaneous comparison and contrast.

(12) Giving oneself spaced, rather than massed, practice when trying to learn the difference between two species of mushrooms, so as to increase the impact on learning of each presentation.

(13) Purposefully exposing oneself to different speakers and syllables when trying to learn a difficult speech sound discrimination such as high-rising versus low-dipping tones in Mandarin for native English speakers or/r/versus/l/for native Japanese speakers.

(14) Training baseball batters to read numbers painted on baseballs to improve their ball tracking ability.

(15) Placing paintings on the walls of a baby's room if one wishes for the baby to later have an easier time identifying and distinguishing the paintings.

(D) Creating new perceptual objects to emphasize important properties

(16) Using Venn diagrams to determine the different possible combinations for three binary variables.

(17) Rewriting a math equation, spacing notational element further apart if they have a relatively low order of precedence, to promote correctly solving it.

(18) Drawing a graph to better understand the nature of a three-way interaction from a psychology experiment.

(E) Creating physical tools to allow us to perceive better

(19) Creating a telescope to view other planets.

(20) Putting ink on a ball before rolling it, so as to better inspect its trajectory.

(21) Creating a cloud chamber to view the trajectories of sub-atomic particles.

(22) Installing a cochlear implant to restore hearing to a deaf individual.

To be sure, not all of these examples are violations of cognitive impenetrability. Examples 11–13 are cases of an observer's goals influencing their perceptual categorizations. It could be argued that they are not relevant, though, to cognitive penetration because the goals are long-term rather than acting on-line during the processing of a single stimulus. We would argue, however, that these kinds of perceptual changes are more influential exactly because they are long-term and chronic, and the perceptual change becomes automatic once acquired (Shiffrin and Lightfoot, 1997). If we restrict the influence of goals to only interactive and on-line influences, then we systematically ignore the large class of situations in which we change the feed-forward characteristics of a perceptual system to make it more efficient for meeting our goals.

Examples like 9 are interesting because motion perception has been singled out as one of the strongest cases for a modularized perceptual system, with well-defined computational accounts (Ullman, 1979) and localized brain regions (e.g., area MT). The fact that one's goals can change the motion that is subjectively perceived is compatible with motion perception being highly

modular. Either people can systematically adjust the inputs to their perceptual apparatus to alter the computation of motion, or the parameters governing the computation of the object correspondences underlying motion perception (Dawson, 1991) can themselves be tuned by goals. There are numerous examples of such tuning being necessary to account for the influences of knowledge and context on motion perception (Palmer, 1999). Just because something is highly modular does not mean that it performs its function without variation or context-sensitivity. In the same way that a function or subroutine can take arguments that affect the computations performed within it, the computations within even a completely opaque perceptual black-box can be modulated, and if the perceptual module is to be responsive and robust, it must be.

The mechanisms described above for changing perceptions have been organized into five categories. The intention is not so much to draw sharp distinctions between these categories as to draw parallels across the categories. For example, we suspect that few people would naturally consider the mechanisms of (E) (except 21) to be perceptual changes at all. However, we see these mechanisms to be comparable to some of the mechanisms of (A). Cupping one's hands over one's ears seems importantly similar to building a telescope. They both extend the normal range of one's sensory organ. It seems less important that one extension is achieved by natural, bodily means, while the other by an inorganic tool. Likewise, there are strong parallels between the mechanisms of (D) and (E). We believe that creating perceptual tools like Venn and Feynman diagrams can be understood as deeply related to creating physical tools that extend our sensory organs (Landy and Goldstone, 2005). A powerful new spatial representation changes how things look just as surely as a microscope does. Compelling examples have been empirically described for how diagrams help thinking by promoting new ways of perceiving. Providing a static diagram may help people see what two seemingly dissimilar instantiations of a "convergence schema" share (Gick and Holyoak, 1983), and if a dynamic animation showing convergence is provided, then even greater transfer is achievable (Pedone et al., 2001). Cheng's (2002) analysis of diagrams points to a suite of desirable properties of diagrams that allow them to serve as effective "cognitive prostheses": (1) they combine globally homogeneous with locally heterogeneous representations of concepts, (2) they integrate alternative perspectives, (3) they allow for expressions to be easily manipulated, and (4) they support compact and uniform procedures.

A new spatial representation does not always need to be physically instantiated to prove effective, once it has been internalized. The benefits of Venn diagrams, once understood, can be secured even when they are only internally generated. As useful as it is to offload cognitive tasks onto the environment (Clark, 2009), it is often equally useful to internalize physical transformations. For example, one of the striking effects of learning the formalisms and diagrams for Signal Detection Theory is that they can become so well internalized that their possessor spontaneously sees connections between doctors diagnosing cancers and farmers determining which melons to ship, even when the learner does not prepare any external representation (Son and Goldstone, 2009). More generally, one of the best hopes for schooling is that students will learn new, habitual ways of seeing their world as a result of their formal

education. Students will learn to see their world through the tools they have acquired.

The term "myopic flailing" is meant to be contrasted with the "blind flailing" of genetic algorithms and natural evolution. Myopic flailing conveys that people can educate their perceptions more efficiently than expected via pure random variation, even though their manipulations are less direct and straightforward than they would be if they could access and manipulate all aspects of the perceptual module. The classic chicken sexing expertise study by Biederman and Shiffrar (1987) provides a good context for appreciating myopic flailing. In this study, novice participants were given a single page of instructions on how to categorize day-old baby chickens that elevated their performance at chick sexing with photographs from slightly above chance to approximately that of experts with 24 years of experience sexing chicks. The novice's impressive improvement with less than an hour of training is striking. It is highly unlikely that the novice's improvement is mostly due to perceptual learning. Most cases of perceptual learning are characterized by slow and protracted learning over the course of weeks or years (Shiffrin and Lightfoot, 1997; Goldstone, 1998). Perceptual learning is an example *par excellence* of the adage that "wisdom can not be taught." One cannot simply read a text-based book that has no illustrations if one wishes to become an expert dog show judge, gymnastics coach, wine taster, or umpire. One needs experiences to change one's perceptual system.

However, it would also be a mistake to completely ignore the beneficial influence of instructional words and verbal justifications. In the case of the chick sexing study, the rapidity of learning suggests that the novice participants already had the perceptual building blocks firmly in place needed to understand and follow the instructions, which featured phrases such as "look for two large cylindrical side lobes near the bottom of each picture" and "Male chicken genitals tend to look round and foolish like a ball or watermelon." This is a case of adaptation that is so clear-sighted that it does not qualify as "flailing" at all. In many cases of perceptual training, the accompanying words are not so directly actionable, but neither are they completely irrelevant. These are the cases where perceptual adaptation is best understood as operating via myopic flailing. Consider, for example, a radiologist instructing her students on how to distinguish between sarcoidosis and pulmonary alveolar proteinosis by looking for fissural beading versus a diffuse mosaic ground glass paving pattern without fibrosis. These perceptual features require months/years of training to develop. It is unlikely that a simple page of imageless instructions will ever suffice for their instruction, and medical schools have converged on training disease identification through a combination of describing bodily appearances and explaining causal bodily mechanisms. Features like "ground glass," "fibrotic," "paving pattern," "mossy," "ulcerated," and "pustulated" are not immediately understandable, and developing an operational understanding of them practically necessitates undergoing perceptual training by witnessing cases. However, the words are nonetheless useful for focusing one's attention on different aspects of a disease, such as its spatial distribution, color, arrangement, tactile feel, and texture. The words do not directly alter the internal workings of perceptual modules, but they do lead to more effective learning than pure random selection. They provide myopic support for tuning perceptions.

A characteristic of many forms of expertise is that the expert has both a highly precise verbal vocabulary and an ability to perceptually parse objects from their domain in a coherent and expressive manner. These two characteristics are correlated because, we believe, each informs the development of the other. In most cases, words cannot *replace* experience for teaching perceptual skills, but they can facilitate perceptual skill learning, as anybody who has tried to learn to distinguish poisonous from edible mushrooms in a completely word-less, instruction-less, and inductive fashion would attest (in the unlikely event that they lived long enough to do so).

CONCLUSION

There is little, if any, gap between perception and high-level cognition because perceptual systems adapt to fit the needs of high-level cognition. These adaptations may be either the result of random variation or more directed tuning. A person gaining experience with the world also acquires more knowledge about how low-level, physical transformations affect high-level cognitive outcomes. For this reason, blind flailing generally gives way to varying degrees of guided tuning through learning. Babies have difficulty even tracing the edges of a high-contrast object with their eyes. A psychophysicist studying color can separately isolate the saturation and brightness levels of an object. Most adults fall somewhere in between these two points, having intermediate-level access to visual properties. Once a visual property has been isolated, it can then be strategically tuned. Before a person has learned to isolate saturation from brightness, it is difficult or impossible for them to selectively attend to just one of these dimensions (Goldstone and Steyvers, 2001). Afterward, they have strategic control over which dimensions they will use for a particular purpose. Thus, people not only learn to attend to perceptual dimensions to address their needs; they also learn how to learn to attend to dimensions. This meta-learning represents the transition from a relatively uncontrolled, random search for a method to improve perceptual processing to a relatively controlled and guide done.

Perceptual learning, and perceptual learning learning, serve to increase the sophistication of our perceptual processes. The result is that people's perceptual processes can support what appear to be long-distance connections requiring formal abstractions. The primary advantage of long-distance connections that are based on perceptual rather than formal symbolic processes is that they are more likely to exist! Formalisms provided by mathematics and logic are typically cognitively inert unless they are grounded in perceptual processes. They are inert in the sense that people are unlikely to realize that two situations are governed by the same formalism, unless they are given a hint to connect the situations. As such, these connections are not likely to be made through application of formalisms.

The promise of making connections based on learned perceptual properties is that the connections can be automatically forged because they are perceptual, but they can nonetheless be sophisticated because they are learned. Strategies and goals shape perceptual learning via "myopic flailing" (see Materials and Methods 10–15), but importantly, once the learning has transpired,

it is automatically deployed during perception. Even when perceptions cannot be semi-permanently changed via learning, the other methods describe ways of manipulating perceptions so as to overcome some of their limitations.

This perspective on achieving sophisticated reasoning through perceptual manipulation can be contrasted with the Quine an approach of trumping perception by higher level reasoning, rules, and the application of definitions. In practice, both kinds of processes must occur. Determining the causes and consequences of each process would constitute a fertile research program, with perhaps even neural correlates. For example, for cases in which perceptual processes are trumped by rules, we might expect frontal cortex to exhibit heightened activity, and to actively inhibit more posterior perceptual regions. In contrast, when perceptual processes are adapted to subserve formal thought, then posterior cortical regions may assume particular importance. This decomposition into modules is roughly compatible with empirically observed neural supercessions –cases in which controlled, initial performance is governed by different neural populations than subsequent automatic processing (Procyk et al., 2000). For example, when a monkey first learns to associate a novel stimulus with a response, some cells with in the supplementary eye fields (SEF) of the dorsal–medial surface of the frontal lobe are highly active, but become decreasingly active with repetition of the stimulus. Other cells show the opposite tendency, becoming increasingly active as the response to a novel stimulus is learned (Chen and Wise, 1995). This pattern of complimentary controlled and automatic processes fits the above developed account in that we have argued that controlled processes operate to make themselves obsolescent by modifying perception over a protracted course of training. While our account is similar in some ways to theories positing a split between rule-based versus automatic, association-based reasoning (Sloman, 1996), our account focuses on the development of new perceptual processes rather than simply association learning, and points to ways in ways in which our rule-based system guides and informs the construction of perceptual processes.

One advantage of training over trumping perception is that the opportunities provided by rich and nuanced interpretations available from a highly evolved and trained perceptual system are not relinquished. As a consequence of the automatic and strategic changes to perception, people can perceive connections between balls and city growth (they can both be resonance systems), toilets and hare–lynx populations (they are both negative feedback systems), and a general surrounding a fortress and removing a tumor by concentrating multiple lasers at the tumor (they are both example of converging forces to overcome an entity). Once bridges are built between these *prime facie* distant but deeply related situations, knowledge and inferences can freely move from one to the other. Connections can be made between situations that, at first, may not appear related at all, because trained appearances can go far beyond first appearances.

ACKNOWLEDGMENTS

REFERENCES

Barsalou, L. W. (2008). Grounded cognition. *Annu. Rev. Psychol.* 59, 617–645.

Biederman, I., and Shiffrar, M. M. (1987). Sexing day-old chicks: a case study and expert systems analysis of a difficult perceptual-learning task. *J. Exp. Psychol. Learn. Mem. Cogn.* 13, 640–645.

Burke, E. K., Hyde, M. R., Kendall, G., Ochoa, G., Ozcan, E., and Woodward, J. R. (2009). "Exploring hyper-heuristic methodologies with genetic programming," in *Computational Intelligence: Collaboration, Fusion and Emergence*, eds C. Mumford and L. Jain (New York: Springer), 177–201.

Byrge, L. A., and Goldstone, R. L. (2011). "Distinguishing levels of grounding that underlie transfer of learning," in *Proceedings of the Thirty-Third Annual Conference of the Cognitive Science Society* (Boston, MA: Cognitive Science Society), 2818–2823.

Cajori, F. (1928). *A History of Mathematical Notations*. La Salle, IL: Open Court Publishing Company.

Chen, L. L., and Wise, S. P. (1995). Neuronal activity in the supplementary eye field during acquisition of conditional oculomotor associations. *J. Neurophysiol.* 73, 1101–1121.

Cheng, P. C. H. (2002). Electrifying diagrams for learning: principles for complex representational systems. *Cogn. Sci.* 26, 685–736.

Chi, M. T. H., Feltovich, P., and Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cogn. Sci.* 5, 121–152.

Clark, A. (2009). *Supersizing the Mind*. Oxford, UK: Blackwell.

Dawson, M. R. (1991). The how and why of what went where in apparent motion: modeling solutions to the motion correspondence problem. *Psychol. Rev.* 98, 569–603.

Day, S., and Goldstone, R. L. (2011). Analogical transfer from a simulated physical system. *J. Exp. Psychol. Learn. Mem. Cogn.* 37, 551–567.

Detterman, D. R. (1993). "The case for prosecution: transfer as an epiphenomenon," in *Transfer on Trial: Intelligence, Cognition, and Instruction*, eds D. K. Detterman and R. J. Sternberg (Westport, CT: Ablex), 17–38.

Edelman, G. (1987). *Neural Darwinism: The Theory of Neuronal Group Selection*. New York: Basic Books.

Fodor, J. A. (1983). *Modularity of Mind: An Essay on Faculty Psychology*. Cambridge, MA: MIT Press.

Galison, P. L. (1997). *Image and Logic: A Material Culture of Microphysics*. Chicago: University of Chicago Press.

Gentner, D. (1988). Metaphor as structure mapping: the relational shift. *Child Dev.* 59, 47–59.

Gick, M. L., and Holyoak, K. J. (1983). Schema induction and analogical transfer. *Cogn. Psychol.* 15, 1–38.

Gillebert, C. R., Op de Beeck, H. P., Panis, S., and Wagemans, J. (2008). Subordinate categorization enhances the neural selectivity in human object-selective cortex for fine shape discriminations. *J. Cogn. Neurosci.* 21, 1054–1064.

Goldstone, R. L. (1998). Perceptual learning. *Annu. Rev. Psychol.* 49, 585–612.

Goldstone, R. L., and Hendrickson, A. T. (2010). Categorical perception. *Wiley Interdiscip. Rev. Cogn. Sci.* 1, 65–78.

Goldstone, R. L., Landy, D. H., and Son, J. Y. (2010). The education of perception. *Top. Cogn. Sci.* 2, 265–284.

Goldstone, R. L., and Sakamoto, Y. (2003). The transfer of abstract principles governing complex adaptive systems. *Cogn. Psychol.* 46, 414–466.

Goldstone, R. L., and Steyvers, M. (2001). The sensitization and differentiation of dimensions during category learning. *J. Exp. Psychol. Gen.* 130,116–139.

Goldstone, R. L., and Wilensky, U. (2008). Promoting transfer through complex systems principles. *J. Learn. Sci.* 17, 465–516.

Goodman, N. (1972). "Seven strictures on similarity," in *Problems and Projects*, ed. N. Goodman (New York: The Bobbs-Merrill Co.), 437–447.

Gross, W. L., and Greene, A. J. (2007). Analogical inference: the role of awareness in abstract learning. *Memory* 15, 838–844.

Hills, T. T., Todd, P. M., and Goldstone, R. L. (2008). Search in external and internal spaces. *Psychol. Sci.* 19, 676–682.

Hills, T. T., Todd, P. M., and Goldstone, R. L. (2010). Priming a central executive search process: exploration and exploitation in generalized cognitive search processes. *J. Exp. Psychol. Gen.* 139, 560–609.

Holland, J. H. (1975). *Adaptation in Natural and Artificial Systems*. Ann Arbor: University of Michigan Press.

Kokinov, B., and Petrov, A. (2001). "Integration of memory and reasoning in analogy-making: the AMBR model," in *The Analogical Mind: Perspectives from Cognitive Science*, eds D. Gentner, K. Holyoak, and B. Kokinov (Cambridge, MA: MIT Press).

Lakoff, G., and Nuñez, R. E. (2000). *Where Mathematics Comes From*. New York: Basic Books.

Landy, D., and Goldstone, R. L. (2005). How we learn about things we don't already understand. *J. Exp. Theor. Artif. Intell.* 17, 343–369.

Landy, D. H., Jones, M. N., and Goldstone, R. L. (2008). "How the appearance of an operator affects its formal precedence," in *Proceedings of the Thirtieth Annual Conference of the Cognitive Science Society* (Washington, DC: Cognitive Science Society), 2109–2114.

Leech, R., Mareschal, D., and Cooper, R. (2008). Analogy as relational priming: a developmental and computational perspective on the origins of a complex cognitive skill. *Behav. Brain Sci.* 31, 357–378.

Markman, A. B., and Brendl, C. M. (2005). Constraining theories of embodied cognition. *Psychol. Sci.* 16, 6–10.

Mitchell, M. (1996). *An Introduction to Genetic Algorithms*. Cambridge, MA: MIT Press.

Newell, A., and Simon, H. A. (1976). Computer science as empirical enquiry: symbols and search. *Commun. ACM* 19, 113–126.

Palmer, S. E. (1999). *Vision science: From Photons to Phenomenology*. Cambridge, MA: Bradford Books/MIT Press.

Pedone, R., Hummel, J. E., and Holyoak, K. J. (2001). The use of diagrams in analogical problem solving. *Mem. Cognit.* 29, 214–221.

Procyk, E., Tanaka, Y. L., and Joseph, J. P. (2000). Anterior cingulate activity during routine and non-routine sequential behaviors in macaques. *Nat. Neurosci.* 3, 502–508.

Pylyshyn, Z. W. (2006). *Seeing and Visualizing: It's Not What You Think*. Cambridge, MA: MIT Press.

Quine, W. V. (1977). "Natural kinds," in *Naming, Necessity, and Natural Kinds*, ed. S. P. Schwartz (Ithaca, NY: Cornell University Press), 155–175.

Ross, B. H. (1987). This is like that: the use of earlier problems and the separation of similarity effects. *J. Exp. Psychol. Learn. Mem. Cogn.* 13, 629–639.

Scheepers, C., Sturt, P., Martin, C., Myachykov, A., Teevan, K., and Viskupova, I. (2011). Structural priming across cognitive domains: from simple arithmetic to relative-clause attachment. *Psychol. Sci.* 22, 1319–1326.

Schunn, C., and Dunbar, K. (1996). Priming, analogy, and awareness in complex reasoning. *Mem. Cognit.* 24, 271–284.

Seitz, A., and Watanabe, T. (2005). A unified model for perceptual learning. *Trends Cogn. Sci. (Regul. Ed.)* 9, 329–334.

Shiffrin, R. M., and Lightfoot, N. (1997). "Perceptual learning of alphanumeric-like characters," in *The Psychology of Learning and Motivation*, Vol. 36, eds R. L. Goldstone, P. G. Schyns, and D. L. Medin (San Diego: Academic Press), 45–82.

Shiu, L. P., and Pashler, H. (1992). Improvement in line orientation discrimination is retinally local but dependent on cognitive set. *Percept. Psychophys.* 52, 582–588.

Sloman, S. A. (1996). The empirical case for two systems of reasoning. *Psychol. Bull.* 119, 3–22.

Smith, L. B., and Thelen, E. (1993). *A Dynamic Systems Approach to Development: Applications*. Cambridge, MA: MIT Press.

Son, J. Y., and Goldstone, R. L. (2009). Contextualization in perspective. *Cogn. Instr.* 27, 51–89.

Son, J. Y., Smith, L. B., and Goldstone, R. L. (2008). implicity and generalization: short-cutting abstraction in children's object categorizations. *Cognition* 108, 626–638.

Spellman, B. A., Holyoak, K. J., and Morrison, R. G. (2001). Analogical priming via semantic relations. *Mem. Cognit.* 29, 383–393.

Ullman, S. (1979). *The Interpretation of Visual Motion*. Cambridge, MA: MIT Press.

Watanabe, T., Náñez, J. E., and Sasaki, Y. (2001). Perceptual learning without perception. *Nature* 413, 844–848.

How category structure influences the perception of object similarity: The atypicality bias

James William Tanaka[1]*, Justin Kantner[1] and Marni Bartlett[2]

[1] Department of Psychology, University of Victoria, Victoria, BC, Canada
[2] Institute for Neural Computation, University of California, San Diego, CA, USA

Why do some faces appear more similar than others? Beyond structural factors, we speculate that similarity is governed by the organization of faces located in a multi-dimensional face space. To test this hypothesis, we morphed a typical face with an atypical face. If similarity judgments are guided purely by their physical properties, the morph should be perceived to be equally similar to its typical parent as its atypical parent. However, contrary to the structural prediction, our results showed that the morph face was perceived to be more similar to the atypical face than the typical face. Our empirical studies show that the atypicality bias is not limited to faces, but extends to other object categories (birds) whose members share common shape properties. We also demonstrate atypicality bias is malleable and can change subject to category learning and experience. Collectively, the empirical evidence indicates that perceptions of face and object similarity are affected by the distribution of stimuli in a face or object space. In this framework, atypical stimuli are located in a sparser region of the space where there is less competition for recognition and therefore, these representations capture a broader range of inputs. In contrast, typical stimuli are located in a denser region of category space where there is increased competition for recognition and hence, these representation draw a more restricted range of face inputs. These results suggest that the perceived likeness of an object is influenced by the organization of surrounding exemplars in the category space.

Keywords: face perception, object perception, categorization, morphing, perceptual similarity

INTRODUCTION

In **Figure 1**, we see the faces of two well known actors (Robert Downey Jr. and George Clooney) and the morph face created by averaging the two parent images together. The morph face presents a curious puzzle: although the morph "child" face receives equal contributions from its celebrity parents, it bears a stronger likeness to one parent than to the other. In this example, most people would agree that the morph face looks more like Robert Downey Jr. If the image is a 50/50 morph of George Clooney and Robert Downey Jr., what are the factors that drive perceptions toward one interpretation of morph face toward one parent image over the other?

In this paper, we propose that perceptions of the averaged morph can be affected by the distinctiveness of its parents. According to the atypicality bias account, when pitted against one another, the atypical parent will exert more influence on the morphed image than the typical parent. In our framework, we conceptualize face representations as attractor fields that compete for activation of the face input. Activation of a face representation is determined by the degree of fit between the face stimulus and face representation. We hypothesize that due to their location in face space, atypical faces have broader attractor fields than typical faces. In this paper, we examine the atypicality bias generalizes to other categories besides faces. Finally, we also explore the learning trajectory of the atypicality

bias and investigate how our perceptions of what's typical and what's atypical change as function of learning and category experience.

FACES IN EUCLIDEAN FACE SPACE

By the time most people reach adulthood, they are able to recognize upwards of thousands if not tens of thousands of faces. A critical question is how are face representations organized in memory to support recognition that is fast, effortless, and relatively error-free. It has been proposed that faces are organized in a hyper-dimensional, Euclidean coordinate system, or "face space" as defined by the perceptual dimensions of the face stimulus (Valentine, 1991). The precise dimensions of face space are not explicitly specified, but presumably correspond to the physical properties of the face, such as the length and width of the face contour, size, and shape of its eyes, nose, and mouth features, etc. Faces that share similar values on a given dimension are clustered closer together and faces with disparate values farther apart. A particular face is localized at a specific coordinate location in face space based on its dimensional values. Although conceptually separable, face dimensions are assumed to be perceptually integrated (Garner, 1974) or holistic (Tanaka and Farah, 2003), such that it is difficult to attend to one dimension (e.g., distance between the eyes) without attending to the other dimensions (e.g., shape of the eyes, spacing between the nose and mouth).

FIGURE 1 | Images of movie actors Robert Downey Jr. (far left photo) and George Clooney (far right photo) and their 75%/25% (left middle), 50% /50% (middle) and 25%/75% (right middle) morph images.

At the origin of face space lies the prototypical face that represents the mean values along these dimensions. Evidence for prototype abstraction comes from empirical studies showing that participants are more likely to falsely recognize a never-before-seen prototype dot pattern (Posner and Keele, 1968) or face (Cabeza et al., 1999) than a previously studied pattern or face. Face typicality varies as a function of distance from the origin of the space. In this coordinate system, typical or average looking faces are located near the prototypical face at the center and less typical or more unusual looking faces are located in more peripheral regions A critical assumption of the face space model is that face representations are normally distributed such that there is a higher density of typical face exemplars at the origin and a decreasing proportion of atypical face exemplars with increasing distance from the origin. Consequently, typical faces are assumed to be less distinctive in memory due to heightened competition from neighboring typical face representations. In contrast, atypical faces have fewer neighbors, are exposed to less competition, and are therefore more distinctive.

The empirical results are consistent with the face space account of distinctiveness. For example, Bartlett et al. (1984) and Going and Read (1974) demonstrated that highly distinctive faces are recognized more accurately than faces rated low in distinctiveness. Whereas atypical faces are faster to recognize due to their distinctiveness, they are slower to be categorized as "faces" due to their deviation from the face category prototype (Johnston and Ellis, 1995). The atypicality face advantage has been demonstrated for the recognition of newly familiarized faces (Light et al., 1979; Bartlett et al., 1984; Vokey and Read, 1992) and famous faces (Valentine and Bruce, 1986). Rhodes et al. (1997) showed that face caricatures, in which the features of a face are exaggerated relative to the average face prototype, are recognized more readily than anti-caricatures, in which distinctive features of a face are diminished relative to the prototype. This view is consistent with Lewis' (2004) Face-Space-R model, where the recognition of a specific face is determined by the number and proximity of neighboring face exemplars in the space. In summary, the collective evidence supports the predictions of the face space model where the clustering of representations in face space renders typical faces less memorable and atypical faces more memorable in recognition.

The face space concept has been explored in aftereffects paradigms where prolonged exposure to an adapting face stimulus causes a systematic bias in the perception of a subsequent test face (for a recent review, see Webster and MacLeod, 2011). For example, adapting to a contracted face causes a normal face to appear expanded (Webster and MacLin, 1999), an Asian face to appear more Caucasian (Webster et al., 2004) and a male face to appear more female (Rhodes et al., 2004; Webster et al., 2004). Identity-specific aftereffects have been demonstrated in which exposure to an individual face enhances recognition of the face lying directly opposite to adapting face in face space (i.e., its anti-face; Leopold et al., 2001). For instance, adapting to a face with a narrow eyes and a small, pointed chin (e.g., Fred) will enhance the perception of a face with broad eyes and a large rounded chin (e.g., anti-Fred). To account for face-specific adaptation effects, it has been proposed the faces lie on the trajectory defined by individual face, the prototypical face, and the "anti-faces" on opposite side of face space and this family of faces is systematically affected by the adaptation manipulation, such that adapting to anti-Fred renders the prototypical face to appear more like Fred. Critically, adaptation effects do not generalize to other faces that are proximal to the target face in face space, but are located off the trajectory (Leopold et al., 2001; Rhodes and Jeffery, 2006). Collectively, results from face aftereffects studies provide compelling evidence that faces are stored in a multi-dimensional face space with the prototypical face lying at its origin.

A second prediction of the face space model is that the similarity of a face stimulus is a linear function of its distance in Euclidean space (Shepard, 1964). That is, the closer a face input is to a stored face representation, the more likely it is to be identified with that face. Face morphs provide a good test of the Euclidean prediction. In the morphing process, two faces are graphically averaged to produce a "child" morph face. The physical face morph varies as a function of the relative contributions of the parent images. For example, an 80/20 morph that receives an 80% contribution from Parent A and 20% contribution from Parent B will lie closer to Parent A than Parent B in face space and should therefore bear a stronger resemblance to the Parent A image. The 50/50 morph face (i.e., a morph face that receives equal contribution from both parent faces) presents an interesting test of recognition because it lies an equal distance from both parent faces and, therefore, should be perceived as equally similar to both parents. As discussed in the next section, although the Euclidean predictions of similarity are straightforward, the empirical evidence suggests that the nearest neighbor is not the only factor that determines how similar (or different) two faces appear.

FACES AND THE ATYPICALITY BIAS

Tanaka et al. (1998) used a delayed match-to-sample task to test the assumption of the attractor field model that distinctive category members possess larger attractor fields than typical members. Tanaka et al. (1998) identified a set of eight typical and eight distinctive faces through pilot testing and morphed each typical face with each distinctive face of the same gender (**Figure 2**). To construct a morph of the atypical and typical parent faces, corresponding control points were identified on the two parent images (e.g., the corner of the left eye on Parent Face Image 1 and Parent Face Image 2). The number of control points for facial features were kept constant, with 12 points on the mouth, 7 points on each eye, 6 points on the nose, 5 points on each eyebrow, and 22 points for the outline of the face. According to the level of morphing, new control points for the morph face were defined by moving the specified distance along the vector connecting the control points in parent images. The locations of intervening pixels were linearly interpolated across the surface based on the position of the nearest control point (Wolberg, 1990). A fade process was then employed in which the brightness values for each corresponding pixel were weighted according to the contribution of each parent image.

This procedure yielded eights morphs, each containing an equal physical contribution from its two "parents" (one typical, one distinctive). The Tanaka et al. (1998) paradigm tested whether the morph face would be judged more similar to the atypical parent, as would be predicted by the attractor field model. On each trial, a pair of parent faces was presented on opposing sides of the computer screen for 2.5 s. The parents were replaced with the morph of the pair for 1 s, after which the morph disappeared, and participants indicated whether it more closely resembled the parent

FIGURE 2 | Examples of the atypical and typical female and male faces and their 50/50 morph faces used in the Tanaka et al. (1998) **study.**

presented on the right or the parent presented on the left. The measure of interest was the percentage of trials on which the atypical parent was chosen. In the most straightforward version of the task (Tanaka et al., 1998, Experiment 1), participants indicated that the atypical parent was more similar to the morph on 60% of trials, significantly greater than the 50% (chance) level that would be expected if the morph appeared equally similar to the typical and atypical parents.

Was the atypical parent chosen more often than the typical parent because it was perceived to be more similar to the morph or simply because it was more memorable than the typical parent? To address this question, Tanaka et al., 1998, Experiment 2) added an "unrelated" condition in which two parent stimuli were followed by the morph of a different set of parents. When participants viewed related morphs, they selected the atypical parent on 63% of trials, replicating the atypicality bias. When viewing unrelated morphs, however, they chose the *typical* parent on 60% of trials. This result suggested that the atypicality bias observed in Experiment 1 and the related condition of Experiment 2 were not due to preferential choosing of the more memorable parent; if memorability alone drove responses, an atypicality bias would be expected in the unrelated condition as well. Instead, a typicality bias was observed, a sensible result given that a randomly selected morph is likely to be more similar to a typical face than an atypical face. Experiment 3 demonstrated an atypicality bias when the morph and parents were presented simultaneously, providing further evidence that the phenomenon is perceptually driven rather than memory-driven.

A second goal of Experiment 3 was to examine the perceptual locus of the atypicality bias. Given that typical faces are assumed to be recognized on the basis of configural properties while atypical faces are often perceived as such based on a particular distinctive feature, inversion provides a means of measuring the relative contributions of each type of processing to the atypicality bias. Because inversion is thought to disrupt configural more than featural processing, the atypicality bias should be attenuated with inverted stimuli to the extent that configural processing underlies the effect. Tanaka et al. (1998) added test blocks in which the stimuli were inverted and observed an atypicality bias that was significant (55%) but diminished relative to the magnitude of the effect for upright faces (62%). Tanaka et al. (1998) concluded that both configural and featural processing play a role in the perception that the morph is more similar to the atypical parent.

In the attractor field model, the attractor basin surrounding each stimulus is demarcated by a boundary signifying a point in similarity space at which a stimulus input will activate either of two representations with equal probability. The disparity in the similarities of the morph to its typical and atypical parents suggests that the boundary between the parents is not located at their midpoint. Where, then, does the boundary or point of subjective equality (PSE) between a typical and an atypical exemplar lie? Tanaka et al., 1998, Experiment 4) explored this issue by creating morphs with unequal contributions from the typical and atypical parent (e.g., 55% typical, 45% atypical, 60/40%, 65/35%; see **Figure 3**). The combination at which a morph is judged equally similar to the typical and atypical parent provides an indication of the relative distance of the attractor boundary from each parent.
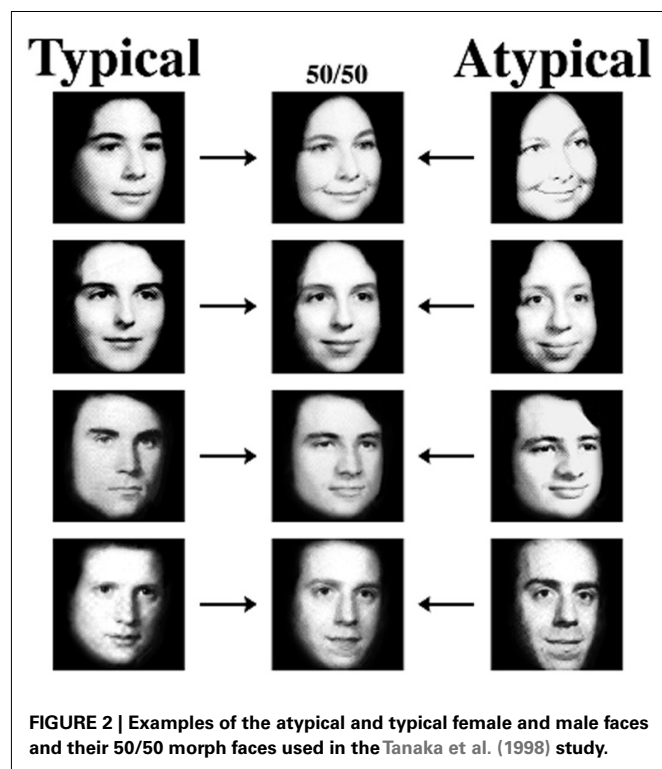
FIGURE 3 | Example of the continuum of morph faces produced by image averaging a typical and atypical face pair. Morph faces in the continuum were produced in 5% intervals ranging from 35% contribution from the atypical (typical) parent to 65%.

Interpolating from the atypicality bias observed in the various morph combination conditions, Tanaka et al. (1998) concluded that a morph containing a 63% contribution from its typical parent and a 37% contribution from its atypical parent is the PSE where the morph is perceived to bear equal resemblance to both parents. This outcome suggests a boundary at roughly two-thirds of the distance between the atypical and the typical parent for the face stimuli used by Tanaka and colleagues.

Tanaka and Corneille (2007) tested a simple yet important prediction of the attractor field model: if a morph is considered more similar to its atypical parent than its typical parent, it should be more difficult to discriminate from its atypical parent than its typical parent. Because 50/50 morphs are highly distinct from both parents and can be identified as such with near-ceiling accuracy, the comparisons of interest concerned morphs with unequal contributions. For example, a 70% atypical, 30% typical morph should be more difficult to discriminate from its atypical parent than a 30% atypical, 70% typical morph should be to discriminate from its typical parent. This prediction falls out of the assumption that atypical exemplars possess the largest attractor fields: the probability that a morph positioned some distance from its atypical parent will fall within the large field of that atypical parent is greater than the probability that a morph positioned the same distance from its typical parent will fall within that typical parent's smaller field.

Tanaka and Corneille, 2007, Experiment 1 tested this prediction with a sequential same-different task in which a parent face was presented for 2 s, a blank screen appeared for 1 s, and the same parent ("same" trials) or a morph with 50, 60, 70, or 80% contribution from that parent ("different" trials) was presented for 1 s. Participants judged whether the second face was the same as or different than the first. Consistent with the attractor field model hypothesis, participants were more accurate in saying "different" when typical parents were paired with 60, 70, and 80% typical morphs than when atypical parents were paired with 60, 70, and 80% atypical morphs. A simultaneous version of the task (Tanaka and Corneille, 2007, Experiment 2) produced similar results.

NON-EUCLIDEAN FACE SPACE AND KRUMHANSL'S DISTANCE-DENSITY HYPOTHESIS

The foregoing results (Tanaka et al., 1998; Tanaka and Corneille, 2007) emphasize an important distinction between *physical* similarity and *psychological* similarity in face perception. Faces that are physically equivalent to each other as measured by their Euclidean distance in face space are not necessarily perceived as

psychologically equivalent. In our work, we found that the atypical face, located in a sparser region of face space with fewer and more distant neighbors, exerts more influence on the morph than the typical face situated in a densely populated region with many close-by neighbors. The main message of the atypicality findings is that the surrounding category context in which the faces are found contributes to their perceived similarity.

Krumhansl's (1978) distance-density hypothesis provides a parsimonious account of how perceived similarity is affected by physical and category factors. In her approach, similarity is determined by two components: (1) the inter-point distance between two exemplars and (2) the spatial density of representations surrounding the exemplars. As applied to faces and face space, these independent contributions are conceptualized in the equation below,

$$\text{Similarity}(\text{face}_1, \text{face}_2) = d(\text{face}_1, \text{face}_2) + \alpha\delta(\text{face}_1) + \beta\delta(\text{face}_2)$$

where the psychological similarity is calculated based on the Euclidean distance between face_1 and face_2, $d(\text{face}_1, \text{face}_2)$, the spatial density of exemplars surrounding face_1, $\delta(\text{face}_1)$, and face_2, $\delta(\text{face}_2)$, and the relative weights, α and β, assigned to those densities. Density (δ) is calculated as the summed difference between the target exemplar and all exemplars in the stimulus domain such that near neighbors will have a larger impact on density than far neighbors. The distance-density hypothesis predicts that if two pairs of faces are equivalent to one another in their Euclidean distance d, the pair located in the sparser, less dense region of the face space will be *perceptually* more similar to each other than the pair situated in a denser region. The morphing evidence confirms this prediction. Although the 50/50 morph is equidistant from the atypical and typical face in similarity space, subjects judged the morph face as bearing a stronger resemblance to the atypical parent than the typical face.

ATTRACTOR FIELDS, FACE SPACES, AND RECOGNITION

To account for the atypicality bias, we conceptualized face representations as attractor fields (Hopfield, 1992). According to this scheme, memory representations, such as familiar faces, form stable points in the face space that carve valleys or basins in its landscape (Tank and Hopfield, 1987). In recognition, analogous to a raindrop moving downhill to minimize its gravitational energy, the to be recognized face stimulus decreases its computational energy by following the path that leads to the nearest attractor basin in representational space. Following this approach, the stimulus need not be a perfect fit with its underlying representation,

only a close enough approximation to fall within the boundaries of the attractor field's basin. The potential activation of any given representation will therefore be directly proportional to the size or span of its attractor field (e.g., representations with large attractor fields will capture a broader range of stimulus inputs).

The attractor basins of visually similar faces are spatially close together in face space whereas faces that are distinct are far apart. The size of the attractor basin determines its influence on face space inputs. Face representations with relatively large attractor fields exert a broad influence on face space and have the potential to capture face inputs possess. Representations with smaller attractor fields have a more limited influence over face inputs. As shown in **Figure 4**, the Euclidean relationships between faces are preserved in the model where the morph face is shown as being equidistant from its typical and atypical parent faces. However, the atypicality bias is the result of the broader attractor field of the atypical face whose influence extends over a greater region of face space. Thus, a 50/50 morph that is situated between a typical and atypical parent face is more likely to fall into the basin of the atypical parent. As mentioned above, the attractor influence of the atypical parent extended to about 37/63% boundary in one of our
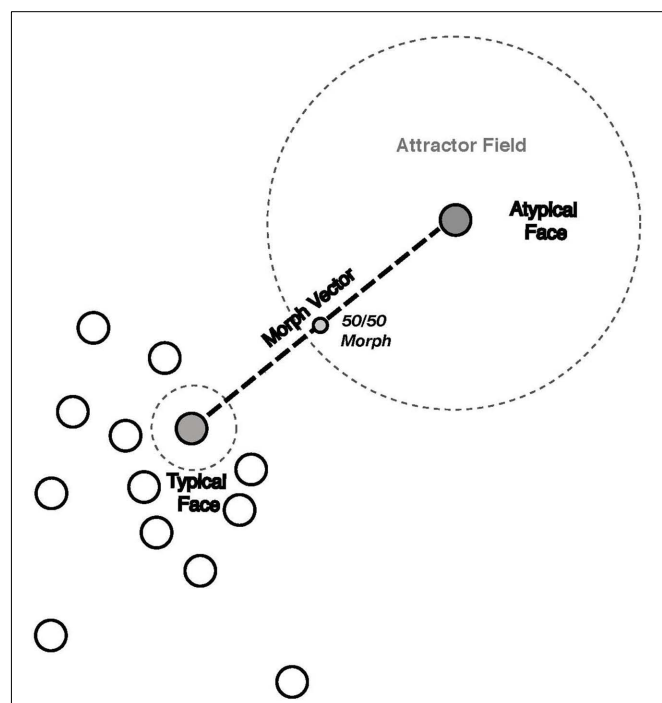
FIGURE 4 | Diagram of the attractor field model. The atypical and typical representations of interest are depicted as filled gray circles. Located in a sparse sub-region of face space, the atypical face representation has a larger attractor field relative to the smaller attractor field of the typical representation situated in a denser sub-region. The morph vector is indicated by the dashed line connecting the atypical and typical parent face representations and morph representations are located along the vector. The 50/50 morph is located at the midpoint of the vector and is equal in its physical distance from its atypical and typical parent representations. The atypicality bias is the result of the 50/50 morph lying closer to attractor boundary of the atypical representation than the typical representation.

experiments, suggesting a point of so-called subjective equality where a morph is equally likely to categorized as more similar to its typical parent or its atypical parent.

ATYPICALITY BIAS FOR NON-FACE OBJECTS

There has been considerable debate as to whether the cognitive processes and neural substrates for face recognition are specific to faces or whether these processes are employed in the recognition of other expert objects (Kanwisher, 2000; Tarr and Gauthier, 2000). A related question is whether the organization of face representations in face space significantly differs from the organization of objects in object space. Structurally, all faces share the same internal features (i.e., eyes, nose, and mouth) arranged in a similar spatial configuration (i.e., the eyes are above the nose which is above the mouth). Therefore, recognition of an individual face depends on the fine-tuned discrimination of its facial features and their spatial configuration. If the typicality effect is dependent on the arrangement of face representations in memory at this subordinate level of visual analysis, it is plausible that atypicality effect may be exclusive to faces.

Although most objects (e.g., "chair," "bird," "car") are fastest to be recognized at the basic level of categorization (Rosch et al., 1976; Grill-Spector and Kanwisher, 2005; Mack et al., 2009), objects are identifiable at the specific, subordinate level (e.g., "rocking chair," "sparrow," "Volkswagon") on the basis of their structural properties. There is sufficient structural variation within an object category such that some members are regarded as bearing a stronger resemblance to the prototypical category shape than other members (Jolicoeur et al., 1984; Murphy and Brownell, 1985). The structural typicality is reflected in category responses where the typical exemplars (e.g., robin, sparrow) are faster to verify as category members (e.g., bird) than less typical exemplars (e.g., penguins, ostriches; Jolicoeur et al., 1984; Murphy and Brownell, 1985). Hence, the grain of resolution is sufficient to arrange non-face objects in a similarity space that is equivalent to faces. Moreover, the organization of object spaces may follow a similar pattern of distribution as faces where there is a higher density of typical objects located at the origin of the space and sparser density of atypical objects at the periphery.

To assess the structure of non-face object spaces, typicality ratings were collected for the car and bird exemplars to determine the typical and atypical members of these categories. The typical and atypical exemplars were then morphed together (see **Figure 5**). A naive group of participants judged whether the morphed object more closely resembled the atypical object parent or the typical object parent. The prediction was that if the typicality effect is special to faces, participants should be just as likely to judge the bird or automobile morph more similar to its typical or atypical parent. On the other hand, if the perception of objects, like faces, is influenced by structural typicality, participants should demonstrate an atypicality bias for the birds and automobiles. Tanaka and Corneille, 2007, Experiment 3) found a significant atypicality bias for both birds (66%) and cars (58%), extending the phenomenon to non-face objects. These findings indicate that the notion of a similarity space applies to birds and cars (and, presumably, to many other object classes beyond faces) and suggests that the attractor field model is applicable to non-face object
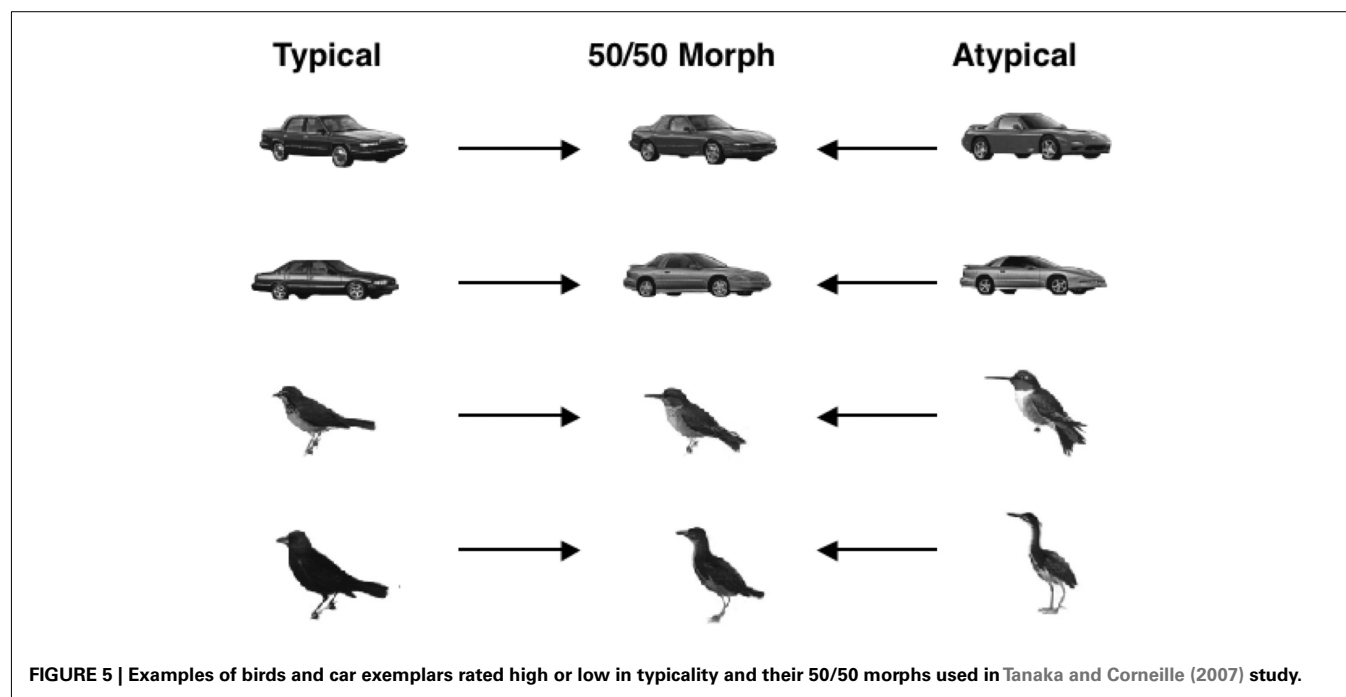
FIGURE 5 | Examples of birds and car exemplars rated high or low in typicality and their 50/50 morphs used in Tanaka and Corneille (2007) **study.**

classes in which a given exemplar may be placed along a continuum of typicality.

ACQUISITION OF THE ATYPICALITY BIAS

How do atypical exemplars come to attract a wider range of inputs than typical exemplars? The dynamic systems approach was intuitively appealing to us for several reasons. First, it stresses the qualities of the time-dependent, ever changing representational states of the perceptual system that respond spontaneously to the environmental inputs (Beer, 2000; Spencer and Schöner, 2003). Second, dynamic systems are also stable in that their behaviors are consistent and predictable and flexible in that they can accommodate and change in response to new learning experiences. This characterization seems well suited to describe the face recognition system that is constantly warped and molded to encode and incorporate new face memories. Third, dynamics systems preserve the physical and psychological properties of face space. Kantner and Tanaka (in press) proposed that the physical structure of items alone cannot drive the atypicality bias because their distinctiveness is a function of their physical relationship to other category members. Until the category structure is learned, then, there should be little or no basis for perceiving certain items as less typical than others.

Kantner and Tanaka tested the hypothesis that the atypicality bias emerges only after category learning has occurred. Because natural categories such as those used in previous studies of the atypicality bias are known to participants from everyday experience, Kantner and Tanaka created artificial stimuli called "blobs" whose category structure could be learned within an experimental session. Each blob category was formed around a central prototype; typical category exemplars were created by making small deviations on the prototype while atypical members were composed of larger deviations. Each category consisted of four typical

and four atypical exemplars, thus equating frequency across levels of typicality and varying only structural atypicality (in natural categories, by contrast, frequency, and typicality are often confounded).

Pilot testing using the preference task described above (e.g., Tanaka et al., 1998) established two blob categories eliciting a statistically negligible level of bias. In order to test the effects of learning on the atypicality bias, Kantner and Tanaka gave a preference task before and after participants trained to categorize members of the two categories to criterion (Experiment 1). Consistent with their hypothesis, participants showed a small, non-significant atypicality bias (52%) before category training but a significant post-training bias (58%) that was consistent in magnitude with those of experiments using natural categories.

In Experiment 2, Kantner and Tanaka demonstrated that category training *per se* is not necessary to induce an atypicality bias within an experimental session. They replaced the training phase in Experiment 1 with a simple pleasantness rating task and observed similar results: a small, non-significant atypicality bias (53%) before the ratings task and a significant atypicality bias after (nearly 60%). This result suggests that participants do not need be engaged in an explicit categorization task to apprehend the normative appearance of members within a category, a finding consistent with past research on implicit category learning. Further, they suggest that this form of learning is sufficient to produce an atypicality bias.

A developmental test of the atypicality effect for faces and objects

The finding that the atypicality bias can be shown to accrue with experience using artificial stimuli raises a related hypothesis concerning the developmental trajectory of the bias. Tanaka et al. (2011) predicted that children who have not gained sufficient experience with natural categories to possess a fully formed sense

of the associated category structure should show a diminished or non-existent bias in perceiving members of those categories. Tanaka et al. (2011) tested this hypothesis by presenting faces, birds, and cars in a preference task to children aged 3–4, 5–6, 7–8, 9–10, and 11–12 as well as college-aged adults. Surprisingly, they found a substantial atypicality bias for faces and birds that was statistically invariant across age groups. Although the bias was stronger for birds (71%) than for faces (63%), both biases were evident from the earliest age group tested. An additional unexpected finding concerned the car stimuli, which did not produce an atypicality bias in any of the age groups. Tanaka et al. (2011) speculated that their car stimuli (normed in 1997) may not have conformed to current standards of typicality. The bird and face results, however, suggested that the mental representation of these category structures is well established even in very young children.

MODELING ATTRACTOR FIELDS

Can the predictions of the distance-density hypotheses be tested in a neural network model? In previous simulations, neural networks have been applied to study caricature recognition (Tanaka and Simon, 1996) and atypicality effects in face recognition processes (Burton et al., 1990; Valentine and Ferrara, 1991). Connectionist implementation has been shown to preserve the Euclidean properties of face space where structurally similar faces are located closer together in face space and structurally dissimilar face are farther apart (Burton et al., 1990; Valentine and Ferrara, 1991; Tanaka and Simon, 1996; Lewis and Johnston, 1998). In this approach, the features of a face, such as its outline, internal features, and spatial configuration are abstractly represented as feature units in a face vector. The similarity between any two face vectors can be expressed by the angle of their dot product such that similar faces will form a relatively small angle and dissimilar faces will have a relatively large angle.

As a test of the atypicality bias, Tanaka et al. (1998) constructed a neural model composed of three layers: a 10-unit input layer, a 5-unit middle (hidden) layer, and a 4-unit output layer. The typical vectors were constructed such that they differed from each other by two feature units (i.e., [1, 0, 1, 0, 1, 0, 1, 0, 1, 0], [1, 1, 0, 0, 1, 0, 1, 0, 1, 0], [1, 0, 0, 0, 1, 1, 1, 0, 1, 0]) whereas the one atypical vector differed from the typical vectors by six feature units (e.g., [0, 0.5, 0.5, 1, 0, 0.5, 1, 0, 0.5, 1]. By virtue of their vector structures, the typical vectors are located nearer the origin of face space whereas the atypical vector is located further away from the origin. In the simulation, the typical and atypical feature vectors provided the inputs to the model with each vector associated with a unique unit at the output. Training continued for 100 epochs. After learning, morph vectors were created by averaging the values of the atypical face vector with and the values of the typical feature vectors. For example, the features of Typical Feature Vector 1 [1, 0, 1, 0, 1, 0, 1, 0, 1, 0] and Atypical Feature Vector 4 [0, 0.5, 0.5, 1, 0, 0.5, 1, 0, 0.5, 1] were averaged to produce the Morph 1–4 vector of [0.5, 0.25, 0.75, 0.5, 0.5, 0.25, 1, 0, 0.75, 0.5].

The critical test of the atypicality bias was the network's response to the morph vector. When the three typical-atypical morph vectors were presented to the network, stronger activation was produced in the atypical face output unit than of the typical face output units. Consistent with the empirical findings, the

neural network showed a bias for atypical representations over typical representations. The simulation demonstrates that neural network recognition is sensitive to distribution of representations in the similarity space. These findings provide computational support for the claim that recognition is determined by the match between the input and the associated output and the density of the surrounding representations.

Bartlett and Tanaka (1998) extended the predictions of an attractor network model using real female face images. In their simulation, a 1000-dimensional Hopfield network was trained on 32 patterns consisting of Gabor filter outputs for eight typical faces and eight atypical faces, and their mirror reversed images. The network was tested by presenting the image representation of 32 female 50/50 morph faces (16 combinations of 4 typical and 4 atypical faces, plus mirror reversed images). The sustained activity pattern for each of the morph images was compared to that for its two parent images. The simulation results showed that the morphed face fell into the basin of attraction for the atypical parent face six times, whereas it never fell into the basin of attraction for the typical parent face. The remaining 26 morphed faces settled to a distinct pattern of sustained activity that differed from both parents. Of these, 22 were closer to the atypical than typical parent, two were equidistant, and two were closer to the typical than atypical parent. Overall, the attractor network would have selected the atypical parent 87.5% of the time based on Euclidean distance of the sustained activity patterns to the parent patterns. The attractor network model with real face images exhibited an atypicality bias in the sustained activity patterns for the morphed faces.

INFORMATION MAXIMIZATION HYPOTHESIS

Although a several computational models demonstrated an atypicality bias, the most parsimonious account is provided by information maximization. This account reduces the model to a simple goal of optimizing information transfer in sensory coding given the statistics of face images in the visual environment. The information maximization account is compatible with Krumhansl's distance-density hypothesis, and generalizes this idea to a computational principle from information theory. Although the back-propagation model illustrated the general concept that an atypicality bias can emerge from the statistics of the training data, information maximization is an unsupervised learning strategy that is a more biologically plausible model of learning in the brain. Information maximization also provides a more parsimonious account than the attractor network model, which assumes that coding takes the form of sustained patterns of activity, and can take many iterations to settle. Moreover in a Hopfield attractor network, decorrelation is a component of the learning rule that is necessary in order to encode highly similar patterns (i.e., faces) as distinct patterns of sustained activity (Kanter and Sompolinsky, 1987). Decorrelation is closely related to information maximization (Bell and Sejnowski, 1997), and hence the atypicality bias that emerged in the attractor network could be related to the underlying principle of information maximization.

Information maximization is a principle from information theory (Shannon, 1948) describing a coding strategy for maximizing the information transfer capacity of a communication system by ensuring that all response levels are used with equal frequency.

More response levels are allocated for high probability ranges of the input signal, and fewer response levels are allocated for low probability ranges of the input signal. In a system with limited dynamic range, information transfer is maximized by matching the response function to the cumulative probability density of the input signal.

This concept is illustrated in **Figure 6**, adapted from Laughlin (1981). The upper curve shows a probability density function for the magnitude of a stimulus property in the world, such as contrast level. The lower curve shows the response function that implements the information maximization strategy. In this example the output has 10 response states, corresponding to 10 "just noticeable differences" (JND) in response. The response function insures that the interval between each response level encompasses an equal area under the probability distribution for the stimulus property, so that each state is used with equal frequency. In the limit where the states are vanishingly small this response function corresponds to the cumulative probability function for stimulus intensities, which is the probability that the stimulus takes on value *x* or less. Note that the slope is steep where the probability density is high, and shallow where the probability density is low. As a consequence, equal changes in stimulus intensity have different effects on the response. The slope of the response function is given by the probability of the stimulus, which can lead to a number of perceptual effects, all stemming from a basic sensory coding goal of optimal information transfer.
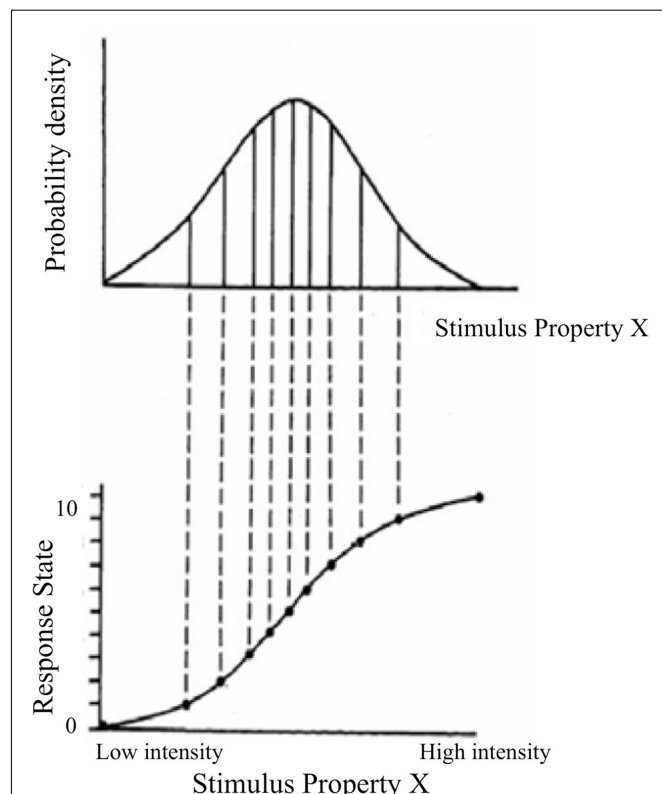
A body of research in computational neuroscience explores how information maximization may be central principles in sensory coding early in the visual system (see Simoncelli and Olshausen, 2001 for a review). For example, Laughlin (1981) showed that luminance contrast coding in the blowfly visual system is consistent with information maximization, where the cumulative probability density of contrasts in the blowfly environment predicted depolarization of a class of cells.

These principles may be relevant to how we think about perceptual coding in higher visual processes such as face recognition as well (see Bartlett, 2007 for a review). Perceptual responses, as revealed by similarity measures and JND, may be influenced by information maximization in sensory coding. For example, face discrimination is superior for same-race than other-race faces (Walker and Tanaka, 2003), which is consistent with a perceptual transfer function that is steeper for face properties in the high density portion of the distribution in an individual's perceptual experience (i.e., same-race faces) than for face properties in the low-density portion of the distribution (other-race faces). These concepts are also explored in Webster and MacLeod (2011).

Here we show how the information maximization principle of optimal perceptual coding would account for face typicality effects, and in particular the atypicality bias. We begin with the assumption that typical faces are from high density regions of the probability distribution for a set of physical characteristics, and that atypical faces have characteristics with lower probability on at least some dimensions, as illustrated in **Figure 7A**. For example, a particular physical characteristic might be distance between the eyes. **Figure 7B** illustrates the shape of the perceptual response function under the information maximization principle. In our example, this response function could be JND for facial identity as a function of interocular distance. The typical face is mapped to a region of the response function that is steep, and the atypical face is mapped to a region of the response function that is shallow. Due to the shape of the response function, the physical 50% morph between the typical and atypical face is projected closer to the atypical parent in the perceptual response, as shown by the dashed line onto *Y*-axis.
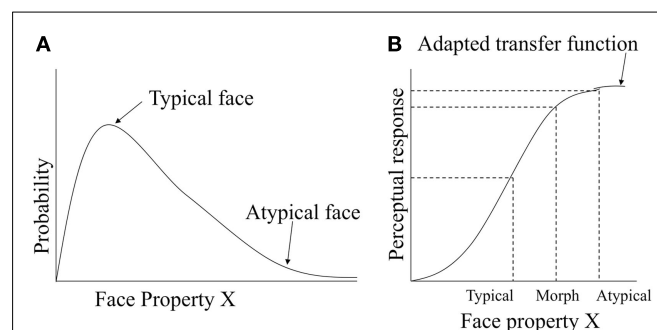

FIGURE 6 | Illustration of information maximization principle. Adapted from Laughlin (1981).

FIGURE 7 | Information maximization account of atypicality bias. (A) Probability density function for typical and atypical faces. (B) Perceptual response for typical, atypical and the typical-atypical morph faces as predicted by adaptive transfer function.

Bartlett and Tanaka (1998) tested the assumption that images of faces rated as typical are from a higher density region of the probability distribution than images of faces rated as atypical. 62 female face images were convolved with a set of Gabor filters, which are an approximate model of the receptive fields of simple cells in primary visual cortex (Daugman, 1988). The image graylevels were passed through a bank of Gabor filters at four spatial scales (32, 26, 8, and 4 pixels per cycle) and four orientations, sampled at 255 spatial locations from the original 120×120 pixel images. The outputs of sine and cosine Gabor filters were squared and summed, and then the contrast was normalized by dividing by the activity across all orientations and scales at each spatial location. Such contrast normalization has been described in primary visual cortex (Heeger, 1991). Typical faces were closer to the mean face than atypical faces, when faces are represented as a bank of Gabor filter outputs. The origin of the face space was estimated as the mean across the set of 62 female faces. There was no significant difference in the distance to the mean face for the graylevel images. However, for the normalized Gabor representation, the typical faces were closer to the mean face [$t(14) = 2.3$, $p < 0.05$]. We also found that both for the original graylevel images, and for the Gabor representations of those images, there was indeed a higher density near faces rated as typical than faces rated as atypical. The mean distance of each face to its nearest neighbor in face space was compared for the eight faces rated most typical and the eight faces rated most atypical. There was a shorter distance to the first neighbor of a typical face than an atypical face for raw graylevel images [$t(14) = 2.8$, $p < 0.05$], and the difference was more pronounced for the normalized Gabor representation [$t(14) = 4.3$, $p < 0.001$]. Therefore, the optimal transfer function predicted by information maximization would have a steeper slope near the typical faces than near atypical faces, supporting the model illustrated in **Figure 7B**.

The infomax account makes an additional prediction. Although it is well known that faces rated as "atypical" tend to also be easier to recognize, this model predicts that subjects will be *less* sensitive to small perturbations in the physical properties of atypical faces. This prediction was born out by the same-different discrimination experiments of Tanaka and Corneille (2007). Sensitivity to small perturbations in the face created by morphing was significantly lower for atypical than typical faces.

A similar account applies to perceptions of morph between same-race and other-race faces. A morph stimulus that is halfway on a physical continuum between a same-race face and an other-race face is typically perceived as more similar to the other-race face (Kaping et al., 2002). This effect follows from the model shown in **Figure 7B**, if we assume a higher probability density of physical properties of same-race faces in an individuals perceptual experience than for the values these properties take on for other-race faces. Again, the 50% physical morph is mapped to a shallow-sloped region of the perceptual response function, and is closer to the response for other-race faces than same-race faces.

GENERAL DISCUSSION

In this review, we describe a simple but powerful test of structural and psychological accounts of perceptual similarity. In our paradigm, two parent faces – an atypical parent face and typical parent face – are morphed together to form a child morph face that constitutes the physical average between its parents. The morph face receives a 50–50 contribution from both parents, and therefore, based on its structural (i.e., physical) properties, it should equally resemble the atypical and typical parent face. However, contrary to the structural view, participants systematically judge the morph face as bearing a stronger resemblance to the atypical parent than the typical parent. The atypicality bias indicate that perceptual similarity is not solely determined by its structural inputs, but by psychological factors related to context of those inputs and the experience of the observer.

We propose that the bias for the atypical parent can be explained by Krumhansl's Distance-Density hypothesis in which the perceived similarity between two faces depends on their physical resemblance as well as their relative locations in the face space. According to Distance-Density hypothesis, perception of the morph stimulus is more heavily weighted toward the atypical face by virtue of its location in face space. An atypical face resides in a sparser region of face space where there is less competition for recognition than typical faces. We propose that atypical faces have broad attractor fields that allow them to capture a wider range of perceptual inputs than typical faces. Further tests of the attractor field hypothesis have shown that the atypical bias does not reflect a memory bias for more atypical, distinctive items (Tanaka et al., 1998, Experiment 3) but indicates a perceptual insensitivity to changes in an atypical face relative to changes in a typical face (Tanaka and Corneille, 2007, Experiment 1). Nor is atypicality bias "special" to faces, but extends to the perception of non-face objects, such as birds (Tanaka and Corneille, 2007, Experiment 2) and these "object" spaces are established relative early on in development (Tanaka et al., 2011).

Although the developmental evidence suggests that the structure of face space structure is established relatively early on in development (Nishimura, the category structure for non-face objects is more malleable and responsive to the effects of perceptual learning. In a recent study, we (Kantner and Tanaka, in press), created two category prototypes of polygon shapes (i.e., blobs) and a family of exemplars for each prototype (Curran et al., 2002); some blob exemplars were created such that they display modest variation from the family prototype (typical exemplars) whereas other blob exemplars showed greater variation (atypical exemplars). Participants were asked to judge where the morphs of the typical and atypical exemplars showed a stronger resemblance to the typical parent or the atypical parent before and after category training. Whereas participants did not show a preference for the atypical exemplar before category learning, they demonstrated a reliable atypicality bias after category learning (Kantner and Tanaka, in press). The influence of typicality bias in perceptual learning was further validated in several neural network simulations and is a general perceptual principle that can be described by information maximization theory (Bartlett and Tanaka, 1998; Bartlett, 2007). After learning atypical and typical inputs, the neural network simulations showed that atypical-typical morph inputs elicited a greater response in the atypical output unit than the typical output units. Collectively, the empirical and simulation results indicate that perceptual similarity is influenced by the physical properties, category structure, and learning histories of the stimuli.

In this paper, we argue that psychological similarity in object perception can be dissociated from physical similarity. To support this claim, we present evidence in which a morphed object that is physically equivalent to a typical and atypical parent object is perceived to be more psychologically similar to the atypical object than the typical object. These results suggest that beyond physical structure, the perceived likeness of an object is influenced by the organization of surrounding exemplars in the category space. Atypical objects exert more influence on judgments of perceived similarity than typical objects. Of course, whether an object is regarded as typical or atypical is not *a priori* pre-determined, but depends on the perceptual experience of the observer Thus, the atypicality bias exemplifies the constant refashioning of our perceptions through learning, and, more broadly, the influence of higher order cognitive process on the perceptual apparatus.

ACKNOWLEDGMENTS

REFERENCES

Bartlett, J. C., Hurry, S., and Thorley, W. (1984). Typicality and familiarity of faces. *Mem. Cognit.* 12, 219–228.

Bartlett, M. S. (2007). Information maximization in face processing. *Neurocomputing* 70, 2204–2217.

Bartlett, M. S., and Tanaka, J. W. (1998). An attractor field model of face representation: effects of typicality and image morphing. *Paper Presented at the Psychonomics Society Satellite Symposium on Object Perception and Memory (OPAM),* Dallas, TX.

Beer, R. D. (2000). Dynamical approaches to cognitive science. *Trends Cogn. Sci. (Regul. Ed.)* 4, 91–99.

Bell, A. J., and Sejnowski, T. J. (1997). The independent components of natural scenes are edge filters. *Vision Res.* 37, 3327–3338.

Burton, M., Bruce, V., and Johnston, R. A. (1990). Understanding face recognition with an interactive activation model. *Br. J. Psychol.* 81, 361–380.

Cabeza, R., Bruce, V., Kato, T., and Oda, M. (1999). The prototype effect in face recognition: extension and limits. *Mem. Cognit.* 27, 139–151.

Curran, T., Tanaka, J. W., and Weiskopf, D. (2002). An electrophysiological comparison of visual categorization and recognition memory. *Cogn. Affect. Behav. Neurosci.* 2, 1–18.

Daugman, J. (1988). Complete discrete 2-D Gabor transforms by neural networks for image analysis and compression. *IEEE Trans. Acoust.* 36, 1169–1179.

Garner, W. R. (1974). *The Processing of Information and Structure*. Potomac, MD: Erlbaum.

Going, M., and Read, J. D. (1974). Effects of uniqueness, sex of subject, and sex of photograph on facial recognition. *Percept. Mot. Skills* 39, 109–110.

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.

Heeger, D. (1991). "Nonlinear model of neural responses in cat visual cortex," in *Computational Models of Visual Processing*, eds M. Landy and J. Movshon (Cambridge, MA: MIT Press), 119–133.

Hopfield, J. J. (1992). Neural networks and physical systems with emergent collective computational abilities. *Proc. Natl. Acad. Sci. U.S.A.* 79, 2554–2558.

Johnston, R. A., and Ellis, H. D. (1995). Age effects in the processing of typical and distinctive faces. *Q. J. Exp. Psychol. A* 48, 447–465.

Jolicoeur, P., Gluck, M. A., and Kosslyn, S. M. (1984). Pictures and names: making the connection. *Cogn. Psychol.* 16, 243–275.

Kanter, I., and Sompolinsky, H. (1987). Associative recall of memory without errors. *Phys. Rev. A* 35, 380–392.

Kantner, J., and Tanaka, J. W. (in press). Experience produces the atypicality bias in object perception. *Perception*.

Kanwisher, N. (2000). Domain specificity in face perception. *Nat. Neurosci.* 3, 759–763.

Kaping, D., Duhamel, P., and Webster, M. A. (2002). Adaptation to natural face categories. *J. Vis.* 2, 128.

Krumhansl, C. L. (1978). Concerning the applicability of geometric models to similarity data: the interrelationship between similarity and spatial density. *Psychol. Rev.* 85, 445–463.

Laughlin, S. (1981). A simple coding procedure enhances a neuron's information capacity. *Z. Naturforsch. C* 36, 910–912.

Leopold, D. A., O'Toole, A. J., Vetter, T., and Blanz, V. (2001). Prototype-referenced shape encoding revealed by high-level after effects. *Nat. Neurosci.* 4, 89–94.

Lewis, M. B. (2004). Face-space-R: towards a unified account of face recognition. *Vis. Cogn.* 11, 29–69.

Lewis, M. B., and Johnston, R. A. (1998). Understanding caricatures of faces. *Q. J. Exp. Psychol.* 51(2), 321–346.

Light, L. L., Kayra-Stuart, F., and Hollander, S. (1979). Recognition memory for typical and unusual faces. *J. Exp. Psychol. Hum. Learn.* 5, 212–228.

Mack, M. L., Wong, A. C., Gauthier, I., Tanaka, J. W., and Palmeri, T. J. (2009). Time course of visual object categorization: fastest does not necessarily mean first. *Vision Res.* 49, 1961–1968.

Murphy, G., and Brownell, H. (1985). Category differentiation in object recognition: typicality constraints on the basic category advantage. *J. Exp. Psychol. Learn. Mem. Cogn.* 11, 70–84.

Posner, M. I., and Keele, S. W. (1968). On the gensis of abstract ideas. *J. Exp. Psychol.* 77, 353–363.

Rhodes, G., Byatt, G., Tremewan, T., and Kennedy, A. (1997). Facial distinctiveness and the power of caricatures. *Perception* 26, 207–223.

Rhodes, G., and Jeffery, L. (2006). Adaptive norm-based coding of facial identity. *Vision Res.* 46, 2977–2987.

Rhodes, G., Jeffery, L., Watson, T. L., Jaquet, E., Winkler, C., and Clifford, C. W. G. (2004). Orientation-contingent face aftereffects and implications for face-coding mechanisms. *Curr. Biol.* 14, 2119–2123.

Rosch, E., Mervis, C., Gray, W., Johnson, D., and Boyes-Braem, P. (1976). Basic objects in natural categories. *Cogn. Psychol.* 8, 382–439.

Shannon, C. E. (1948). A mathematical theory of communication. *Bell Syst. Tech. J.* 27, 379–423, 623–656.

Shepard, R. N. (1964). Attention and the metric structure of the stimulus space. *J. Math. Psychol.* 1, 54–87.

Simoncelli, E. P., and Olshausen, B. A. (2001). Natural image statistics and neural representation. *Annu. Rev. Neurosci.* 24, 1193–1216.

Spencer, J. P., and Schöner, G. (2003). Developmental science. *Dev. Sci.* 6, 392–412.

Tanaka, J., and Simon, V. (1996). Caricature recognition in a neural network. *Vis. cogn.* 3, 301–320.

Tanaka, J. W., and Corneille, O. (2007). Typicality effects in face and object perception: further evidence for the attractor field model. *Percept. Psychophys.* 69, 619–627.

Tanaka, J. W., and Farah, M. J. (2003). "Holistic face recognition," in *Analytic and Holistic Processes in the Perception of Faces, Objects and Scenes*, Vol. 2, eds M. Peterson and G. Rhodes (New York: University Oxford Press), 53–91.

Tanaka, J. W., Giles, M., Kremen, S., and Simon, V. (1998). Mapping attractor fields in face space: the atypicality bias in face recognition. *Cognition* 68, 199–220.

Tanaka, J. W., Meixner, T. L., and Kantner, J. (2011). Exploring the perceptual spaces of faces, cars and birds in children and adults. *Dev. Sci.* 14, 762–768.

Tank, D. W., and Hopfield, J. J. (1987). Collective computation in neuron like circuits. *Sci. Am.* 257, 104–114.

Tarr, M. J., and Gauthier, I. (2000). FFA: a flexible fusiform area for subordinate-level visual processing automatized by expertise. *Nat. Neurosci.* 3, 764–769.

Valentine, T. (1991). A unified account of the effects of distinctiveness, inversion, and race in face recognition. *Q. J. Exp. Psychol.* 43A, 161–204.

Valentine, T., and Bruce, V. (1986). The effect of race inversion and encoding activity upon face recognition. *Acta Psychol. (Amst.)* 61, 259–273.

Valentine, T., and Ferrara, A. (1991). Typicality in categorization, recognition and identification: evidence from face recognition. *Br. J. Psychol.* 82, 87–102.

Vokey, J. R., and Read, J. D. (1992). Familiarity, memorability, and the effect of typicality on the recognition of faces. *Mem. Cognit.* 20, 291–302.

Walker, T. M., and Tanaka, J. W. (2003). An encoding advantage for own-race versus other-race faces. *Perception* 23, 1117–1125.

Webster, M. A., Kaping, D., Mizokami, Y., and Duhamel, P. (2004). Adaptation to natural facial categories. *Nature* 428, 557–561.

Webster, M. A., and MacLeod, D. I. (2011). Visual adaptation and face perception. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 366, 1702–1725.

Webster, M. A., and MacLin, O. H. (1999). Figural aftereffects in the

perception of faces. *Psychon. Bull. Rev.* 6, 647–653.

Wolberg, G. (1990). *Digital Image Warping.* Los Alamitos, CA: IEEE Computer Science Society Press.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Micro-valences: perceiving affective valence in everyday objects

Sophie Lebrecht[1], **Moshe Bar**[2,3], **Lisa Feldman Barrett**[2,3,4] **and Michael J. Tarr**[5]*

[1] Department of Cognitive, Linguistic and Psychological Sciences, Brown University, Providence, RI, USA
[2] Department of Psychiatry, Massachusetts General Hospital/Harvard Medical School, Boston, MA, USA
[3] Martinos Center for Biomedical Imaging, Massachusetts General Hospital/Harvard Medical School, Boston, MA, USA
[4] Department of Psychology, Northeastern University, Boston, MA, USA
[5] Department of Psychology, Center for the Neural Basis of Cognition, Carnegie Mellon University, Pittsburgh, PA, USA

Perceiving the affective valence of objects influences how we think about and react to the world around us. Conversely, the speed and quality with which we visually recognize objects in a visual scene can vary dramatically depending on that scene's affective content. Although typical visual scenes contain mostly "everyday" objects, the affect perception in visual objects has been studied using somewhat atypical stimuli with strong affective valences (e.g., guns or roses). Here we explore whether affective valence must be strong or overt to exert an effect on our visual perception. We conclude that everyday objects carry subtle affective valences – "micro-valences" – which are intrinsic to their perceptual representation.

Keywords: affective valence, visual object perception, object recognition, micro-valence, object preference

INTRODUCTION

While grabbing a coffee mug from the cupboard a phone call diverts your attention and, in the split second before answering the call, you select a seemingly random mug from your collection. What factors guide you to choose one mug over the others? Here we answer this question by proposing that *the majority of physical objects* around us possess a subtle valence – a "micro-valence" – which ranges in magnitude but is always present. Although it is certainly the case that others have made similar proposals (Lang et al., 1998; Cunningham et al., 2004; Barrett and Bar, 2009; Colibazzi et al., 2010), the majority of valence studies have focused on objects with strong valence (e.g., guns or roses). More importantly, to the extent that everyday objects automatically evoke some perception of valence, valence can be considered a higher-level object property that connects vision to behavior. In this sense, as with other higher-order properties of objects, we suggest that valence is automatically perceived and incorporated into the mental representations of objects (**Figure 1**). One implication of this claim is that "standard" models of object perception may need to be rethought (Ungerleider and Mishkin, 1982; Serre et al., 2007). Moreover, by exploring high-level, functional object properties such as valence, we gain a better understanding of how our perceptual system translates visual information into tangible cues for behavior.

Assigning affective values to mental constructs is not a new concept. Social psychologists have been investigating "attitude objects" for decades (Fazio et al., 1986, 1995; Bargh et al., 1992; Greenwald et al., 1998; Giner-Sorolla et al., 1999). Attitude objects can be conceptualized as a person, place, concept, word, or thing that is readily associated with a particular attitude from memory (e.g.,

associating "fox hunting" with cruelty, or "yoga" with peacefulness). Although related, our conceptualization of micro-valence differs in that we are not focused on individuals' cognitive attitudes to concepts, but rather on the mechanisms underlying how the visual system computes affective valence from perceptual representations. Thus, valence is not a label or judgment applied to the object postrecognition, but rather an integral component of mental object representations. While one might argue that valence is too abstract to be considered an object property, we should note that higher-order properties such as material (Cant and Goodale, 2007; Cant et al., 2008), category (Grill-Spector and Kanwisher, 2005), and function (Beauchamp et al., 2002) are also inferred from multi-modal and non-perceptual sources.

Research investigating the links between affect and perception most commonly relies on object, face, or scene stimuli that generate strong, well-defined valences (Greenwald et al., 1998; Avero and Calvo, 2006; Calvo and Avero, 2008; Rudrauf et al., 2008; Colibazzi et al., 2010; Weierich et al., 2010). In contrast, few studies have examined how more subtle valences are perceived in common objects (e.g., lamps, clocks, or coffee cups; McManus, 1980; Giner-Sorolla et al., 1999; Rentschler et al., 1999; Duckworth et al., 2002; Bar et al., 2006; Bar and Neta, 2007). Of late, studies examining weaker valences in the form of the value placed on everyday objects, such as snack foods (Constantino and Daw, 2010; Krajbich et al., 2010; Litt et al., 2011), have been arisen due to interest in "neuroeconomics." Still, studies relying on extreme responses to highly affective stimuli remain our primary source for understanding the role of affect in perception and cognition. Although critical for outlining the core components of affective processing,

we suggest that the field should take a more fine-grained look at how subtle affective valences influence our everyday perceptions and interactions with the world.
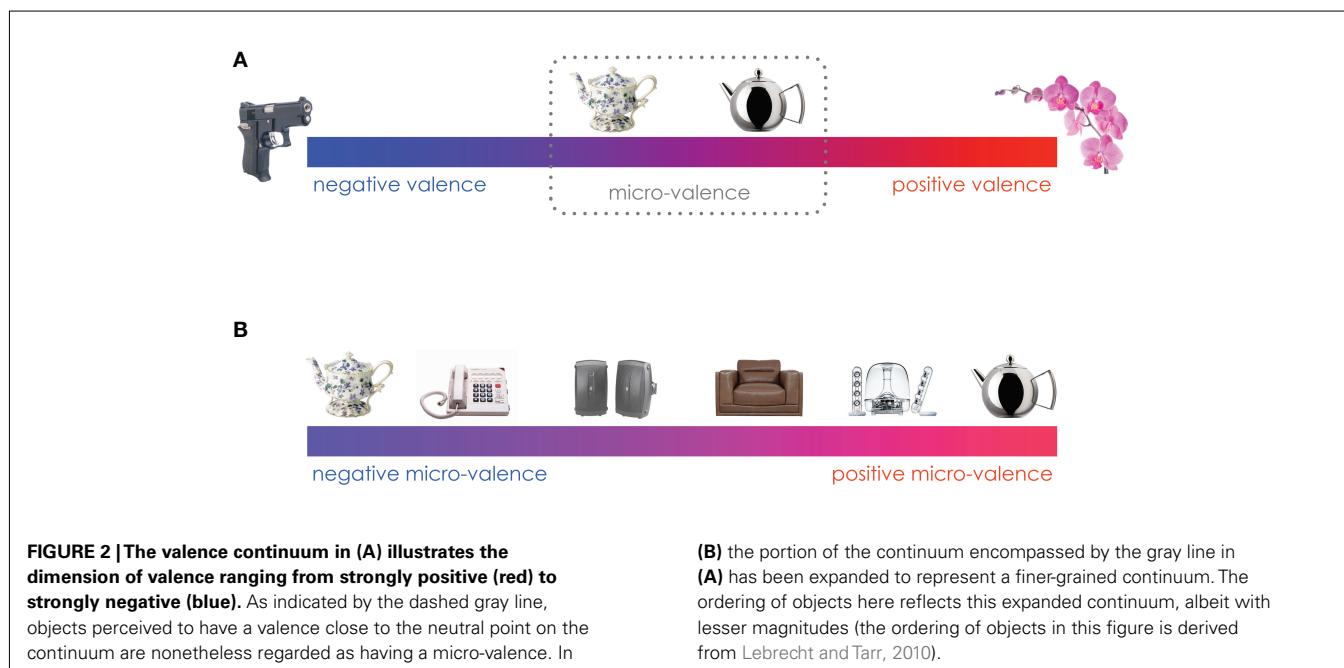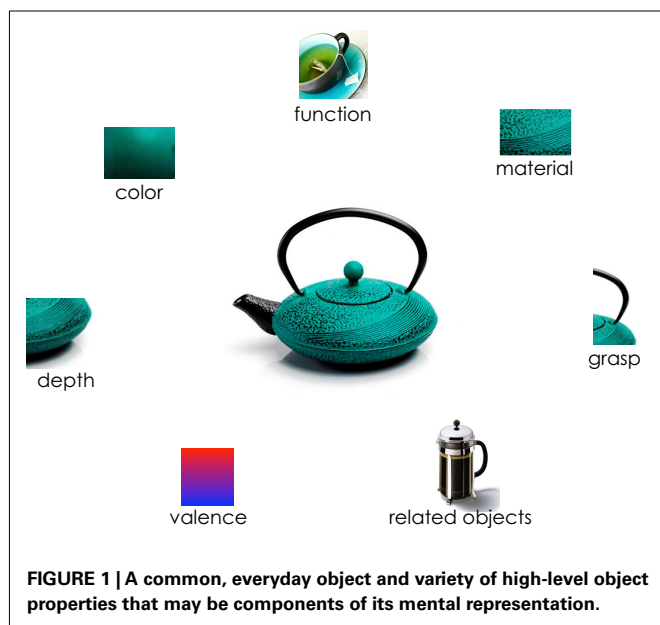
Affect itself is commonly defined along two continuous dimensions: valence (pleasantness) and arousal (activation; Russell, 1980; Russell and Carroll, 1999; Barrett, 2006). With respect to micro-valences we are primarily focused on the single dimension of valence. Although this approach assumes that valence is continuous, we should emphasize that there are still notable differences in the intensity of and the variance in one's responses to strongly and weakly valenced objects, for example, a bloody weapon and a coffee mug. A bloody weapon will prompt a highly intense affective

FIGURE 1 | A common, everyday object and variety of high-level object properties that may be components of its mental representation.

response in most individuals and the between-subject variance for these strong responses is expected to be relatively low. That is, the vast majority of us will experience the same highly negative feeling when viewing such a negative stimulus. In contrast, a coffee mug will likely generate a weaker or more subtle response, which we refer to as that object's micro-valence. This valence is described as "micro" because the intensity of the response is less than the bloody weapon or other similarly strongly affective objects. However, this weak intensity should not be confused with a weak *effect*. There are many small, yet robust effects, for example Sternberg's (1966) classic digit memory search exhibited an effect of less than 40 ms per an item in memory (Sternberg, 1966).

THE PREVALENCE OF VALENCE

Affective valence is highly prevalent in our visual environment. That is, objects in a scene prompt the perception of valence that varies in strength and intensity along the valence continuum (**Figure 2A**). As mentioned, while construing valence as a continuous dimension is not a new idea (Russell, 1980; Russell and Carroll, 1999; Colibazzi et al., 2010), many theorists have focused on strongly valenced objects at the extreme ends of the continuum[1]. Here we focus on the subtle differences in micro-valence for objects nearer to the center of the valence continuum. We posit that observers perceive a subtle valence for objects even in the region immediately surrounding neutral. That is, everyday objects such as chairs and clocks possess a micro-valence and so are either slightly preferred or anti-preferred. Indeed, when the

---

[1] By way of comparison, when researchers working within the framework of attitude objects refer to a continuum they are typically referring to the continuous relationship between associated strength and automaticity (Fazio et al., 1986). This means that attitude objects vary in the level of associated *strength between* the attitude and the "object." This continuous relationship is then used to predict the speed that an attitude can be automatically activated from memory.

FIGURE 2 | The valence continuum in (A) illustrates the dimension of valence ranging from strongly positive (red) to strongly negative (blue). As indicated by the dashed gray line, objects perceived to have a valence close to the neutral point on the continuum are nonetheless regarded as having a micro-valence. In **(B)** the portion of the continuum encompassed by the gray line in **(A)** has been expanded to represent a finer-grained continuum. The ordering of objects here reflects this expanded continuum, albeit with lesser magnitudes (the ordering of objects in this figure is derived from Lebrecht and Tarr, 2010).

scale at the center of the continuum is expanded we expect it to be organized much like the larger continuum, the only difference being the overall magnitude of valence is weaker at the center than at the extreme ends (**Figure 2B**; Lebrecht and Tarr, 2010). Note that this model does not preclude the presence of some objects happening to fall exactly at the center point of the continuum and therefore giving rise to a truly neutral perception of valence.

While micro-valence only accounts for a small absolute region of the valence continuum, it is integral in the automatic perception of objects. As such, "micro" valences are a big issue for understanding visual perception. That is, while it is undoubtedly the case that explicating the processing of strongly affective objects is useful in anchoring how affect and perception might interact, we rarely encounter blocks of gold or blood stained weapons in everyday life. Even if we are to acknowledge that valence functions along a continuum whereby subtle valences are possible, further research is necessary to elucidate the characteristics of the subtle valences that populate the central region of this space.

UNDERSTANDING THE ORIGINS OF MICRO-VALENCE

Understanding how everyday objects acquire a micro-valence forms a critical aspect of our theory. During perception visual objects evoke a rich set of information (not necessarily affective), including properties that are not an intrinsic feature of the percept or the percept's larger category. For example, seeing an object can automatically activate an entire network of information from memory (Bar, 2007). This network is derived from both contextual experiences (Bar et al., 2008) as well as conceptual or semantic knowledge (Martin et al., 1996; Haxby et al., 2001; Patterson et al., 2007). An operational example of this idea can be seen in the efforts by historians to tell the history of the world in only 100 objects (Schama, 2010): by selecting and describing key objects that elicit a wide range of associations, historians were able to evoke critical periods in history. For example, one particular object, the Roman Samian Bowl (Balmuildy Fort, Antonine Wall – second century AD), was bright red, engraved with human, animal, and floral figures. The bowl, made in a workshop in France, was thought to have been used at banquets for Roman officers and other high status individuals. The information attributed to this object told a story of craftsmanship and Roman rule in France in 200 AD. Historians regard objects as loci of archival information acting as an access point for vast quantities of knowledge and experience. In much the same way, we perceive a richer network of information than that immediately available in the percept. Our perception of any common object automatically accesses a large network of information in our memories – valence is only one component of a wide array of information, including semantic and conceptual information that arises during object perception.

It is these associations, spontaneously reactivated during perception, that help form what becomes an object's micro-valence. For example, the principle that the contextual experience with an object informs the micro-valence is evident in the micro-valence of telephones: phones that resembled those used in the bedroom were rated as more positive in valence than phones were used in the office (Lebrecht and Tarr, 2010). Moreover, these and other micro-valence ratings obtained in our studies have been found to be consistent both within and across observers, thereby indicating that micro-valence is more than random variation around a neutral mean (Lebrecht and Tarr, 2010).

What makes the derivation of micro-valence much more complex is that we have typically seen an object in many different contexts and with a variety of individuals, all of which will contribute to the affective memories or associations that become automatically reactivated during visual object perception. Micro-valence is the aggregate of this information plus, as discussed below, valence-relevant perceptual object properties. This cumulative valence – the contextual component being accrued over our personal experiences – determines, in part, whether a given object is ultimately preferred (having positive micro-valence) or anti-preferred (having negative micro-valence).

Supporting this framework, there is strong evidence suggesting that individuals rapidly attribute valence information to objects, which then influences subsequent perceptions (Zajonc and Markus, 1982; Murphy and Zajonc, 1993; Duckworth et al., 2002; Ghuman and Bar, 2006; Bliss-Moreau et al., 2008). Such findings indicate that individuals are able to associate affective information with objects, which can then bias perception at a later point in time.

THE PERCEPTUAL COMPONENTS OF MICRO-VALENCE

Although micro-valences appear to be "high-level," there is evidence suggesting that low-level visual properties contribute to an object's perceived valence. That is, visual features such as shape, curvature, color, material, and symmetry may add to an object being perceived as positive or negative independent of affective associations.

Experimentally, this question has been addressed by examining novel objects for which there are few pre-existing associations. For example, when making rapid "gut reaction" judgments participants consistently prefer curved over sharp or jagged objects for both familiar and novel objects (Bar et al., 2006). Other studies have observed that participants can make valence judgments on simple shapes (McManus, 1980; Rentschler et al., 1999; Bar et al., 2006). At the same time, several studies report more reliable ratings for real-world images as compared to abstract shapes (Vessel and Rubin, 2010), indicating that experience-based associations are dominant in forming valences.

More plausibly micro-valences arise from an integration of visual properties and learned associations. Moreover, these two attributes may potentially interact in that it may be easier to form positive associations with objects already possessing "positive" perceptual features. Consider that an observer might more readily generate positive associations with a shiny, curved, symmetrical teapot, whereas the same observer might more readily generate negative associations with a dull, angular, asymmetric teapot. Conversely, there is some evidence that this interaction between perceptual features and associations may also function in reverse: data to suggest that color preference might sometimes arise from the degree to which an individual prefers an object with a particular color, so that participants would be more likely to prefer green to yellow if they prefer apples to bananas (Palmer and Schloss, 2010).

BEYOND VALENCE

It is our view that valence should be construed as a property of object representations. As such, valence contributes to and is correlated with, but is not equivalent to, either aesthetics or preference. With respect to the former, object valence is closely related to visual aesthetics. Whereas we are focused on the singular dimension of valence in perception, aesthetics can be broadly defined as a discourse of how works of art are judged to be beautiful across different eras (Baumgarten, 1750; Ramachandran and Hirstein, 1999; Kirk et al., 2009). Thus, how our visual system evaluates objects or images as positive or negative may contribute to aesthetic experience, but cannot account for its entirety. Valence therefore can be considered but one aspect of aesthetic experience. More closely related to micro-valence is the emerging field of everyday aesthetics (Mandoki, 2007; Saito, 2007), which addresses the ramifications of making aesthetic judgments on everyday environments, for example, the ecological impact of maintaining lush green lawns because they are aesthetically pleasing.

As with aesthetics, it is our perspective that valence is but one component of preference, and not actually preference *per se*. Preference is defined at the level of behavior (Lichtenstein and Slovic, 2006). For example, if you select a cappuccino over an espresso you are held to have a preference for cappuccinos. By the time you reach this level of behavior an array of mental processes have been deployed: the object has been perceived, the relative valences of objects in the scene have been computed, these valences have been compared one's current goals and motivations in the specific context (Constantino and Daw, 2010; Krajbich et al., 2010; Litt et al., 2011), and these factors are combined to produce the current "value" of the objects (Rangel et al., 2008; Grabenhorst and Rolls, 2011). Only then can an individual make a decision and in doing so exhibit preference. As such, perceiving valence in objects is only one component of preference formation and, although valence and preference are often highly correlated, they are not the same mental construct. For example, a person may perceive one cell phone to possess a more positive micro-valence than the other, yet may ultimately buy the cell phone with the negative micro-valence because it has better features. Within our framework, valence is a positive or negative property of object representations that is computed automatically by our visual systems, whereas preference is an active choice behavior.

DISCUSSION

The functional significance of objects with strong valence is highly intuitive; we dislike objects that indicate danger, threat, or disgust, such as angry tigers or moldy food, and we like objects that indicate sustenance or pleasure, such as appetizing food or attractive mates. But why have we also come to automatically evaluate the valence of everyday objects that appear neither threatening nor life promoting?

We contend that micro-valences function to optimize our ability to either select or orientate toward objects with a positive micro-valence and away from those with a negative micro-valence. Throughout the day we make multiple unconscious decisions: what mug to use for our morning coffee, what pen to sign with, and what bottle of water to purchase. We suggest that these decisions are facilitated by micro-valences computed during perception, which we can use to reduce uncertainty and/or to orientate toward some objects and away from others.

Not surprisingly, these issues have not been lost on the product design and marketing communities. For example, Donald Norman, an expert in the psychology of product design, has argued that affective properties, or in our terms, micro-valence, enhance the usability of particular objects (Norman, 2003).

In summary, we contend that our perception of the world is always colored by our experiences and predispositions. We are social creatures that, through a variety of contextual experiences, create a visual world animated with affect. As observers we must decode the multitude of perceptual, affective, and semantic information presented to our senses. To solve the affect part of this equation we evaluate the valence of all visual objects across the scene. Much in the same way that we automatically perceive the shape, size, or color of objects, we cannot help but perceive the valence in objects. In this sense, valence is not a label applied after the fact to perceptual entities, but rather is an intrinsic element of visual perception with the same mental status as other object properties.

REFERENCES

Avero, P., and Calvo, M. G. (2006). Affective priming with pictures of emotional scenes: the role of perceptual similarity and category relatedness. *Span. J. Psychol.* 9, 10–18.

Bar, M. (2007). The proactive brain: using analogies and associations to generate predictions. *Trends Cogn. Sci. (Regul. Ed.)* 11, 280–289.

Bar, M., Aminoff, E., and Schacter, D. L. (2008). Scenes unseen: the parahippocampal cortex intrinsically subserves contextual associations, not scenes or places per se. *J. Neurosci.* 28, 8539–8544.

Bar, M., and Neta, M. (2007). Visual elements of subjective preference modulate amygdala activation. *Neuropsychologia* 45, 2191–200.

Bar, M., Neta, M., and Linz, H. (2006). Very first impressions. *Emotion* 6, 269–278.

Bargh, J. A., Chaiken, S., Govender, R., and Pratto, F. (1992). The generality of the automatic attitude activation effect. *J. Pers. Soc. Psychol.* 62, 893–912.

Barrett, L. F. (2006). Valence is a basic building block of emotional life. *J. Res. Pers.* 40, 35–55.

Barrett, L. F., and Bar, M. (2009). See it with feeling: affective predictions during object perception. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 364, 1325–1334.

Baumgarten, A. G. (1750). *Aesthetica.* Impens. Hildesheim: I. C. Kleyb.

Beauchamp, M. S., Lee, K. E., Haxby, J. V., and Martin, A. (2002). Parallel visual motion processing streams for manipulable objects and human movements. *Neuron* 34, 149–159.

Bliss-Moreau, E., Barrett, L. F., and Wright, C. I. (2008). Individual differences in learning the affective value of others under minimal conditions. *Emotion* 8, 479–493.

Calvo, M. G., and Avero, P. (2008). Affective priming of emotional pictures in parafoveal vision: left visual field advantage. *Cogn. Affect. Behav. Neurosci.* 8, 41.

Cant, J. S., and Goodale, M. A. (2007). Attention to form or surface properties modulates different regions of human occipitotemporal cortex. *Cereb. Cortex* 17, 713–731.

Cant, J. S., Large, M. E., McCall, L., and Goodale, M. A. (2008). Independent processing of form, colour, and texture in object perception. *Perception* 37, 57–78.

Colibazzi, T., Posner, J., Wang, Z., Gorman, D., Gerber, A., Yu, S., and Peterson, B. S. (2010). Neural systems subserving valence and arousal during the experience of induced emotions. *Emotion* 10, 377–389.

Constantino, S. M., and Daw, N. D. (2010). A closer look at choice. *Nat. Neurosci.* 13, 1153–1154.

Cunningham, W. A., Raye, C. L., and Johnson, M. K. (2004). Implicit and explicit evaluation: fMRI correlates of valence, emotional intensity, and control in the processing of attitudes. *J. Cogn. Neurosci.* 16, 1717–1729.

Duckworth, K. L., Bargh, J. A., Garcia, M., and Chaiken, S. (2002).

The automatic evaluation of novel stimuli. *Psychol. Sci.* 13, 513–519.

Fazio, R. H., Jackson, J. R., Dunton, B. C., and Williams, C. J. (1995). Variability in automatic activation as an unobtrusive measure of racial attitudes: a bona fide pipeline? *J. Pers. Soc. Psychol.* 69, 1013–1027.

Fazio, R. H., Sanbonmatsu, D. M., Powell, M. C., and Kardes, F. R. (1986). On the automatic activation of attitudes. *J. Pers. Soc. Psychol.* 50, 229–238.

Ghuman, A. S., and Bar, M. (2006). The influence of nonremembered affective associations on preference. *Emotion* 6, 215–223.

Giner-Sorolla, R., García, M. T., and Bargh, J. A. (1999). The automatic evaluation of pictures. *Soc. Cogn.* 17, 76–96.

Grabenhorst, F., and Rolls, E. T. (2011). Value, pleasure and choice in the ventral prefrontal cortex. *Trends Cogn. Sci. (Regul. Ed.)* 15, 56–67.

Greenwald, A. G., McGhee, D. E., and Schwartz, J. L. (1998). Measuring individual differences in implicit cognition: the implicit association test. *J. Pers. Soc. Psychol.* 74, 1464–1480.

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.

Haxby, J. V., Gobbini, M. I., Furey, M. L., Ishai, A., Schouten, J. L., and Pietrini, P. (2001). Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science* 293, 2425–2430.

Kirk, U., Skov, M., Hulme, O., Christensen, M. S., and Zeki, S. (2009). Modulation of aesthetic value by semantic context: an fMRI study. *Neuroimage* 44, 1125–1132.

Krajbich, I., Armel, C., and Rangel, A. (2010). Visual fixations and the computation and comparison of value in simple choice. *Nat. Neurosci.* 13, 1292–1298.

Lang, P. J., Bradley, M. M., Fitzsimmons, J. R., Cuthbert, B. N., Scott, J. D., Moulder, B., and Nangia, V. (1998). Emotional arousal and activation of the visual cortex: an fMRI analysis. *Psychophysiology* 35, 199–210.

Lebrecht, S., and Tarr, M. (2010). Defining an object's micro-valence through implicit measures. *J. Vis.* 10, 966.

Lichtenstein, S., and Slovic, P. (2006). *The Construction of Preference.* Cambridge, NY: Cambridge University Press.

Litt, A., Plassmann, H., Shiv, B., and Rangel, A. (2011). Dissociating valuation and saliency signals during decision-making. *Cereb. Cortex* 21, 95–102.

Mandoki, K. (2007). *Everyday Aesthetics: Prosaics, the Play of Culture and Social Identities.* Burlington, VT: Ashgate.

Martin, A., Wiggs, C. L., Ungerleider, L. G., and Haxby, J. V. (1996). Neural correlates of category-specific knowledge. *Nature* 379, 649–652.

McManus, I. C. (1980). The aesthetics of simple figures. *Br. J. Psychol.* 71(Pt 4), 505–524.

Murphy, S. T., and Zajonc, R. B. (1993). Affect, cognition, and awareness: affective priming with optimal and suboptimal stimulus exposures. *J. Pers. Soc. Psychol.* 64, 723–739.

Norman, D. A. (2003). *Emotional Design: Why We Love (or Hate) Everyday Things,* 1st Edn. New York: Basic Books.

Palmer, S. E., and Schloss, K. B. (2010). An ecological valence theory of human color preference. *Proc. Natl. Acad. Sci. U.S.A.* 107, 8877–8882.

Patterson, K., Nestor, P. J., and Rogers, T. T. (2007). Where do you know what you know? The representation of semantic knowledge in the human brain. *Nat. Rev. Neurosci.* 8, 976–987.

Ramachandran, V. S., and Hirstein, W. (1999). The science of art: a neurological theory of aesthetic experience. *J. Conscious. Stud.* 6, 15–51.

Rangel, A., Camerer, C., and Montague, P. R. (2008). A framework for studying the neurobiology of value-based decision making. *Nat. Rev. Neurosci.* 9, 545–556.

Rentschler, I., Jüttner, M., Unzicker, A., and Landis, T. (1999). Innate and learned components of human visual preference. *Curr. Biol.* 9, 665–671.

Rudrauf, D., David, O., Lachaux, J. P., Kovach, C. K., Martinerie, J., Renault, B., and Damasio, A. (2008). Rapid interactions between the ventral visual stream and emotion-related structures rely on a two-pathway architecture. *J. Neurosci.* 28, 2793–2803.

Russell, J. A. (1980). A circumplex model of affect. *J. Pers. Soc. Psychol.* 39, 1161–1178.

Russell, J. A., and Carroll, J. M. (1999). On the bipolarity of positive and negative affect. *Psychol. Bull.* 125, 3–30.

Saito, Y. (2007). *Everyday Aesthetics.* Oxford, NY: Oxford University Press.

Schama, S. (2010). What objects say about our times. *Financial Times.*

Serre, T., Wolf, L., Bileschi, S., Riesenhuber, M., and Poggio, T. (2007). Robust object recognition with cortex-like mechanisms. *IEEE Trans. Pattern Anal. Mach. Intell.* 29, 411–426.

Sternberg, S. (1966). High-speed scanning in human memory. *Science* 153, 652.

Ungerleider, L. G., and Mishkin, M. (1982). "Two cortical visual systems," in *Analysis of Visual Behavior,* eds D. J. Ingle, M. A. Goodale, and R. J. W. Mansfield (Cambridge, MA: The MIT Press), 549–586.

Vessel, E. A., and Rubin, N. (2010). Beauty and the beholder: highly individual taste for abstract, but not real-world images. *J. Vis.* 10, 18.1–18.14.

Weierich, M. R., Wright, C. I., Negreira, A., Dickerson, B. C., and Barrett, L. F. (2010). Novelty as a dimension in the affective brain. *Neuroimage* 49, 2871–2878.

Zajonc, R. B., and Markus, H. (1982). Affective and cognitive factors in preferences. *J. Consum. Res.* 9, 123–131.

Synesthesia, sensory-motor contingency, and semantic emulation: how swimming style-color synesthesia challenges the traditional view of synesthesia

Aleksandra Mroczko-Wąsowicz[1,2]* and Markus Werning[2]

[1] Institute of Philosophy of Mind and Cognition, National Yang-Ming University, Taipei, Taiwan
[2] Department of Philosophy II, Ruhr University of Bochum, Bochum, Germany

Synesthesia is traditionally regarded as a phenomenon in which an additional non-standard phenomenal experience occurs consistently in response to ordinary stimulation applied to the same or another modality. Recent studies suggest an important role of semantic representations in the induction of synesthesia. In the present proposal we try to link the empirically grounded theory of sensory-motor contingency and mirror system based embodied simulation/emulation to newly discovered cases of swimming style-color synesthesia. In the latter color experiences are evoked only by showing the synesthetes a picture of a swimming person or asking them to think about a given swimming style. Neural mechanisms of mirror systems seem to be involved here. It has been shown that for mirror-sensory synesthesia, such as mirror-touch or mirror-pain synesthesia (when visually presented tactile or noxious stimulation of others results in the projection of the tactile or pain experience onto oneself), concurrent experiences are caused by overactivity in the mirror neuron system responding to the specific observation. The comparison of different forms of synesthesia has the potential of challenging conventional thinking on this phenomenon and providing a more general, sensory-motor account of synesthesia encompassing cases driven by semantic or emulational rather than pure sensory or motor representations. Such an interpretation could include top-down associations, questioning the explanation in terms of hard-wired structural connectivity. In the paper the hypothesis is developed that the wide-ranging phenomenon of synesthesia might result from a process of hyperbinding between "too many" semantic attribute domains. This hypothesis is supplemented by some suggestions for an underlying neural mechanism.

Keywords: synesthesia, sensory-motor contingency, simulation, hyperbinding, emulative semantics, frames, embodied cognition, mirror neuron system

INTRODUCTION: SYNESTHESIA

Even though we are equipped with similar sensory mechanisms and cognitive functions, the way we perceive the world around us differs between subjects. An attempt to realize these individual differences as well as the commonalities in epistemic abilities makes us more sensitive to the problem of understanding the human mind. One of such extraordinary phenomena is synesthesia in which the stimulation of one sensory or cognitive pathway (the inducer) elicits stable sensory experiences (the concurrent) in the same or another modality (Baron-Cohen and Harrison, 1997; Mattingley et al., 2001; Ramachandran and Hubbard, 2001a,b). Some synesthetes have color sensations (e.g., red) seeing an alphanumeric symbol (e.g., an "A" or a "2"), individuals with another kind of synesthesia perceive colored spatial moving patterns when hearing music or, more generally, sounds. Other forms of synesthesia relate tastes, smells, visual, or tactile experiences to one another in almost any combination. Potentially, a huge number of different forms of synesthesia exist. The prevalence of the phenomenon depends on the particular type of synesthetic association with grapheme-color synesthesia being the

most frequent one (Cytowic and Wood, 1982; Mroczko-Wąsowicz and Nikolić, in press). The most characteristic feature of genuine synesthetic experiences is that they are non-acquired and involuntary conscious perception-like experiences which remain remarkably constant throughout the lifespan (Baron-Cohen et al., 1987). These consistent inducer-concurrent pairings may easily be differentiated from other, non-synesthetic associations. First, because they are elicited by a stimulus that normally is not associated with this additional, internally generated experience (Treisman, 2005). Second, although the associated experience, frequently color sensation, is as vivid and realistic as the veridical perception of the ink color, synesthetes always know which colors are ink and which are synesthetic. This exhibits the opacity of synesthetic phenomenal experiences (see, e.g., Metzinger, 1999). Simple consistent shape–color pairings that arise as a result of learning and memory associations are not likely to produce synesthesia. Such pairings occur often in our everyday life (e.g., the shape of a lemon is associated with the color yellow), but these matches do not convert into permanent synesthetic associations.

PERSPECTIVE: COGNITIVE ACCOUNTS IN SYNESTHESIA RESEARCH

One of the pressing questions in current research on synesthesia is the distinction between "low-" and "high-level" synesthesia, i.e., the problem of how to determine whether it is a sensory or semantic/conceptual phenomenon. At first, mostly the low-level explanation in terms of cross-talk of the senses has accumulated supporting evidence and dominated the debate. However, this appears to be merely a partial truth, since in many forms of synesthetic association only the concurrents exhibit purely perceptual features and inducers seem either to have a semantic nature or at least also to involve cognitive aspects, linking these types of synesthesia to high-level cognitive phenomena. Together with other authors (Dixon et al., 2006; Simner, 2007; Jürgens and Nikolić, 2012) we propose that a full account of the currently investigated phenomenon should transcend the traditional view. The traditional view is captured by the original compound "syn" + "aesthesia" (Greek for union of the senses) and takes synesthesia to be a solely perceptual phenomenon, such that a sensory stimulus of one modality elicits an additional perception in the same or another modality (e.g., Baron-Cohen and Harrison, 1997; Ramachandran and Hubbard, 2001b; Cytowic, 2002a,b). A more adequate understanding of the phenomenon should also take into account phenomena that do not merely involve sensations. For that purpose Nikolić et al. (2011) have coined the term *ideaesthesia*, *idea + aesthesia*, meaning sensing ideas, sensing concepts, and referring to the conceptual processing in synesthesia with perceptual concurrents and conceptual inducers joined together (see also Jürgens and Nikolić, 2012; Gómez Milán et al., in press). Such a phenomenon can be explained by high-level semantic mechanisms that guide the assignment and evocation of low-level sensory concurrents. Another analysis relates synesthesia to *the unity of cognition* and demonstrates that this phenomenon exhibits certain holistic epistemic ability integrating different mental faculties into a hypercoherently unified conscious experience (Mroczko-Wąsowicz, in press). In recent work Werning (2012) construes semantic representation as a form of non-symbolic compositional emulation, where the content objects, properties, and situations are emulated by brain mechanisms. Synesthesia very often involves inducers that are not strictly sensory, e.g., words, numbers, time units, musical notes, or personalities. Synesthetes may exhibit inducer-concurrent pairs with a cognitive representation of an abstract concept as an inducer and a sensory experience as a concurrent: Thinking about the number three – irrespectively of how the number is graphically represented, as "3," "III," or "…" – may produce a concrete color experience. Synesthetic associations are thus not merely cross-modal, but cross-domain, where the domains may not only involve the various sensual domains, but also the domains of bodily and emotional states as well as domains of abstract, conceptually represented entities like numbers or time units. Moreover, as we will highlight in this paper, synesthesia may even cross the motoric and sensory domains.

Evidence for this alternative perspective on synesthesia is getting stronger. Recent studies suggest an important role of semantic representations in the induction of synesthesia. The term "synesthetic conception" introduced by Grossenbacher and Lovelace

(2001) accounts for the conceptual aspects involved in synesthesia. In the above mentioned time unit-space synesthesia (e.g., Smilek et al., 2007; Jarick et al., 2008; Mann et al., 2009) subjects experience units of time, typically hours, days of the week, and months, as being placed at specific spatial locations in reference to their body. Semantic aspects of synesthesia can furthermore be identified in the "tip of the tongue" phenomenon or generally in lexical-gustatory forms of synesthesia. Here the verbalization of the stimulus is not necessary for the sensation of taste and the activation of the respective concept is sufficient (Simner and Ward, 2006). Cases of synesthesia that transcend traditionally denoted sensory modalities can also be found in reading musical notes, calculating, imagining, or just thinking of a stimulus (Frith and Paulesu, 1997; Grossenbacher, 1997; Dixon et al., 2000, 2006; Ramachandran and Hubbard, 2001a; Cytowic, 2002a; Rich et al., 2005; Ward et al., 2006). Synesthesia is also prevalent as an intralinguistic phenomenon in so-called synesthetic metaphors: "loud colors," "itching tunes," "cold smell." Here a concept from some sensory modality is used to modify a concept from another sensory modality. Interestingly, synesthetic metaphors can be communicated and are understood across subjects simply on the basis of shared semantic knowledge and without the use of the relevant sensory information channels (Williams, 1976; Yu, 2003; Werning et al., 2006). Also, a substitution of familiar graphemes with exotic, newly learned ones with the same meaning – the letter "A," e.g., now being replaced by a Glagolitic grapheme – can result in a transfer of synesthetic color experiences in less than 10 minutes of training (Mroczko et al., 2009). Therefore, synesthesia seems to rely essentially on a certain interpretation of the stimulus and the meaning that it has for the subject. To account for these phenomena we have to assume that the meaning of the inducing stimulus for the subject has to be read off before the concurrent experience can occur. We can no longer maintain that the synesthetic association is caused solely by low-level hard-wired preexisting connections between sensory areas (Dixon et al., 2000; Nikolić et al., 2011; Jürgens and Nikolić, 2012).

THEORY AND HYPOTHESIS: SENSORY-MOTOR CONTINGENCY, EMULATION, AND SYNESTHESIA

In philosophy of mind and cognitive neuroscience advocates of embodied cognition propose that many features of cognition are causally or constitutively associated with the physical body and bodily actions of an agent. According to this line of thinking the sensorimotor account of perception (known also as enactivism) states that our senses mediated by different forms of sensory-motor contingency explore the environment. The mind is constituted by the sensory-motor contingency between the agent and the world (Noë, 2005). The sensorimotor theory is a high-level cognitive model of conscious experience in which sensory experience results through the subject's cognitive access to a sensory-motor skill (O'Regan and Noë, 2001a). According to this model, consciousness arises from representation of interactions between organism and environment such as sensory changes induced by different motor actions, i.e., sensory-motor contingencies. Thus, phenomenal differences between various perceptual experiences can be accounted for by different actions when using

different sensory modalities; transformations in qualitative experience may well be explained in terms of a dynamic model of interdependence between sensory or semantic inputs and embodied activity (Hurley and Noë, 2003). Hence, a distinctive pattern of sensory-motor contingency conditions the subject to act in a manner such that the ways things appear to the subject are affected. This results in a matching between bodily and environmental features (O'Regan and Noë, 2001a,b; O'Regan et al., 2006). The theory is supported by empirical findings regarding effective sensory substitution, sensory-motor adaptation for color perception or for touch as found for instance in the rubber hand illusion and mirror therapy reducing phantom limb pain (Ramachandran and Rogers-Ramachandran, 1996; Botvinick and Cohen, 1998; Bompas and O'Regan, 2006; Proulx and Störig, 2006). Another sensory-motor contingency may be observed in the matching between different modalities and domains. The theory emphasizing these issues, called embodied social cognition or emulation/simulation theory, explains the phenomenon of intersubjectivity as intercorporeity or visuo-tactile matching in various positive symptoms employing mirror neuron systems and other mirroring mechanisms in our brain (Gallese and Goldman, 1998; Rizzolatti and Sinigaglia, 2010). Mirroring others' experiences requires a mapping between the self and others. This, indeed, may involve some of the brain mechanisms underlying social competencies. Mirror systems may have evolved as an adaptation for interpersonal understanding (Gallese and Goldman, 1998). They refer to cortical areas that respond both to observing another person's state and being in that same state oneself (Gallese, 2003). Mirror systems have been reported in humans not only for actions (Rizzolatti et al., 1996), but also for emotions (Bastiaansen et al., 2009; disgust – Wicker et al., 2003; facial expressions – Carr et al., 2003), and sensations (pain – Avenanti et al., 2005; touch – Blakemore et al., 2005). Hence, these mechanisms map the sensory representations of the sensation, emotion, or action of an observed person onto the perceiver's own somatosensory, viscero-motor, or motor representations of the sensation, emotion, or action. Such a mapping enables the observer of another person's sensation, emotion, or action, to feel as if he were experiencing that sensation or emotion or performing that action himself. The mirroring mechanisms represent sub-personal instantiations of embodied emulation, i.e., respective neural activations for brain-body states associated with sensations, emotions, and actions evoked in the observer while observing social stimuli as if he were undergoing a similar experience.

Generally speaking, an emulator is a device that mimics the state transitions of a target system due to some structural mapping (total or partial isomorphism, homomorphism, etc.) between the states and transitions of the device and the states and transitions of the target system. Emulators may thus have representational content due to their structure and their transitional performance. Being dynamical systems with representational powers, they may also serve as forward models and generate predictions (Grush, 2004; Bartels, 2006). Emulation is a subcategory of simulation, but typically contrasts with high-level simulation which might, e.g., be based on global mathematical equations. In the cognitive science literature the notions "emulation" and "simulation" are often used interchangeably.

Embodied social emulations exhibit the relations that our brain bears to other persons by mapping others' sensations and emotions onto our cerebral somatosensory and viscero-motor states, and others' actions onto our cortical motor states (Gallese and Metzinger, 2003; Gallese, 2007; Gallese and Sinigaglia, 2011). These brain states are themselves representational and represent features, evaluations, and affordances of objects, events, and situations in the world as well as states of one's body and potential actions resulting therefrom. Emulation plays a central role in acquiring knowledge about our environment (Gordon, 1986, 1995; Goldman, 1989, 1992; Metzinger and Gallese, 2003; Thomas et al., 2006). Effective perception or action requires the capacity of emulation in order to predict impending sensory events or consequences of action (Grush, 2004). Emulation is not restricted to the somatosensory or motor domain. In the theory of emulative semantics (Werning, 2012), emulations are interpreted as semantic values that can be linked to each other and thus become constitutive for concepts (see below).

Neural mechanisms of mirror systems seem to be involved in synesthesia. It is particularly vivid in forms of mirror-sensory synesthesia, e.g., mirror-touch or mirror-pain synesthesia (Blakemore et al., 2005; Banissy and Ward, 2007; Fitzgibbon et al., 2010; Keysers et al., 2010). This implies that many aspects of everyday social cognition can be conceptualized as synesthesia-like processing. The already illustrated phenomenon of sensory-motor contingency recalls other perceptual phenomena and mental processes in human cognition supported by multisensory, sensory-motor, or cross-activation mechanisms similar to those in synesthesia. In the present paper, and especially in this section, we want to point out the omnipresence of synesthesia-like processing and the relevance of an emulation theory of cognition in explaining such experiences. The reason for doing this is motivated by numerous studies on different multimodal, cross-activation phenomena emphasizing the inductive role of sensory-motor processing and semantic representations of the stimulus, such as mirroring one's own body or self and projecting it outside the body via the imposition of multisensory conflicts using congruent and incongruent visuo-tactile inputs (Metzinger, 2005; Lenggenhager et al., 2007; Ionta et al., 2011; Ferri et al., 2012), autoscopic symptoms that occur when patients hallucinate their mirror image (Zamboni et al., 2005), undergo out-of-body experiences (Blanke and Metzinger, 2009), double delusions (Brugger, 2002), or feel the sense of presence in the widowhood hallucinations (Rees, 1971).

Inter-modal analogies outside canonical synesthesia are quite universal. We can agree on such commonly shared associations like experiencing higher pitch as lighter and smaller. Comparable evidence of correspondences between different domains exist for vision and touch (see, e.g., mirror-touch synesthesia, rubber hand and full body illusions) and is based on vicarious activation, activation of a brain region that is typically involved in processing the observer's own actions and sensations, but that is now activated by seeing similar actions or sensations in another person. The observation of touch has long been considered a solely visual event. However, recent studies suggest the involvement of

vicarious activations in somatosensation, i.e., that a somatosensory component is also activated in the observer. Observed touch is processed in visual parts of the brain and in somatosensory areas (Ebisch et al., 2008, 2011; Keysers et al., 2010). Watching another person being touched usually activates a neural circuit similar to that of the actual touch. Only "similar" because at the neural level the overlap is not total between one's own experiences and the experiences observed in someone else. In the former case both primary and secondary somatosensory cortices (SI and SII) are activated. However, while observing the touch only vicarious activation in the secondary somatosensory cortex occurs (Keysers et al., 2010). Despite this vicarious SII activation, in daily life we are not confused about who is being touched. This is because the primary somatosensory cortex including the Brodmann's area [BA] 3, is only recruited when we ourselves are being touched. This fact seems to shed some light on the functional foundation of the neural mechanisms underlying mirror-sensory synesthesia, explaining why certain people have conscious somatosensory experience during the observation of similar stimulation applied to another person. Here, the inducer is the "observed bodily touch," rather than vision *per se*. Depending on whether this stimulus is applied to a synesthete or to a non-synesthete, an observation of touched body parts affects somatosensory activation and experiences to a different extent. A recent fMRI study with non-synesthetic healthy subjects shows that some brain regions involved in first-person sensory experiences appear to actively distinguish between self and other conditions, in the sense that they are positively modulated (activated) for first-person sensory experiences, but negatively modulated (deactivated) when observing these experiences in others. The specific pattern of negative modulation has been detected in the posterior insula for subjects watching affective social touch applied to other individuals. In contrast to this vicarious activation, a positive modulation in somatosensory regions has been found for the actual tactile experience. The posterior insula, embedded in a distributed network grounding a sense of the bodily self, seems to differentiate between self and others when affective experiences are implicated (Ebisch et al., 2011). The mentioned partial overlap and deactivation processes, may be responsible in the general population for not experiencing others' experiences during social perception. They may explain why usually no synesthetic experiences occur when we simulate or mirror others' bodily experiences. Alterations at this level could result in sensory experiences also when perceiving them in others as this is the case in experimentally or neurologically induced illusory situations (e.g., out-of-body experiences) or specific individuals such as mirror-sensory synesthetes.

Amputation or paralysis of a limb is frequently accompanied by tactile, painful, or motoric sensations in the location of the missing limb, a so-called phantom limb. Ramachandran and Rogers-Ramachandran (1996) used mirrors to create a duplicate image of the amputated arm based on a reflection of the patient's existing arm. Observed movement or touch at the intact arm produced the identical proprioceptive sensations at the phantom limb. This synesthesia-like processing can be conceived of as a behaviorally induced form of synesthesia or a temporal sensory substitution. The phenomenon of synesthesia as well as others including cross-activation and mirroring mechanisms distinctly exhibit a

multimodal integration and multisensory awareness of selfhood. Both of them seem to be necessary conditions for pre-reflective bodily self-consciousness.

The construction of one's own body image requires an integration of multimodal information from such different sources as visual, haptic, and proprioceptive perception. Interactions between these different domains may also yield specific self-identification and embodiment illusions like the enfacement illusion (Sforza et al., 2010), the rubber hand illusion (Botvinick and Cohen, 1998) or the full body illusion induced on purpose in the lab with a virtual reality apparatus (Lenggenhager et al., 2007) or due to some neurological origin (Blanke et al., 2004). Studying the mentioned illusions and alterations of the body image based on a synesthesia-like processing caused by the interpersonal multisensory stimulation (felt touch and vision of touch) provides a promising tool for learning more about the neural representations of stimulated body parts as well as about the attribution, localization, and ownership of the bodily self (Tsakiris, 2008). Studies measuring visual-somatosensory spatial congruency effects for the full body and subjective changes in global bodily self-consciousness have shown that during incongruent trials, i.e., conflicting visuo-tactile input, tactile stimuli were mislocalized toward the observed body. Since no such tests have been done with mirror-somatosensory synesthetes, aside from studies revealing that these subjects have difficulties with distinguishing between actual and synesthetic touch (Banissy et al., 2009), future research should concentrate on discriminating differential brain activations related to cross-modal conflicts also in specific individuals (Aspell et al., 2009).

The difference between the above mentioned autoscopic phenomena and shared multisensory experiences in the mirror-(somato)sensory synesthesia such as mirror-touch or mirror-pain synesthesia is that in the former case the body is visually mirrored, and in the latter one – visually presented tactile or noxious stimulation of others results in a somatosensory activation in oneself, i.e., in the projection of the conscious tactile or pain experience onto the respective part of the observer's body (Blakemore et al., 2005; Banissy et al., 2009). Mirror-sensory synesthetes feel on their bodies what observed people are feeling when they are really touched. Also in this form of synesthesia concurrents can be experienced even by simply thinking of the inducers, e.g., when imagining another person in pain (Fitzgibbon et al., 2012) or anticipating such an experience. It has been shown that anticipation of somatosensation can increase activation in the primary somatosensory cortex without actual stimulation (Carlsson et al., 2000). However, inducer-concurrent correspondences are not so individual here as in other forms. They are much more regular and dependent on the stimulus. For instance, the intensity of the felt touch is stronger when observed touch is applied to real bodies than to dummy bodies. This feature of synesthetic mirror-touch might be related to a more general differential susceptibility to the involvement of somatosensory and interoceptive cortices into embodied simulation found in fMRI studies with non-synesthetic healthy subjects (Ebisch et al., 2008, 2011). This suggests that top-down processes may also modulate the intensity of the synesthetic mirror-touch experience (Fitzgibbon et al., 2010; Holle et al., 2011). Different potential

mechanisms of mirror-sensory synesthesia are discussed in the literature. The most widely proposed underlying mechanism is an over-activation of the mirror neuron system for somatosensation, significantly beyond the threshold for consciousness. Both mirror-touch synesthetes and non-synesthetes when experiencing and observing touch activate visuo-tactile mirroring mechanisms, i.e., similar brain areas (partially overlapping) in premotor and parietal regions, primary and secondary somatosensory cortices (Blakemore et al., 2005). Nevertheless, when observing touch on others, fMRI has revealed the difference in synesthetic as compared to non-synesthetic mirror-touch, namely a greater activation within somatosensory cortices (including BA 3) as well as bilateral activation of the anterior insula (Blakemore et al., 2005; Keysers et al., 2010). The latter has been related to self-awareness (Critchley et al., 2004) and processing one's awareness of others (Lamm and Singer, 2010). The amount of neural activity within the insula has been shown to increase by directing attention to one's own emotions when viewing affective stimuli (Straube and Miltner, 2011). Thus, the insular cortex is likely to be a candidate for modulating distinctions between the sources of one's own and another's tactile conscious experience within the mirror-touch system as well as an important component for constructing a self-model from sensory-motor emulation based mechanisms of social cognition.

These issues, however, have not been investigated thoroughly and systematically enough by neuroimaging studies. Since it seems that other mechanisms also may well be involved in the induction of synesthesia-like processing, other hypotheses regarding the functional basis of the mirror-sensory phenomenon have come up. Another potential underpinning considered in this context is associative learning and heightened empathy (Fitzgibbon et al., 2010). These processes are actually not mutually exclusive or contradictory to the hyperactivity of the mirror neuron system. Thus, their strengths should be pulled together in representing a more complex interplay of relevant processes providing an appropriate and full explanation of these cross-activated multisensory perceptions. The theory of mirror-sensory synesthesia as a learned association or a result of associative learning is encouraged by effects of sensory-motor modulation of mirror areas found in studies demonstrating that, e.g., ballet dancers exhibit an increased mirror neuron activity when observing ballet compared to dancers of other styles (Calvo-Merino et al., 2005). This may suggest the involvement of associative learning in the development of mirror systems. If so, mirror-sensory synesthesia would be the result of a learned association through sensory-motor experience mediated by mirroring mechanisms. This is not irreconcilable with the suggestion that mirror systems are over-activated in mirror-sensory synesthesia leading to a conscious somatic sensation. Studies on mirror-sensory experiences, especially those acquired in mirror-pain synesthesia in amputees, suggest that they may come about through disinhibition of systems involved in empathy for pain and as such they may be understood as a result of enhanced empathic capacity (Fitzgibbon et al., 2010). Visuo-tactile interaction as in mirror-touch synesthesia – and generally multisensory integration – enables the implementation of a number of social skills related to empathy and reading other minds, e.g., understanding intentions, feelings, and emotional states of other people.

In the embodied simulation and sensory-motor approach to the theory of mind and social cognition it has been postulated that this interpersonal competence is not just an effect of rational reasoning about mental states of others. It seems to depend on a special human faculty of perspective taking by simulating other people's aims and actions using mirror neuron systems (Gallese and Goldman, 1998; Gallese, 2007). Social interactions based on shared feelings and simulated experiences may be seen as a specific kind of synesthesia-like processing. Thus, sensory-motor exchange when observing others' behavior may induce a concurrent motor reaction, e.g., automatic mimicking of someone's facial expression (Dimberg, 1982) or sensory–sensory interactions when the empathic experience of someone's pain elicits qualitatively similar somatosensory experience in the observer. Cross-modal associations exhibiting synesthesia-like processing are inbuilt both into veridical and illusory perceptions as well as into human (social) cognition. Integrating information from cognitive, motor, and perceptual domain, across different sensory modalities and diverse reference frames (space, time, object, and subject) is our fundamental ability used every day to make sense of the reality that surrounds us (see Goldstone and Barsalou, 1998; Sagiv et al., 2011).

SWIMMING STYLE-COLOR SYNESTHESIA OR A NEW VARIANT OF MIRROR-SENSORY SYNESTHESIA. FUTURE DIRECTIONS FOR INTERPRETING THE STATE OF THE ART IN SYNESTHESIA RESEARCH

A novel form of synesthesia, swimming style-color synesthesia, has been discovered in two known grapheme-color synesthetes, who are semi-professional swimmers. The visual experience or imagination of four different swimming styles (breaststroke, butterfly, crawl, and backstroke) is synesthetically associated by them with four different colors (Nikolić et al., 2011). The induction of this kind of synesthesia took place exclusively under laboratory conditions and did not require any measurements in a swimming pool. All what subjects had to do was to take a look at photographs of other people swimming (e.g., **Figure 1**) or to think about/mentally visualize a given swimming style. This was sufficient to elicit their color experience.

FIGURE 1 | An example photograph shown subjects in the study of Nikolić et al. (2011).

FIGURE 2 | Stimuli used in the Stroop test: Example pictures of a person swimming butterfly stroke, painted either in a subject's synesthetic color (congruent) or in one of his non-synesthetic colors (incongruent). Reprint from Nikolić et al. (2011).

Not only subjects' reports were taken into account, but also two standard tests for synesthesia, the consistency test, and the Stroop task, provided objective evidence for the existence of swimming style-color synesthesia (see Baron-Cohen et al., 1987; Odgaard et al., 1999; Mroczko et al., 2009). Swimming style-color synesthetes reported significantly more consistent colors than control non-synesthetes trained in these associations. Furthermore, in a Stroop task the color naming times accelerated when the color of the colored photograph was congruent with the synesthetic color of the presented swimming style and slowed down when the color was incongruent (**Figure 2**).

Hence, swimming style-color synesthesia is a genuine form of synesthesia similar to other common types of the phenomenon, such as grapheme-color synesthesia or colored hearing, although mostly related to mirror-sensory variants. As such swimming style-color synesthesia might be understood as mirror-movement associated with color experience. According to our hypothesis further brain imaging studies should reveal the respective activation of mirror systems after such multimodal, or even multi-domain, visuo-motor stimulation. Interestingly, the current findings demonstrate clearly that inducers of swimming style-color synesthesia do not have to be modality-dependent sensations directly; the activation of kinesthetic simulation does suffice. The mirror based embodied emulation of another's action eliciting synesthetic concurrents can be interpreted in terms of cross-domain integration which fits well with the above mentioned cognitive account of synesthesia as an overall epistemic ability, *the unity of cognition* (Mroczko-Wąsowicz, in press). The original study showed that this kind of synesthesia may occur even when direct sensory or motor inputs are absent and only indirect ones exist. This allows us to conclude on the complex nature of the phenomenon. It is evidence for an extended case of embodied emulation in which the synesthetic concurrent is elicited without any direct motoric stimulation, only through kinesthetic emulation activated by the mental visualization of a certain swimming

style. This means, the motor imagery of a certain swimming style seems to produce motor and somatosensory activations in form of the mirror system based motor emulations. The latter ones, based on the individual connections, lead to synesthetic color sensations. This form of synesthesia has some special features allowing the separation of direct sensory inputs, i.e., motor and proprioceptive inputs during the act of swimming, from those that evoke mental emulations, i.e., pictures, thoughts, or words related to respective swimming activities. Hence, the observation of a picture representing a swimming action may induce the creation of the sensory-motor emulation on a higher conceptual level. This again supports the possibility of a form of semantically driven synesthesia. Obviously, the fact that the synesthetes did not have to undergo any physical exercise does not imply that their motor system was entirely silent. Regardless whether it was a mirror based embodied simulation, a motor imagery based embodied emulation, or a result of the action concept activation, there is a motor neural representation of action even in the absence of any explicit motor behavior.

What enables the direct matching between the visual representation of an action, its motor representation, and further synesthetic color representation is still a matter of questions and speculations (see Rizzolatti et al., 2001). Early forms of automatic imitation of adult facial and manual gestures by human neonates (Meltzoff and Moore, 1977, 1983, 1994) seem to suggest some hard-wired mechanisms coupling action observation with action execution. The traditional view on synesthesia as a form of cross-wiring between senses, would suggest that the above described phenomenon could arise only by providing corresponding motor inputs directly. To the contrary, the referred study indicated that no direct motor or somatosensory stimulation is necessary. To validate experiences characteristic for synesthesia from the third person perspective, it was sufficient to activate the respective emulations by showing pictures of swimming persons. Therefore, we can conclude that visualization of a given swimming style, be it

perceptual or imaginary may induce a corresponding synesthetic color experience in the absence of any overt muscular activation. On the basis of these results as well as of findings concerning the inducing role of imagined graphemes in grapheme-color synesthesia (Frith and Paulesu, 1997; Ramachandran and Hubbard, 2001a) and sensory-motor contingency in observing dancers of ballet and capoeira (Calvo-Merino et al., 2005, 2006) the following generalization to swimming style-color synesthesia seems to be legitimate: kinesthetic emulation triggers concurrent color sensations, much like the original motor input itself. If the activation of motor emulations is sufficient for the concurrent to be evoked, other cases of synesthesia and synesthesia-like processes may be revealed in which particular body movements serve as inducers, e.g., different styles and techniques in sports. This refers not only to motion and behavior. Since producing vicarious activations enables the observer of another person's sensation, emotion, or action to feel as if he were experiencing that sensation, emotion, or performing that action himself, triggering synesthetic experiences may be mediated by the activation of a specific model that constitutes an internal, brain-based emulation of the perceived event, and of organism-environment interactions. The hypothesis can immediately be applied to the category of emotions – already known cases of emotion driven synesthesia (e.g., emotion to color synesthesia, personalities inducing colors, smells, or touch; see Ward, 2004; Sinha, 2009) might well be extended to inducers in form of observed emotions experienced by other people.

Also, distinctive features of swimming style-color synesthesia may explain why, in synesthetes, only some types of modality/domain related emulations produce synesthesia. Both synesthetes have been active swimmers since early childhood. This supports the hypothesis that at that time, when synesthesias are known to develop (Baron-Cohen et al., 1987; Harrison and Baron-Cohen, 1997), the categories of inducers that are especially disposed to

acquire synesthesia are those that play a central role in the child's life. Synesthesia seems to develop most easily for activities and emulations of them that children spend most time with and that possess predominant representational contents in the course of learning and playing. Therefore, theoretically there is no reason to exclude any kind of mental representation from being a possible inducer of concurrent synesthetic experiences. The key aspect seems to be the frequency with which this potential trigger is being employed by a young synesthete. The present results imply that the initiation of synesthetic associations may be regulated by the ability of eliciting emulations of regularly experienced events, properties, or situations.

HYPERBINDING, EMULATIVE SEMANTICS, AND THE THEORY OF NEURO-FRAMES

Swimming style-color synesthesia may also be viewed as a nonstandard binding pattern between the neurobiologically realized attributes of color and bodily motion. Binding patterns between intra- and cross-modal as well as cross-domain attributes, including i.a. perceptual, proprioceptive, emotional, numerical, and motor attributes, in ordinary cases play an important role in the formation of concepts. As an underlying theoretical framework, the theory of neuro-frames (Werning and Maye, 2005, 2007; Werning, 2012) has been developed. This neuro-cognitive model of situated conceptualization (Barsalou, 2008) postulates neuro-frames as neuronal bases for concepts. A frame is defined for a large domain of things and contains a fixed set of attributes, each of which allows for a number of different values (Barsalou, 1992). The attributes in question are not constrained to perceptual modalities, but may involve motor attributes as well as further attributes (in this paper we are focusing on the interrelation between perceptual and motor attributes). Frames can be nested hierarchically and mutual constraints between attributes (e.g., between states of an
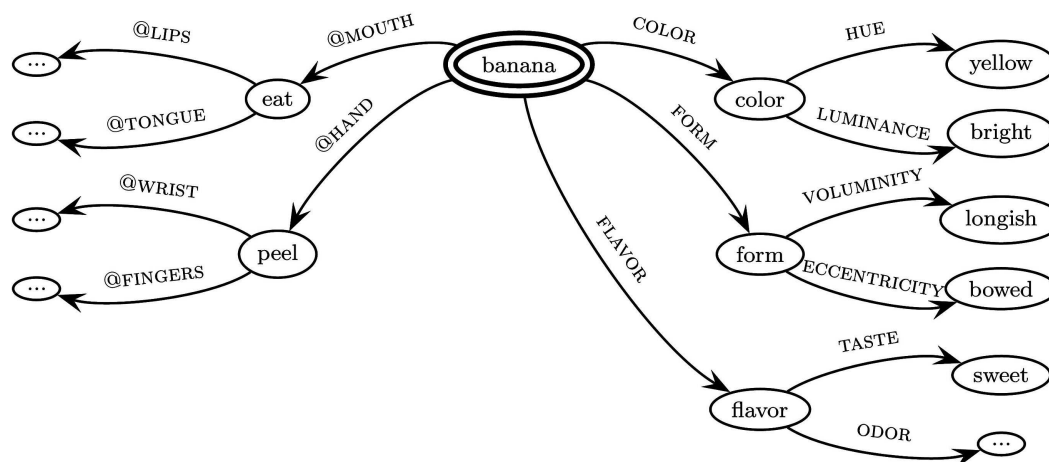

FIGURE 3 | Hypothetical fragment of the frame for the concept [banana]. The substance concept to be decomposed is marked by a double-circle as the referring node of the frame. The labeled arrows denote attributes, the nodes their values. Nodes are themselves regarded as concepts and thus as conceptual parts of the central concept. In English, feature attributes (shown on the right) are frequently lexicalized – their arguments typically enter possessive constructions like *The color of the*

banana is yellow or *The banana has the color yellow*. Based on linguistic and neurobiological evidence, we assume that affordances often relate to body parts and hence use the convention "@ + body part." Formally, attributes are mappings from domains of some type into domains of some other type. Petersen and Werning (2007) provide an explicit account of frames using a calculus of typed feature hierarchies and incorporating typicality effects.

object and actions directed to it) and between larger frames can be incorporated (see **Figure 3**).

For many attributes involved in perceptual processing one can anatomically identify cortical correlates. Those areas often exhibit a twofold topological structure and justify the notion of a feature map: (i) a receptor topology (e.g., retinotopy in vision, somatotopy in touch): neighboring regions of neurons code for neighboring regions of the receptor; and (ii) a feature topology: neighboring regions of neurons code for similar features. With respect to the monkey, more than 30 cortical areas forming feature maps are experimentally known for vision alone (Felleman and van Essen, 1991).

Motor attributes may also be parts of frames and appear to have cortical correlates, predominantly in the premotor and motor cortex (Werning, 2010). The cortical organization of motor control with regard to the effectors follows similar topological principles as the cortical organization in perception with regard to the receptors. The discovery of the so-called canonical motor neurons in the mirror neuron system, activated by the sight of an object to which a certain action is applicable (Rizzolatti and Luppino, 2001; Rizzolatti and Craighero, 2004), may provide a basis to integrate affordances (Gibson, 1977) – specific qualities of the object that allow the agent to perform particular actions upon it – into frames. **Figure 4** shows a number of neural maps that relate to various attributes of frames.

Canonical neurons are involved in mechanisms for recognizing object affordances and contribute to the semantic knowledge about the object (Sahin and Erdogan, 2009). Hence, the activation of the mirror system brings its multimodal neurons to respond not only to action performance, but also to visual, auditory, somatosensory, and proprioceptive signals. This suggests that related processes are grounded functionally by multimodal circuits (Gallese and Lakoff, 2005; Rizzolatti and Sinigaglia, 2010). In particular, the intraparietal sulcus and inferior parietal lobule are involved in multisensory integration and vicarious sensory-motor activations (Bremmer et al., 2001; Rozzi et al., 2006; Ishida et al., 2010; Rizzolatti and Sinigaglia, 2010). These regions, able to receive visual input, are directly connected with each other and with the somatosensory cortex (i.e., BA 2; Pons and Kaas, 1986; Lewis and van Essen, 2000) integrating tactile and proprioceptive stimuli as well as containing shared sensory-motor representations (Keysers et al., 2010). These multimodal circuits exhibit some basic semantic features. The activation of a specific action concept, e.g., expressing an affordance or any other motor attribute, induces the activation of the multimodal neural circuits (Pulvermüller and Fadiga, 2010). In swimming style-color synesthesia the functioning of such a multimodal mechanism might take place when synesthetes swim, see other person swimming, when they see a swimming pool and recognize this object affordances or use a concrete action concept such as for example [breaststroke].

The fact that values of different attributes may be instantiated by the same object, but are processed in distinct regions of cortex is a version of the binding problem (Treisman, 1993; Tacca, 2010): how is this information integrated in an object-specific way? How can the color and taste of a banana be represented in distinct regions of cortex, although they are part of the representation of one and the same object? A prominent and experimentally
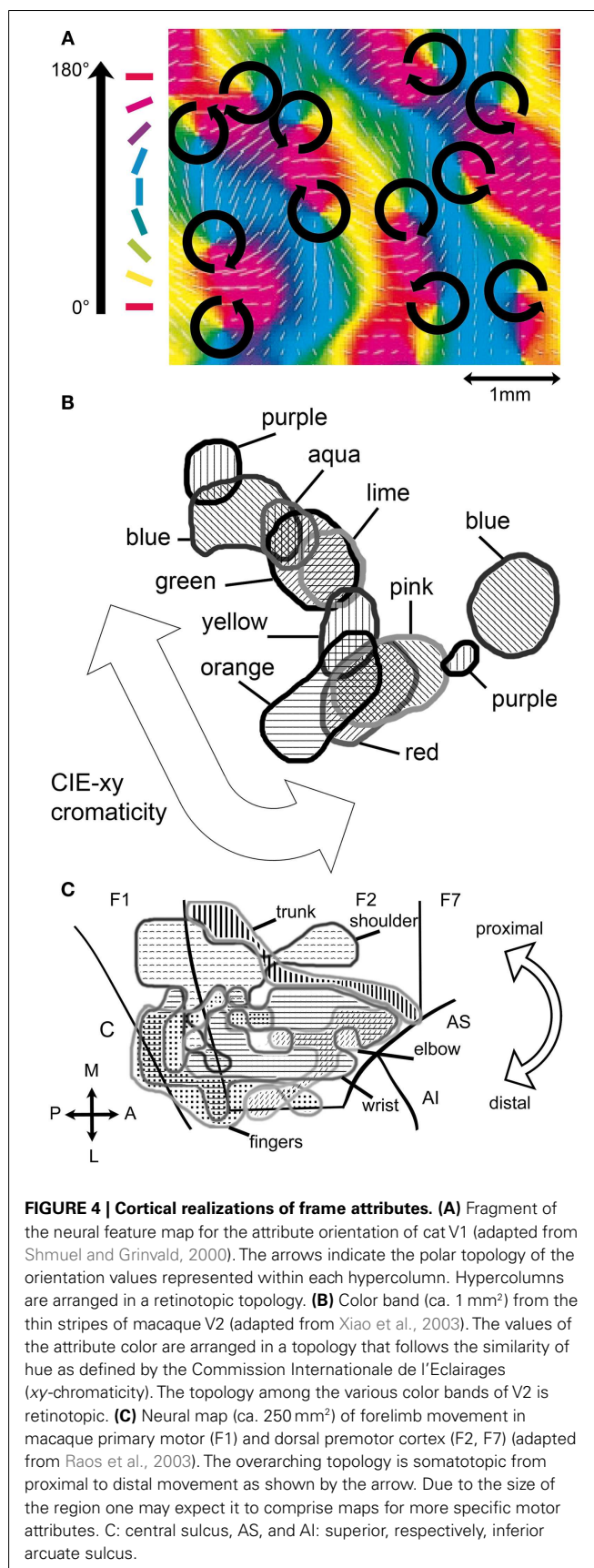

FIGURE 4 | Cortical realizations of frame attributes. (A) Fragment of the neural feature map for the attribute orientation of cat V1 (adapted from Shmuel and Grinvald, 2000). The arrows indicate the polar topology of the orientation values represented within each hypercolumn. Hypercolumns are arranged in a retinotopic topology. **(B)** Color band (ca. 1 mm²) from the thin stripes of macaque V2 (adapted from Xiao et al., 2003). The values of the attribute color are arranged in a topology that follows the similarity of hue as defined by the Commission Internationale de l'Eclairages (*xy*-chromaticity). The topology among the various color bands of V2 is retinotopic. **(C)** Neural map (ca. 250 mm²) of forelimb movement in macaque primary motor (F1) and dorsal premotor cortex (F2, F7) (adapted from Raos et al., 2003). The overarching topology is somatotopic from proximal to distal movement as shown by the arrow. Due to the size of the region one may expect it to comprise maps for more specific motor attributes. C: central sulcus, AS, and AI: superior, respectively, inferior arcuate sulcus.

well supported solution postulates oscillatory neural synchronization as a mechanism of binding: Clusters of neurons that are indicative of different properties sometimes show synchronous oscillatory activity, but only when the properties indicated are instantiated by the same object in the perceptual field; otherwise they are firing asynchronously. Synchronous oscillation, thus, might be regarded as fulfilling the task of binding various property representations together to form the representation of an object having these properties (Singer, 1999). Using oscillatory networks as biologically motivated models, it could be demonstrated how the topological organization of information in the cortex by mechanisms of synchronization may yield a logically structured semantics of concepts (Maye and Werning, 2004; Werning and Maye, 2007). Compositionality theorems have been provided (Werning, 2005). Oscillation functions play the role of object concepts. Clusters of feature sensitive neurons play the role of attributive concepts. The experimental findings by Schnitzler et al. (2006) on the essential role of neural synchronization for action control may justify the extension of the synchrony-based neuro-frame approach from features to affordances. It should be noted that the envisaged semantics is one of emulation: the neuronal structure is partially isomorphic to a (model-theoretic) model of the representational content. A concept like [banana] thus interrelates i.a. sensoric and motoric emulations: Having the concept [banana] means being able to emulate what a banana would look, taste, feel, and smell like and being able to emulate actions afforded by a banana. Triggering the concept activates the respective sensoric and motoric cerebral regions for the purpose of emulation even in the absence of a real banana. The neuro-frame captures how the various sensoric and motoric emulations are linked to each other. Emulative semantics is a non-symbolic, embodied, but still compositional semantics and might be used to link conceptual resources employed in perception and motor planning to linguistic meaning (Werning, 2012).

Support for the theory of neuro-frames also comes from a number of neuro-linguistic studies. Based on a review of neurobiological data, Pulvermüller et al. (1999) suggests that neural assemblies that pertain to the sensory-motor cortices and are bound by neural synchronization play an important role in understanding the meanings of words and sentences. These cortical sensory-motor action and perception circuits are interdependent in language comprehension. Neuroimaging investigations have shown that perception and understanding of stimuli depend on motor circuits, i.e., specific motor activations can be found when subjects understand speech sounds, word meanings, semantic categories, and sentence structures (Pulvermüller and Fadiga, 2010). FMRI studies (Pulvermüller, 2005) regarding the understanding of verbs, e.g., hint at a differential top-down activation of motor and pre-motors areas. We know that the understanding of concrete nouns like *hammer*, for which not only features, but also affordances are salient, results in an activity distributed over the premotor and the visual cortex (Martin et al., 1996; Martin, 2007). The hypothesis that words for substance concepts arouse more widely distributed activity than words for attributive concepts has also been supported by EEG studies (Rappelsberger et al., 2000). Brain areas involved in motor control contribute to neural networks in which verb representations are grounded, e.g., studies on motor deficits such as Parkinson disease reveal impairment of

patients' action naming (Rodríguez-Ferreiro et al., 2009). Higher-order abilities such as thinking or linguistic concept use are based in sensory-motor abilities.

Unlike in normal concept formation, where perceptual and motor attributes forming a concept are bound together into a frame (e.g., the concept [banana] in **Figure 3**), synesthesia may generally be regarded as a case of hyperbinding (Emrich et al., 2002, 2004; Sagiv and Robertson, 2005; Mroczko-Wąsowicz, in press). Attributes that do not form a sensible concept frame are bound together. This is especially striking in the case of visuo-motor or mirror-movement to color synesthesia as observed in swimming style-color synesthesia. Certain attributes of bodily motion are contingently linked to the attribute color. An additional synesthetic attribute of the concurrent is not only bound additively to the attributes of the inducer, it is experienced integratively. Also, one and the same attribute of the synesthetic inducer may be integrated with two values, the ordinary and the synesthetic one. E.g., in grapheme-color synesthesia the letter "A" is bound to the values "black" and "red" of the attribute color. A plausible hypothesis for the wide-ranging phenomenon of synesthesia thus might be that it results from a process of hyperbinding. If one assumes the neurobiological hypothesis that binding is achieved by some sort of neural synchronization between neurons that code for attribute values (perceptual features, motor affordances, etc.) then hyperbinding might be neurobiologically manifested by "too much" synchrony between "too many" neural feature maps and clusters in "too many" cortical regions. Alternative hypothesis of how binding is achieved might lead to analogous predictions.

CONCLUSIONS

This paper undertakes the broader attempt of understanding the role of sensory-motor processes in synesthesia as part of a theory of mental representation as emulation. This regards the involvement of emulations in higher-level cognitive functions such as visual and motor imagery, object and action recognition, iconic memory, the representation of object and action concepts and language comprehension (Rumiati and Caramazza, 2005; Rumiati et al., 2010). The reconsideration of mental processes involved in synesthesia-like experiences proposed here will lead to vital implications not only for synesthesia research, but also generally for theories of perception and cognition. The wide spectrum of synesthesia-like processing extends itself from multimodal sensory or sensory-motor phenomena, through universal cross-modal or cross-domain correspondences to linguistic metaphors (Day, 1996; Martino and Marks, 2001; Sagiv and Ward, 2006), linking sensory-motor contingency and emulation with synesthetic associations, language comprehension, and social competencies. The interpersonal experience of mirror-sensory synesthesia is thought to rely on hyperactivity of the same brain mechanisms that we all use when observing another person's experiences. It therefore offers insights into how our brain shapes interpersonal representations between self and others.

The comparison between different psychological phenomena employing synesthesia-like processing as well as the use of a peculiar type of synesthesia have the potential of challenging conventional thinking on this phenomenon and the existing interpretations. If swimming style as a kinesthetic simulation can

induce synesthetic color perception, then interpretations in terms of bottom-up hard-wired structural connectivity, postulated not only between word/grapheme form and color areas in grapheme-color synesthesia, but generally in synesthesia research, may be challenged. Interpretations in terms of more top-down associations that are related to complex semantic representations might fit better.

REFERENCES

Aspell, J. E., Lenggenhager, B., and Blanke, O. (2009). Keeping in touch with one's self: multisensory mechanisms of self-consciousness. *PLoS ONE* 4, e6488. doi:10.1371/journal.pone.0006488

Avenanti, A., Bueti, D., Galati, G., and Aglioti, S. M. (2005). Transcranial magnetic stimulation highlights the sensorimotor side of empathy for pain. *Nat. Neurosci.* 8, 955–960.

Banissy, M. J., Cohen Kadosh, R., Maus, G., Walsh, V., and Ward, J. (2009). Prevalence, characteristics and a neuro-cognitive model of mirror-touch synaesthesia. *Exp. Brain Res.* 198, 261–272.

Banissy, M. J., and Ward, J. (2007). Mirror-touch synaesthesia is linked with empathy. *Nat. Neurosci.* 10, 815–816.

Baron-Cohen, S., and Harrison, J. E. (ed.). (1997). *Synaesthesia: Classic and Contemporary Readings.* Cambridge: Blackwell.

Baron-Cohen, S., Wyke, M. A., and Binnie, C. (1987). Hearing words and seeing colours: an experimental investigation of a case of synaesthesia. *Perception* 16, 761–767.

Barsalou, L. W. (1992). "Frames, concepts, and conceptual fields," in *Frames, Fields, and Contrasts: New Essays in Semantic and Lexical Organization*, eds E. Kittay and A. Lehrer (Hillsdale, NJ: Lawrence Erlbaum Associates), 21–74).

Barsalou, L. W. (2008). Grounded cognition. *Annu. Rev. Psychol.* 59, 617–645.

Bartels, A. (2006). Defending the structural concept of representation. *Theoria.* 21, 7–19.

Bastiaansen, J. A. C. J., Thious, M., and Keysers, C. (2009). Evidence for mirror systems in emotion. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 364, 2391–2404.

Blakemore, S. J., Bristow, D., Bird, G., Frith, C., and Ward, J. (2005). Somatosensory activations following the observation of touch and a case of vision-touch synaesthesia. *Brain* 128, 1571–1583.

Blanke, O., Landis, T., Spinelli, L., and Seeck, M. (2004). Out-of-body experience and autoscopy of neurological origin. *Brain* 127, 243–258.

Blanke, O., and Metzinger, T. (2009). Full body illusion and minimal phenomenal selfhood. *Trends Cogn. Sci. (Regul. Ed.)* 13, 7–13.

Bompas, A., and O'Regan, J. K. (2006). More evidence for sensorimotor adaptation in color perception. *J. Vis.* 6, 142–153.

Botvinick, M., and Cohen, J. (1998). Rubber hands "feel" touch that eyes see. *Nature* 391, 756.

Bremmer, F., Schlack, A., Shah, N. J., Zafiris, O., Kubischik, M., Hoffmann, K.-P., Zilles, K., and Fink, G. R. (2001). Polymodal motion processing in posterior parietal and premotor cortex: a human fMRI study strongly implies equivalencies between humans and monkeys. *Neuron* 29, 287–296.

Brugger, P. (2002). Reflective mirrors: perspective transformation in autoscopic phenomena. *Cogn. Neuropsychiatry* 7, 179–194.

Calvo-Merino, B., Glaser, D. E., Grezes, J., Passingham, R. E., and Haggard, P. (2005). Action observation and acquired motor skills: an FMRI study with expert dancers. *Cereb. Cortex* 15, 1243–1249.

Calvo-Merino, B., Grezes, J., Glaser, D. E., Passingham, R. E., and Haggard, P. (2006). Seeing or doing? Influence of visual and motor familiarity in action observation. *Curr. Biol.* 16, 1905–1910.

Carlsson, K., Petrovic, P., Skare, S., Petersson, K. M., and Ingvar, M. (2000). Tickling expecta-tions: neural processing in anticipation of a sensory stimulus. *J. Cogn. Neurosci.* 12, 691–703.

Carr, L., Iacoboni, M., Dubeau, M., Mazziotta, J. C., and Lenzi, G. (2003). Neural mechanisms of empathy in humans: a relay from neural systems for imitation to limbic areas. *Proc. Natl. Acad. Sci. U.S.A.* 100, 5497–5502.

Critchley, H. D., Wiens, S., Rotshstein, P., Öhman, A., and Dolan, R. J. (2004). Neural systems supporting interoceptive awareness. *Nat. Neurosci.* 7, 189–195.

Cytowic, R. E. (2002a). *Synesthesia: A Union of the Senses*, 2nd Edn. Cambridge: MIT Press.

Cytowic, R. E. (2002b). Touching tastes, seeing smells – and shaking up brain science. *Cerebrum* 4, 7–26.

Cytowic, R. E., and Wood, F. B. (1982). Synesthesia I: a review of major theories and their brain basis. *Brain Cogn.* 1, 23–35.

Day, S. (1996). Synaesthesia and synaesthetic metaphors. *Psyche (Stuttg)* 2, 1–19.

Dimberg, U. (1982). Facial reactions to facial expressions. *Psychophysiology* 19, 643–647.

Dixon, M. J., Smilek, D., Cudahy, C., and Merikle, P. M. (2000). Five plus two equals yellow. *Nature* 406, 365–365.

Dixon, M. J., Smilek, D., Duffy, P. L., Zanna, M. P., and Merikle, P. M. (2006). The role of meaning in grapheme-colour synaesthesia. *Cortex* 42, 243–252.

Ebisch, S. J., Ferri, F., Salone, A., Perrucci, M. G., D'Amico, L., Ferro, F. M., Romani, G. L., and Gallese, V. (2011). Differential involvement of somatosensory and interoceptive cortices during the observation of affective touch. *J. Cogn. Neurosci.* 23, 1808–1822.

Ebisch, S. J. H., Perrucci, M. G., Ferretti, A., Gratta, C. D., Romani, G. L., and Gallese, V. (2008). The sense of touch: embodied simulation in a visuotactile mirroring mechanism for observed animate or inanimate touch. *J. Cogn. Neurosci.* 20, 1611–1623.

Emrich, H. M., Zedler, M., and Schneider, U. (2002). Bindung und Hyperbindung in intermodaler Wahrnehmung und Synästhesie. *Z. Semiotik* 24, 31–37.

Emrich, H. M., Zedler, M., and Schneider, U. (2004). *Welche Farbe hat der Montag. Synästhesie: Das Leben mit verknüpften Sinnen,* 2. Aufl. Stuttgart: Hirzel.

Felleman, D. J., and van Essen, D. C. (1991). Distributed hierarchical processing in the primate cerebral cortex. *Cereb. Cortex* 1, 1–47.

Ferri, F., Frassinetti, F., Ardizzi, M., Costantini, M., and Gallese, V. (2012). A sensorimotor network for the bodily self. *J. Cogn. Neurosci.* 24, 1584–1595.

Fitzgibbon, B. M., Enticott, P. G., Rich, A. N., Giummarra, M. J., Georgiou-Karistianis, N., and Bradshaw, J. L. (2012). Mirror-sensory synaesthesia: exploring "shared" sensory experiences as synaesthesia. *Neurosci. Biobehav. Rev.* 36, 645–657.

Fitzgibbon, D. M., Giummarra, M. J., Georgiou-Karistianis, N., Enticott, P. G., and Bradshaw, J. L. (2010). Shared pain: from empathy to synaesthesia. *Neurosci. Biobehav. Rev.* 34, 500–512.

Frith, C. D., and Paulesu, E. (1997). "The physiological basis of synaesthesia," in *Synaesthesia. Classic and Contemporary Readings*, eds S. Baron-Cohen and J. E. Harrison (Oxford: Blackwell), 123–137.

Gallese, V. (2003). The roots of empathy: the shared manifold hypothesis and the neural basis of intersubjectivity. *Psychopathology* 36, 171–180.

Gallese, V. (2007). Before and below "theory of mind": embodied simulation and the neural correlates of social cognition. *Philos. Trans. R. Soc. B Biol. Sci.* 362, 659–669.

Gallese, V., and Goldman, A. (1998). Mirror neurons and the simulation theory of mind-reading. *Trends Cogn. Sci. (Regul. Ed.)* 2, 493–501.

Gallese, V., and Lakoff, G. (2005). The Brain's Concepts: the Role of the Sensory-Motor System in Reason and Language. *Cogn. Neuropsychol.* 22, 455–479.

Gallese, V., and Metzinger, T. (2003). Motor ontology: the representational reality of goals, actions, and selves. *Philos. Psychol.* 13, 365–388.

Gallese, V., and Sinigaglia, C. (2011). What is so special with embodied simulation. *Trends Cogn. Sci. (Regul. Ed.)* 15, 512–519.

Gibson, J. J. (1977). "The theory of affordances," in *Perceiving, Acting, and Knowing: Toward an Ecological Psychology*, eds R. Shaw and J. Bransford (Hillsdale, NJ: Lawrence Erlbaum), 67–82.

Goldman, A. (1989). Interpretation psychologized. *Mind Lang.* 4, 161–185.

Goldman, A. (1992). In defense of the simulation theory. *Mind Lang.* 7, 104–119.

Goldstone, R., and Barsalou, L. W. (1998). Reuniting perception and conception. *Cognition* 65, 231–262.

Gómez Milán, E., Iborra, O., de Córdoba, M. J., Juárez-Ramos, V., Rodríguez Artacho, M. A., and Rubio, J. L. (in press). The Kiki-Bouba effect: a case of personification and ideaesthesia. *J. Conscious. Stud.*

Gordon, R. M. (1986). Folk psychology as simulation. *Mind Lang.* 1, 158–171.

Gordon, R. M. (1995). "Simulation without introspection or inference from me to you," in *Mental Simulation*, eds M. Davies and T. Stone (Oxford: Blackwell), 53–67.

Grossenbacher, P. G. (1997). "Perception and sensory information in synaesthetic experience," in *Synaesthesia: Classic and Contemporary Readings*, eds S. Baron-Cohen and J. Harrison (Oxford: Blackwell), 148–172.

Grossenbacher, P. G., and Lovelace, C. T. (2001). Mechanisms of synesthesia: cognitive and physiological constraints. *Trends Cogn. Sci. (Regul. Ed.)* 5, 36–41.

Grush, R. (2004). The emulation theory of representation: motor control, imagery, and perception. *Behav. Brain Sci.* 27, 377–442.

Harrison, J. E., and Baron-Cohen, S. (1997). "Synaesthesia: an introduction," in *Synaesthesia. Classic and Contemporary Readings*, eds S. Baron-Cohen and J. E. Harrison (Oxford: Blackwell), 3–16.

Holle, H., Banissy, M. J., Wright, T., Bowling, N., and Ward, J. (2011). "That's not a real body": identifying stimulus qualities that modulate synaesthetic experiences of touch. *Conscious. Cogn.* 20, 720–726.

Hurley, S., and Noë, A. (2003). Neural plasticity and consciousness. *Biol. Philos.* 18, 131–168.

Ionta, S., Gassert, R., and Blanke, O. (2011). Multisensory and sensorimotor foundation of bodily self-consciousness – an interdisciplinary approach. *Front. Psychol.* 2:383. doi:10.3389/fpsyg.2011.00383

Ishida, H., Nakajima, K., Inase, M., and Murata, A. (2010). Shared mapping of own and others' bodies in visuotactile bimodal area of monkey parietal cortex. *J. Cogn. Neurosci.* 22, 83–96.

Jarick, M., Dixon, M., Maxwell, E., and Smilek, D. (2008). Time-space associations in synaesthesia: when input modality matters. *J. Vis.* 8, 525.

Jürgens, M. U., and Nikolić, D. (2012). Ideaesthesia: conceptual processes assign similar colours to similar shapes. *Transl. Neurosci.* 3, 22–27.

Keysers, C., Kaas, J. H., and Gazzola, V. (2010). Somatosensation in social perception. *Nat. Rev. Neurosci.* 11, 417–428.

Lamm, C., and Singer, T. (2010). The role of anterior insular cortex in social emotions. *Brain Struct. Funct.* 214, 579–591.

Lenggenhager, B., Tadi, T., Metzinger, T., and Blanke, O. (2007). Video ergo sum: manipulating bodily self-consciousness. *Science* 317, 1096–1099.

Lewis, J. W., and van Essen, D. C. (2000). Corticocortical connections of visual, sensorimotor, and multimodal processing areas in the parietal lobe of the macaque monkey. *J. Comp. Neurol.* 428, 112–137.

Mann, H., Korzenko, J., Carriere, J. S. A., and Dixon, M. J. (2009). Time-space synaesthesia – a cognitive advantage? *Conscious. Cogn.* 18, 619–627.

Martin, A. (2007). The representation of object concepts in the brain. *Annu. Rev. Psychol.* 58, 25–45.

Martin, A., Wiggs, C. L., Ungerleider, L., and Haxby, J. V. (1996). Neural correlates of category specific knowledge. *Nature* 379, 649–652.

Martino, G., and Marks, L. (2001). Synesthesia: strong and weak. *Curr. Dir. Psychol. Sci.* 10, 61–65.

Mattingley, J. B., Rich, A. N., Yelland, G., and Bradshaw, J. L. (2001). Unconscious priming eliminates automatic binding of colour and alphanumeric form in synaesthesia. *Nature* 410, 580–582.

Maye, A., and Werning, M. (2004). Temporal binding of non-uniform objects. *Neurocomputing* 58–60, 941–948.

Meltzoff, A. N., and Moore, M. K. (1977). Imitation of facial and manual gestures by human neonates. *Science* 198, 75–78.

Meltzoff, A. N., and Moore, M. K. (1983). Newborn infants imitate adult facial gestures. *Child Dev.* 54, 702–809.

Meltzoff, A. N., and Moore, M. K. (1994). Imitation, memory, and the representation of persons. *Infant Behav. Dev.* 17, 83–99.

Metzinger, T. (1999). *Subjekt und Selbstmodell: die Perspektivität phänomenalen Bewußtseins vor dem Hintergrund einer naturalistischen Theorie mentaler Repräsentation*. Paderborn: Mentis.

Metzinger, T. (2005). Out-of-body experiences as the origin of the concept of a soul. *Mind Matter* 3, 57–84.

Metzinger, T., and Gallese, V. (2003). The emergence of a shared action ontology: building blocks for

a theory. *Conscious. Cogn.* 12, 549–571.

Mroczko, A., Metzinger, T., Singer, W., and Nikolić, D. (2009). Immediate transfer of synesthesia to a novel inducer. *J. Vis.* 9, 25, 1–8.

Mroczko-Wąsowicz, A. (in press). *The Unity of Consciousness and the Phenomenon of Synesthesia* [*Die Einheit des Bewusstseins und das Phänomen der Synästhesie*]. Doctoral dissertation, Johannes Gutenberg, University of Mainz, Available at: http://ubm.opus.hbz-nrw.de/newest.php

Mroczko-Wąsowicz, A., and Nikolić, D. (in press). "Coloured alphabets in bi-lingual synaesthetes," in *Oxford Handbook of Synaesthesia*, eds J. Simner and E. Hubbard (Oxford University Press).

Nikolić, D., Jürgens, U. M., Rothen, N., Meier, B., and Mroczko, A. (2011). Swimming-style synesthesia. *Cortex* 47, 874–879.

Noë, A. (2005). *Action in Perception*. Cambridge: MIT Press.

Odgaard, E. C., Flowers, J. H., and Bradman, H. L. (1999). An investigation of the cognitive and perceptual dynamics of a color-digit synesthete. *Perception* 28, 651–664.

O'Regan, J. K., Myin, E., and Noë, A. (2006). Skill, corporality and alerting capacity in an account of sensory consciousness. *Prog. Brain Res.* 150, 55–68.

O'Regan, J. K., and Noë, A. (2001a). A sensorimotor account of vision and visual consciousness. *Behav. Brain Sci.* 24, 883–917.

O'Regan, J. K., and Noë, A. (2001b). Authors' response. Acting out our experience. *Behav. Brain Sci.* 24, 955–976.

Petersen, W., and Werning, M. (2007). "Conceptual fingerprints: lexical decomposition by means of frames – a neuro-cognitive model," in *Conceptual Structures: Knowledge Architectures for Smart Applications (LNAI 4604)*, eds U. Priss, S. Polovina, and R. Hill (Berlin: Springer), 415–428.

Pons, T. P., and Kaas, J. H. (1986). Corticocortical connections of area 2 of somatosensory cortex in macaque monkeys: a correlative anatomical and electrophysiological study. *J. Comp. Neurol.* 248, 313–335.

Proulx, M. J., and Störig, P. (2006). Seeing sounds and tingling tongues: qualia in synaesthesia and sensory substitution. *Anthropol. Philos.* 7, 135–150.

Pulvermüller, F. (2005). Brain mechanisms linking language and action. *Nat. Rev. Neurosci.* 6, 576–582.

Pulvermüller, F., and Fadiga, L. (2010). Active perception: sensorimotor circuits as a cortical basis for language. *Nat. Rev. Neurosci.* 11, 351–360.

Pulvermüller, F., Lutzenberger, W., and Preissl, H. (1999). Nouns and verbs in the intact brain: evidence from event-related potentials and high-frequency cortical responses. *Cereb. Cortex* 9, 497–506.

Ramachandran, V. S., and Hubbard, E. M. (2001a). Synaesthesia: a window into perception, thought and language. *J. Conscious. Stud.* 8, 3–34.

Ramachandran, V. S., and Hubbard, E. M. (2001b). Psychophysical investigations into the neural basis of synaesthesia. *Proc. R. Soc. Lond. B Biol. Sci.* 268, 979–983.

Ramachandran, V. S., and Rogers-Ramachandran, D. (1996). Synaesthesia in phantom limbs induced with mirrors. *Proc. R. Soc. Lond. B Biol. Sci.* 263, 377–386.

Raos, V., Franchi, G., Gallese, V., and Fogassi, L. (2003). Somatotopic organization of the lateral part of area f2 (dorsal premotor cortex) of the macaque monkey. *J. Neurophysiol.* 89, 1503–1518.

Rappelsberger, P., Weiss, S., and Schack, B. (2000). "Coherence and phase relations between EEG traces recorded from different locations," in *Time and the Brain*, ed. R. Müller (Amsterdam: Harwood Academic Publishers), 297–330.

Rees, W. D. (1971). The hallucinations of widowhood. *Br. Med. J.* 4, 37–41.

Rich, A. N., Bradshaw, J. L., and Mattingley, J. B. (2005). A systematic, large-scale study of synaesthesia: implications for the role of early experience in lexical-colour associations. *Cognition* 98, 53–84.

Rizzolatti, G., and Craighero, L. (2004). The mirror-neuron system. *Annu. Rev. Neurosci.* 27, 169–192.

Rizzolatti, G., Fadiga, L., Fogassi, L., and Gallese, V. (1996). Premotor cortex and the recognition of motor actions. *Cogn. Brain Res.* 3, 131–141.

Rizzolatti, G., Fogassi, L., and Gallese, V. (2001). Neurophysiological mechanisms underlying the understanding and imitation of action. *Nat. Rev. Neurosci.* 2, 661–670.

Rizzolatti, G., and Luppino, G. (2001). The cortical motor system. *Neuron* 31, 889–901.

Rizzolatti, G., and Sinigaglia, C. (2010). The functional role of the parieto-frontal mirror circuit: interpretations and misinterpretations. *Nat. Rev.* 11, 264–274.

Rodríguez-Ferreiro, J., Menéndez, M., Ribacoba, R., and Cuetos, F. (2009). Action naming is impaired in Parkinson disease patients. *Neuropsychologia* 47, 3271–3274.

Rozzi, S., Calzavara, R., Belmalih, A., Borra, E., Gregoriou, G. G., Matelli, M., and Luppino, G. (2006). Cortical connections of the inferior parietal cortical convexity of the macaque monkey. *Cereb. Cortex* 16, 1389–1417.

Rumiati, R. I., and Caramazza, A. (2005). The multiple functions of sensory-motor representations: an introduction. *Cogn. Neuropsychol.* 22, 259–261.

Rumiati, R. I., Papeo, L., and Corradi-Dell'Acqua, C. (2010). Higher-level motor processes. *Ann. N. Y. Acad. Sci.* 1191, 219–241.

Sagiv, N., Ilbeigi, I., and Ben-Tal, O. (2011). Reflections on synaesthesia, perception, and cognition. *Intellectica* 55, 81–94.

Sagiv, N., and Robertson, L. C. (2005). "Synesthesia and the Binding Problem," in *Synesthesia: Perspectives from Cognitive Neuroscience*, eds L. C. Robertson and N. Sagiv (New York: Oxford University Press), 90–107.

Sagiv, N., and Ward, J. (2006). "Cross-modal interactions: lessons from synaesthesia," in *Visual Perception – Fundamentals of Awareness: Multi-Sensory Integration of High-Order Perception.* Progress in Brain Research 155, eds S. Martinez-Conde, S. L. Macknik, L. M. Martinez, J. M. Alonso, and P. U. Tse (London: Elsevier), 263–275.

Sahin, E., and Erdogan, S. T. (2009). "Towards linking affordances with mirror/canonical neurons," in *International Symposium on Computer and Information Sciences*, Guzelyurt, 397–404.

Schnitzler, A., Timmermann, L., and Gross, J. (2006). Physiological and pathological oscillatory networks in the human motor system. *J. Physiol.* 99, 3–7.

Sforza, A., Bufalari, I., Haggard, P., and Aglioti, S. M. (2010). My face in yours: visuo-tactile facial stimulation influences sense of identity. *Soc. Neurosci.* 5, 148–162.

Shmuel, A., and Grinvald, A. (2000). Coexistence of linear zones and pinwheels within orientation maps in cat visual cortex. *Proc. Natl. Acad. Sci. U.S.A.* 97, 5568–5573.

Simner, J. (2007). Beyond perception: synaesthesia as a psycholinguistic phenomenon. *Trends Cogn. Sci. (Regul. Ed.)* 11, 23–29.

Simner, J., and Ward, J. (2006). The taste of words on the tip of the tongue. *Nature* 444, 438.

Singer, W. (1999). Neuronal synchrony: a versatile code for the definition of relations? *Neuron* 24, 49–65.

Sinha, J. (ed.). (2009). *Synästhesie der Gefühle.* Hannover: Synaisthesis.

Smilek, D., Callejas, A., Dixon, M. J., and Merikle, P. M. (2007). Ovals of time: time-space associations in synaesthesia. *Conscious. Cogn.* 16, 507–519.

Straube, T., and Miltner, W. H. R. (2011). Attention to aversive emotion and specific activation of the right insula and right somatosensory cortex. *NeuroImage* 54, 2534–2538.

Tacca, M. C. (2010). *Seeing Objects: The Structure of Visual Representation.* Padeborn: Mentis.

Thomas, R., Press, C., and Haggard, P. (2006). Shared representations in body perception. *Acta Psychol. (Amst.)* 121, 317–330.

Treisman, A. (1993). "The perception of features and objects," in *Attention: Selection, Awareness and Control. A Tribute to Donald Broadbent*, eds A. Baddeley and L. Weiskrantz (Oxford: Clarendon Press University), 5–35.

Treisman, A. (2005). "Synaesthesia: implications for attention, binding and consciousness – a commentary," in *Synesthesia: Perspectives from Cognitive Neuroscience*, eds L. Robertson and N. Sagiv (Oxford: Oxford University Press), 239–254.

Tsakiris, M. (2008). Looking for myself: current multisensory input alters self-face recognition. *PLoS ONE* 3, e4040. doi:10.1371/journal.pone.0004040

Ward, J. (2004). Emotionally mediated synaesthesia. *Cogn. Neuropsychol.* 21, 761–772.

Ward, J., Tsakanikos, E., and Bray, A. (2006). Synaesthesia for reading and playing musical notes. *Neurocase* 12, 27–34.

Werning, M. (2005). The temporal dimension of thought: cortical foundations of predicative representation. *Synthese* 146, 203–224.

Werning, M. (2010). Complex first? On the evolutionary and developmental priority of semantically thick words. *Philos. Sci.* 77, 1096–1108.

Werning, M. (2012). "Non-symbolic compositional representation and its neuronal foundation: towards an emulative semantics," in *The Oxford Handbook of Compositionality*, eds M. Werning, W. Hinzen, and M. Machery (Oxford: Oxford University Press), 635–656.

Werning, M., Fleischhauer, J., and Beseoglu, H. (2006). "The cognitive accessibility of synaesthetic metaphors," in *Proceedings of the Twenty-Eighth Annual Conference of the Cognitive Science Society*, eds R. Sun and N. Miyake (London: Lawrence Erlbaum), 2365–2370.

Werning, M., and Maye, A. (2005). "Frames, coherency chains and hierarchical binding: the cortical implementation of complex concepts," in *Proceedings of the Twenty-Seventh Annual Conference of the Cognitive Science Society*, eds B. G. Bara, L. Barsalou, M. Bucciarelli (Mahwah, NJ: Erlbaum), 2347–2352.

Werning, M., and Maye, A. (2007). The cortical implementation of complex attribute and substance concepts: synchrony, frames, and hierarchical binding. *Chaos Complexity Lett.* 2, 435–452.

Wicker, B., Keysers, C., Plailly, J., Royet, J. P., Gallese, V., and Rizzolatti, G. (2003). Both of us disgusted in my insula: the common neural basis of seeing and feeling disgust. *Neuron* 40, 655–664.

Williams, J. M. (1976). Synesthetic adjectives: a possible law of semantic change. *Language* 52, 461–478.

Xiao, Y., Wang, Y., and Felleman, D. J. (2003). A spatially organized representation of colour in macaque cortical area V2. *Nature* 421, 535–539.

Yu, N. (2003). Synesthetic metaphor: a cognitive perspective. *J. Literary Semantics* 32, 19–34.

Zamboni, G., Budriesi, C., and Nichelli, P. (2005). "Seeing oneself": a case of autoscopy. *Neurocase* 11, 212–215.