

FEEDFORWARD AND FEEDBACK PROCESSES IN VISION

EDITED BY : Hulusi Kafaligonul, Bruno G. Breitmeyer and Haluk Öğmen
PUBLISHED IN: Frontiers in Psychology



frontiers

Frontiers Copyright Statement

© Copyright 2007-2015 Frontiers Media SA. All rights reserved.

All content included on this site, such as text, graphics, logos, button icons, images, video/audio clips, downloads, data compilations and software, is the property of or is licensed to Frontiers Media SA ("Frontiers") or its licensees and/or subcontractors. The copyright in the text of individual articles is the property of their respective authors, subject to a license granted to Frontiers.

The compilation of articles constituting this e-book, wherever published, as well as the compilation of all other content on this site, is the exclusive property of Frontiers. For the conditions for downloading and copying of e-books from Frontiers' website, please see the Terms for Website Use. If purchasing Frontiers e-books from other websites or sources, the conditions of the website concerned apply.

Images and graphics not forming part of user-contributed materials may not be downloaded or copied without permission.

Individual articles may be downloaded and reproduced in accordance with the principles of the CC-BY licence subject to any copyright or other notices. They may not be re-sold as an e-book.

As author or other contributor you grant a CC-BY licence to others to reproduce your articles, including any graphics and third-party materials supplied by you, in accordance with the Conditions for Website Use and subject to any copyright notices which you include in connection with your articles and materials.

All copyright, and all rights therein, are protected by national and international copyright laws.

The above represents a summary only. For the full conditions see the Conditions for Authors and the Conditions for Website Use.

ISSN 1664-8714

ISBN 978-2-88919-594-7

DOI 10.3389/978-2-88919-594-7

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

FEEDFORWARD AND FEEDBACK PROCESSES IN VISION

Topic Editors:

Hulusi Kafaligonul, Bilkent University, Turkey

Bruno G. Breitmeyer, University of Houston, USA

Haluk Öğmen, University of Houston, USA

The visual system consists of hierarchically organized distinct anatomical areas functionally specialized for processing different aspects of a visual object (Felleman & Van Essen, 1991). These visual areas are interconnected through ascending feedforward projections, descending feedback projections, and projections from neural structures at the same hierarchical level (Lamme et al., 1998). Accumulating evidence from anatomical, functional and theoretical studies suggests that these three projections play fundamentally different roles in perception. However, their distinct functional roles in visual processing are still subject to debate (Lamme & Roelfsema, 2000).

The focus of this Research Topic is the roles of feedforward and feedback projections in vision. Even though the notions of feedforward, feedback, and reentrant processing are widely accepted, it has been found difficult to distinguish their individual roles on the basis of a single criterion. We welcome empirical contributions, theoretical contributions and reviews that fit into any one (or a combination) of the following domains: 1) their functional roles for perception of specific features of a visual object 2) their contributions to the distinct modes of visual processing (e.g., pre-attentive vs. attentive, conscious vs. unconscious) 3) recent techniques/methodologies to identify distinct functional roles of feedforward and feedback projections and corresponding neural signatures. We believe that the current Research Topic will not only provide recent information about feedforward/feedback processes in vision but also contribute to the understanding fundamental principles of cortical processing in general.

Citation: Kafaligonul, H., Breitmeyer, B. G., Öğmen, H., eds. (2015). Feedforward and Feedback Processes in Vision. Lausanne: Frontiers Media. doi: 10.3389/978-2-88919-594-7

Table of Contents

- 04 Feedforward and feedback processes in vision**
Hulusi Kafaligonul, Bruno G. Breitmeyer and Haluk Öğmen
- 07 A feed-forward spiking model of shape-coding by IT cells**
August Romeo and Hans Supèr
- 16 Early recurrent feedback facilitates visual object recognition under challenging conditions**
Dean Wyatte, David J. Jilk and Randall C. O'Reilly
- 26 The temporal window of individuation limits visual capacity**
Andreas Wutz and David Melcher
- 39 Neural dynamics of feedforward and feedback processing in figure-ground segregation**
Oliver W. Layton, Ennio Mingolla and Arash Yazdanbakhsh
- 59 Adaptive learning in a compartmental model of visual cortex—how feedback enables stable category learning and refinement**
Georg Layher, Fabian Schrodtt, Martin V. Butz and Heiko Neumann
- 78 Inter-element orientation and distance influence the duration of persistent contour integration**
Lars Strother and Danila Alferov
- 86 Combined contributions of feedforward and feedback inputs to bottom-up attention**
Peyman Khorsand, Tirin Moore and Alireza Soltani
- 97 Limits to the usability of iconic memory**
Ronald A. Rensink
- 106 Reentrant processing mediates object substitution masking: comment on Pöder (2013)**
Vincent Di Lollo
- 111 The changing picture of object substitution masking: reply to Di Lollo (2014)**
Endel Pöder
- 114 A computational investigation of feedforward and feedback processing in metacontrast backward masking**
David N. Silverstein
- 128 Contributions of cortical feedback to sensory processing in primary visual cortex**
Lucy S. Petro, Luca Vizioli and Lars Muckli
- 136 Visual crowding illustrates the inadequacy of local vs. global and feedforward vs. feedback distinctions in modeling visual perception**
Aaron M. Clarke, Michael H. Herzog and Gregory Francis
- 148 A hidden ambiguity of the term “feedback” in its use as an explanatory mechanism for psychophysical visual phenomena**
Talis Bachmann

Feedforward and feedback processes in vision

Hulusi Kafaligonul^{1*}, Bruno G. Breitmeyer^{2,3} and Haluk Öğmen^{3,4}

¹ National Magnetic Resonance Research Center (UMRAM), Bilkent University, Ankara, Turkey, ² Department of Psychology, University of Houston, Houston, TX, USA, ³ Center for Neuro-Engineering and Cognitive Science, University of Houston, Houston, TX, USA, ⁴ Department of Electrical and Computer Engineering, University of Houston, Houston, TX, USA

Keywords: vision, visual system, feedforward, feedback, mechanisms

Hierarchical processing is key to understanding vision. The visual system consists of hierarchically organized distinct anatomical areas functionally specialized for processing different aspects of a visual object (Felleman and Van Essen, 1991). These visual areas are interconnected through ascending feedforward projections, descending feedback projections, and projections from neural structures at the same hierarchical level (Lamme et al., 1998). Even though accumulating evidence suggests that these three projections play fundamentally different roles in perception, their distinct functional roles in visual processing are still subject to debate (Lamme and Roelfsema, 2000). The focus of this Research Topic was the roles of feedforward and feedback projections in vision. In fact, our motivation to edit this Research Topic was threefold: (i) to provide current views on the functional roles of feedforward and feedback projections for the perception of specific visual features, (ii) to invite recent views on how these functional roles contribute to the distinct modes of visual processing, (iii) to provide recent methodological views to identify distinct functional roles of feedforward and feedback projections and corresponding neural signatures. As summarized below, these aims are largely achieved thanks to fourteen contributions to this issue.

OPEN ACCESS

Edited and reviewed by:

Philippe G. Schyns,
University of Glasgow, UK

*Correspondence:

Hulusi Kafaligonul,
hulusi@bilkent.edu.tr

Specialty section:

This article was submitted to
Perception Science, a section of the
journal *Frontiers in Psychology*

Received: 22 February 2015

Accepted: 25 February 2015

Published: 12 March 2015

Citation:

Kafaligonul H, Breitmeyer BG and
Öğmen H (2015) Feedforward and
feedback processes in vision.
Front. Psychol. 6:279.
doi: 10.3389/fpsyg.2015.00279

Feedforward and Feedback Projections for Different Aspects of a Visual Object

The cortical areas and the way they connect with each other lead to distinct pathways functionally specialized for processing different aspects of a visual object (Van Essen and Gallant, 1994). For example, the ventral processing stream has been associated with object recognition and identification. Romeo and Supér (2014) have constructed a feedforward spiking hierarchical model for simulating IT cortex along the ventral stream. The simulation results indicate that figure-ground segregation occurs at an earlier level of processing relative to the level at which shape selection takes place. Wyatte et al. (2014) propose that object recognition requires more than feedforward processing. By reviewing a number of studies, they first differentiate two types of additional processing along the ventral stream: (i) early, short-distance (local) recurrent processes, and (ii) late, long-distance feedback processes related to attention. They further propose that early local recurrent feedback plays a functionally distinct role in attention-independent stimulus disambiguation, since it facilitates object recognition well before the onset of any attentional influences. Wutz and Melcher (2014) provide a review on temporal window for object recognition and individuation. They propose that mid-level vision adopts a temporal window whose duration is short enough for picking out separate objects (without appreciable smearing of their retinal images when they move), while simultaneously being long enough to integrate sufficient sensory information for accurate detection. Based on psychophysical and neurophysiological data, they suggest that phase synchronization plays a key role in this process

by coordinating feedforward and feedback involved in complex and dynamic visual scenes. Several studies in this collection emphasize the role of feedback projections at different levels of processing within the ventral stream. Layton et al. (2014) propose a dynamic hierarchical model which can effectively perform figure-ground segregation in visual scenes with multiple objects. Their results indicate that the inhibitory feedback sharpens the population activity in the “lower stage” and that the dynamic balancing of feedforward signals with specific feedback mechanisms is crucial to identifying figural region. Furthermore, Layher et al. (2014) describe a model architecture to investigate the role of feedback mechanisms in learning new categories of visual objects. They basically use two types of feedback mechanisms to achieve seamless and automatic acquisition of category representation by an unsupervised learning mechanism integrated into a recurrent network architecture. Hence, they not only address the classic stability/plasticity dilemma but also elucidate how the predictive power of feedback mechanisms together with the feedforward sweep realize associative memory. Contour integration has been considered to be another crucial stage of visual object recognition. By varying the inter-element properties in a perceptual fading paradigm, Strother and Alferov (2014) focus on the individual roles of bottom-up feedforward and top-down feedback processing in such integration. In agreement with previous reports, their findings highlight the importance of feedforward processes in primary visual cortex (V1) and shape-related feedback from higher-tier visual cortical areas for contour integration.

Roles of Feedforward and Feedback Projections in Different Modes of Visual Processing

Accumulating evidence from modeling and experimental studies indicates that feedforward and feedback projections play important roles in different modes of visual processing and attention. However, their distinct contributions are still controversial. Khorsand et al. (2015) set the stage for feedforward and feedback contributions to the exogenous attentional selection. Bottom-up exogenous attention has been considered to rely only on feedforward processing of the external inputs. However, Khorsand et al. (2015) review recent experimental and theoretical studies supporting the view that stimulus dependent processing involves feedback connections and signals running in top-down direction of the hierarchy as well. Their review raises an important conceptual issue and provides an account of feedforward and feedback contributions to exogenous attentional shifts. In another study, Rensink (2014) identifies different levels of processing for iconic memory by using a modified visual search paradigm. Besides feedforward processing, he highlights the importance of two types of feedback projections (due to horizontal connections within a level as well as links between different levels) for iconic memory. He further characterizes “iconic,” “preattentive,” and “attentive” representations within this framework. As briefly mentioned above, based

on the literature about visual object recognition, Wyatte et al. (2014) dissociate the late top-down processing originating from frontoparietal areas from early recurrent local projections within the ventral processing stream. They also review some studies emphasizing that this late top-down processing to striate cortex provides attentional support for salient or behaviorally-relevant features.

Explaining Various Visual Phenomena by Feedforward and Feedback Processes

The notions of feedforward and feedback processing have been extensively used to explain various visual phenomena. Di Lollo et al. (2014) hypothesizes that reentrant (feedback) processing gives the best account for a form of visual masking called object substitution masking (OSM). On the other hand, Pöder et al. (2014) presents the contrasting view that reentrant processing is not necessary to explain OSM and that the attentional gating model is the simplest and most reasonable explanation for OSM results. Silverstein (2015) takes an interesting approach to examine the roles of feedforward and feedback processes in visual backward masking. Using a biophysical model of V1 and V2, he explains visual processing in terms of interacting cortical attractors. The simulation results indicate that both feedforward and feedback processes predict several aspects of backward masking. Additionally, Petro et al. (2014) focus on the functional role of cortical feedback projections on V1. By reviewing the most current theory and experimental data, they discuss how top-down feedback signals conveying information from higher-processing stages (e.g., prediction, reward, memory and behavioral context) are involved in shaping sensory processing in V1 and hence, explain recent experimental findings along this direction.

A contrasting view is provided by Clarke et al. (2014), who argue against the usefulness of making feedforward and feedback distinctions for explaining experimental results. They tested three existing models with different local/global and feedback/feedforward characteristics to see whether they can account for some recent findings on visual crowding. All three models failed to predict the results even qualitatively. Clarke et al. (2014) discuss these model failures within the context of a broader view and suggest that the dichotomies such as feedforward/feedback and local/global may not be useful for scientists designing experiments to understand vision. Bachmann (2014) argues another interesting point. He basically posits that experimental findings that have been proposed to support models of specific top-down re-entrant processing could equally support those with a generic, non-specific feedback loop.

Taken together, the research topic presents a timely addition to the field of vision research and to understanding the functional principles of brain in general. It provides an update on the roles of feedforward and feedback projections in several but not all types of visual processing. For example, an update about the roles of feedforward and feedback projections in motion processing (mostly carried out by the dorsal pathway) is missing. The advent of optogenetics and neuroimaging has provided

additional remarkable investigative tools. How these recent techniques will contribute to the prevailing arguments of feedforward and feedback projections in vision is still open. We hope this issue will inspire the readers and act as a catalyst for future work on the issues of feedforward and feedback processes in vision.

References

- Bachmann, T. (2014). A hidden ambiguity of the term “feedback” in its use as an explanatory mechanism for psychophysical visual phenomena. *Front. Psychol.* 5:780. doi: 10.3389/fpsyg.2014.00780
- Clarke, A. M., Herzog, M. H., and Francis, G. (2014). Visual crowding illustrates the inadequacy of local vs. global and feedforward vs. feedback distinctions in modeling visual perception. *Front. Psychol.* 5:1193. doi: 10.3389/fpsyg.2014.01193
- Di Lollo, V. (2014). Reentrant processing mediates object substitution masking: comment on Pöder (2013). *Front. Psychol.* 5:819. doi: 10.3389/fpsyg.2014.00819
- Felleman, D. J., and Van Essen, D. C. (1991). Distributed hierarchical processing in the primate cerebral cortex. *Cereb. Cortex* 1, 1–47. doi: 10.1093/cercor/1.1.1
- Khorsand, P., Moore, T., and Soltani, A. (2015). Combined contributions of feedforward and feedback inputs to bottom-up attention. *Front. Psychol.* 6:155. doi: 10.3389/fpsyg.2015.00155
- Lamme, V. A. F., and Roelfsema, P. R. (2000). The distinct modes of vision offered by feedforward and recurrent processing. *Trends Neurosci.* 23, 571–579. doi: 10.1016/S0166-2236(00)01657-X
- Lamme, V. A. F., Supér, H., and Spekreijse, H. (1998). Feedforward, horizontal, and feedback processing in the visual cortex. *Curr. Opin. Neurobiol.* 8, 529–535. doi: 10.1016/S0959-4388(98)80042-1
- Layher, G., Schrod, F., Butz, M. V., and Neumann, H. (2014). Adaptive learning in a compartmental model of visual cortex—how feedback enables stable category learning and refinement. *Front. Psychol.* 5:1287. doi: 10.3389/fpsyg.2014.01287
- Layton, O. W., Mingolla, E., and Yazdanbakhsh, A. (2014). Neural dynamics of feedforward and feedback processing in figure-ground segregation. *Front. Psychol.* 5:972. doi: 10.3389/fpsyg.2014.00972
- Petro, L. S., Vizioli, L., and Muckli, L. (2014). Contributions of cortical feedback to sensory processing in primary visual cortex. *Front. Psychol.* 5:1223. doi: 10.3389/fpsyg.2014.01223
- Pöder, E. (2014). The changing picture of object substitution masking: reply to Di Lollo (2014). *Front. Psychol.* 5:1004. doi: 10.3389/fpsyg.2014.01004
- Rensink, R. A. (2014). Limits to the usability of iconic memory. *Front. Psychol.* 5:971. doi: 10.3389/fpsyg.2014.00971
- Romeo, A., and Supér, H. (2014). A feed-forward spiking model of shape-coding by IT cells. *Front. Psychol.* 5:481. doi: 10.3389/fpsyg.2014.00481
- Silverstein, D. N. (2015). A computational investigation of feedforward and feedback processing in metacontrast backward masking. *Front. Psychol.* 6:6. doi: 10.3389/fpsyg.2015.00006
- Strother, L., and Alferov, D. (2014). Inter-element orientation and distance influence the duration of persistent contour integration. *Front. Psychol.* 5:1273. doi: 10.3389/fpsyg.2014.01273
- Van Essen, D. C., and Gallant, J. L. (1994). Neural mechanisms of form and motion processing in the primate visual system. *Neuron* 13, 1–10. doi: 10.1016/0896-6273(94)90455-3
- Wutz, A., and Melcher, D. (2014). The temporal window of individuation limits visual capacity. *Front. Psychol.* 5:952. doi: 10.3389/fpsyg.2014.00952
- Wyatte, D., Jilk, D. J., and O'Reilly, R. C. (2014). Early recurrent feedback facilitates visual object recognition under challenging conditions. *Front. Psychol.* 5:674. doi: 10.3389/fpsyg.2014.00674

Acknowledgments

We would like to thank all the contributors, reviewers and the Frontiers staff for helping us make this Research topic possible. HK was supported by a Co-Funded Brain Circulation Fellowship (TUBITAK 112C010).

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.

Copyright © 2015 Kafaligonul, Breitmeyer and Ögmen. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



A feed-forward spiking model of shape-coding by IT cells

August Romeo¹ and Hans Supér^{1,2,3*}

¹ Department of Basic Psychology, Faculty of Psychology, University of Barcelona, Barcelona, Spain

² Institute for Brain, Cognition and Behavior (IR3C), Barcelona, Spain

³ Catalan Institution for Research and Advanced Studies (ICREA), Barcelona, Spain

Edited by:

Hulusi Kafaligonul, Bilkent
University, Turkey

Reviewed by:

Ozgur Yilmaz, Turgut Ozal
University, Turkey
Saumil Surendra Patel, Baylor
College of Medicine, USA

*Correspondence:

Hans Supér, Department of Basic
Psychology, Faculty of Psychology,
University of Barcelona, Pg. Vall
d'Hebron 171, 08035 Barcelona,
Spain
e-mail: hans.super@icrea.cat

The ability to recognize a shape is linked to figure-ground (FG) organization. Cell preferences appear to be correlated across contrast-polarity reversals and mirror reversals of polygon displays, but not so much across FG reversals. Here we present a network structure which explains both shape-coding by simulated IT cells and suppression of responses to FG reversed stimuli. In our model FG segregation is achieved before shape discrimination, which is itself evidenced by the difference in spiking onsets of a pair of output cells. The studied example also includes feature extraction and illustrates a classification of binary images depending on the dominance of vertical or horizontal borders.

Keywords: spiking model, feed-forward, shape, classifiers, IT

INTRODUCTION

Neurons in the inferior temporal cortex (IT) have been linked to visual shape representation and object recognition (Rolls et al., 1977; Logothetis et al., 1995; DiCarlo and Maunsell, 2000; Riesenhuber and Poggio, 2000; Rollenhagen and Olson, 2000). Lesions in this area result in visual agnosia (Farah, 1990). fMRI studies in humans show how objects activate this part of the cortex and how restricted spots of it are driven by specific classes of stimuli (Desimone, 1991; Malach et al., 1995; Tanaka, 1996). Individual IT cells discriminate, in particular, the shape or color of the stimulus or both parameters (Desimone et al., 1985). Their selective responses are maintained across changes in the size or location on the retina. Actually, in Baylis and Driver's paper (Baylis and Driver, 2001), the visual shape preferences of IT neurons of monkeys were also invariant under two stimulus transformations. The stimuli were different polygon displays and the correlated transforms consisted of either a change in the contrast polarity between the figure and the background or a mirror image. That form of invariance or symmetry is often referred to as "generalization" and its degree of exactness is typically subject to some amount of elasticity.

The exact computational process by which the IT region represents shape remains controversial (Peterson et al., 1991). A central mechanism herein is figure-ground (FG) segmentation, or the segregation of visual information into objects and their surrounding regions (Rubin, 1958). If this task were performed by the brain solely through the contours distinguishing the input displays, then generalization under FG reversal would be expected as well. However, it was absent from Baylis and Driver's results (Baylis and Driver, 2001). Thus, shape coding is not exclusively based on the processing of contour features. For explaining such results, some type of segregation has to be included.

Similarly, psychological findings on human visual shape judgments indicate that one-sided assignment of edges plays a crucial role (Baylis and Driver, 1995a,b; Nakayama et al., 1995; Rubin,

2001). Such an assignment means that the border is "owned" by the side which is imagined "in front," and regarded as "figure." Since the dividing curve is the same, the background shares the same informative contour as the original figure, and has its "profile" embedded. Even so, humans typically rate a mirror image of a figure as more similar to the original than the background in isolation (Hoffman and Richards, 1984). Likewise, IT cell responses generalize more strongly across mirror imaging than across FG reversal. That is, they are activated by shape components only after FG assignment (Baylis and Driver, 1995c, see also Hulleman et al., 2005). Apparently, the shape of an object is then coded after the perception of it as a separate entity (however, this issue was contended for a long time and other alternatives were offered, e.g., by Peterson et al., 1991).

We have already favored the idea that the visual system uses one-sided edge assignment to figures (Supér et al., 2010). In fact, we developed a spiking model which by means of surround inhibition gave FG responses. We concluded that feed-forward connections contribute to the neural mechanisms underlying FG organization, namely, that the phenomenon arises from the computations that happen in earlier stages. Feedback merely controls FG segregation by influencing the neural firing patterns of feed-forward projecting neurons (Supér and Romeo, 2011). Motivated by all the above observations, we have constructed a network structure, based on our previous work, which explains both the suppression of responses to FG reversed stimuli and the possibility of achieving shape selectivity for the other transformations.

In summary, when an IT cell is selective to a certain shape, the fact that this shape is presented as figure or as ground does matter. We shall be upholding the hypothesis that FG segregation takes place before feature extraction and further processing (alternative hypotheses admitted that shape recognition was possible before FG relationships were determined—Peterson et al., 1991). The present work includes these specific elements: (1) A proposed mechanism for figure segregation: local excitation and global

inhibition leading to rebound spiking on regions of smallest area, already introduced by Supèr et al. (2010), and (2) An additional structure for extracting and processing features which, if applied to the considered image type, classifies shapes by vertical/horizontal edge dominance and reproduces the observed weakening in the response when the shape goes into the background.

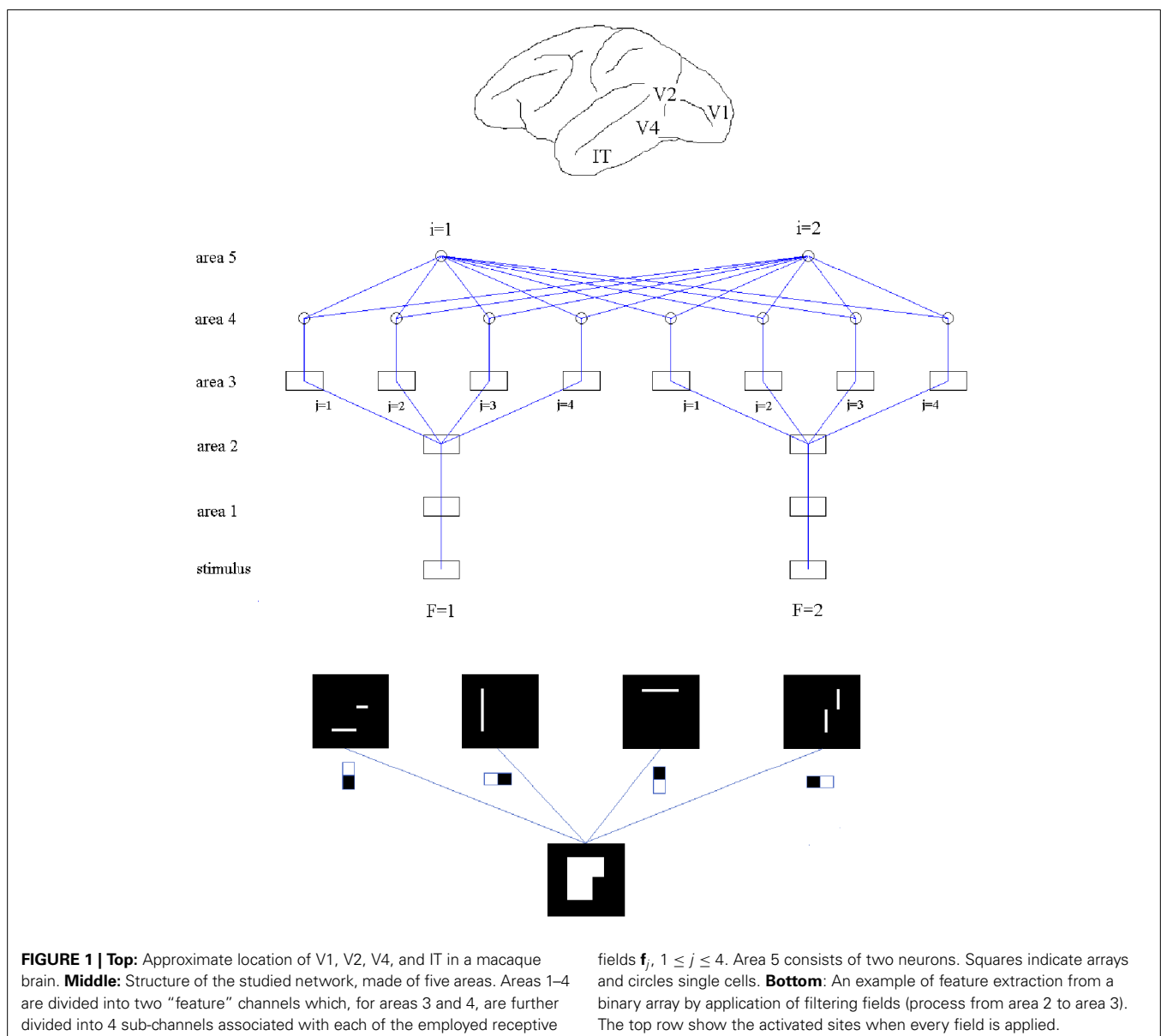
MATERIALS AND METHODS

Our network consists of five areas made of Izhikevich's neurons (Izhikevich, 2003, 2007). The dynamics of that neural model is explained in the Supplementary Material. Of the five areas forming the network, areas 1–4 are divided into two feature channels labeled by F , and in areas 3 and 4 each channel is further divided into 4 sub-channels associated with the 4 employed receptive fields labeled by j . Area 5 consists of two cells, indicated by i , for classification (see Figure 1, middle).

The shapes used as stimuli are polygons made of straight frame edges at the top, bottom and along one side, and a “profile” line—possibly but not necessarily curved—on the other side (Baylis and Driver, 2001). When that profile runs between mid-points of opposed frame sides, the total length of the present borders is the same for the original and for the three transformations (see Figure 2).

A combination of local excitation and global inhibition on area 2 is meant to cause the rebound spiking effects described in Supèr et al. (2010). In area 1 the images are accurately represented, as the two-channel input is mapped onto this layer. Only the neurons at the locations of white regions are firing spikes, while those on black regions are quiescent.

Neurons in area 2 receive spiking input from area 1. Each cell gets retinotopic excitatory input and global inhibitory input. For the channel receiving the region of smallest area, the spatial



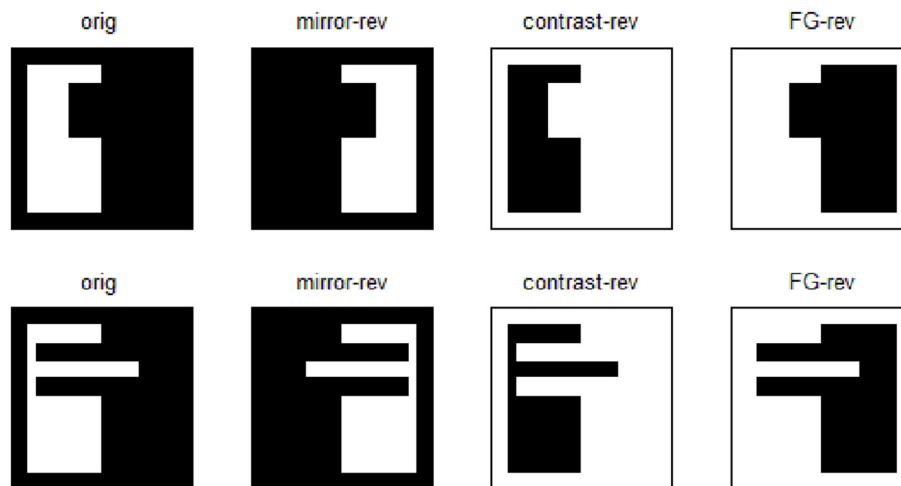


FIGURE 2 | Chosen images and their mirror-reversals, contrast-reversals, and figure-ground reversals. Note that within each row, the total length of the existing borders for every image is the

same. The two originals have inner size $n = 64$ without margins, outer size $N = 76$ including margins, and an equal area ratio of 0.42 without frame, 0.30 including frame.

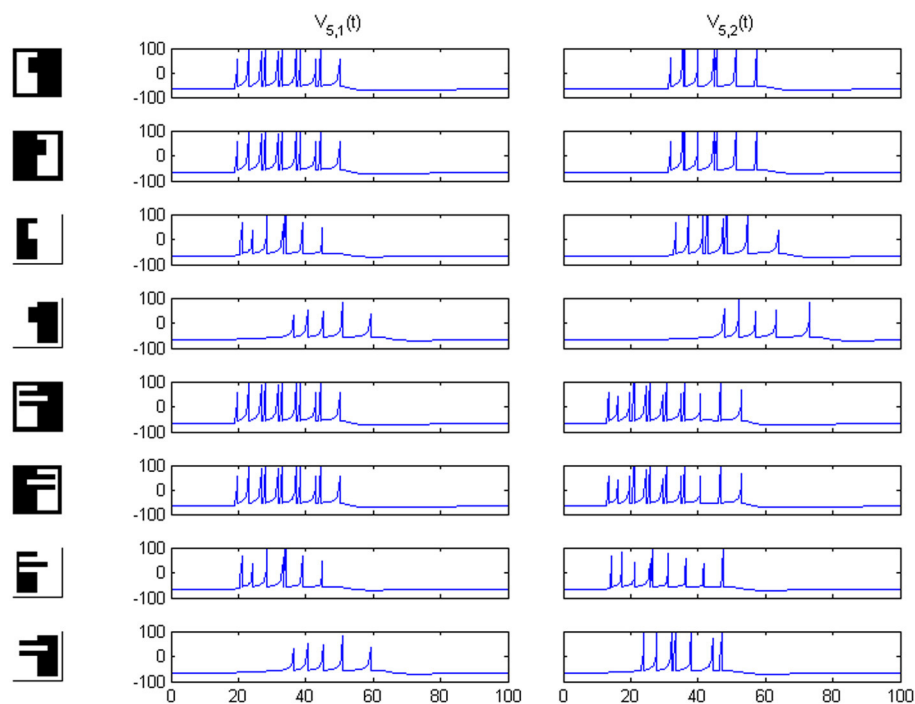


FIGURE 3 | Network responses on area 5 for the image sets of Figure 2, employing the $w_{s,i}$ weights quoted in the text. Times are given in ms and potentials in mV. For figure-ground reversal the responses are suppressed

while, for the other three cases, the firing order of cells 1 and 2 on area 5 signals the pertinence to one of two possible object categories (second and third columns).

pattern of spiking activity reproduces the excitatory input pattern. On the contrary, for the channel receiving the region of largest area, the spatial activity pattern is the reversal of the input pattern, signaling the complementary region. That change is explained by rebound spiking after a strong inhibition in the smallest region. For neurons on the largest region, global inhibition is partly compensated by retinotopic excitation. However, for cells on the smallest region, that inhibition is the only input and gives rise to a

strong a rapid hyperpolarization which provokes rebound spiking of these cells.

The new parts are added “on top” of the previous structure. In area 3, features are extracted by applying a non-linear function—in fact, a step function with given threshold—to convolutions of spike maps and filters (see **Figure 1**, bottom). The signals produced by application of the different filter types are fed into separate sub-channels. Area 4 collects spatial integrations of

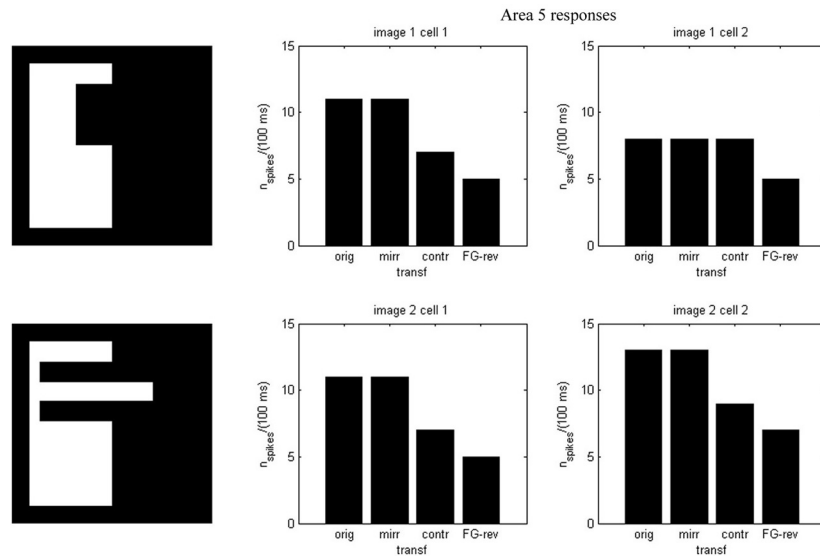


FIGURE 4 | Spike counts for the example of Figure 2. Each plot corresponds to an image set and an area 5 cell. In every case there are fewer spikes for FG-reversal.

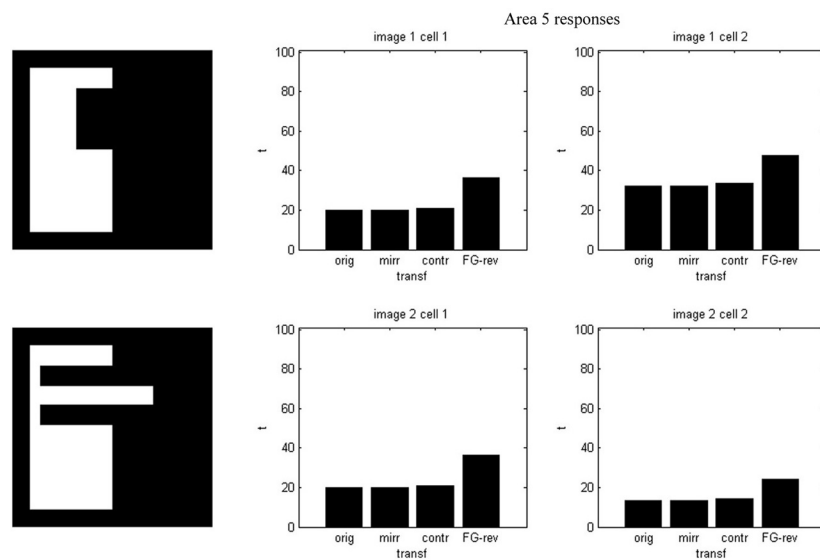


FIGURE 5 | Firing onset times—i.e., first spike times—for the example of Figure 2. Each plot is associated with an image set and an area 5 cell. In every set the spiking starts later when FG-reversal is applied.

the obtained detections within each sub-channel. Finally, area 5, which contains several output units, receives combinations of area 4 signals, including, in principle, all channels and sub-channels. Hypothetically there are as many output units as categories for classification (in our particular example, 2).

The numerical values of our inputs are set by the following rules:

$$I_{1F} = w_1 T_F, \quad F = 1, 2$$

$$I_{2F} = w_{2e} S_{1F} - |w_{2i}| \overline{S_{1F}} \mathbf{1}, \quad \overline{S_{1F}} \equiv \frac{1}{N^2} \sum_{k,l} (S_{1F})_{kl}, \quad F = 1, 2$$

$$I_{3Fj} = w_3 \Theta(S_{2F} * \mathbf{f}_j - 1), \quad F = 1, 2, \quad 1 \leq j \leq 4$$

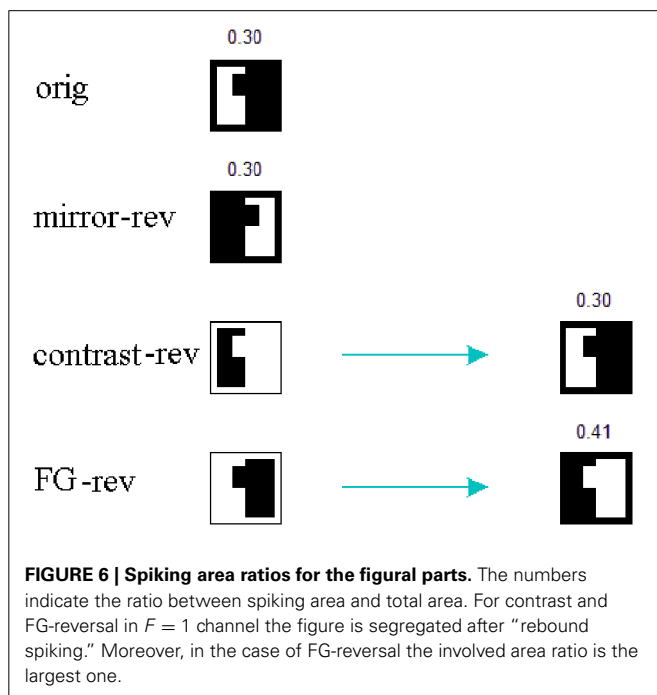
$$I_{4Fj} = w_4 \overline{S_{3Fj}}, \quad \overline{S_{3Fj}} \equiv \frac{1}{N^2} \sum_{k,l} (S_{3Fj})_{kl}, \quad F = 1, 2, \quad 1 \leq j \leq 4$$

$$I_{5i} = \sum_{F=1}^2 \sum_{j=1}^4 w_{5iFj} S_{4Fj}, \quad i = 1, 2.$$

T_F , $F = 1, 2$, stand for original stimulus ($F = 1$) and its contrast-reversed version ($F = 2$). Since the inhibitory weight w_{2i} is negative, we have written it as $w_{2i} = -|w_{2i}|$. Concerning the

inputs themselves, \mathbf{I}_{1F} , \mathbf{I}_{2F} , $F = 1, 2$ and \mathbf{I}_{3Fj} , $F = 1, 2$, $1 \leq j \leq 4$, are $N \times N$ matrices; \mathbf{I}_{4Fj} , $F = 1, 2$, $1 \leq j \leq 4$, and \mathbf{I}_{5i} , $i = 1, 2$, are scalars. An analogous convention is employed to indicate the binary (0,1) spike maps: \mathbf{S}_{1F} denotes the spike map produced by the potentials on area 1 channel F , and so on. Thus, \mathbf{S}_{1F} , \mathbf{S}_{2F} , $F = 1, 2$, and \mathbf{S}_{3Fj} , $F = 1, 2$, $1 \leq j \leq 4$, are $N \times N$ matrices, while \mathbf{S}_{4Fj} , $F = 1, 2$, $1 \leq j \leq 4$, are scalars. For $I = 1, 2$, every w_{5i} can be regarded as a matrix of two rows, labeled by F , and four columns, labeled by j . The $\mathbf{1}$ symbol indicates an $N \times N$ matrix whose coefficients are all them equal to one. Array convolution product is denoted by the “*” symbol, and Θ indicates the step function $\Theta(x) = 1$ if $x = 0$ and 0 otherwise. The feature-selective \mathbf{f}_j filters are given by:

$$\mathbf{f}_1 = \begin{pmatrix} -1 \\ 1 \end{pmatrix} \quad \mathbf{f}_2 = \begin{pmatrix} -1 & 1 \end{pmatrix} \quad \mathbf{f}_3 = \begin{pmatrix} 1 \\ -1 \end{pmatrix} \quad \mathbf{f}_4 = \begin{pmatrix} 1 & -1 \end{pmatrix}$$



In the studied set-up we adopt $w_1 = 10$, $w_{2e} = 400$, $w_{2i} = -750$, $w_3 = 500$, $w_4 = 5.0$, all of them in μA . The considered images (Figure 2) are squares of side $n = 64$ pixels when margins are not included. As margins are 6 pixels wide, $N = 76$ pixels. The number of white pixels is the same in the two original images, and they yield an area ratio of 0.42 without frame, or 0.30 including frame.

The ability to classify will depend on the particular form of the w_5 matrices. On area 5, cell $i = 1|2$ has to show preference for image 1|2. The question can be addressed by considering the role of the j indices, initially labeling the applied filters. For cell 1, limitation to vertical contrast takes place by setting non-zero values in even columns only. Analogously, horizontal contrast for cell 2 is obtained by adopting non-zero values just in the odd columns. Figure 7 illustrates that the strongest signal from FG-reversal goes through $F = 2$, related to the second row of w_{5i} . Because this signal should yield the weakest output, the remaining non-zero coefficients in the second rows have to be smaller than those in the first rows. A solution meeting this requirement in terms of only two non-zero constants A , B is

$$w_{51} = \begin{pmatrix} 0 & A & 0 & A \\ 0 & B & 0 & B \end{pmatrix}, \quad w_{52} = \begin{pmatrix} A & 0 & A & 0 \\ B & 0 & B & 0 \end{pmatrix}$$

with B smaller than A . In practice, satisfactory performance is obtained for $A = 100 \mu\text{A}$, $B = 5 \mu\text{A}$.

In agreement with Baylis and Driver's results (Baylis and Driver, 2001) and our previous proposals, FG discrimination is achieved already in area 2, long before shape recognition, and rests on one-sided edge assignment to figures. The shape-selective responses of area 5, identified as IT, depend mainly on the w_{5i} matrices, which—hypothetically—would consist of a group of learned weights. Shape-coding is evidenced by the difference in spiking onsets for the output units. Cells in V4 code diagnostic boundary features at specific locations, already ascribed to the object figure, which represent through their population response the complete shape. This matches with the findings by Patsupathy and Connor (2002).

RESULTS

The described model processes sets of figures consisting of original, mirror-reversed, contrast-reversed, and FG-reversed versions

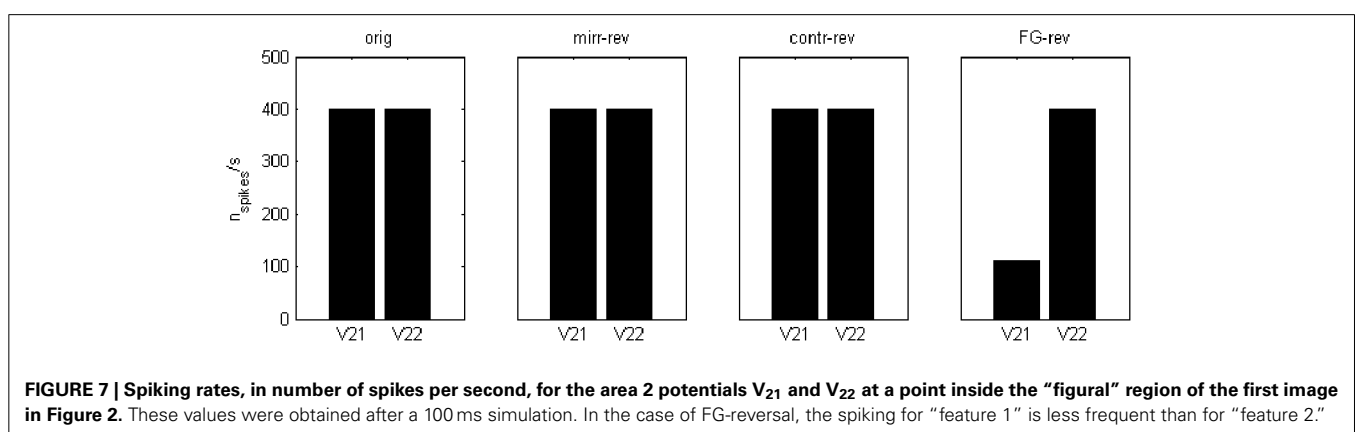


FIGURE 7 | Spiking rates, in number of spikes per second, for the area 2 potentials V_{21} and V_{22} at a point inside the “figural” region of the first image in Figure 2. These values were obtained after a 100 ms simulation. In the case of FG-reversal, the spiking for “feature 1” is less frequent than for “feature 2.”

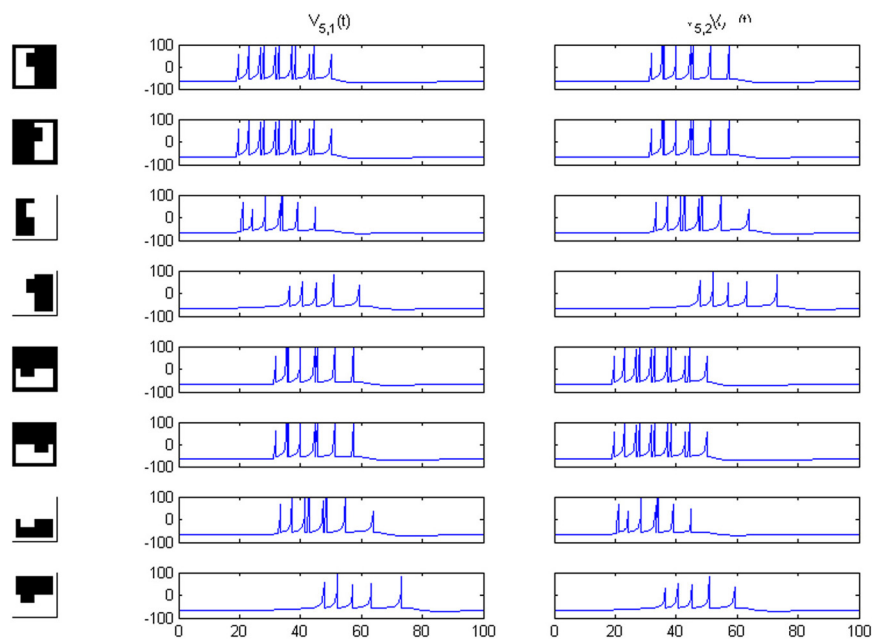


FIGURE 8 | Potentials on area 5 for the first image set of Figure 2 and its own rotated version. Cell 1 and cell 2 responses are interchanged.

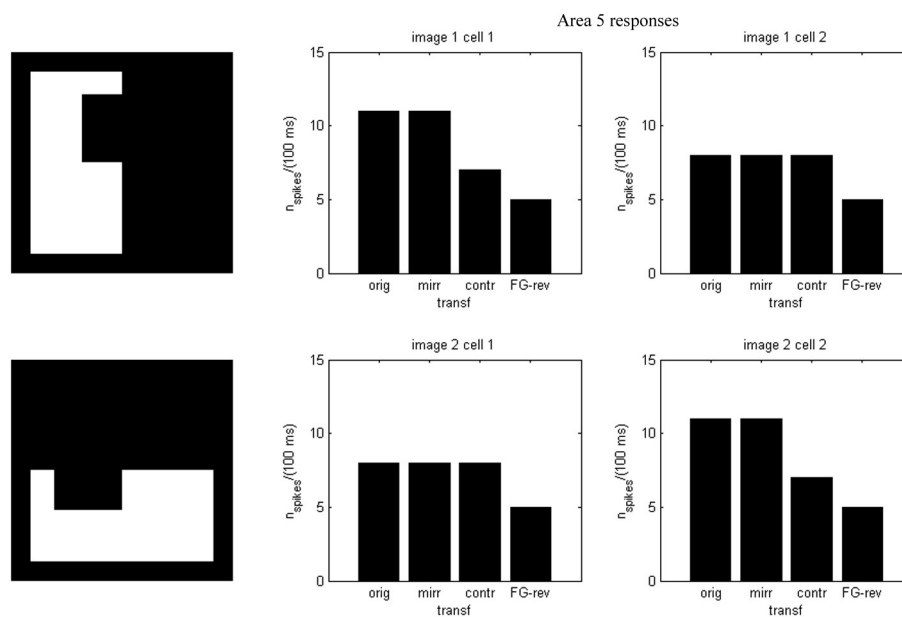


FIGURE 9 | Spike counts for the images of Figure 8. Cell 1 and cell 2 counts are interchanged.

of the original one. Depending on the lengths of horizontal and vertical borders, the different activity of the output units classifies the elements of these sets. In addition, responses are similar for original, mirror-reversed and contrast-reversed transformations of the same image, and significantly decrease for the FG-reverse version.

Results of running the network with our particular matrices are shown in **Figure 3**. On area 5, cell 1 spikes earlier than

cell 2 for image 1 and cell 2 spikes sooner than cell 1 for image 2. Since the non-zero columns of matrices $w_{51}|w_{52}$ correspond to vertical|horizontal contrast features, the employed solution is valid for any case in which the predominance of vertical|horizontal borders can be a distinctive criterion. Moreover, within each image set, responses to FG-reversed images are the lowest because row 2 (which weights the inputs from “ $F = 2$ ” channel) has smaller coefficients than row 1 (which multiplies

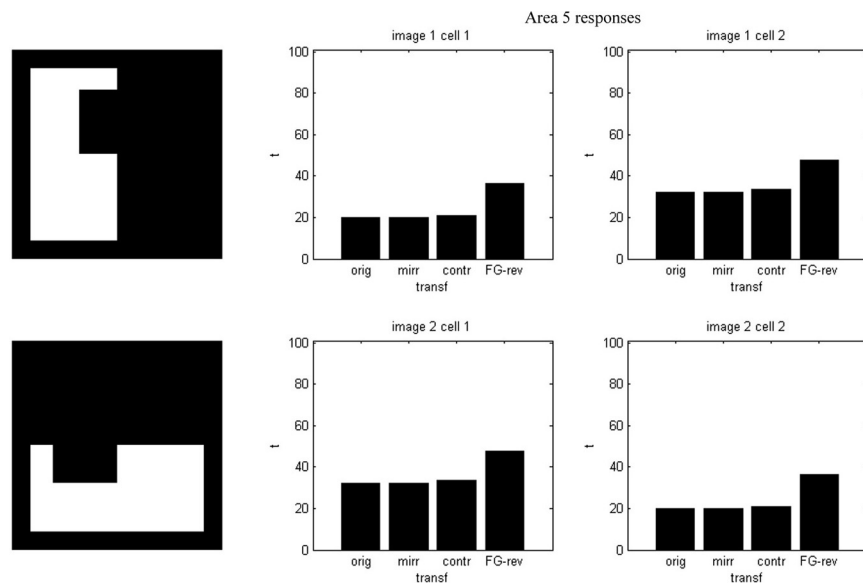


FIGURE 10 | Firing onset times for the images of Figure 8. Cell 1 and cell 2 times are interchanged.

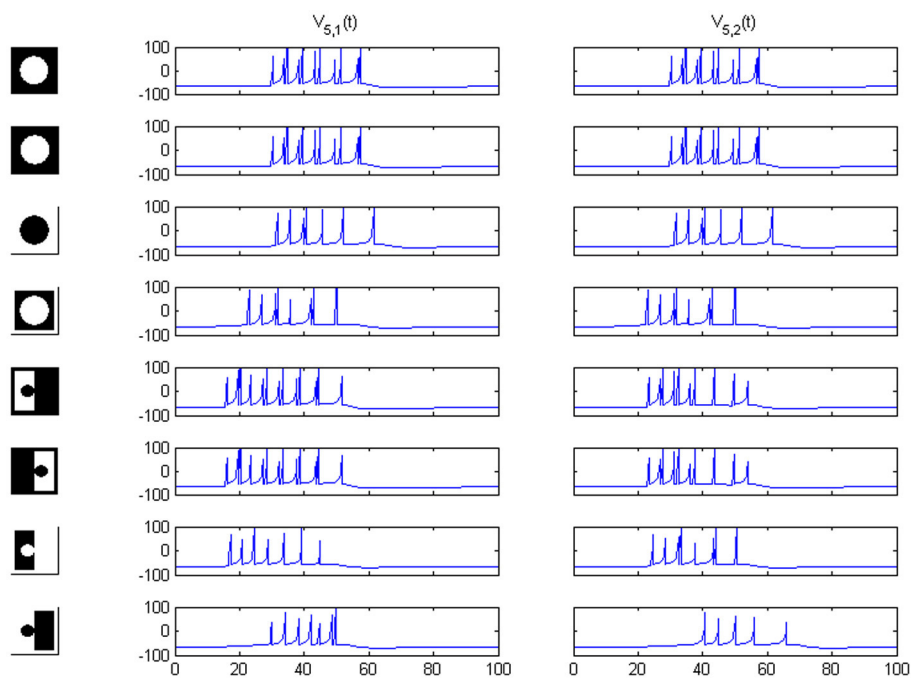


FIGURE 11 | Rows 1–4: network responses on area 5 for a circle disconnected from a hypothetical frame. Rows 5–8: responses for a circle connected to the frame by the mid-points of opposed sides (preserving the border length, as required).

the “ $F = 1$ ” channel signals). Indeed the spike counts shown in **Figure 4** indicate that there are fewer spikes for the FG-reversal of every image. Furthermore, the produced spike bursts start later when applying FG-reversal, as can be seen in **Figure 5**. On the whole, firing onset times are a better criterion than spike counts.

The applied mechanism may be understood in terms of spiking area ratios for figural parts because, in the end, the number of spikes relative to the total area has a decisive contribution to the excitation-inhibition balance. For the case of contrast and FG-reversal in $F = 1$ channel, the figural part is not segregated until “rebound spiking” takes place on area 2 (rebound spiking occurs

after a strong inhibition, even in the absence of excitation—see Izhikevich, 2003, 2007 or Supèr et al., 2010). For FG-reversal the involved area is the largest (see **Figure 6**) and the resulting inhibition, which is proportional to the spiking area, turns out to be somewhat stronger (**Figure 7**).

Because our criterion rests on differences in length between vertical and horizontal borders, the system distinguishes an image from its own rotated version, as can be seen in **Figures 8–10**. Predictably, for area 4, responses in sub-channels with even and odd indices are interchanged, and for area 5, the 1 and 2 cell responses are swapped as well.

In the considered image realm profiles should run between mid-points in opposite frame sides (see lower part of **Figure 1** in Baylis and Driver, 2001) in order to preserve the total length of all the boundaries. Going out of this image class we can imagine the case of a disconnected circle. Then, the weakest signal is the “contrast reversed” one, while the “FG-reversed” version produces a higher response (see **Figure 11**, upper part) caused by the existence of a longer boundary. For this example the third transformation must be simply ignored, because it just amounts to the reversal of an unconnected frame, while the only reasonable analog to FG-reversal is now the contrast reversal itself. Examination of the numerical output reveals that it starts spiking marginally later than the original and mirror-reversal (by 1.25 ms) and with fewer spikes (7 instead of 11). Thus, the result is not inconsistent. When the circular shape is connected to the frame and the overall area ratio correctly set, normal working is restored (**Figure 11**, lower part).

DISCUSSION

We have been able to design a network structure which models the suppression of responses to FG reversed stimuli, and shows the possibility of producing selective outputs that generalize across mirror reversed and contrast reversed stimuli. Although the model was not meant for complex images and had no pretence to describe state-of-the-art knowledge on IT processing, it is quite coherent as its outcome fits our previous findings, was constructed using similar values to our forerunning model (Supèr et al., 2010; Supèr and Romeo, 2011) and yields invariance in the pattern of responses across a variety of stimuli and their transformations.

An essential ingredient was the dual pathway for the given figure and its own contrast-reversed version, which represents the existence of two input preferences (Supèr et al., 2010). Although the incoming signals for these two channels are different, the spiking parts in area 2 eventually highlight a single region, identified as “figure.” Despite the space coincidence, the strengths of these signals may still vary, showing a sizable difference for the FG-reversal case. Later, the obtained figural part undergoes a multiple feature extraction process. Spatially-averaged results of that feature detection procedure are then fed into cells mimicking IT neurons. By virtue of the devised scheme, which benefits from the linear character of the I_{5i} inputs, our IT cells are in fact selective for two image categories. The nature of the performed selection is determined by the weight choice.

A correspondence between model architecture and visual system can be depicted as follows: The first area transforms the input

into a spiking train like the Ganglion cell area of the retina, the second area then would be V1, assuming that the LGN (lateral geniculate nucleus) merely relays sensory information. Areas 3–4 may be assimilated to connections occurring both in V2 and in V4, while area 5 would be analogous to IT.

The remarked dependency on orientation can be viewed as the consequence of “experience” (contained in the values of the w_{5i} weights) that causes the system to perform holistic processing. In the case of the rotated image, the features or components are processed in the same way as in the original (by V4 neurons). If there were edge detectors for enough different orientations and all their outputs could be integrated in a rotationally-invariant fashion, responses for an image and its own rotated version ought to be equal. In our case the limited “experience” implicit in the weights does not suffice for obtaining this symmetry. An implication is that in our model both sorts of information are explicitly encoded as suggested by Schwaneinger et al. (2002).

Another consequence would be that our memory of a category has a specific orientation, the usual one in the type of stimulus processed. A well-known example of this affirmation is the Thatcher illusion, where the eyes and mouth of a face are turned upside down (see Thompson, 1980). When the whole image is subsequently inverted the grotesque appearance vanishes. In the context of our model implications, the component representations would then be normal and thus could be matched with the output of the holistic process.

At least for polygons of the studied type, our model bears out the view offered by Baylis and Driver (2001) and provides a computational scheme explaining their observations. FG discrimination is achieved in an area which becomes active before shape selection takes place, and is based on one-sided edge assignments. Such a mechanism, which accounts for the observed generalization, operates by a purely feed-forward process.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://www.frontiersin.org/journal/10.3389/fpsyg.2014.00481/abstract>

REFERENCES

- Baylis, G. C., and Driver, J. (1995a). One-sided edge-assignment in vision: 1 figure-ground segmentation and attention to objects. *Curr. Dir. Psychol. Sci.* 4, 140–146. doi: 10.1111/1467-8721.ep10772580
- Baylis, G. C., and Driver, J. (1995b). One-sided edge-assignment in vision: 2 part decomposition, shape description, and attention to objects. *Curr. Dir. Psychol. Sci.* 4, 201–206. doi: 10.1111/1467-8721.ep10772580
- Baylis, G. C., and Driver, J. (1995c). Obligatory edge-assignment in vision: the role of figure and segmentation in short-term visual matching. *J. Exp. Psychol. Hum. Percept. Perform.* 6, 1323–1342. doi: 10.1037/0096-1523.21.6.1323
- Baylis, G. C., and Driver, J. (2001). Shape-coding in IT cells generalizes over contrast and mirror reversal, but not figure-ground reversal. *Nat. Neurosci.* 4, 937–942. doi: 10.1038/nn0901-937
- Desimone, R. (1991). Face-selective cells in the temporal cortex of monkeys. *J. Cogn. Neurosci.* 3, 1–8. doi: 10.1162/jocn.1991.3.1.1
- Desimone, R., Schein, S. J., Moran, J., and Ungerleider, L. G. (1985). Contour, color and shape analysis beyond the striate cortex. *Vision Res.* 24, 441–452. doi: 10.1016/0042-6989(85)90069-0
- DiCarlo, J. J., and Maunsell, J. H. R. (2000). Form representation in monkey inferotemporal cortex is virtually unaltered by free viewing. *Nat. Neurosci.* 3, 814–821. doi: 10.1038/77722

- Farah, M. J. (1990). *Visual Agnosia*. Cambridge, MA: MIT Press.
- Hoffman, D. D., and Richards, W. A. (1984). Parts of recognition. *Cognition* 18, 65–96. doi: 10.1016/0010-0277(84)90022-2
- Hulleman, J., Gedamke, L., and Humphreys, G. W. (2005). The time course of figure-ground reversal. *Percept. Psychophys.* 67, 595–605. doi: 10.3758/BF03193517
- Izhikevich, E. M. (2003). Simple model of spiking neurons. *IEEE Trans. Neural Netw.* 14, 1569–1572. doi: 10.1109/TNN.2003.820440
- Izhikevich, E. M. (2007). *Dynamical Systems in Neuroscience: The Geometry of Excitability and Bursting*. Cambridge, MA: MIT Press.
- Logothetis, N. K., Pauls, J., and Poggio, T. (1995). Shape recognition in the inferior temporal cortex of monkeys. *Curr. Biol.* 5, 552–563. doi: 10.1016/S0960-9822(95)00108-4
- Malach, R., Reppas, J. B., Benson, R. R., Kwong, K. K., Jiang, H., Kennedy, W. A., et al. (1995). Object-related activity by functional magnetic imaging in human occipital cortex. *Proc. Natl. Acad. Sci. U.S.A.* 92, 8135–8139. doi: 10.1073/pnas.92.18.8135
- Nakayama, K., He, Z. J., and Shimojo, S. (1995). “Visual surface representation: a critical link between lower-level and higher-level vision,” in *Visual Cognition*, eds S. M. Kosslyn and D. N. Osherson (Cambridge, MA: MIT Press), 1–70.
- Patsupathy, A., and Connor, C. E. (2002). Population coding of shape in area V4. *Nat. Neurosci.* 5, 1332–1338. doi: 10.1038/nn972
- Peterson, M. A., Harvey, E. M., and Weidenbacher, H. J. (1991). Shape recognition contributions to figure-ground reversal: which route counts? *J. Exp. Psychol.* 17, 1075–1089.
- Riesenhuber, M., and Poggio, T. (2000). Models of object recognition. *Nat. Neurosci.* 3, 1199–1204. doi: 10.1038/81479
- Rollenhagen, J. E., and Olson, C. R. (2000). Mirror-image confusion in single neurons of the macaque inferotemporal cortex. *Science* 287, 1506–1508. doi: 10.1126/science.287.5457.1506
- Rolls, E. T., Judge, S. J., and Sanghera, M. K. (1977). Activity of neurons in the inferotemporal cortex of the alert monkey. *Brain Res.* 130, 229–238. doi: 10.1016/0006-8993(77)90272-4
- Rubin, E. (1958). “Figure and ground,” in *Readings in Perception*, eds D. Bredsdlee and M. Wertheimer (Princeton, NJ: Van Nostrand), 35–101.
- Rubin, N. (2001). Figure and ground in the brain. *Nat. Neurosci.* 4, 857–858. doi: 10.1038/nn0901-857
- Schwaninger, A., Lobmaier, J. S., and Collishaw, S. M. (2002). Role of featural and configural information in familiar and unfamiliar face recognition. *Lect. Notes Comput. Sci.* 2525, 643–650. doi: 10.1007/3-540-36181-2_64
- Supèr, H., and Romeo, A. (2011). Feedback enhances feedforward figure-ground segmentation by changing firing mode. *PLoS ONE* 6:e21641. doi: 10.1371/journal.pone.0021641
- Supèr, H., Romeo, A., and Keil, M. (2010). Feed-forward segmentation of figure-ground and assignment of border-ownership. *PLoS ONE* 5:e10705. doi: 10.1371/journal.pone.0010705
- Tanaka, K. (1996). Inferotemporal cortex and object vision. *Annu. Rev. Neurosci.* 19, 109–139. doi: 10.1146/annurev.ne.19.030196.000545
- Thompson, P. (1980). Margaret Thatcher: a new illusion. *Perception* 9, 483–484. doi: 10.1068/p090483

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.

Received: 20 February 2014; accepted: 02 May 2014; published online: 27 May 2014.
 Citation: Romeo A and Supèr H (2014) A feed-forward spiking model of shape-coding by IT cells. *Front. Psychol.* 5:481. doi: 10.3389/fpsyg.2014.00481
 This article was submitted to Perception Science, a section of the journal *Frontiers in Psychology*.
 Copyright © 2014 Romeo and Supèr. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Early recurrent feedback facilitates visual object recognition under challenging conditions

Dean Wyatte¹, David J. Jilk^{2*} and Randall C. O'Reilly¹

¹ Department of Psychology and Neuroscience, University of Colorado Boulder, Boulder, CO, USA

² eCortex, Inc., Boulder, CO, USA

Edited by:

Hulusi Kafaligonul, Bilkent University, Turkey

Reviewed by:

Arjen Alink, Medical Research Council, UK

Radoslaw Martin Cichy, Massachusetts Institute of Technology, USA

*Correspondence:

David J. Jilk, eCortex, Inc., c/o Reliascent, 5710 Flatiron Parkway, Ste. B, Boulder, CO 80301, USA
e-mail: dave@jilk.com

Standard models of the visual object recognition pathway hold that a largely feedforward process from the retina through inferotemporal cortex leads to object identification. A subsequent feedback process originating in frontoparietal areas through reciprocal connections to striate cortex provides attentional support to salient or behaviorally-relevant features. Here, we review mounting evidence that feedback signals also originate within extrastriate regions and begin during the initial feedforward process. This feedback process is temporally dissociable from attention and provides important functions such as grouping, associational reinforcement, and filling-in of features. Local feedback signals operating concurrently with feedforward processing are important for object identification in noisy real-world situations, particularly when objects are partially occluded, unclear, or otherwise ambiguous. Altogether, the dissociation of early and late feedback processes presented here expands on current models of object identification, and suggests a dual role for descending feedback projections.

Keywords: object recognition, feedback, top-down attention, illusory contours, amodal completion

INTRODUCTION

Visual object recognition has traditionally been described as a largely feedforward process that operates independently of and prior to top-down signals that reflect strategic processing or attentional effects. This standard model of object recognition is supported by research that spans multiple levels of analysis, including single- and multi-unit recording, computational modeling, and behavioral experiments, all of which have been discussed in detail in recent reviews (e.g., Serre et al., 2007; DiCarlo et al., 2012). Feedback projections, nearly equal in density to feedforward neurons throughout the ventral visual stream (Felleman and Van Essen, 1991; Sporns and Zwi, 2004), are commonly thought to subserve slower, attention-mediated processing that happens after recognition processes are complete, but not the core object recognition processing itself (Hochstein and Ahissar, 2002).

The proposal advanced in this paper is that these local, recurrent feedback connections also provide an avenue for rapid top-down signals that influence object recognition-related processing *as it is being carried out*—well before the slower attention-mediated processes. The theory is inspired by the pioneering work of Dehaene et al. (2006) and Lamme (2003, 2006) in identifying the neural correlates of consciousness. Both of these researchers' theories dissociate between local recurrent processing and top-down signals from frontoparietal areas in terms of the effects that they have on awareness. The present work draws a similar distinction between top-down, attention-mediated processing, and local recurrent processing between hierarchically adjacent areas within the ventral stream.

We support this distinction by first providing evidence that there are two temporally dissociable processes operating on these feedback projections; and second by presenting results showing an important functional role for the earlier, local recurrent processing.

EVIDENCE FOR A TEMPORAL DISSOCIATION OF LOCAL RECURRENT AND TOP-DOWN PROCESSING

Top-down attention is known to be a consciously generated, executive signal originating in frontal and parietal areas (Thompson et al., 2005; Bressler et al., 2008). Signals reflecting these strategic processes do not manifest in early visual areas until 150–170 ms after stimulus onset at the earliest, with most reported effects occurring within the range of 200–300 ms (Mehta et al., 2000a,b; Martinez et al., 2001; Noesselt et al., 2002). The relatively long latency of attentional effects in early visual areas is thought to arise from top-down signals that target late stages of the ventral stream and then progress backward toward V1 (Buffalo et al., 2010). Local recurrent processing can also be thought of as a top-down process, except that the signal originates from within the ventral stream itself, as opposed to frontal or parietal areas. Local recurrent processing is completely involuntary and does not require conscious execution, evidenced by its observation in recordings from anesthetized animals (Roland et al., 2006; Roland, 2010) and is simply a consequence of signal propagation through recurrent corticocortical connectivity. Specifically, as soon as a given area responds, signals are routed both to higher-level and lower-level connected areas. Feedback to immediately lower levels occurs with very short latencies—as quickly as 10 ms after the initial feedforward responses (Hupé et al.,

2001; Pascual-Leone and Walsh, 2001)—and thus could plausibly be underway after initial feedforward IT neural responses (ca. 80–100 ms) but before the completion of the categorization process (ca. 150 ms).

Recent research using methods that temporarily interfere with cortical processing have revealed strong evidence that recurrent feedback circuits are engaged during the first 80–150 ms of visual processing. One line of evidence comes from experiments that use transcranial magnetic stimulation (TMS) to temporarily prevent a targeted brain area from responding. In a recent study, Koivisto et al. (2011) used fMRI-localized TMS to selectively inactivate V1/V2 while subjects categorized images according to whether they contained an animal. The authors found that applying TMS to V1/V2 with stimulus onset asynchronies (SOAs) of 90–210 ms impaired categorization performance and subjective perception of stimuli. Camprodon et al. (2010) observed similar results with TMS applied over V1 only, but found that there were actually two windows of impairment with SOAs of 100 and 220 ms. Earlier work from Corthout et al. also found an early window of activity with an SOA of around 100 ms during which applying TMS over V1 impairs letter recognition (Corthout et al., 1999a,b). Collectively, these experiments show that disruption of processing in early visual areas around 100 ms after stimulus presentation impairs visual recognition. Importantly, this time window occurs after the earliest contributions of IT neurons, opening up the possibility that the impairment is due to the disruption of feedback to lower-level areas in influencing the quality of object representations. Furthermore, several of these studies found a second, later time window around 200 ms during which TMS also impaired recognition. This later time window coincides with the latency of spatial attention-mediated processing (Mehta et al., 2000a,b; Martinez et al., 2001; Noesselt et al., 2002; Buffalo et al., 2010), providing a temporal dissociation from the rapid recurrent processing effects that are of interest here.

Visual backward masking experiments have also identified a similar time window for recurrent processing around 100 ms after stimulus onset (Fahrenfort et al., 2007, 2008; Boehler et al., 2008). In backward masking experiments, a first stimulus (the target) is followed by a second stimulus (the mask) at a particular latency. At very short latencies, backward masking can impair recognition of the target and in some cases, prevent it from reaching awareness (Macknik and Livingstone, 1998). While the effect of backward masking was initially accounted for with a feedforward explanation (Breitmeyer and Ganz, 1976), modern theories of backward masking emphasize recurrent processing between higher-level and lower-level areas (Enns and Di Lollo, 2000; Lamme and Roelfsema, 2000; Wyatte et al., 2012a). Specifically, if information about a target stimulus being processed in higher-level areas is fed back down to lower areas, but a masking stimulus is simultaneously being processed at that lower level, there will be a fundamental mismatch in the information being processed at each level (Lamme and Roelfsema, 2000). This mismatch causes a decoupling in the functional connectivity (i.e., co-activation) between the visual areas involved in processing the stimulus, which has the psychological effect of greatly reduced perceptual visibility (Dehaene et al., 2001; Haynes et al., 2005).

Boehler et al. (2008) combined a backward masking paradigm with magnetoencephalography (MEG) recording to determine the time course of recurrent feedback to V1 during a recognition task. On trials where subjects correctly recognized the target stimulus (i.e., no impairment from the mask), there was modulation of the V1 MEG signal from 100 to 120 ms. This modulation occurred soon after (ca. 27 ms) the initial V1 signals and almost immediately after (ca. 11 ms) extrastriate generated signals, in strong accordance with being driven by rapid recurrent feedback from extrastriate areas to V1. Again, these rapid recurrent processing effects were dissociable from slower attentional modulation, which manifested 250–300 ms after stimulus presentation and only when subjects attended to the same region of the display that the target appeared in. In contrast, modulation from rapid recurrent processing occurred regardless of where subjects directed attention. Similar results have been demonstrated when combining backward masking with electroencephalography (EEG) recording with both rapid recurrent and slower attentional modulation, but with less emphasis on the specific neural generators of effects given the relatively poor spatial resolution of EEG (Fahrenfort et al., 2007, 2008).

Together, TMS and backward masking experiments provide strong support for the idea that recurrent visual processing engages striate and extrastriate areas around 100 ms after stimulus onset during visual recognition tasks. This local rapid recurrent processing is dissociable from attention-mediated or strategic processing both in terms of where the signals originate (within the ventral stream vs. frontal and parietal areas) and in terms of their relative time courses (ca. 100 ms vs. 150–170 ms at the earliest). Attention has long been known to modulate early EEG responses such as the P1 (first positive deflection, ca. 100 ms) and N1 (first negative deflection, ca. 150–200 ms) (Luck et al., 1990a; Hillyard and Anllo-Vento, 1998). Given the data reviewed here, it seems plausible that the P1 indexes recurrent feedback generated within the ventral stream while the N1 reflects the first influences of frontal and parietal attentional signals that progress backwards through visual areas toward V1 (Buffalo et al., 2010; see also Luck et al., 1990b).

While TMS impairment around 100 ms is consistent with the disruption of recurrent processing, it cannot rule out the possibility that the TMS is actually disrupting delayed feedforward responses. Specifically, low-level image properties such as local contrast can affect the temporal order of feedforward spikes, with lower contrast image regions exhibiting delay relative to more salient image regions (VanRullen and Thorpe, 2001, 2002). However, the information content of these regions is much lower than the salient regions that exhibit fast responses and thus it is unlikely that disrupting their contribution to object recognition processing will negatively impact recognition ability for relatively unambiguous images. More importantly, backward masking experiments that target impairment of recurrent processing provide additional constraints in interpreting TMS effects. Finally, 90–110 ms post stimulus onset is hypothesized to be the time at which peak feedback signals arrive at V1 from extrastriate areas (Roland et al., 2006; Roland, 2010).

Overall, it seems clear that recurrent processing operates within the established time course of object recognition, which

spans the first 150 ms of visual processing. The data reviewed in this section are summarized in **Table 1** with a rough sketch of overall feedforward and feedback events shown in **Figure 1**. Having established support for the idea that recurrent feedback occurs rapidly beginning around 100 after stimulus presentation, this paper now turns to discussion of its function.

EVIDENCE FOR A DISTINCT FUNCTIONAL ROLE FOR LOCAL RECURRENT SIGNALS

There is considerable evidence that local recurrent processing is important when stimuli are degraded, partial, or otherwise ambiguous, and we hypothesize that this is one important functional role for the dissociated process described in the previous section. The basic logic behind this proposal is that degrading a stimulus has been shown to weaken the initial responses

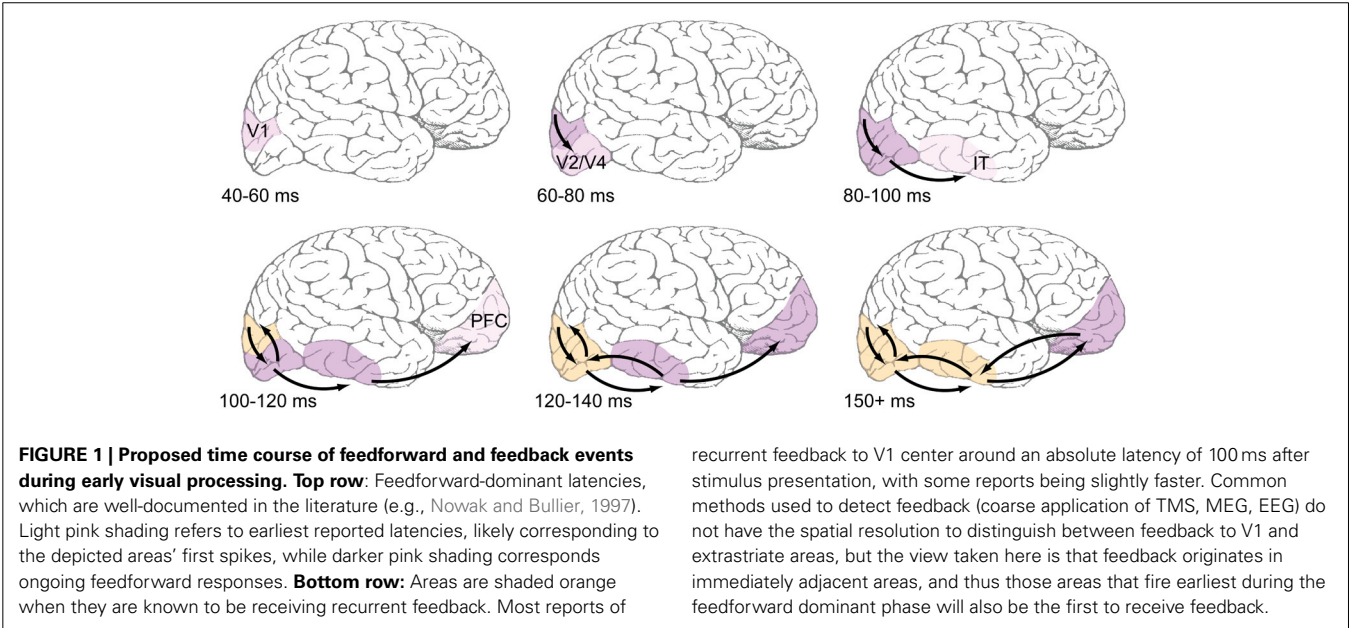
in object-selective areas (Sclar et al., 1990; Kovacs et al., 1995; Nielsen et al., 2006; Williford and Maunsell, 2006), but recurrent processing over time can strengthen responses back to near undegraded levels and preserve selectivity via top-down reinforcement. Consistent with this idea, object-selective responses in IT cortex remain intact when stimuli are occluded, but take significantly more time to manifest than when stimuli are unoccluded (around an extra 50 ms on average, Kovacs et al., 1995; Nielsen et al., 2006).

Single-unit recordings that use reversible cooling to temporarily inactivate a particular brain area provide further support for our hypothesis. Hupé et al. (1998) applied cooling to area V5/MT, a visual area in the superior temporal sulcus of the monkey brain that sends feedback projections to areas V1, V2, and V3. Recordings from V1 through V3 indicated that

Table 1 | Summary of data that suggest a temporal dissociation of local recurrent and top-down attentional processing.

Reference	Milliseconds after Stimulus Onset																														Comments
	10	20	30	40	50	60	70	80	90	100	110	120	130	140	150	160	170	180	190	200	210	220	230	240	250	260	270	280	290	300	
Koivisto et al., 2011																															TMS over human V1. Scene categorization impairment at indicated times
Corthout et al., 1999a																															TMS over human V1. Letter recognition impairment at indicated times
Corthout et al., 1999b																															TMS over human V1. Letter recognition impairment at indicated times
Camprodon et al., 2010																															TMS over human V1. Scene categorization impairment at indicated times
Boehler et al., 2008																															MEG in human subjects. Effects of masking over V1 at indicated times
Fahrenfort et al., 2007																															EEG in human subjects. Effect of masking over occipital channels at indicated times
Fahrenfort et al., 2008																															EEG in human subjects. Effect of masking over occipital channels at indicated times
Martinez et al., 2001																															EEG/fMRI in human subjects. V1 modulation with attention at indicated times
Noesselt et al., 2002																															EEG/MEG/fMRI in human subjects. V1 modulation with attention at indicated times
Mehta et al., 2000a, 2000b																															LFP in awake behaving monkey. V1 modulation with attention at indicated times

Putative feedforwardProposed local recurrentPutative attentional



responses to moving bar stimuli were vastly weakened (fewer spikes observed per second) when V5/MT was inactive compared to control experiments in which it was active. This attenuation of lower-level responses was most dramatic in low salience conditions, such as when the bar had a very low contrast, a point that will be discussed in further detail later in this section. These results suggest that when higher-level visual areas are active, they provide additional excitatory input to lower levels. Similar effects have been shown for other recurrent circuits in other mammalian species such as those involving middle suprasylvian (MS) cortex and V1 (Galuske et al., 2002) as well as V2 and V1 (Sandell and Schiller, 1982; Mignard and Malpeli, 1991), suggesting that top-down amplification is a highly generic mechanism that occurs between any two recurrently connected areas.

Top-down amplification promotes visual awareness (Lamme, 2003, 2006; Dehaene et al., 2006), and some data indicate that amplification is a simple contrast gain operation, as some have suggested is implemented by attention (e.g., Reynolds et al., 2000; Reynolds and Heeger, 2009). However, there is mounting evidence that recurrent amplification also plays an important functional role in visual object recognition when stimuli are degraded or ambiguous, by promoting a complex grouping and “filling-in” process.

Wyatte et al. (2012a) degraded visual object stimuli using visual occlusion and contrast degradation and used backward masking to control whether recurrent processing mechanisms were available (Enns and Di Lollo, 2000; Lamme and Roelfsema, 2000). When relatively clear stimuli were masked using a relatively long latency 100 ms SOA pattern mask, there was little impairment in recognition performance. However, when heavily occluded or low contrast stimuli were masked, the mask had a much larger effect, suggesting that recurrent processing was crucial in resolving object identity in these conditions. Simulations using a computational model of object recognition that included recurrent feedback between hierarchically adjacent layers (O'Reilly et al., 2013) showed that responses in both lower layers (corresponding to striate/extrastriate regions) and upper layers (corresponding to IT cortex) strengthened over time when objects were occluded. Backward masking selectively interfered with this strengthening process, which was crucial when stimulus signals were weak due to degradation. Furthermore, the strengthening dynamic was found to be specifically due to recurrent feedback—purely feedforward versions of the model exhibited asymptotic response levels across areas.

One possibility for the mechanism underlying these recognition performance differences is a grouping and “filling-in” process similar to what is observed in the figure-ground literature in V1 (Figure 2A), but repeated between higher levels of the visual hierarchy. As an illustration, consider a population of IT neurons that respond to bicycle stimuli (Figures 2B,C). If a bicycle stimulus is occluded and only the wheels are visible, some members of this population will become active (specifically, those corresponding to wheel-like features), but the selective response across the full population will be unavailable. The partial responses, however, will be propagated back to earlier visual areas, which will drive neurons that are sensitive to visual features that are known to

co-occur with bicycle wheels, such as a bicycle's frame, handlebars, and saddle. Importantly these responses occur in the absence of these features in the actual stimulus. These “illusory” responses in turn provide new driving potential to IT neurons, ultimately evoking the selective response corresponding to the unoccluded stimulus across the full IT population responsive to bicycles. The IT response is “object complete,” meaning that there is little-to-no difference between the response to the partially occluded object and the complete object—the brain has filled in the missing information.

Computationally, recurrent processing's amplification effect is capable of supporting a grouping or surface-based encoding. The most convincing demonstrations of these computations are found in the figure-ground processing literature, where the term “contextual modulation” is used to describe them (Zipser et al., 1996; Lamme et al., 1998). In contextual modulation, neurons with non-overlapping receptive fields such as those found in V1 are capable of modulating and reinforcing each other by virtue of shared connections through higher levels in the visual hierarchy where receptive fields do tend to overlap. This extra modulation has the effect of grouping together figural elements of a display and enhancing their activity relative to background elements effectively spreading activation throughout the figure interior and “filling” it in as a perceptually salient surface (Figure 2A). The models suggest that contextual modulation is driven by recurrent feedback, because lesions of feedback from extrastriate and dorsal structures to V1 obliterate the surface filling effect. They further illustrate that the timing of contextual modulation to area V1 would be on the order of 80–100 ms after stimulus presentation, coinciding with the known time course of feedback to striate areas during visual processing. Finally, contextual modulation is dissociable from slower top-down attentional effects, not just with respect to time course but also because its surface filling computations are retained even when attention is deployed away from the target stimulus (Poort et al., 2012).

There are two phenomena in the experimental literature that support the grouping and filling-in roles of recurrent processing during object recognition. The first is the perception of illusory contours, such as in displays containing Kanizsa shapes (Figure 3). V1 neurons have been shown to respond to the illusory contours that compose Kanizsa shapes, such as the edges of the illusory square in Figure 3. Multi-unit recordings have indicated that these responses occur beginning around 100 ms after stimulus presentation, which is shortly after the V1 responses to a physical contour with the same orientation and location, suggesting a role for feedback in their encoding (Lee and Nguyen, 2001; Seghier and Vuilleumier, 2006). Specifically, recurrent feedback from extrastriate areas could support the perception of illusory contours in the Kanizsa illusion by grouping similarly oriented contours at the V1 level that fall within the shape's receptive field; this would cause the shape to be perceived as perceptually salient surface similar to the way texture-defined shapes are perceived (Figure 3B). As such, a recent experiment has indicated that global contour information emerges in V1 responses shortly after the first V4 responses, implicating recurrent feedback in this grouping process (Chen et al., 2014).

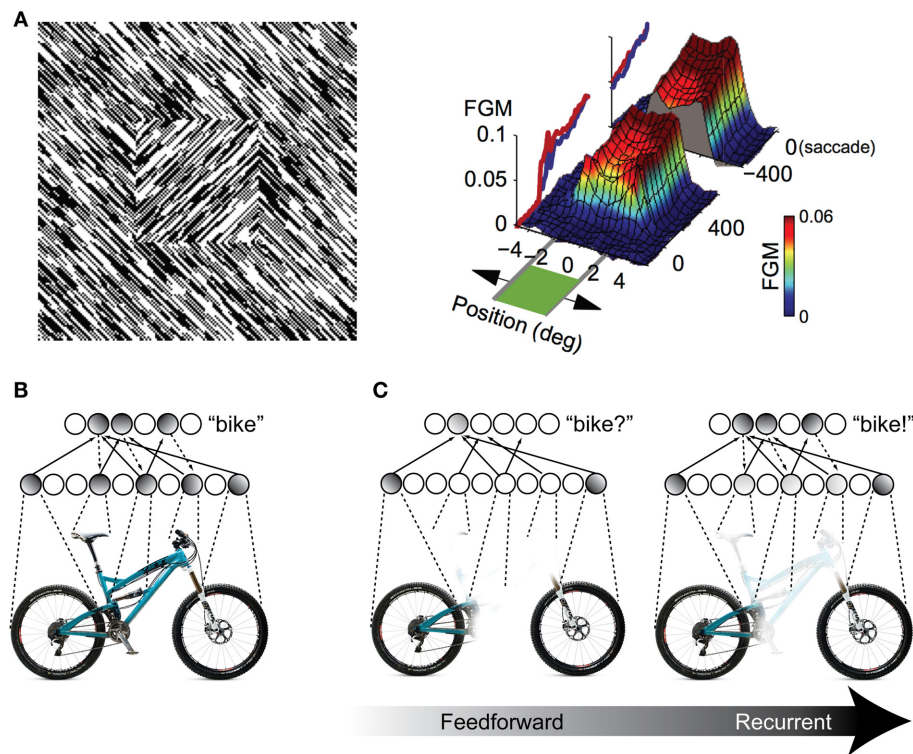


FIGURE 2 | Illustration of recurrent processing's filling-in computations during figure-ground processing and object recognition. (A) Processing of an orientation-defined square stimulus results in enhancement of the figural elements compared to the background elements. This enhancement comes in the form of recurrent feedback that groups together common image elements and spreads activation throughout the interior of the square, effectively "filling" it in as a perceptually salient surface. FGM, Figure Ground Modulation, i.e., difference between figure and background responses. Adapted from Lamme et al. (1998) and Poort et al. (2012). **(B,C)** The same

feedback-based "filling-in" principle can be applied to object recognition processing when stimuli are occluded. When object features are occluded, only a partial representation is elicited by the first feedforward responses. However, recurrent feedback (e.g., between IT and extrastriate areas) propagates these partial responses back to early visual areas, driving neurons that respond to co-occurring features that might be occluded in the physical stimulus. This recurrent processing between hierarchically adjacent visual areas can effectively "fill in" the occluded features in the object representation.

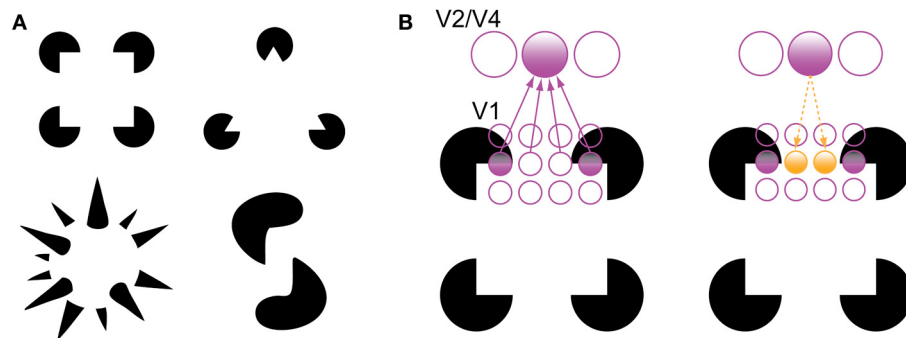


FIGURE 3 | Illusory contour perception in Kanizsa shapes (Kanizsa, 1979). (A) Traditional and more complex Kanizsa shapes that evoke strong illusory contour percepts. Complex shapes courtesy of Steven Lehar (<http://cns-alumni.bu.edu/~slehar/Lehar.html>). **(B)** Perception of illusory contours has been suggested to arise

by virtue of recurrent feedback from extrastriate areas to V1. Specifically, feedback from an extrastriate neuron drives neurons that code similarly oriented contours that fall within its receptive field, spreading activation across the gap in the Kanizsa shape. Adapted from Lee (2003).

The second supportive phenomenon is an actual object completion effect, which has gained support from fMRI studies that show little-to-no difference in the activation levels of occluded and unoccluded stimuli in object-selective regions of cortex

(Lerner et al., 2002). Intact activation, however, could simply reflect increased gain of the encoded object fragments without a more complex completion process. To differentiate between these two possibilities, one can use an fMRI adaptation paradigm,

which depends on neural mechanisms that decrease response levels for repeated stimuli that are perceived as the same. This method gives an experimenter an index of how perceptually similar two experimental conditions are. For example, Kourtzi and Kanwisher (2001) presented observers with images that contained occluding bars either in front of or behind target objects (in which case, the targets were effectively unoccluded). The experiment measured the hemodynamic response in the lateral occipital cortex (LOC), which has been strongly suggested as the human homolog of IT cortex in monkey (Grill-Spector et al., 2001; Orban et al., 2004). The results indicated that there was no significant change in hemodynamic response when subjects were presented with two identical objects in sequence, as well as when subjects were presented with occluded and unoccluded versions of an object in sequence. Thus, at the level of LOC, there is little difference in the way that unoccluded and occluded versions of the same object are represented. More recent techniques such as representational similarity (Kriegeskorte et al., 2008a,b) or decoding analyses (Tang et al., in press) might further illuminate how occluded objects are represented in various regions of cortex.

While the perception of illusory contours has been linked to recurrent feedback (Lee, 2003), this explanation has not been explored as extensively in the object completion literature, likely due to most studies using relatively coarse measures like fMRI (e.g., the aforementioned studies that rely on fMRI adaptation). Computationally, illusory contour perception and object completion could be implemented by the same mechanism, whereby higher-level neurons with overlapping receptive fields feed responses back to lower-level neurons in the absence of the visual information itself and produce the perception of illusory object features. According to this view, when operating between extrastriate levels and V1, the mechanism produces illusory contours; when operating between IT cortex and extrastriate areas, it produces more complex illusory object features. There is some support for this idea in the literature. For example, Rauschenberger et al. (2006) demonstrated object completion effects in LOC as well as in extrastriate areas when stimuli were presented for longer durations, suggesting that there is a “temporal unfolding” of object completion from higher levels of the ventral stream to lower-level areas.

However, illusory contour stimuli evoke a perceptually salient completion phenomenon, whereas the filling-in of objects does not. These processes have been distinguished in the literature as “modal” and “amodal” completion, respectively (Johnson and Olshausen, 2005; Seghier and Vuilleumier, 2006). Modal completion has been shown to elicit illusory responses in V1 (Lee and Nguyen, 2001), supporting the idea that whatever representation is present in V1 is what we “perceive” (Bullier, 2001). It is unclear whether amodal completion processing also reaches back to the level of V1. Some studies indicate that V1 represents completed shapes (Rauschenberger et al., 2006), whereas others show that the complete representation is only present in extrastriate and higher-level areas (Weigelt et al., 2007). More recently, Emmanouil and Ro (2014) showed that object completion can occur rapidly and without visual awareness, further supporting the dissociation of object completion from top-down attention.

If our proposal is correct, the time course of object completion effects should agree with the time course of recurrent processing as described above. Some studies show object completion effects beginning to manifest over temporal and parietal sites (as indexed by EEG scalp recordings) around 130 ms at the earliest and continuing to evolve until around 200 ms into processing (Johnson and Olshausen, 2005; Chen et al., 2009). These data are consistent with the explanation of object completion rapidly engaging recurrent processing with striate and extrastriate areas, assuming the 50 ms delay typically observed when the brain is processing occluded object stimuli (Kovacs et al., 1995; Nielsen et al., 2006).

However, other studies have suggested a much later time course for object completion effects, beginning around 200 ms and completing around 400 ms (Doniger et al., 2000; Sehatpour et al., 2006, 2008). One consistent characteristic of these latter studies is that they use fragmented line drawings of objects, whereas studies that associate an early time course with object completion have used photorealistic images of objects. It is unclear whether this late temporal correlate of object completion is due to relatively slow, attention-mediated processing, or due to a fundamentally different type of processing. For example, photorealistic occlusion might recruit the surface-coded computations associated with recurrent processing since there are explicitly depicted depth planes (an occluder and an object) whereas resolving contour fragmentation might rely on a completely different computation since depth planes are less well-defined in line drawings. Furthermore, the studies that associate the later time course with object completion have not used a paradigm such as response adaptation that crucially allows inference about whether an unoccluded and occluded object are represented similarly.

In summary, it seems clear that recurrent processing promotes signal amplification between reciprocally connected brain regions. There is substantial evidence that this is not a simple multiplicative gain operation, but a considerably more complex grouping or surface-based computation that spreads activation between related object features. This idea has been well-studied in the literature on illusory contour perception and the data support the explanation that illusory contour perception is due to V1 neurons receiving recurrent feedback from extrastriate regions. The same idea can be applied to object completion effects in IT cortex, predicting that they are due to feedback-rectified signals from extrastriate regions. This recurrent processing-based explanation has received little attention in the literature, but is generally supported by the timing of object completion effects.

SUMMARY AND FUTURE RESEARCH

Over the last 5–10 years, evidence has accumulated that local recurrent signals are an integral part of early visual processing. TMS studies have indicated that recurrent processing engages striate and extrastriate areas during visual recognition tasks in as little as 100 ms (Camprodon et al., 2010; Koivisto et al., 2011) and theories of backward masking have provided additional accordant timing data as well as suggested a general theory of how corticocortical interactions support visual perception (Fahrenfort et al., 2007, 2008; Boehler et al., 2008). Surprisingly though, relatively little work has focused on synthesizing these ideas with theories of visual object recognition, which is commonly held

to be primarily a feedforward process (DiCarlo et al., 2012). Instead, theories of recurrent processing have focused on the role of interactions between brain areas in promoting visual awareness (Lamme, 2003, 2006; Dehaene et al., 2006). Object perception has long been known to benefit from top-down signals that reflect attention or strategic processing, but its time course has been considered to be too slow to support the initial rapid recognition processes (Hochstein and Ahissar, 2002; VanRullen, 2007).

This paper has attempted to map out the time course of feedforward- and feedback-based events during the first 150 ms of visual processing and establish the function that rapid recurrent processing between brain areas plays within this time frame. Specifically, we propose the following overall process: A feedforward-dominant wave of activation flows up to IT in the first 80–100 ms after stimulus presentation, quickly evoking object-selective responses, while, simultaneously, activation is also feeding backward through this pathway. In the following 20 ms (an absolute latency of 100–120 ms), prefrontal areas that support the actual object categorization decision receive their first feedforward responses from IT neurons, while simultaneously, recurrent feedback from extrastriate areas has had sufficient time to more fully engage V1 populations. Recurrent feedback to V1 amplifies neurons' initial responses by grouping the responses to similar object features and enhancing them relative to other responses (Zipser et al., 1996; Lamme et al., 1998; Poort et al., 2012). In some cases, these grouping computations can cause the perception of illusory contours and surfaces (Lee and Nguyen, 2001; Seghier and Vuilleumier, 2006), but they also seem to be important when objects are degraded in order to rectify signals (Hupé et al., 1998). At an absolute latency of 120–140 ms after the initial stimulus presentation, the now extensive recurrent processing between IT and extrastriate areas can cause the representation of more complex illusory features that support object completion, by propagating these illusory responses back toward IT populations. We have recently developed a biologically-based computational model that exhibits just these dynamics (O'Reilly et al., 2013), and can provide a platform for integrating the various data cited here, while generating further testable predictions.

It is unlikely that object completion in IT cortex is a sole function of rectified responses from extrastriate areas being propagated forward in the range of 120–140 ms (or 170–190 ms, assuming the 50 ms delay observed when the brain processes occluded object stimuli; see Kovacs et al., 1995; Nielsen et al., 2006). Object completion likely also benefits from the first recurrent responses from prefrontal areas that arrive shortly after this time frame. This feedback from prefrontal areas could reflect top-down predictions that constrain the space of potential object representations in IT cortex (Bar et al., 2006; Kveraga et al., 2007), which might also have the effect of filling in visual information when it is missing from the physical stimulus. It is also plausible that lateral interactions within IT cortex itself could support object completion by enforcing statistical co-occurrences and mutual exclusions between object features (Akrami et al., 2009; Daelli and Treves, 2010). It would not be surprising if a combination of rectified feedforward responses, feedback from prefrontal areas, and lateral interactions within IT cortex itself support object completion by bringing the brain as a whole into

an attractor that combines bottom-up sensory information with top-down task demands and appropriate local constraints (e.g., Spivey, 2008). Future research that uses sophisticated techniques to rapidly and systematically disable feedforward, recurrent and lateral connectivity (e.g., optogenetics, Deisseroth, 2011) might be necessary to disentangle the relative contributions of each of these influences. Nevertheless, any contribution to object completion from local recurrent processes is supportive of the distinct functional role in resolving degraded or ambiguous stimuli proposed here.

One remaining question concerns whether recurrent processing is necessary for recognizing relatively unambiguous stimuli. "Core object recognition" (DiCarlo et al., 2012) of stimuli that vary in terms of their spatial position, scale, pose, and illumination can be rapidly decoded from the first IT responses (Hung et al., 2005). Early IT responses are also known to exhibit invariance to limited clutter (Missal et al., 1997; Zoccolan et al., 2005), suggesting that the bulk of object recognition is solved by a largely feedforward process. Importantly, these data are not fundamentally incompatible with the theory proposed here. Feedback acts on immediately lower areas with latencies as short as 10 ms (Hupé et al., 2001; Pascual-Leone and Walsh, 2001) and might be important for the Winner-Take-All (WTA) or "max" computations (Riesenhuber and Poggio, 1999; Wyatte et al., 2012b; O'Reilly et al., 2013) that have been suggested to contribute to core object recognition. Our theory has focused on recurrent processing under challenging object recognition conditions such as when stimuli are occluded or otherwise degraded. However, more substantial variability in the spatial properties of inputs might also benefit from recurrent processing. A variant of the "animal/no animal" recognition task used in many studies has shown that increasing target viewing distance in the stimulus causes backward masking to have a greater effect (Serre et al., 2007, supporting information), implicating recurrent processing for robust recognition under these conditions (Wyatte et al., 2012a). Further research with stimuli whose spatial properties can be manipulated parametrically (DiCarlo et al., 2012; Cadieu et al., 2013) combined with methods like TMS and backward masking will be necessary to determine the exact conditions under which recurrent processing is necessary.

If the theory proposed here is true, the standard description of object recognition as a feedforward process is somewhat misleading. Simply put, there is always ongoing brain activity that must be combined with new incoming sensory information, so that the notion of a strictly "feedforward sweep" is fundamentally ill-conceived (Arieli et al., 1996; Tsodyks et al., 1999). Ongoing activity could be used to establish moment-to-moment constraints that effectively guide coherent perception via recurrent processing mechanisms. While the seminal research on object recognition often focused on simple spike counts of anesthetized animals to map out the receptive field characteristics of neurons throughout the ventral stream in a well-controlled manner, future research should emphasize more complex corticocortical interactions in the awake, behaving brain to determine how neural interactions involving feedforward, lateral, and recurrent processing mechanisms combine to give rise to the visual system's robust perceptual abilities even in difficult stimulus conditions.

ACKNOWLEDGMENTS

The authors would like to thank Tim Curran and Albert Kim for helpful comments and feedback on early drafts. This work was supported by ONR grants N00014-13-1-0067, N00014-10-1-0177, and D00014-12-C-0638.

REFERENCES

- Akrami, A., Liu, Y., Treves, A., and Jagadeesh, B. (2009). Converging neuronal activity in inferior temporal cortex during the classification of morphed stimuli. *Cereb. Cortex* 19, 760–776. doi: 10.1093/cercor/bhn125
- Arieli, A., Sterkin, A., Grinvald, A., and Aertsen, A. (1996). Dynamics of ongoing activity: explanation of the large variability in evoked cortical responses. *Science* 273, 1868–1871. doi: 10.1126/science.273.5283.1868
- Bar, M., Kassam, K., Ghuman, A., Boshyan, J., and Schmidt, A. (2006). Top-down facilitation of visual recognition. *Proc. Natl. Acad. Sci. U.S.A.* 103, 449–454. doi: 10.1073/pnas.0507062103
- Boehler, C. N., Schoenfeld, M. A., Heinze, H. J., and Hopf, J. M. (2008). Rapid recurrent processing gates awareness in primary visual cortex. *Proc. Natl. Acad. Sci. U.S.A.* 105, 8742–8747. doi: 10.1073/pnas.0801999105
- Breitmeyer, B. G., and Ganz, L. (1976). Implications of sustained and transient channels for theories of visual pattern masking, saccadic suppression, and information processing. *Psychol. Rev.* 83, 1–36. doi: 10.1037/0033-295X.83.1.1
- Bressler, S. L., Tang, W., Sylvester, C. M., Shulman, G. L., and Corbetta, M. (2008). Top-down control of human visual cortex by frontal and parietal cortex in anticipatory visual spatial attention. *J. Neurosci.* 28, 10056–10061. doi: 10.1523/JNEUROSCI.1776-08.2008
- Buffalo, E. A., Fries, P., Landman, R., Liang, H., and Desimone, R. (2010). A backward progression of attentional effects in the ventral stream. *Proc. Natl. Acad. Sci. U.S.A.* 107, 361–365. doi: 10.1073/pnas.0907658106
- Bullier, J. (2001). Feedback connections and conscious vision. *Trends Cogn. Sci.* 5, 369–370. doi: 10.1016/S1364-6613(00)01730-7
- Camprodon, J. A., Zohary, E., Brodbeck, V., and Pascual-Leone, A. (2010). Two phases of v1 activity for visual recognition of natural images. *J. Cogn. Neurosci.* 22, 1262–1269. doi: 10.1162/jocn.2009.21253
- Cadiou, C. F., Hong, H., Yamins, D., Pinto, N., Majaj, N. J., and DiCarlo, J. J. (2013). “The neural representation benchmark and its evaluation on brain and machine,” in *International Conference on Learning Representations (ICLR)* (Scottsdale, AZ).
- Chen, J., Liu, B., Chen, B., and Fang, F. (2009). Time course of amodal completion in face perception. *Vision Res.* 49, 752–758. doi: 10.1016/j.visres.2009.02.005
- Chen, M., Yan, Y., Gong, X., Gilbert, C., Liang, H., and Li, W. (2014). Incremental integration of global contours through interplay between visual cortical areas. *Neuron* 82, 682–694. doi: 10.1016/j.neuron.2014.03.023
- Corthout, E., Uttl, B., Walsh, V., Hallett, M., and Cowey, A. (1999a). Timing of activity in early visual cortex as revealed by transcranial magnetic stimulation. *Neuroreport* 10, 2631–2634. doi: 10.1097/00001756-199908200-00035
- Corthout, E., Uttl, B., Ziemann, U., Cowey, A., and Hallett, M. (1999b). Two periods of processing in the (circum)striate visual cortex as revealed by transcranial magnetic stimulation. *Neuropsychologia* 37, 137–145. doi: 10.1016/S0028-3932(98)00088-8
- Daelli, V., and Treves, A. (2010). Neural attractor dynamics in object recognition. *Exp. Brain Res.* 203, 241–248. doi: 10.1007/s00221-010-2243-1
- Dehaene, S., Changeux, J. P., Naccache, L., Sackur, J., and Sergent, C. (2006). Conscious preconscious and subliminal processing: a testable taxonomy. *Trends Cogn. Sci.* 10, 204–211. doi: 10.1016/j.tics.2006.03.007
- Dehaene, S., Naccache, L., Cohen, L., Le Bihan, D., Mangin, J. F., Poline, J. B., et al. (2001). Cerebral mechanisms of work masking and unconscious repetition priming. *Nat. Neurosci.* 4, 752–758. doi: 10.1038/89551
- Deisseroth, K. (2011). Optogenetics. *Nat. Methods* 8, 26–29. doi: 10.1038/nmeth.f.324
- DiCarlo, J. J., Zoccolan, D., and Rust, N. C. (2012). How does the brain solve visual object recognition? *Neuron* 73, 415–434. doi: 10.1016/j.neuron.2012.01.010
- Doniger, G. M., Foxe, J. J., Murray, M. M., Higgins, B. A., Snodgrass, J. G., Schroeder, C. E., et al. (2000). Activation timecourse of ventral visual stream object-recognition areas: high density electrical mapping of perceptual closure processes. *J. Cogn. Neurosci.* 12, 615–621. doi: 10.1162/089892900562372
- Emmanouil, T. A., and Ro, T. (2014). Amodal completion of unconsciously presented objects. *Psychon. Bull. Rev.* doi: 10.3758/s13423-014-0590-9. [Epub ahead of print].
- Enns, J. T., and Di Lollo, V. (2000). What’s new in visual masking? *Trends Cogn. Sci.* 4, 345–352. doi: 10.1016/S1364-6613(00)01520-5
- Fahrenfort, J. J., Scholte, H. S., and Lamme, V. A. F. (2007). Masking disrupts reentrant processing in human visual cortex. *J. Cogn. Neurosci.* 19, 1488–1497. doi: 10.1162/jocn.2007.19.9.1488
- Fahrenfort, J. J., Scholte, H. S., and Lamme, V. A. F. (2008). The spatiotemporal profile of cortical processing leading up to visual perception. *J. Vision* 8, 12.1–12.12. doi: 10.1167/8.1.12
- Felleman, D. J., and Van Essen, D. C. (1991). Distributed hierarchical processing in the primate cerebral cortex. *Cereb. Cortex* 1, 1–47. doi: 10.1093/cercor/1.1.1
- Galuske, R. A. W., Schmidt, K. E., Goebel, R., Lomber, S. G., and Payne, B. R. (2002). The role of feedback in shaping neural representations in cat visual cortex. *Proc. Natl. Acad. Sci. U.S.A.* 99, 17083–17088. doi: 10.1073/pnas.2423.99199
- Grill-Spector, K., Kourtzi, Z., and Kanwisher, N. (2001). The lateral occipital complex and its role in object recognition. *Vision Res.* 41, 1409–1422. doi: 10.1016/S0042-6989(01)00073-6
- Haynes, J. D., Driver, J., and Rees, G. (2005). Visibility reflects dynamic changes of effective connectivity between v1 and fusiform cortex. *Neuron* 46, 811–821. doi: 10.1016/j.neuron.2005.05.012
- Hillyard, S. A., and Anlo-Vento, L. (1998). Event-related brain potentials in the study of visual selective attention. *Proc. Natl. Acad. Sci. U.S.A.* 95, 781–787. doi: 10.1073/pnas.95.3.781
- Hochstein, S., and Ahissar, M. (2002). View from the top: hierarchies and reverse hierarchies in the visual system. *Neuron* 36, 791–804. doi: 10.1016/S0896-6273(02)01091-7
- Hung, C. P., Kreiman, G., Poggio, T., and DiCarlo, J. J. (2005). Fast readout of object identity from macaque inferotemporal cortex. *Science* 310, 863–866. doi: 10.1126/science.1117593
- Hupé, J. M., James, A. C., Girard, P., and Bullier, J. (2001). Response modulations by static texture surround in area v1 of the macaque monkey do not depend on feedback connections from v2. *J. Neurophysiol.* 85, 146–163.
- Hupé, J. M., James, A. C., Payne, B. R., Lomber, S. G., Girard, P., and Bullier, J. (1998). Cortical feedback improves discrimination between figure and background by v1 v2 and v3 neurons. *Nature* 394, 784–787. doi: 10.1038/29537
- Johnson, J. S., and Olshausen, B. A. (2005). The recognition of partially visible natural objects in the presence and absence of their occluders. *Vision Res.* 45, 3262–3276. doi: 10.1016/j.visres.2005.06.007
- Kanizsa, G. (1979). *Organization in Vision: Essays on Gestalt Perception*. Westport, CT: Praeger Publishers.
- Koivisto, M., Railo, H., Revonsuo, A., Vanni, S., and Salminen-Vaparanta, N. (2011). Recurrent processing in v1/v2 contributes to categorization of natural scenes. *J. Neurosci.* 31, 2488–2492. doi: 10.1523/JNEUROSCI.3074-10.2011
- Kourtzi, Z., and Kanwisher, N. (2001). Representation of perceived object shape by the human lateral occipital complex. *Science* 293, 1506–1509. doi: 10.1126/science.1061133
- Kovacs, G., Vogels, R., and Orban, G. A. (1995). Selectivity of macaque inferior temporal neurons for partially occluded shapes. *J. Neurosci.* 15, 1984–1997.
- Kriegeskorte, N., Mur, M., and Bandettini, P. (2008a). Representational similarity analysis – connecting the branches of systems neuroscience. *Front. Syst. Neurosci.* 2:4. doi: 10.3389/neuro.06.004.2008
- Kriegeskorte, N., Mur, M., Ruff, D. A., Kiani, R., Bodurka, J., Esteky, H., et al. (2008b). Matching categorical object representations in inferior temporal cortex of man and monkey. *Neuron* 60, 1126–1141. doi: 10.1016/j.neuron.2008.10.043
- Kveraga, K., Ghuman, A., and Bar, M. (2007). Top-down predictions in the cognitive brain. *Brain Cogn.* 65, 145–168. doi: 10.1016/j.bandc.2007.06.007
- Lamme, V., and Roelfsema, P. (2000). The distinct modes of vision offered by feedforward and recurrent processing. *Trends Neurosci.* 23, 571–579. doi: 10.1016/S0166-2236(00)01657-X

- Lamme, V., Super, H., and Spekreijse, H. (1998). Feedforward, horizontal, and feedback processing in the visual cortex. *Curr. Opin. Neurobiol.* 8, 529–535. doi: 10.1016/S0959-4388(98)80042-1
- Lamme, V. A. (2003). Why visual attention and awareness are different. *Trends Cogn. Sci.* 7, 12–18. doi: 10.1016/S1364-6613(02)00013-X
- Lamme, V. A. F. (2006). Towards a true neural stance on consciousness. *Trends Cogn. Sci.* 10, 494–501. doi: 10.1016/j.tics.2006.09.001
- Lee, T. S. (2003). Computations in the early visual cortex. *J. Physiol.* 97, 121–139. doi: 10.1016/j.jphysparis.2003.09.015
- Lee, T. S., and Nguyen, M. (2001). Dynamics of subjective contour formation in the early visual cortex. *Proc. Natl. Acad. Sci. U.S.A.* 98, 1907–1911. doi: 10.1073/pnas.98.4.1907
- Lerner, Y., Hendler, T., and Malach, R. (2002). Object-completion effects in the human lateral occipital cortex. *Cereb. Cortex* 12, 163–177. doi: 10.1093/cercor/12.2.163
- Luck, S. J., Heinze, H. J., Mangun, G. R., and Hillyard, S. A. (1990a). Visual event-related potentials index focused attention within bilateral stimulus arrays. I. Evidence for early selection. *Electroencephalogr. Clin. Neurophysiol.* 75, 511–527. doi: 10.1016/0013-4694(90)90139-B
- Luck, S. J., Heinze, H. J., Mangun, G. R., and Hillyard, S. A. (1990b). Visual event-related potentials index focused attention within bilateral stimulus arrays. II. Functional dissociation of P1 and N1 components. *Electroencephalogr. Clin. Neurophysiol.* 75, 528–542. doi: 10.1016/0013-4694(90)90139-B
- Macknik, S. L., and Livingstone, M. S. (1998). Neuronal correlates of visibility and invisibility in the primate visual system. *Nat. Neurosci.* 1, 144–149. doi: 10.1038/393
- Martinez, A., Di Russo, F., Anlo-Vento, L., Sereno, M. I., Buxton, R. B., and Hillyard, S. A. (2001). Putting spatial attention on the map: timing and localization of stimulus selection processes in striate and extrastriate visual areas. *Vision Res.* 41, 1437–1457. doi: 10.1016/S0042-6989(00)00267-4
- Mehta, A. D., Ulbert, I., and Schroeder, C. E. (2000a). Intermodal selective attention in monkeys. I: distribution and timing of effects across visual areas. *Cereb. Cortex* 10, 343–358. doi: 10.1093/cercor/10.4.343
- Mehta, A. D., Ulbert, I., and Schroeder, C. E. (2000b). Intermodal selective attention in monkeys. II: physiological mechanisms of modulation. *Cereb. Cortex* 10, 359–370. doi: 10.1093/cercor/10.4.359
- Mignard, M., and Malpeli, J. G. (1991). Paths of information flow through visual cortex. *Science* 251, 1249–1251. doi: 10.1126/science.1848727
- Missal, M., Vogels, R., and Orban, G. A. (1997). Responses of macaque inferior temporal neurons to overlapping shapes. *Cereb. Cortex* 7, 758–767. doi: 10.1093/cercor/7.8.758
- Nielsen, K. J., Logothetis, N. K., and Rainer, G. (2006). Dissociation between local field potentials and spiking activity in macaque inferior temporal cortex reveals diagnosticity-based encoding of complex objects. *J. Neurosci.* 26, 9639–9645. doi: 10.1523/JNEUROSCI.2273-06.2006
- Noesselt, T., Hillyard, S. A., Woldorff, M. G., Schoenfeld, A., Hagner, T., Jäncke, L., et al. (2002). Delayed striate cortical activation during spatial attention. *Neuron* 35, 575–587. doi: 10.1016/S0896-6273(02)00781-X
- Nowak, L., and Bullier, J. (1997). “The timing of information transfer in the visual system,” in *Cerebral Cortex Vol. 12: Extrastriate Cortex in Primates*, eds K. S. Rockland, J. H. Kaas, and A. Peters (New York, NY: Plenum), 205–241.
- Orban, G. A., Van Essen, D., and Vanduffel, W. (2004). Comparative mapping of higher visual areas in monkeys and humans. *Trends Cogn. Sci.* 8, 315–324. doi: 10.1016/j.tics.2004.05.009
- O'Reilly, R., Wyatte, D., Herd, S., Mingus, B., and Jilk, D. (2013). Recurrent processing during object recognition. *Front. Psychol.* 4:124. doi: 10.3389/fpsyg.2013.00124
- Pascual-Leone, A., and Walsh, V. (2001). Fast backprojections from the motion to the primary visual area necessary for visual awareness. *Science* 292, 510–512. doi: 10.1126/science.1057099
- Poort, J., Raudies, F., Wannig, A., Lamme, V. A. F., Neumann, H., and Roelfsema, P. R. (2012). The role of attention in figure-ground segregation in areas v1 and v4 of the visual cortex. *Neuron* 75, 143–156. doi: 10.1016/j.neuron.2012.04.032
- Rauschenberger, R., Liu, T., Slotnick, S. D., and Yantis, S. (2006). Temporally unfolding neural representation of pictorial occlusion. *Psychol. Sci.* 17, 358–364. doi: 10.1111/j.1467-9280.2006.01711.x
- Reynolds, J., Pasternak, T., and Desimone, R. (2000). Attention increases sensitivity of v4 neurons. *Neuron* 26, 703–714. doi: 10.1016/S0896-6273(00)81206-4
- Reynolds, J. H., and Heeger, D. J. (2009). The normalization model of attention. *Neuron* 61, 168–185. doi: 10.1016/j.neuron.2009.01.002
- Riesenhuber, M., and Poggio, T. (1999). Hierarchical models of object recognition in cortex. *Nat. Neurosci.* 2, 1019–1025. doi: 10.1038/14819
- Roland, P. (2010). Six principles of visual cortical dynamics. *Front. Syst. Neurosci.* 4:28. doi: 10.3389/fnsys.2010.00028
- Roland, P. E., Hanazawa, A., Undeman, C., Eriksson, D., Tompa, T., Nakamura, H., et al. (2006). Cortical feedback depolarization waves: a mechanism of top-down influence on early visual areas. *Proc. Natl. Acad. Sci. U.S.A.* 103, 12586–12591. doi: 10.1073/pnas.0604925103
- Sandell, J. H., and Schiller, P. H. (1982). Effect of cooling area 18 on striate cortex cells in the squirrel monkey. *J. Neurophysiol.* 48, 38–48.
- Sclar, G., Maunsell, J. H., and Lennie, P. (1990). Coding of image contrast in central visual pathways of the macaque monkey. *Vision Res.* 30, 1–10. doi: 10.1016/0042-6989(90)90123-3
- Seghier, M. L., and Vuilleumier, P. (2006). Functional neuroimaging findings on the human perception of illusory contours. *Neurosci. Biobehav. Rev.* 30, 595–612. doi: 10.1016/j.neubiorev.2005.11.002
- Sehatpour, P., Molholm, S., Javitt, D. C., and Foxe, J. J. (2006). Spatiotemporal dynamics of human object recognition processing: an integrated high-density electrical mapping and functional imaging study of “closure” processes. *Neuroimage* 29, 605–618. doi: 10.1016/j.neuroimage.2005.07.049
- Sehatpour, P., Molholm, S., Schwartz, T. H., Mahoney, J. R., Mehta, A. D., Javitt, D. C., et al. (2008). A human intracranial study of long-range oscillatory coherence across a frontal-occipital-hippocampal brain network during visual object processing. *Proc. Natl. Acad. Sci. U.S.A.* 105, 4399–4404. doi: 10.1073/pnas.0708418105
- Serre, T., Kreiman, G., Kouh, M., Cadieu, C., Knoblich, U., and Poggio, T. (2007). A quantitative theory of immediate visual recognition. *Prog. Brain Res.* 165, 33–56. doi: 10.1016/S0079-6123(06)65004-8
- Spivey, M. J. (2008). *The Continuity of Mind*. New York, NY: Oxford University Press.
- Sporns, O., and Zwi, J. D. (2004). The small world of the cerebral cortex. *Neuroinformatics* 2, 145–162. doi: 10.1385/NI:2:2:145
- Tang, H., Buia, C., Madhavan, R., Madsen, J., Anderson, W., Crone, N., et al. (in press). Spatiotemporal dynamics underlying object completion in human ventral visual cortex. *Neuron*.
- Thompson, K. G., Biscoe, K. L., and Sato, T. R. (2005). Neuronal basis of covert spatial attention in the frontal eye field. *J. Neurosci.* 25, 9479–9487. doi: 10.1523/JNEUROSCI.0741-05.2005
- Tsodyks, M., Kenet, T., Grinvald, A., and Arieli, A. (1999). Linking spontaneous activity of single cortical neurons and the underlying functional architecture. *Science* 286, 1943–1946. doi: 10.1126/science.286.5446.1943
- VanRullen, R. (2007). The power of the feed-forward sweep. *Adv. Cogn. Psychol.* 3, 167–176. doi: 10.2478/v10053-008-0022-3
- VanRullen, R., and Thorpe, S. J. (2001). Rate coding versus temporal order coding: what the retinal ganglion cells tell the visual cortex. *Neural Comput.* 13, 1255–1284. doi: 10.1162/08997660152002852
- VanRullen, R., and Thorpe, S. J. (2002). Surfing a spike wave down the ventral stream. *Vision Res.* 42, 2593–2615. doi: 10.1016/S0042-6989(02)00298-5
- Weigelt, S., Singer, W., and Muckli, L. (2007). Separate cortical stages in amodal completion revealed by functional magnetic resonance adaptation. *BMC Neurosci.* 8:70. doi: 10.1186/1471-2202-8-70
- Williford, T., and Maunsell, J. H. R. (2006). Effects of spatial attention on contrast response functions in macaque area v4. *J. Neurophysiol.* 96, 40–54. doi: 10.1152/jn.01207.2005
- Wyatte, D., Curran, T., and O'Reilly, R. (2012a). The limits of feedforward vision: recurrent processing promotes robust object recognition when objects are degraded. *J. Cogn. Neurosci.* 24, 2248–2261. doi: 10.1162/jocn_a_00282
- Wyatte, D., Herd, S., Mingus, B., and O'Reilly, R. (2012b). Competitive inhibition and top-down feedback in binding during object recognition. *Front. Psychol.* 3:182. doi: 10.3389/fpsyg.2012.00182
- Zipser, K., Lamme, V. A. F., and Schiller, P. H. (1996). Contextual modulation in primary visual cortex. *J. Neurosci.* 16, 7376–7373.
- Zoccolan, D., Cox, D. D., and DiCarlo, J. J. (2005). Multiple object response normalization in monkey inferotemporal cortex. *J. Neurosci.* 25, 8150–8164. doi: 10.1523/JNEUROSCI.2058-05.2005

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.

Received: 14 May 2014; accepted: 10 June 2014; published online: 01 July 2014.

Citation: Wyatte D, Jilk DJ and O'Reilly RC (2014) Early recurrent feedback facilitates visual object recognition under challenging conditions. *Front. Psychol.* 5:674. doi: 10.3389/fpsyg.2014.00674

This article was submitted to Perception Science, a section of the journal Frontiers in Psychology.

Copyright © 2014 Wyatte, Jilk and O'Reilly. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



The temporal window of individuation limits visual capacity

Andreas Wutz* and David Melcher

Active Perception Laboratory, Center for Mind/Brain Sciences, University of Trento, Rovereto, Italy

Edited by:

Haluk Ogmen, University of Houston,
USA

Reviewed by:

Niko Busch,
Charité–Universitätsmedizin Berlin,
Germany
Aaron Michael Clarke, École
Polytechnique Fédérale de Lausanne,
Switzerland

*Correspondence:

Andreas Wutz, Active Perception
Laboratory, Center for Mind/Brain
Sciences, University of Trento, Palazzo
Fedrigotti, Corso Bettini 31, Rovereto
38068, Italy
e-mail: andreas.wutz-1@unitn.it

One of the main tasks of vision is to individuate and recognize specific objects. Unlike the detection of basic features, object individuation is strictly limited in capacity. Previous studies of capacity, in terms of subitizing ranges or visual working memory, have emphasized spatial limits in the number of objects that can be apprehended simultaneously. Here, we present psychophysical and electrophysiological evidence that capacity limits depend instead on time. Contrary to what is commonly assumed, subitizing, the reading-out a small set of individual objects, is not an instantaneous process. Instead, individuation capacity increases in steps within the lifetime of visual persistence of the stimulus, suggesting that visual capacity limitations arise as a result of the narrow window of feedforward processing. We characterize this temporal window as coordinating individuation and integration of sensory information over a brief interval of around 100 ms. Neural signatures of integration windows are revealed in reset alpha oscillations shortly after stimulus onset within generators in parietal areas. Our findings suggest that short-lived alpha phase synchronization (≈ 1 cycle) is key for individuation and integration of visual transients on rapid time scales (< 100 ms). Within this time frame intermediate-level vision provides an equilibrium between the competing needs to individuate invariant objects, integrate information about those objects over time, and remain sensitive to dynamic changes in sensory input. We discuss theoretical and practical implications of temporal windows in visual processing, how they create a fundamental capacity limit, and their role in constraining the real-time dynamics of visual processing.

Keywords: visual capacity, temporal window, oscillatory phase synchrony, individuation, integration

INTRODUCTION – VIRTUAL CONTINUITY AND STABILITY OF PERCEPTUAL SPACE AND TIME

The perception system is faced with the task of transforming continuous sensory input into discrete objects and events. It is critical for survival that the perceptual system is sensitive and quickly responsive to changes in the input over time in order, for example, to detect and interpret signals regarding object or self-motion. However, a primary goal of perceptual systems is also to uncover stability in the identity and location of spatiotemporal objects and to integrate information over extended periods of time in order to understand complex phenomena such as biological motion (Neri et al., 1998) or events (Hasson et al., 2008; Lerner et al., 2011; Zacks and Magliano, 2011). Information must be integrated over time to recover the regularities in the world and to use this perception of order to make predictions about the near future (Nastase et al., 2014). Thus, vision in real-time requires a balance combining information over time (in order to integrate motion signals or to keep track of the same spatiotemporal object) and sensitivity to new information.

A simple example of this challenge for a perceptual system is the task of crossing a busy street. Perceiving and predicting the motion of vehicles requires combining information over 100s of milliseconds or even seconds, often including the combination of motion information across occlusion or changes in retinal position caused by eye movements. On the one hand, combining information over a longer time period would likely lead to the best possible estimate

of all of the features of the oncoming cars. Nonetheless, the visual system must also provide a good enough estimate of the current location of each vehicle in order to support action. Thus, the perceptual system must optimally balance the competing needs of speed and information: more time yields better information but slows down the ability of the organism to react rapidly to the current state of affairs. It seems likely that the brain provides a compromise by utilizing a hierarchy of different temporal integration windows (Pöppel, 1997, 2009; Hasson et al., 2008; Melcher and Colby, 2008; Holcombe, 2009; Lerner et al., 2011; Masquelier et al., 2011) and by alternating periods of feedforward sampling of new information with feedback/re-entrant processes (Di Lollo et al., 2000; Lamme and Roelfsema, 2000) that create a perceptual synthesis of the disparate sensory information into coherent, stable spatio-temporal entities like objects. Indeed, converging evidence suggests that temporal limits on visual processing can be broadly divided into two groups of perceptual mechanisms (Holcombe, 2009). A fast group comprises processes of feedforward feature detection and works on the scale of some 10s of milliseconds. The second group of visual mechanisms is much slower, taking more than at least 100 ms and operates on more high-level properties, like objects that have been selected and individuated.

Here we consider evidence regarding how the temporal window of object individuation might bridge the gap between fast feedforward sampling of information and slower object-based computations. We start with a selective review of the relevant literature on object individuation, its capacity limits and temporal

limits in visual perception. Then, we describe a methodology to experimentally reduce the effective visual persistence of a visual display in order to more closely map out the time course of object individuation processes. We review our recent behavioral studies using this method to show the unfolding of object individuation and working memory over time. Then we present and discuss magnetencephalography (MEG) evidence regarding the neural correlates of this process, including the possibility that neural synchronization patterns can provide useful information about the nature of integration and individuation. Finally, we discuss the implications of these findings for capacity limits in visual cognition, their relationship with natural vision and oscillatory brain dynamics, and point out some open questions and directions for future research.

INDIVIDUATION MEASURES VISUO-SPATIAL OBJECT PROCESSING

INDIVIDUATION: AN INTERMEDIATE STEP BETWEEN SAMPLING FEATURES AND OBJECTS

Although sensory information seems to extend continuously into perceptual space and time, the content of cognitive operations consists of coherent scenes containing a limited number of discrete and invariant objects in any particular instance (Treisman and Gelade, 1980; Tipper et al., 1990; Kahneman et al., 1992; Baylis and Driver, 1993; Scholl et al., 2001). Such parsing of the sensory environment into elemental perceptual units (Spelke, 1988) provides a link between sensation and cognition that couples perception to the external world, free from an infinite regress of referring to semantic categories (Pylyshyn, 2001). Reading-out objects from feedforward sensory input is called individuation and involves selecting features from a crowded scene, binding them into a unitary spatiotemporal entity and segregating this perceptual unit from other individuals in the image (Treisman and Gelade, 1980; Xu and Chun, 2009). The output of this intermediate-level visual analysis is a stable object-based reference frame in which the different features of a specific location in the scene can be bound together.

Object representations at this stage are suggested to be coarse and contain only minimal feature information. In fact such individual entities do not necessarily provide information about object identity, but can be regarded as a spatio-temporal placeholder of the object in focus until feedback processes fill in content. Several theoretical, psychophysical and neuroimaging studies have emphasized the computational importance and necessity of such incremental object representations in intermediate-level vision, with these entities described as visual indexes (Pylyshyn, 1989), proto objects (Rensink, 2000), or object-files (Kahneman et al., 1992; Xu and Chun, 2009). In its essence an object can be defined as an entity whose recent spatio-temporal history can be reviewed and therefore still can be referred to as the same entity despite of changes in its location over time (Kahneman et al., 1992).

Individuation is an intermediate step in object processing between bottom-up feature detection and the recognition of stable and coherent objects. Object representations at this level of processing are commonly measured with an enumeration task that solely requires knowing whether an object is an individual rather

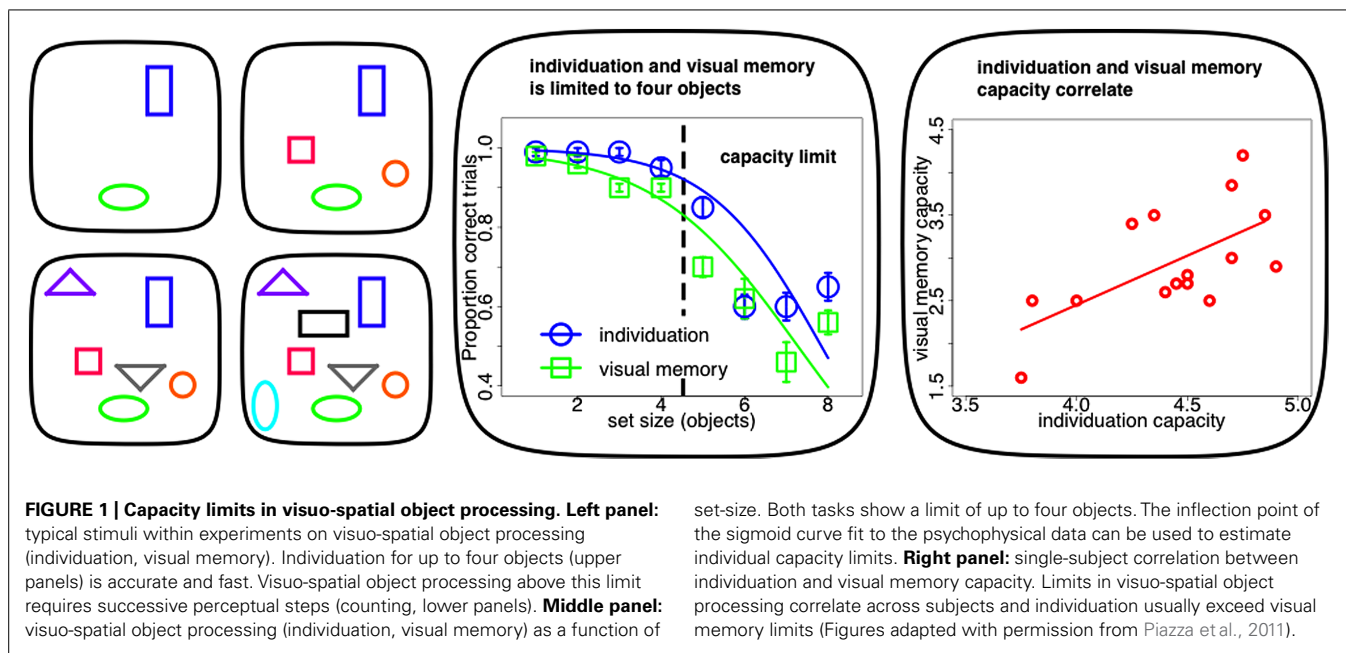
than its identity (which is usually measured with change detection and interpreted as the content of visual working memory). In this review we will map the temporal dynamics of visual object processing as a cascade from (a) sampling a visual signal over (b) a temporal window of ca. 100 ms duration during which a scene is segmented and individuated into (c) stable object-based representations. Our main result characterizes a brief time window of persisting sensory information after stimulus onset that limits object individuation and accounts for capacity limits in visual object processing.

CAPACITY LIMITS IN INDIVIDUATION AND VISUAL MEMORY

Although human cognition is remarkably powerful, its online workspace, working memory, appears to be highly limited in the number of informational units it processes (Miller, 1956; Luck and Vogel, 1997; Cowan, 2000). It is interesting to note that this capacity is linked to cognitive abilities in general. For example, inter-individual variability in measures of fluid intelligence and capacity estimates are highly correlated (Engle et al., 1999; Cowan et al., 2005; Fukuda et al., 2010b) and reduced capacity is often found in patients with neuropsychiatric disorders (Karatekin and Asarnow, 1998; Lee et al., 2010).

Recent evidence suggests that two distinct mechanisms, object individuation and identification, work together in creating these visual object capacity limitations (Xu and Chun, 2009). Individuation appears to be the initial bottleneck in visual object processing from an unlimited in capacity, but fragile, purely bottom-up and in parallel computed sensory representation (iconic memory: Sperling, 1960, 1963; Neisser, 1967) to such a capacity limited, durable and cognitively structured visual store (visual short-term memory: Sperling, 1960, 1963; Phillips and Baddeley, 1971). A subset of these individuated objects are elaborated subsequently during object identification. It is at this stage that identity information becomes available to the observer and the content of the object can be consolidated into durable and reportable representations in visual working memory (Xu and Chun, 2009). As individuation precedes identification, the capacity of the latter has its upper bound in the limit of the former (Melcher and Piazza, 2011; Dempere-Marco et al., 2012; **Figure 1** middle panel). In fact on a single-subject level, estimates of individuation capacity commonly exceed visual memory limits and the two measures tend to be highly correlated (Piazza et al., 2011; **Figure 1** right panel).

It has long been noted that individuation is limited in capacity: we can quickly and effortlessly perceive that there are exactly two items but not that there are exactly eight items (Jevons, 1871; compare **Figure 1** left panel upper row with lower row). Enumeration is equally quick, accurate and effortless within a narrow range of one to four objects. Such small numbers of items are supposedly simultaneously apprehended by a qualitatively distinct mechanism known as “subitizing” (Kaufman et al., 1949). Performance for set-sizes exceeding this range, as measured by reaction time and accuracy, deteriorates with every additional item to be enumerated (**Figure 1** middle panel). This suggests that visual object capacity limits are grounded in this “subitizing” phenomenon and that visual processing beyond this limit has to rely on imprecise estimation or serial and time-consuming counting that requires



successive perceptual steps. In contrast, “subitizing” is thought to measure visuo-spatial object processing within one single feed-forward processing iteration (for review, see Melcher and Piazza, 2011; Piazza et al., 2011).

THEORIES ABOUT OBJECT PROCESSING CAPACITY LIMITS

In light of its importance for cognitive and perceptual functioning, the search for the root of this capacity limitation is fundamental to the study of visual cognition. There are a number of competing theories for why “subitizing,” and individuation in general, is limited to sets of only about three or four items (for review, see Piazza et al., 2011). These theories start with the idea that capacity measures the number of objects individuated “immediately” (Kaufman et al., 1949), as reflected in the root of the word “subitizing” (*subitus*). This capacity is characterized in terms of spatial metaphors such as an index, pointer, or slots. Capacity is thus typically thought of as a limit in spatial resolution, rather than temporal limits. Because of the apparent automaticity and immediateness of processing, several theories assumed an *ad hoc*, direct and continuous indexing between external coordinates and object-files (Pylyshyn, 1989), like focal slots waiting to be filled in with content (Luck and Vogel, 1997; Fukuda et al., 2010a). Since performance tends to deteriorate after around four items (although this does depend on individuals and task), it was proposed that there were four indexes or slots.

Starting with the idea that subitizing is an all-or-none, uniform process might, however, neglect the possibility that capacity is related to the temporal period during which individuation occurs. Individuation is a computationally complex task. Ullman (1984) has characterized vision in terms of serial tasks that involves indexing of salient items, marking previously indexed locations and multiple shifts of the processing focus. In fact, execution of such complex coding in real-time would seem likely to require

the implementation of a specialized routine set-up as a series of elemental operations (Roelfsema et al., 2000). As reviewed in the following section, temporal aspects of visual perception have been studied extensively and show that visual processing is not “immediate” (Kaufman et al., 1949) but always occurs over time. This raises the question of whether these temporal factors, rather than or in addition to spatial factors, might underlie capacity limits.

In terms of time, object individuation is a process that must, as described above, balance between the need for speed and the aim of integrating information over time about salient objects in order to recognize, remember and respond to their properties. This trade-off is apparent in the case of computer vision systems for robotics, in which an exact, metric representation of the environment is computationally expensive and typically too slow to guide behavior in real-time. Computer systems used to drive cars, for example, do not represent in detail the entire visual scene (Bertozzi et al., 2000) because such a complete, metric model cannot be updated in real-time. In the case of the human visual system, one strategy to deal with this trade-off is to individuate and integrate information about a small number of potentially important items within each perceptual cycle.

TEMPORAL RESOLUTION OF VISUO-SPATIAL OBJECT PROCESSING

VISUO-TEMPORAL LIMITS BETWEEN FEATURE DETECTION AND OBJECT-BASED COMPUTATIONS

Temporal resolution refers to the precision of a measurement with respect to time. Estimates of the temporal resolution of vision come from a variety of different tasks but can be divided into two groups of temporal limits: a fast group that operates on the order of 10s of milliseconds and a slower group of visual mechanisms taking more than 100 ms (Holcombe, 2009). The fast temporal limits are usually explained by temporal integration of low-level visual

features (like in the case of flicker fusion or integration masking; Crozier and Wolf, 1941; Kietzman and Sutton, 1968; Scheerer, 1973a,b; Di Lollo and Wilson, 1978; Coltheart, 1980; Enns and Di Lollo, 2000; Breitmeyer and Ögmen, 2006). In contrast, slower temporal limits are usually associated with high-level processing in an object-based frame of reference like in the case of feature conjunctions across space (color-shape: Holcombe and Cavanagh, 2001; or orientation-location: Motoyoshi and Nishida, 2001) or consolidation of objects in visual working memory (Gegenfurtner and Sperling, 1993; Vogel et al., 2006). Unlike the temporal blurring of basic image features, temporal processing limits for this slower group have been suggested to depend on selective attention (Holcombe, 2009). Together these two groups of processes act in concert to create a coherent perceptual impression in time.

Here, we try to combine these two frameworks, temporal resolution and attentional selection. As reviewed above, object individuation appears to be the basic set-up process for object-based representations, introducing selectivity in processing individual properties of a scene. Consistent with this idea recent evidence suggests that “subitizing” and individuation in general, rather than being a pre-attentive indexing mechanism (Trick and Pylyshyn, 1994), requires selective attention (Egeth et al., 2008; Olivers and Watson, 2008; Railo et al., 2008). We show that individuation is limited by temporal integration of sensory information over time and how visual capacity limits arise naturally as a consequence of this integration window. We argue that intermediate-level vision bridges the gap between fast feature detection and slower object-based computations, and that this depends on a temporal integration window that is used to structure and stabilize individual perceptual elements within a sampled sensory image.

TEMPORAL INTEGRATION OF SENSORY PERSISTENCE

Following stimulus onset a briefly presented visual display persists perceptually for a limited temporal window of 80–120 ms (Haber and Standing, 1970; Coltheart, 1980; Di Lollo, 1980). This persisting window acts like a low-pass filter on dynamic aspects of real-time vision, limiting the temporal resolution of perceiving each single visual event. When a second stimulus is presented in rapid succession to a first stimulus, the associated features of both stimulus onsets are partly integrated into a single percept. Such short-lived sensory integration intervals have been described to influence visual perception (Scheerer, 1973a; Enns and Di Lollo, 2000; Breitmeyer and Ögmen, 2006), visual memory (Di Lollo, 1980), and rapid perceptual decision-making (Scharnowski et al., 2009; Rüter et al., 2012).

Important insights into the temporal dynamics of sensory integration have been achieved through the study of visual masking: the reduction of the visibility of one stimulus, called the target, by another stimulus shown before and/or after it, called the mask (Enns and Di Lollo, 2000; Breitmeyer and Ögmen, 2006). It is classically explained in terms of a two-factor theory: integration and interruption masking (Scheerer, 1973a,b). Interruption masking limits more high-level feedback processing after perceptual analysis of the target has largely finished. Integration masking,

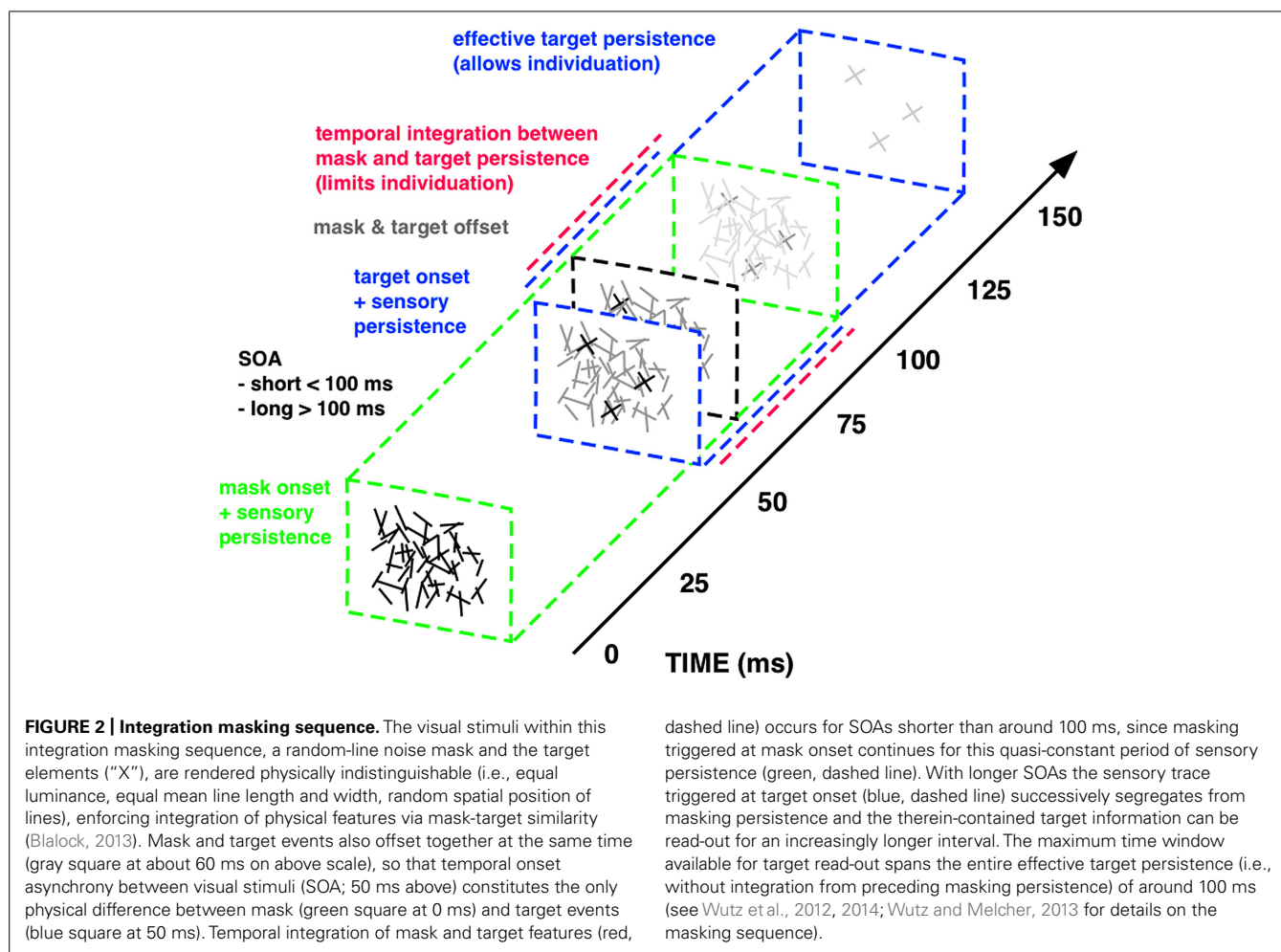
however, results from short-lived temporal collapsing of feedforward sensory signals, as a consequence of the imprecise temporal resolution of the visual system. Integration of sensory persistence between rapid successive stimuli reduces the time to access the sensory trace of each single stimulus. Hence integration masking degrades visual performance by fractionating the sensory persistence of the target display and limiting its effective presentation time. Integration masking is very effectively implemented with a specific forward masking technique that makes it possible to quantitatively change the duration of sensory persistence and the degree of temporal integration by varying the onset asynchrony between the first and second display (Di Lollo, 1980; Wutz et al., 2012; **Figure 2**).

Mask and target elements share the same physical properties, in order to equate stimulus energy from both visual events. The only physical difference between mask and target constitutes their temporal onset asynchrony. Temporal integration of mask and target features occurs for stimulus onset asynchronies (SOAs) shorter than around 100 ms, because of smearing of sensory persistence triggered at each onset. For SOAs exceeding this critical time frame, mask and target persistence segregate in time and the sensory trace of the target display can be read-out. In this way, varying the SOA within this integration masking sequence controls the effective presentation time of a visual display by fractionating its sensory trace. We designed this technique to map the temporal dynamics of successive perceptual processes involved in object processing with identical visual stimuli only varying task demands: from basic detection to subsequent individuation and finally identification and consolidation of objects in visual working memory (**Figure 3**).

INTEGRATION WINDOWS LIMIT INDIVIDUATION CAPACITY INDIVIDUATION CAPACITY INCREASES UNIT BY UNIT WITHIN THE SENSORY WINDOW

Individuation stabilizes visual perception by computing objects. This process is thought to operate within a single glance and is strictly limited in capacity to a small set of around four objects. We tested whether visual object capacity is indeed reached at the very moment a stimulus enters the visual field or instead accumulates with longer viewing time by fractionating a single glance into smaller units. We used an integration masking paradigm (see **Figure 2**) in order to vary the time to access the sensory trace of the to be individuated items and measured individuation performance for different set-sizes. Contrary to what is commonly found in “subitizing” tasks, which has consistently shown highly accurate performance up to around four objects across a wide range of studies (see **Figure 1**), fractionating the sensory persistence of the stimulus with integration masking dramatically reduces individuation capacity. This suggests that reading-out a small set of individual and stable objects is not an instantaneous process (**Figure 3**) but rather evolves over time.

Individuation capacity increases in steps within the lifetime of sensory persistence of the stimulus (**Figure 3**; Wutz et al., 2012). Within integration masking, SOA between mask and target directly reflects effective target persistence and time to read-out individual objects. Temporal integration of visual signals is complete and target information is completely inaccessible if there is

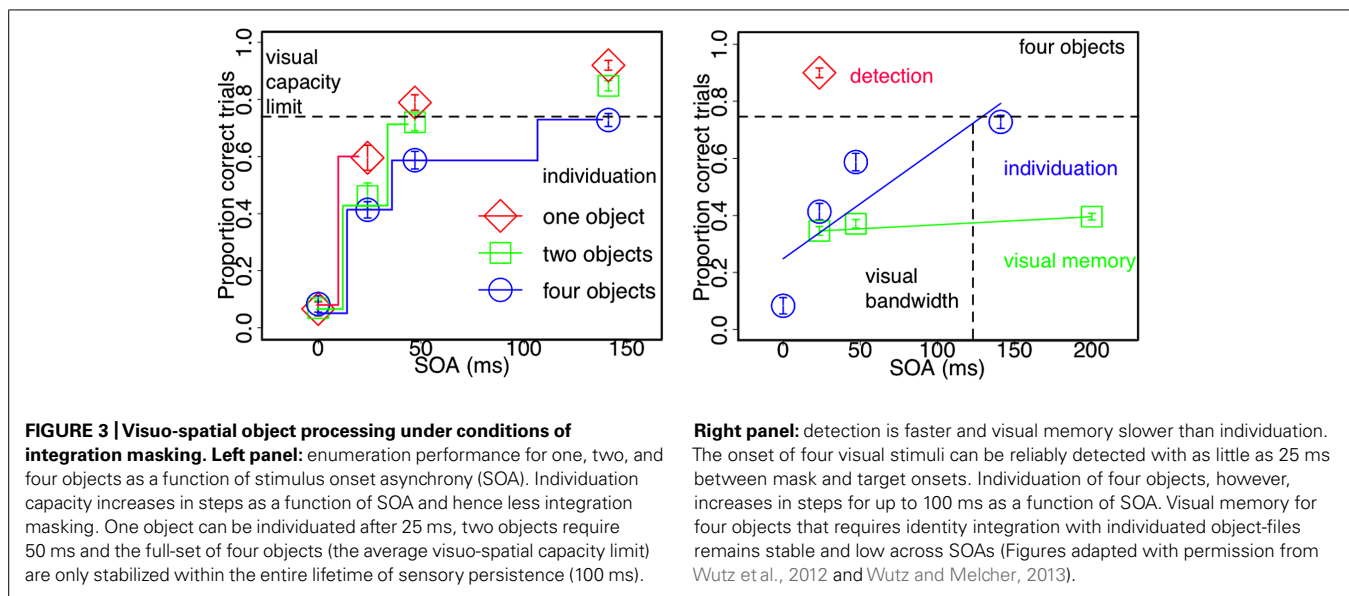


common stimulus onset (SOA = 0 ms). With increasing SOA, visual signals segregate in time and the read-out of each single sensory trace increases correspondingly. The slopes of individuation across read-out time, however, co-vary with the number of individual objects to be processed. Whereas one object is sufficiently stable within 25 ms, two objects require 50 ms to be individuated. Individuation capacity for four objects, which is the average visuo-spatial capacity limit (Figure 1), is asymptotically reached after around 100 ms (Figure 3 left panel; Wutz et al., 2012). Limiting the effective presentation time with integration masking reveals that processing speed and object capacity interact, rather than a uniform individuation improvement across the "subitizing range" with less temporal limitations. Consistent with this result, interactions between perceptual speed and object selection have also been reported for multiple object tracking (Holcombe and Chen, 2013).

Incremental individuation of objects within a stimulus' sensory persistence suggests that this temporary integration buffer is functionally critical for object processing. Such an integration interval might reflect the need to equilibrate read-out of invariant and stable perceptual form and almost simultaneously integrate changes in sensory input into a continuous stream of visual impressions. Sensory images that remain stationary within the first 100 ms after

sampling are successively segmented and structured into objects within its sensory persistence. Consequently, visuo-spatial object capacity limitations arise as a result of the narrow integration window bandwidth (Figure 3).

The speed of stable information accrual, however, is particularly crucial in case of fast changes in the sampled sensory image (<100 ms). When the sensory signal changes faster than the integration window (<100 ms; change, motion, short SOA masking sequence) individuation capacity is reduced as a function of the rate of sensory change, stabilizing only a subset of objects. This drop in visuo-spatial object processing with higher temporal processing demands balances the needs for perceptual stability in space and continuity in time. One object can already be stabilized in some 10s of milliseconds. In this way at least one object can be selected and further tracked for speeds drawing near the upper temporal limit of visual processing (Kietzman and Sutton, 1968). Structuring an entire scene into multiple objects, however, requires processing over an interval of around 100 ms. We argue that vision uses the time window of sensory persistence following stimulus onset to balance the opposing needs of individuating stable objects and maintaining the temporal resolution necessary to track rapidly changing events.



SAMPLING FEATURES IS FASTER, WHILE VISUAL MEMORY IS SLOWER THAN INDIVIDUATION

Temporal buffering of input signals does not necessarily imply that sampling of new information is inhibited completely within this integration interval. In fact, merely detecting a second event requires as little as 25 ms between event onsets (Figure 3 right panel). Despite this remarkable processing speed, the informational content of such fast feedforward sampling is considered to be virtually unlimited in capacity (Wundt, 1899; Sperling, 1960) and can already involve higher-level visual areas, allowing for rapid scene categorization of natural images (Thorpe et al., 1996; Li et al., 2002), basic image grouping (Field et al., 1993; Roelfsema et al., 2000), visual analysis of scene semantics (but not scene syntax, Vö and Wolfe, 2012; Vö and Wolfe, 2013) or computation of global summary statistics of the raw sensory image. For example, the average size of a set of objects can be computed even when the display changes continuously (Albrecht and Scholl, 2010). Thus certain global properties of the sensory image can be read-out during fast sampling, serving as a layout for visual analysis (“the gist”; Rensink, 2000).

Without translation into a perceptually invariant and stable representation, however, information about individual elements within the sensory image is easily over-written by subsequent input (Wundt, 1899; Sperling, 1960, 1963; Breitmeyer and Ögmen, 2006). Hence, selectivity in spatio-temporal processing does not arise from a failure to sample the sensory image, but reflects subsequent structuring and stabilization of individual perceptual elements (Wutz and Melcher, 2013). Accordingly, individuation (but not basic bottom-up detection) of multiple perceptual elements evokes a set-size specific modulation of the N2pc EEG-component that is commonly assumed to index attentional selection (Mazza and Caramazza, 2011).

This coupling of the spatio-temporal coordinates of the sensory signal to a specific object representation enables identity integration between rapidly sampled content and slowly computed structure. Consequently, visual memory for an entire

array of individual elements that requires binding of identity to location remains low throughout the integration bandwidth (Figure 3 right panel; Wutz and Melcher, 2013). Consistent with the idea that individuation precedes identification, visual working memory performance rises gradually to asymptote under the influence of *backward masking* (Gegenfurtner and Sperling, 1993; Vogel et al., 2006). In contrast to the forward masking paradigm described above, backward masking is thought to reflect a disruption of processing after feedforward perceptual analysis is already completed (Scheerer, 1973a,b) but before consolidating information into visual working memory. This distinction in object processing stages between individuation and identification of objects is further fostered by task-specific activation patterns in parietal areas (Xu and Chun, 2006; Xu, 2007). Within such a “neural object-file” framework multiple visual objects are selected and individuated in an initial feedforward operation involving the inferior intra-parietal sulcus (IPS) and only subsequently identified and maintained in visual working memory (within superior IPS; Xu and Chun, 2009).

Step-wise, feedforward individuation of only a limited number of objects within a temporal buffer limits the temporal dynamics of vision. In real-time processing, however, delayed feedback systems (like the visual system; Felleman and Van Essen, 1991) exhibit asymptotic unstable behavior when confronted with signals with different latencies that have to be combined (Sandberg, 1963). Temporal buffering provides a solution to this problem by synchronizing convergent input streams. In this way, feedback processing, like identification, operates upon the outcome of the whole temporal buffer to ensure spatio-temporally coherent vision. This provides a possible solution to the problem of how to carve continuous sensory input into coherent objects, despite the presence of feedback loops. Temporal windows allow for the read-out of individual elements but also the integration of sensory flux into a dynamic stream of visual impressions (Ögmen, 1993; Wutz and Melcher, 2013).

NEURAL MECHANISMS: ALPHA PHASE SYNCHRONIZES INDIVIDUATION AND INTEGRATION

It has been suggested that implementation of integration windows within perceptual processing might involve brain oscillations (Varela et al., 1981; Pöppel and Logothetis, 1986; Dehaene, 1993). Numerous studies have shown that the temporal relation between sensory stimuli and neural oscillations can alter the perceptual outcome. For example, psychophysical threshold estimates have been shown to vary with the phase of ongoing oscillatory activity (Busch et al., 2009; Mathewson et al., 2009) and recent evidence suggests even a causal link between the two (Neuling et al., 2012). Moreover, perceived simultaneity and sequentiality of apparent motion percepts depend on the phase of the occipital alpha rhythm (Varela et al., 1981; Gho and Varela, 1988). Such periodic fluctuations have previously been described as rhythmic background sampling of the sensory surrounding (VanRullen et al., 2007; Busch and VanRullen, 2010). These results suggest that oscillations impose a “perceptual frame” on feedforward processing such that integration and individuation of sensory signals depends on its periodic phase.

One key characteristic of brain oscillations is robust phase synchronization to transient input (Buzsáki and Draguhn, 2004). In addition to effects of ongoing oscillations prior to stimulus onset, stimulus evoked synchronization patterns might reveal how phase information influences perceptual integration. In this view, external stimulation results in a “reset” of functionally relevant oscillatory patterns such that their phase synchronization is locked to stimulus onset. Resets might in particular occur in response to transient sensory change, like saccadic eye movements or real-world transitions (i.e., stimulus onset). In fact, evoked responses to successfully detected and entirely missed stimuli differ extensively (Busch et al., 2009) and alpha phase-locking accounts for individual differences in a rapid visual discrimination task (Hanslmayr et al., 2005). Likewise, reset cyclic patterns in visual task performance have been reported in response to sudden flash events (Landau and Fries, 2012) or auditory sounds (Romei et al., 2012). Moreover, electro-cortical stimulation studies demonstrated a causal link between phase resets and perceptual performance by showing that repetitive transcranial magnetic stimulation (TMS) at 10 Hz synchronizes natural alpha oscillations (Thut et al., 2011) and biases spatial selection in visual tasks (Romei et al., 2010, 2011).

In support for the idea of a link between phase synchronization and temporal integration windows, we have demonstrated that the perceptual outcome of integration masking depends on short-lived alpha phase synchrony over parietal sensors measured with MEG (Wutz et al., 2014). We contrasted trials in which observers accurately individuated low set-sizes of target items (up to 3) from masking persistence with trials in which mask and target elements integrated in time and individuation failed (see **Figure 2**). Correct individuation is accompanied by a reset selectively synchronizing alpha oscillations within a temporal window of around 100 ms (so for approximately one alpha cycle) shortly after onset of the masking sequence (**Figure 4**).

It is important to note that alpha phase synchrony reset by the masking sequence only distinguishes between individuation and integration of visual transients on rapid time scales (<100 ms; short SOA trials). Segregating sensory changes exceeding this critical time frame (long SOA trials) instead depends on slower beta power modulations prior to stimulus onset (Wutz et al., 2014). The time course of the alpha phase synchrony reset (≈ 100 ms; \approx one alpha cycle) is consistent with the perceptual effects of integration masking (Enns and Di Lollo, 2000; Breitmeyer and Öğmen, 2006; Wutz et al., 2012). These results suggest that short-lived alpha synchronization is in particular key for perceptual processing of fast sensory changes. Precise phase coding within this integration cycle (through e.g., eigenfrequency damped oscillations; Buzsáki and Draguhn, 2004) in response to sensory transitions might balance individuation of perceptual elements and integration of sensory flux to guarantee spatio-temporal coherent perceptual outcomes.

IMPLICATIONS AND FUTURE DIRECTIONS

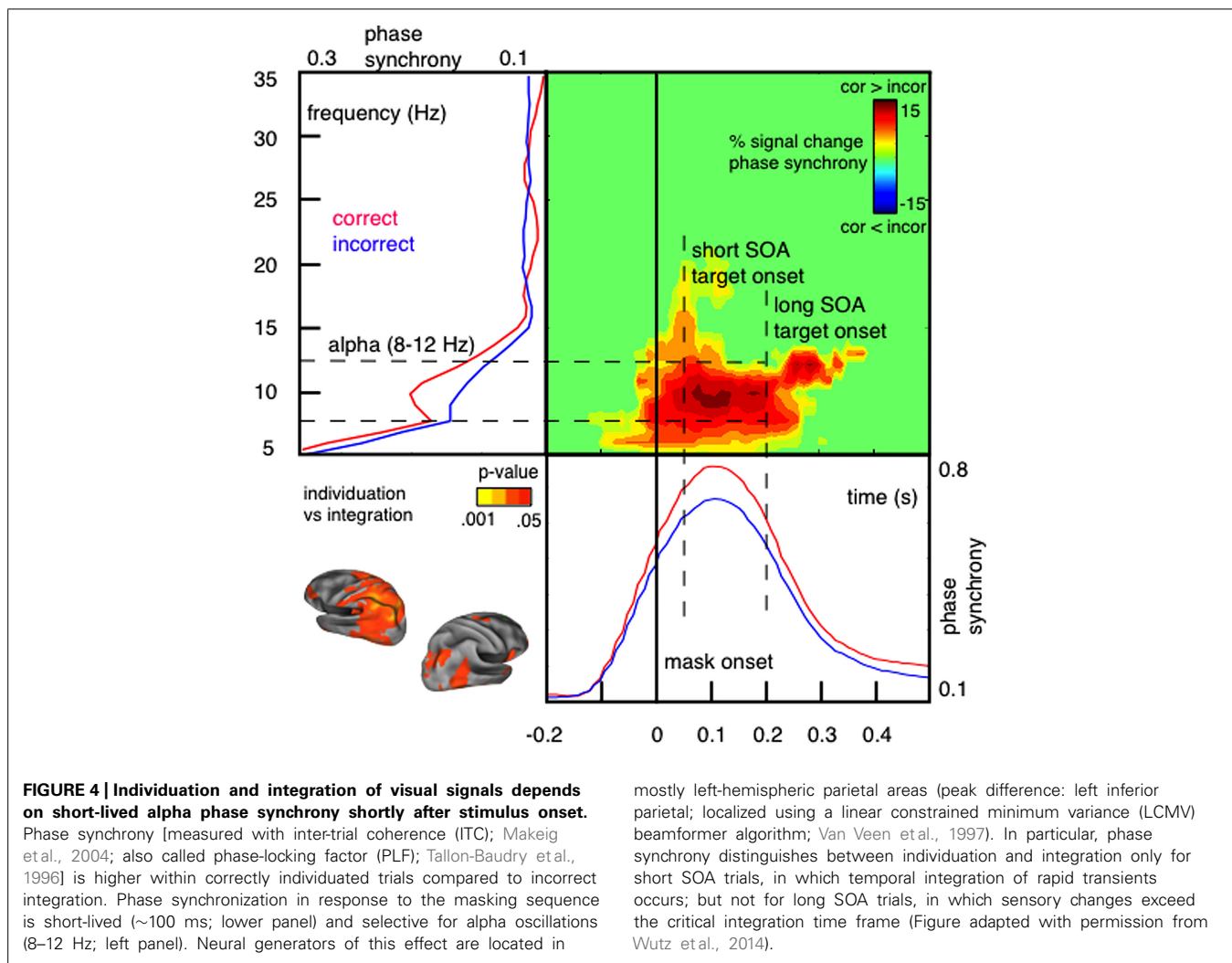
THE MAGIC WINDOW: TIME AND CAPACITY LIMITS

Following Miller's (1956) seminal paper discussing the “magic number” of 5–7 objects, the nature of these capacity limits has been a matter of intensive debate. Although a review of this extensive literature is beyond the scope here (for review see Cowan, 2000), it is important to note that the role of *time* in capacity limits has been almost neglected in any of the major theories. As described above, limits in the capacity of object individuation can be explained by the limited duration of visual persistence and the cycle of feedforward and feedback processing: in other words, temporal, rather than spatial, bandwidth. One advantage of a temporal window explanation of capacity is that capacity limits emerge naturally out of the rate of object individuation within this window of persistence, without the need to posit any *ad hoc* mechanisms.

In terms of neural implementations, the MEG evidence reported here, as well as related neuroimaging studies (Todd and Marois, 2004; Knops et al., 2014) suggest that neurons in posterior parietal cortex (PPC) may be involved in the individuation of objects. Specifically, capacity limits may reflect the spatial and temporal nature of attentional priority (saliency) maps in PPC (Melcher and Piazza, 2011; Franconeri et al., 2013; Knops et al., 2014). Unlike the priority maps in early visual areas (Zhang et al., 2012), attention priority maps in parietal cortex are thought to integrate bottom-up and top-down saliency estimates for objects over time (Bogler et al., 2011), allowing for object information to be accumulated and maintained (Mirpour et al., 2009; van Koningsbruggen et al., 2010). The results reviewed here emphasize the temporal aspects of the individuation process in determining attentional priority and capacity.

INTERACTION WITH NATURAL VISION: RETINOTOPY AND VISUAL STABILITY

A fundamental challenge for the perception of coherent spatiotemporal objects is that objects move and so do our sensory receptors. In retinotopic space, object motion would be expected to create smear within the image plane along the motion path and blurry object representations (the so-called “moving ghost problem”; Öğmen and Herzog, 2010). Whereas motion smear can be reduced



by mechanisms similar to meta-contrast masking (Chen et al., 1995; Purushothaman et al., 1998), the read-out of moving objects would still result in fuzzy perceptual form computations. In order to avoid such “ghost-like” appearances the visual system might rely on motion segmentation when computing non-retinotopic representations (Ögmen and Herzog, 2010). This development of non-retinotopic representations necessitates integration over a temporal interval on the order of 100–150 ms (Ögmen et al., 2006; Ögmen and Herzog, 2010; Otto et al., 2010). Temporal integration of feature persistence over this temporal interval has also been implicated in the use of spatial cues for motion direction in natural images (a “motion streak”; Geisler, 1999). In general, perceptual mechanisms responsible for motion and clear, un-smearred objects share functional characteristics and are capable of analyzing form and motion concurrently (Ramachandran et al., 1974; Burr, 1980; Burr et al., 1986), fostering the close link between object form and motion perception, and temporal integration over an interval of ca. 100 ms of image persistence.

The temporal window of individuation reviewed here might serve as a buffer to translate fast retinotopic representations into stable, but slower non-retinotopic (including spatiotopic,

frame-based or object-based: Melcher, 2008; Lin and He, 2012) representations that are of particular importance when objects move or change quickly. Perceiving an object as an individual within a crowded scene requires the observer to represent an object’s spatiotemporal coordinates distinct from the background and from other individuals in the image. Such a structured perceptual representation contains information about sensory input that is invariant to its absolute retinotopic coordinates and gives rise to non-retinotopic form. Static input remains long enough on a well-defined location in the image, so that its associated features can be firmly attached to this location and capacity limits arise as a function of individuated locations within the image persistence. In case of fast changes in the image plane, however, only a subset of locations can be selected and individuated into non-retinotopic representations. In this way the need for higher temporal resolution balances with limits in the computation of stable non-retinotopic individuals in each single instance. Such an equilibrium might be essential in mediating between stable object and dynamic motion perception with minimal motion smear in the image plane.

Likewise, eye and head movements create a change in the retinal input and thus, potentially, a source of confusion when integrating information over time (for a discussion of the similarity between the effects of object and eye motion, see: Ağaoğlu et al., 2012). Typically, stable eye fixation periods last on the order of 150–300 ms in reading and natural viewing tasks (for review see Rayner, 1998). The external world seems stable despite these dramatic spatio-temporal disruptions in sensory information, perhaps relying on non-retinotopic object representations (Melcher and Colby, 2008; Burr and Morrone, 2011; Melcher, 2011).

We speculate that the visual system might deal with the problems of object and self-motion in a similar way, involving at least two stages of processing (see also Otto et al., 2010). At the first stage, relatively brief visual integration windows, such as those in visual masking studied here, combine information in a retinotopic manner over a time course that allows for feedforward processing. This time window is used to successively individuate spatio-temporal elements and hence stabilize sensory input. It is not coincidental, then, that the most brief eye fixations found in reading and natural viewing and intermediate-level visual integration windows would be of similar minimum durations since the goal of each new fixation is to sample part of the visual scene in order to individuate the most relevant objects. It would not make sense to move the eye before all of the information is sampled up to the level of object individuation, or to “mis-align” this integration window so that the saccade occurs right in the middle (integrating information during individuation from two different spatial locations). Moreover, the complete cycle of feedforward and feedback processing would tend to exhaust all of the useful information available from the fovea, making long fixation durations inefficient unless the information of the retina was dynamic or difficult to resolve.

At the second stage, however, information about the same object should be combined over time, over a longer time window and a non-retinotopic spatial reference frame. Accurate perception of object motion relies on non-retinotopic form computation (Öğmen et al., 2006). Likewise, there are a growing number of examples of spatiotopic perceptual effects across eye movements (for review, see Melcher and Colby, 2008; Burr and Morrone, 2011; Melcher, 2011) and there is converging evidence that this involves time scales of several hundred milliseconds (Zimmermann et al., 2013a,b). Overall, these studies suggest that there are both relatively brief, retinotopic integration windows and longer, spatiotopic windows.

One clear hypothesis from this idea is that retinotopic temporal integration windows should be reset by saccades and aligned to new eye fixations. As described above, it would be problematic if the basic object individuation process combined information from different spatial locations due to a saccadic eye movement changing retinal position during the integration window. Some evidence for a reset in the window of object individuation comes from studies of masking. Visual persistence, as measured by the missing dot task (Di Lollo, 1980), does not continue across saccades (Bridgeman and Mayer, 1983; Jonides et al., 1983) and masking can be disrupted by the intention to make a saccade (De Pisapia et al., 2010). On the other hand, the much longer temporal integration windows involved in apparent motion, over

100s of milliseconds, do not seem to be disrupted by saccades (Fracasso et al., 2010; Melcher and Fracasso, 2012). Further studies are needed to precisely define the relationship between fixation onset and the temporal windows of object individuation. The exact timing of temporal integration windows relative to eye movements might play a critical role for the impression of visual stability on rapid time scales. Such fast, feedforward computations might still involve retinotopic coordinates and therefore require saccadic remapping. However, much of the impression of visual stability might involve longer time windows that are not entirely retinotopic and thus do not require saccadic remapping.

NEURAL SYNCHRONIZATION COORDINATES FEEDFORWARD AND FEEDBACK OBJECT PROCESSING

We have reported that the short-lived alpha phase synchronization reset by stimulus onset predicts perceptual performance on an integration task. Time- and frequency characteristics of this effect (100 ms at 10 Hz) point to an alpha phase reset involved in feedforward individuation of objects. This is in line with classical findings identifying partly reset alpha oscillations in event-related potential (ERP) signatures (especially in the N1 component, Makeig et al., 2002). The functional role of alpha oscillations in perception and cognition are debated. Recent advances, however, have associated alpha phase information with the selection and recognition of object representations (for review see Palva and Palva, 2007). In support of this view, but in contrast to spatial or numerical limits in object segmentation, we propose an account based on temporal bandwidth in which phase-locking couples external signals to alpha integration cycles. Processing limits might then arise as a result of feedforward encoding within one synchronized cycle. A temporal window model based on neural synchronization patterns has several interesting functional characteristics that could coordinate feedforward and feedback object processing.

Synchronous coupling to oscillatory dynamics can structure processing into cyclic time windows for coherent integration of convergent inputs that arrive with different latencies (Buzsáki and Draguhn, 2004). In this way alpha oscillatory cycles might reflect temporal reference frames as elementary building blocks in feedforward processing. In fact, alpha cycles have been previously discussed as segmenting input into discrete snapshots of ~100 ms (VanRullen and Koch, 2003). In line with this view illusory motion reversals in the continuous wagon wheel illusion are most prominent at wheel-motion frequencies around 10 Hz and are correlated with alpha band amplitude in the ongoing EEG trace (VanRullen et al., 2005, 2006).

VanRullen and Koch (2003) also suggested a possible way to read-out object information within such a temporal window that might involve coupled networks of nested oscillatory sub-cycles (coding individual content) within slow-wave carriers (defining the temporal reference frame). Such neural networks are capable of representing individual information by means of frequency-division multiplexing (Lisman and Idiart, 1995). Especially, phase-amplitude coupling between α - and γ -frequency bands could prioritize the selection of multiple visual objects (Jensen et al., 2012, 2014). In this view the selection of individual

items might be regulated via timed release of inhibition within one alpha cycle (Van Rullen and Thorpe, 2001; Klimesch et al., 2007; Jensen et al., 2012). Indeed, neural network dynamics of individuation can be modeled based on inhibition between competing items in a saliency map (Knops et al., 2014). Whereas multiplex coding is a well-established principle of neural function (O'Keefe and Recce, 1993; Kayser et al., 2009; Siegel et al., 2009), future work is needed to determine its functional significance for human visual cognition. Our results support the view that oscillatory synchronization might represent multiplexed phase coding and suggest that object capacity limits can arise, not only by the read-out speed of individual elements, but also from the bandwidth of the carrier function.

Importantly, integration windows can help to coordinate visual processing dynamically, because phase synchronization occurs in response to internal or external changes in input (via phase resetting; Buzsáki and Draguhn, 2004; Buzsáki, 2006; for review see Thut et al., 2012). In this way, brief phase synchronization might contribute to the rapid coordination of distributed neuronal populations (like the retinotopically organized areas along the visual hierarchy; von der Malsburg, 1981; von der Malsburg and Schneider, 1986; Singer and Gray, 1995; Fries, 2005). This might be important in order to cope with the combinatorial complexity of crowded visual scenes that contain individual elements that can consist of a nearly infinite number of feature combinations and can appear at any given moment in time or spatial location. This flexibility in combining arbitrarily complex features over space and time would seem to require neural network communication. In line with this view, phase synchronization has been hypothesized to sub-serve cross-modal integration or feature binding and to gate the information flow between local neuronal ensembles (Singer, 1999; Salinas and Sejnowski, 2001). Consistent with this idea, phase synchrony between distributed processing sites has been demonstrated to predispose visual perception (Hipp et al., 2011), route selective attention (Siegel et al., 2008; for review see Womelsdorf and Fries, 2007), predict individual working memory capacity (Palva et al., 2010) and reflect higher-level temporal processing limits (Gross et al., 2004).

Our results reveal wide spread synchronization patterns in parietal cortices locked to stimulus onset already at the level of object segmentation. We argue that vision makes use of phase synchronization as a temporal reference frame in which distributed processing can be orchestrated and aligned to input transitions. Reset synchronization patterns might therefore coordinate feed-forward and feedback mechanisms involved in encoding complex and dynamic visual scenes with nearly real-time speeds. In this framework, temporal windows might reflect a neural strategy for coherent perception of objects in space and time.

CONCLUSION

As described above, there is accumulating psychophysical and electrophysiological evidence for an intermediate-level temporal window involved in the individuation of a small number of relevant objects in a scene. Individuation capacity increases in steps within the lifetime of visual persistence of the stimulus, suggesting that visual capacity limitations arise as a result of the narrow

temporal window of sensory persistence. In contrast to the main theories based on spatial slots or finite spatial resources, these findings suggest that *time* is the critical factor in the emergence of capacity limits. In this way, capacity limits can be seen as a result of the need of the visual system to coordinate feedforward and feedback processes. The cycle of feedforward and feedback processing reflects a compromise between the competing needs of a perceptual system to integrate information over extended periods of time (to get a better estimate of stable object and event properties) and sensitivity to changes in the environment.

ACKNOWLEDGMENT

This research was supported by a European Research Council Grant (agreement no. 313658).

REFERENCES

- Ağaoğlu, M. N., Herzog, M. H., and Öğmen, H. (2012). Non-retinotopic feature processing in the absence of retinotopic spatial layout and the construction of perceptual space from motion. *Vision Res.* 71, 10–17. doi: 10.1016/j.visres.2012.08.009
- Albrecht, A. R., and Scholl, B. J. (2010). Perceptually averaging in a continuous visual world: extracting statistical summary representations over time. *Psychol. Sci.* 21, 560–567. doi: 10.1177/0956797610363543
- Baylis, G. C., and Driver, J. (1993). Visual attention and objects: evidence for hierarchical coding of location. *J. Exp. Psychol. Hum. Percept. Perform.* 19:451. doi: 10.1037/0096-1523.19.3.451
- Bertozzi, M., Broggi, A., and Fascioli, A. (2000). Vision-based intelligent vehicles: state of the art and perspectives. *Rob. Auton. Syst.* 32, 1–16. doi: 10.1016/S0921-8890(99)00125-6
- Blalock, L. D. (2013). Mask similarity impacts short-term consolidation in visual working memory. *Psychon. Bull. Rev.* 20, 1290–1295. doi: 10.3758/s13423-013-0461-9
- Bogler, C., Bode, S., and Haynes, J. D. (2011). Decoding successive computational stages of saliency processing. *Curr. Biol.* 21, 1667–1671. doi: 10.1016/j.cub.2011.08.039
- Breitmeyer, B. G., and Öğmen, H. (2006). *Visual Masking: Time Slices Through Conscious and Unconscious Vision*. New York, NY: Oxford University Press. doi: 10.1093/acprof:oso/9780198530671.001.0001
- Bridgeman, B., and Mayer, M. (1983). Failure to integrate visual information from successive fixations. *Bull. Psychon. Soc.* 21, 285–286. doi: 10.3758/BF03334711
- Burr, D. (1980). Motion smear. *Nature* 284, 164–165. doi: 10.1038/284164a0
- Burr, D. C., and Morrone, M. C. (2011). Spatiotopic coding and remapping in humans. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 366, 504–515. doi: 10.1098/rstb.2010.0244
- Burr, D. C., Ross, J., and Morrone, M. C. (1986). Seeing objects in motion. *Proc. R. Soc. Lond. B Biol. Sci.* 227, 249–265. doi: 10.1098/rspb.1986.0022
- Busch, N. A., Dubois, J., and Van Rullen, R. (2009). The phase of ongoing EEG oscillations predicts visual perception. *J. Neurosci.* 29, 7869–7876. doi: 10.1523/JNEUROSCI.0113-09.2009
- Busch, N. A., and VanRullen, R. (2010). Spontaneous EEG oscillations reveal periodic sampling of visual attention. *Proc. Natl. Acad. Sci. U.S.A.* 107, 16048–16053. doi: 10.1073/pnas.1004801107
- Buzsáki, G. (2006). *Rhythms of the Brain*. New York, NY: Oxford University Press. doi: 10.1093/acprof:oso/9780195301069.001.0001
- Buzsáki, G., and Draguhn, A. (2004). Neuronal oscillations in cortical networks. *Science* 304, 1926–1929. doi: 10.1126/science.1099745
- Chen, S., Bedell, H. E., and Öğmen, H. (1995). A target in real motion appears blurred in the absence of other proximal moving targets. *Vision Res.* 35, 2315–2328. doi: 10.1016/0042-6989(94)00308-9
- Coltheart, M. (1980). Iconic memory and visible persistence. *Percept. Psychophys.* 27, 183–228. doi: 10.3758/BF03204258
- Cowan, N. (2000). The magical number 4 in short-term memory: a reconsideration of mental storage capacity. *Behav. Brain Sci.* 24, 87–185. doi: 10.1017/S0140525X01003922

- Cowan, N., Elliott, E. M., Scott Saults, J., Morey, C. C., Mattox, S., Hismjatullina, A., et al. (2005). On the capacity of attention: its estimation and its role in working memory and cognitive aptitudes. *Cogn. Psychol.* 51, 42–100. doi: 10.1016/j.cogpsych.2004.12.001
- Crozier, W. J., and Wolf, E. (1941). Theory and measurement of visual mechanisms IV. Critical intensities for visual flicker, monocular and binocular. *J. Gen. Physiol.* 24, 505–534. doi: 10.1085/jgp.24.4.505
- Dehaene, S. (1993). Temporal oscillations in human perception. *Psychol. Sci.* 4, 264–270. doi: 10.1111/j.1467-9280.1993.tb00273.x
- Dempere-Marco, L., Melcher, D., and Deco, G. (2012). Effective visual working memory capacity: an emergent effect from the neural dynamics in an attractor network. *PLoS ONE* 7:e42719. doi: 10.1371/journal.pone.0042719
- De Pisapia, N., Kaunitz, L., and Melcher, D. (2010). Backward masking and unmasking across saccadic eye movements. *Curr. Biol.* 20, 613–617. doi: 10.1016/j.cub.2010.01.056
- Di Lollo, V. (1980). Temporal integration in visual memory. *J. Exp. Psychol. Gen.* 109, 75–97. doi: 10.1037/0096-3445.109.1.75
- Di Lollo, V., Enns, J. T., and Rensink, R. A. (2000). Competition for consciousness among visual events: the psychophysics of reentrant visual processes. *J. Exp. Psychol. Gen.* 129, 481–507. doi: 10.1037/0096-3445.129.4.481
- Di Lollo, V., and Wilson, A. E. (1978). Iconic persistence and perceptual moment as determinants of temporal integration in vision. *Vision Res.* 18, 1607–1610. doi: 10.1016/0042-6989(78)90251-1
- Egeth, H. E., Leonard, C. J., and Palomares, M. (2008). The role of attention in subitizing: is the magical number 1? *Vis. Cogn.* 16, 463–473. doi: 10.1080/13506280801937939
- Engle, R. W., Tuholski, S. W., Laughlin, J. E., and Conway, A. R. (1999). Working memory, short-term memory, and general fluid intelligence: a latent-variable approach. *J. Exp. Psychol. Gen.* 128:309. doi: 10.1037/0096-3445.128.3.309
- Enns, J. T., and Di Lollo, V. (2000). What's new in visual masking? *Trends Cogn. Sci.* 4, 345–352. doi: 10.1016/S1364-6613(00)01520-5
- Felleman, D. J., and Van Essen, D. C. (1991). Distributed hierarchical processing in the primate cerebral cortex. *Cereb. Cortex* 1, 1–47. doi: 10.1093/cercor/1.1.1
- Field, D. J., Hayes, A., and Hess, R. F. (1993). Contour integration by the human visual system: evidence for a local “association field.” *Vision Res.* 33, 173–193. doi: 10.1016/0042-6989(93)90156-Q
- Fracasso, A., Caramazza, A., and Melcher, D. (2010). Continuous perception of motion and shape across saccadic eye movements. *J. Vis.* 10:14. doi: 10.1167/10.13.14
- Franconeri, S. L., Alvarez, G. A., and Cavanagh, P. (2013). Flexible cognitive resources: competitive content maps for attention and memory. *Trends Cogn. Sci. (Regul. Ed.)* 17, 134–141. doi: 10.1016/j.tics.2013.01.010
- Fries, P. (2005). A mechanism for cognitive dynamics: neuronal communication through neuronal coherence. *Trends Cogn. Sci. (Regul. Ed.)* 9, 474–480. doi: 10.1016/j.tics.2005.08.011
- Fukuda, K., Awh, E., and Vogel, E. K. (2010a). Discrete capacity limits in visual working memory. *Curr. Opin. Neurobiol.* 20, 177–182. doi: 10.1016/j.conb.2010.03.005
- Fukuda, K., Vogel, E., Mayr, U., and Awh, E. (2010b). Quantity, not quality: the relationship between fluid intelligence and working memory capacity. *Psychon. Bull. Rev.* 17, 673–679. doi: 10.3758/17.5.673
- Gegenfurtner, K. R., and Sperling, G. (1993). Information transfer in iconic memory experiments. *J. Exp. Psychol. Hum. Percept. Perform.* 19, 845–866. doi: 10.1037/0096-1523.19.4.845
- Geisler, W. S. (1999). Motion streaks provide a spatial code for motion direction. *Nature* 400, 65–69. doi: 10.1038/21886
- Gho, M., and Varela, F. J. (1988). A quantitative assessment of the dependency of the visual temporal frame upon the cortical rhythm. *J. Physiol. (Paris)* 83:95.
- Gross, J., Schmitz, F., Schnitzler, I., Kessler, K., Shapiro, K., Hommel, B., et al. (2004). Modulation of long-range neural synchrony reflects temporal limitations of visual attention in humans. *Proc. Natl. Acad. Sci. U.S.A.* 101, 13050–13055. doi: 10.1073/pnas.0404944101
- Haber, R. N., and Standing, L. G. (1970). Direct estimates of the apparent duration of a flash. *Can. J. Psychol.* 24:216. doi: 10.1037/h0082858
- Hanslmayr, S., Klimesch, W., Sauseng, P., Gruber, W., Doppelmayr, M., Freunberger, R., et al. (2005). Visual discrimination performance is related to decreased alpha amplitude but increased phase locking. *Neurosci. Lett.* 375, 64–68. doi: 10.1016/j.neulet.2004.10.092
- Hasson, U., Yang, E., Vallines, I., Heeger, D. J., and Rubin, N. (2008). A hierarchy of temporal receptive windows in human cortex. *J. Neurosci.* 28, 2539–2550. doi: 10.1523/JNEUROSCI.5487-07.2008
- Hipp, J. F., Engel, A. K., and Siegel, M. (2011). Oscillatory synchronization in large-scale cortical networks predicts perception. *Neuron* 69, 387–396. doi: 10.1016/j.neuron.2010.12.027
- Holcombe, A. O. (2009). Seeing slow and seeing fast: two limits on perception. *Trends Cogn. Sci. (Regul. Ed.)* 13, 216–221. doi: 10.1016/j.tics.2009.02.005
- Holcombe, A. O., and Cavanagh, P. (2001). Early binding of feature pairs for visual perception. *Nat. Neurosci.* 4, 127–128. doi: 10.1038/83945
- Holcombe, A. O., and Chen, W. Y. (2013). Splitting attention reduces temporal resolution from 7 Hz for tracking one object to < 3 Hz when tracking three. *J. Vis.* 13:12. doi: 10.1167/13.1.12
- Jensen, O., Bonnefond, M., and VanRullen, R. (2012). An oscillatory mechanism for prioritizing salient unattended stimuli. *Trends Cogn. Sci. (Regul. Ed.)* 16, 200–206. doi: 10.1016/j.tics.2012.03.002
- Jensen, O., Gips, B., Bergmann, T. O., and Bonnefond, M. (2014). Temporal coding organized by coupled alpha and gamma oscillations prioritize visual processing. *Trends Neurosci.* 37, 357–369. doi: 10.1016/j.tins.2014.04.001
- Jevons, W. S. (1871). The power of numerical discrimination. *Nature* 3, 363–372. doi: 10.1038/003281a0
- Jonides, J., Irwin, D., and Yantis, S. (1983). Failure to integration visual information from successive fixations. *Science* 222:188. doi: 10.1126/science.6623072
- Kahneman, D., Treisman, A., and Gibbs, B. J. (1992). The reviewing of object files: object-specific integration of information. *Cogn. Psychol.* 24, 174–219. doi: 10.1016/0010-0285(92)90007-O
- Karatekin, C., and Asarnow, R. F. (1998). Working memory in childhood-onset schizophrenia and attention-deficit/hyperactivity disorder. *Psychiatry Res.* 80, 165–176. doi: 10.1016/S0165-1781(98)00061-4
- Kaufman, E. L., Lord, M. W., Reese, T., and Volkman, J. (1949). The discrimination of visual number. *Am. J. Psychol.* 62, 496–525. doi: 10.2307/1418556
- Kayser, C., Montemurro, M. A., Logothetis, N. K., and Panzeri, S. (2009). Spike-phase coding boosts and stabilizes information carried by spatial and temporal spike patterns. *Neuron* 61, 597–608. doi: 10.1016/j.neuron.2009.01.008
- Kietzman, M. L., and Sutton, S. (1968). The interpretation of two-pulse measures of temporal resolution in vision. *Vision Res.* 8, 287–302. doi: 10.1016/0042-6989(68)90016-3
- Klimesch, W., Sauseng, P., and Hanslmayr, S. (2007). EEG alpha oscillations: the inhibition-timing hypothesis. *Brain Res. Rev.* 53, 63–88. doi: 10.1016/j.brainresrev.2006.06.003
- Knops, A., Piazza, M., Sengupta, R., Eger, E., and Melcher, D. (2014). A shared, flexible neural map architecture reflects capacity limits in both visual short term memory and enumeration. *J. Neurosci.* 34, 9857–9866. doi: 10.1523/JNEUROSCI.2758-13.2014
- Lamme, V. A. F., and Roelfsema, P. R. (2000). The distinct modes of vision offered by feedforward and recurrent processing. *Trends Neurosci.* 23, 571–579. doi: 10.1016/S0166-2236(00)01657-X
- Landau, A. N., and Fries, P. (2012). Attention samples stimuli rhythmically. *Curr. Biol.* 22, 1000–1004. doi: 10.1016/j.cub.2012.03.054
- Lee, E. Y., Cowan, N., Vogel, E. K., Rolan, T., Valle-Inclán, F., and Hackley, S. A. (2010). Visual working memory deficits in patients with Parkinson's disease are due to both reduced storage capacity and impaired ability to filter out irrelevant information. *Brain* 133, 2677–2689. doi: 10.1093/brain/awq197
- Lerner, Y., Honey, C. J., Silbert, L. J., and Hasson, U. (2011). Topographic mapping of a hierarchy of temporal receptive windows using a narrated story. *J. Neurosci.* 31, 2906–2915. doi: 10.1523/JNEUROSCI.3684-10.2011
- Li, F. F., VanRullen, R., Koch, C., and Perona, P. (2002). Rapid natural scene categorization in the near absence of attention. *Proc. Natl. Acad. Sci. U.S.A.* 99, 9596–9601. doi: 10.1073/pnas.092277599

- Lin, Z., and He, S. (2012). Automatic frame-centered object representation and integration revealed by iconic memory, visual priming, and backward masking. *J. Vis.* 12:24. doi: 10.1167/12.11.24
- Lisman, J. E., and Idiart, M. A. (1995). Storage of 7 +/- 2 short-term memories in oscillatory subcycles. *Science* 267, 1512–1515. doi: 10.1126/science.7878473
- Luck, S. J., and Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature* 390, 279–281. doi: 10.1038/36846
- Makeig, S., Debener, S., Onton, J., and Delorme, A. (2004). Mining event-related brain dynamics. *Trends Cogn. Sci. (Regul. Ed.)* 8, 204–210. doi: 10.1126/science.1066168
- Makeig, S., Westerfield, M., Jung, T. P., Enghoff, S., Townsend, J., Courchesne, E., et al. (2002). Dynamic brain sources of visual evoked responses. *Science* 295, 690–694. doi: 10.1126/science.1066168
- Mathewson, K. E., Gratton, G., Fabiani, M., Beck, D. M., and Ro, T. (2009). To see or not to see: prestimulus alpha phase predicts visual awareness. *J. Neurosci.* 29, 2725–2732. doi: 10.1523/JNEUROSCI.3963-08.2009
- Masquelier, T., Albantakis, L., and Deco, G. (2011). The timing of vision – how neural processing links to different temporal dynamics. *Front. Psychol.* 2:151. doi: 10.3389/fpsyg.2011.00151
- Mazza, V., and Caramazza, A. (2011). Temporal brain dynamics of multiple object processing: the flexibility of individuation. *PLoS ONE* 6:e17453. doi: 10.1371/journal.pone.0017453
- Melcher, D. (2008). Dynamic, object-based remapping of visual features in trans-saccadic perception. *J. Vis.* 8, 2.1–2.17. doi: 10.1167/8.14.2
- Melcher, D. (2011). Visual stability. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 366, 468–475. doi: 10.1098/rstb.2010.0277
- Melcher, D., and Colby, C. L. (2008). Trans-saccadic perception. *Trends Cogn. Sci. (Regul. Ed.)* 12, 466–473. doi: 10.1016/j.tics.2008.09.003
- Melcher, D., and Fracasso, A. (2012). Remapping of the line motion illusion across eye movements. *Exp. Brain Res.* 218, 503–514. doi: 10.1007/s00221-012-3043-6
- Melcher, D., and Piazza, M. (2011). The role of attentional priority and saliency in determining capacity limits in enumeration and visual working memory. *PLoS ONE* 6:e29296. doi: 10.1371/journal.pone.0029296
- Miller, G. A. (1956). The magical number seven, plus or minus two: some limits on our capacity for processing information. *Psychol. Rev.* 63, 81–97. doi: 10.1037/h0043158
- Mirpour, K., Arcizet, F., Ong, W. S., and Bisley, J. W. (2009). Been there, seen that: a neural mechanism for performing efficient visual search. *J. Neurophysiol.* 102, 3481–3491. doi: 10.1152/jn.00688.2009
- Motoyoshi, I., and Nishida, S. Y. (2001). Temporal resolution of orientation-based texture segregation. *Vision Res.* 41, 2089–2105. doi: 10.1016/S0042-6989(01)00096-7
- Nastase, S., Iacovella, V., and Hasson, U. (2014). Uncertainty in visual and auditory series is coded by modality-general and modality-specific neural systems. *Hum. Brain Mapp.* 35, 1111–1128. doi: 10.1002/hbm.22238
- Neisser, U. (1967). *Cognitive Psychology*. New York, NY: Appleton-Century-Crofts.
- Neri, P., Morrone, M. C., and Burr D. C. (1998). Seeing biological motion. *Nature* 395, 894–896. doi: 10.1038/27661
- Neuling, T., Rach, S., Wagner, S., Wolters, C. H., and Herrmann, C. S. (2012). Good vibrations: oscillatory phase shapes perception. *Neuroimage* 63, 771–778. doi: 10.1016/j.neuroimage.2012.07.024
- Ögmen, H. (1993). A neural theory of retino-cortical dynamics. *Neural Netw.* 6, 245–273. doi: 10.1016/0893-6080(93)90020-W
- Ögmen, H., and Herzog, M. H. (2010). The geometry of visual perception: retinotopic and nonretinotopic representations in the human visual system. *Proc. IEEE Inst. Electr. Electron. Eng.* 98, 479–492. doi: 10.1109/JPROC.2009.2039028
- Ögmen, H., Otto, T., and Herzog, M. H. (2006). Perceptual grouping induces non-retinotopic feature attribution in human vision. *Vision Res.* 46, 3234–3242. doi: 10.1016/j.visres.2006.04.007
- O'Keefe, J., and Recce, M. L. (1993). Phase relationship between hippocampal place units and the EEG theta rhythm. *Hippocampus* 3, 317–330. doi: 10.1002/hipo.450030307
- Olivers, C. N., and Watson, D. G. (2008). Subitizing requires attention. *Vis. Cogn.* 16, 439–462. doi: 10.1080/13506280701825861
- Otto, T. U., Ögmen, H., and Herzog, M. H. (2010). Attention and non-retinotopic feature integration. *J. Vis.* 10:8. doi: 10.1167/10.12.8
- Palva, J. M., Monto, S., Kulashekhar, S., and Palva, S. (2010). Neuronal synchrony reveals working memory networks and predicts individual memory capacity. *Proc. Natl. Acad. Sci. U.S.A.* 107, 7580–7585. doi: 10.1073/pnas.0913113107
- Palva, S., and Palva, J. M. (2007). New vistas for α -frequency band oscillations. *Trends Neurosci.* 30, 150–158. doi: 10.1016/j.tins.2007.02.001
- Phillips, W. A., and Baddeley, A. D. (1971). Reaction time and short-term visual memory. *Psychon. Sci.* 22, 73–74. doi: 10.3758/BF03332500
- Piazza, M., Fumarola, A., Chinello, A., and Melcher, D. (2011). Subitizing reflects visuo-spatial object individuation capacity. *Cognition* 121, 147–153. doi: 10.1016/j.cognition.2011.05.007
- Pöppel, E. (1997). A hierarchical model of temporal perception. *Trends Cogn. Sci. (Regul. Ed.)* 1, 56–61. doi: 10.1016/S1364-6613(97)01008-5
- Pöppel, E. (2009). Pre-semantically defined temporal windows for cognitive processing. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 364, 1887–1896. doi: 10.1098/rstb.2009.0015
- Pöppel, E., and Logothetis, N. (1986). Neuronal oscillations in the human brain. *Naturwissenschaften* 73, 267–268. doi: 10.1007/BF0036778
- Purushothaman, G., Ögmen, H., Chen, S., and Bedell, H. E. (1998). Motion deblurring in a neural network model of retino-cortical dynamics. *Vision Res.* 38, 1827–1842. doi: 10.1016/S0042-6989(97)00350-7
- Pylyshyn, Z. W. (1989). The role of location indexes in spatial perception: a sketch of the FINST spatial index model. *Cognition* 32, 65–97. doi: 10.1016/0010-0277(89)90014-0
- Pylyshyn, Z. W. (2001). Visual indexes, preconceptual objects, and situated vision. *Cognition* 80, 127–158. doi: 10.1016/S0010-0277(00)00156-6
- Railo, H., Koivisto, M., Revonsuo, A., and Hannula, M. M. (2008). The role of attention in subitizing. *Cognition* 107, 82–104. doi: 10.1016/j.cognition.2007.08.004
- Ramachandran, V. S., Rao, V. M., and Vidyasagar, T. R. (1974). Sharpness constancy during movement perception – short note. *Perception* 3, 97–98. doi: 10.1068/p030097
- Rayner K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychol. Bull.* 124, 372–422. doi: 10.1037/0033-2909.124.3.372
- Rensink, R. A. (2000). The dynamic representation of scenes. *Vis. Cogn.* 7, 17–42. doi: 10.1080/135062800394667
- Roelfsema, P. R., Lamme, V. A. E., and Spekreijse, H. (2000). The implementation of visual routines. *Vision Res.* 40, 1385–1411. doi: 10.1016/S0042-6989(00)00004-3
- Romei, V., Driver, J., Schyns, P. G., and Thut, G. (2011). Rhythmic TMS over parietal cortex links distinct brain frequencies to global versus local visual processing. *Curr. Biol.* 21, 334–337. doi: 10.1016/j.cub.2011.01.035
- Romei, V., Gross, J., and Thut, G. (2010). On the role of prestimulus alpha rhythms over occipito-parietal areas in visual input regulation: correlation or causation? *J. Neurosci.* 30, 8692–8697. doi: 10.1523/JNEUROSCI.0160-10.2010
- Romei, V., Gross, J., and Thut, G. (2012). Sounds reset rhythms of visual cortex and corresponding human visual perception. *Curr. Biol.* 22, 807–813. doi: 10.1016/j.cub.2012.03.025
- Rüter, J., Marclille, N., Sprekeler, H., Gerstner, W., and Herzog, M. H. (2012). Paradoxical evidence integration in rapid decision processes. *PLoS Comput. Biol.* 8:e1002382. doi: 10.1371/journal.pcbi.1002382
- Salinas, E., and Sejnowski, T. J. (2001). Correlated neuronal activity and the flow of neural information. *Nat. Rev. Neurosci.* 2, 539–550. doi: 10.1038/35086012
- Sandberg, L. W. (1963). Signal distortion in nonlinear feedback systems. *Bell Syst. Tech. J.* 42, 2533–2550. doi: 10.1002/j.1538-7305.1963.tb00976.x
- Scharnowski, F., Rüter, J., Jolij, J., Hermens, F., Kammer, T., and Herzog, M. H. (2009). Long-lasting modulation of feature integration by transcranial magnetic stimulation. *J. Vis.* 9:1. doi: 10.1167/9.6.1
- Scheerer, E. (1973a). Integration, interruption and processing rate in visual backward masking. *Psychol. Forsch.* 36, 71–93. doi: 10.1007/BF00424655
- Scheerer, E. (1973b). Integration, interruption and processing rate in visual backward masking. *Psychol. Forsch.* 36, 95–115. doi: 10.1007/BF00424965
- Scholl, B. J., Pylyshyn, Z. W., and Feldman, J. (2001). What is a visual object? Evidence from target merging in multiple object tracking. *Cognition* 80, 159–177. doi: 10.1016/S0010-0277(00)00157-8
- Siegel, M., Donner, T. H., Oostenveld, R., Fries, P., and Engel, A. K. (2008). Neuronal synchronization along the dorsal visual pathway reflects the focus of spatial attention. *Neuron* 60, 709–719. doi: 10.1016/j.neuron.2008.09.010

- Siegel, M., Warden, M. R., and Miller, E. K. (2009). Phase-dependent neuronal coding of objects in short-term memory. *Proc. Natl. Acad. Sci. U.S.A.* 106, 21341–21346. doi: 10.1073/pnas.0908193106
- Singer, W. (1999). Neuronal synchrony: a versatile code for the definition of relations? *Neuron* 24, 49–65. doi: 10.1016/S0896-6273(00)80821-1
- Singer, W., and Gray, C. M. (1995). Visual feature integration and the temporal correlation hypothesis. *Annu. Rev. Neurosci.* 18, 555–586. doi: 10.1146/annurev.ne.18.030195.003011
- Spelke, E. (1988). “Where perceiving ends and thinking begins: the apprehension of objects in infancy,” in *Perceptual Development in Infancy*, ed. A. Yonas (Hillsdale, NJ: Erlbaum), 197–234.
- Sperling, G. (1960). The information available in brief visual presentations. *Psychol. Monogr. Gen. Appl.* 74, 1–29. doi: 10.1037/h0093759
- Sperling, G. (1963). A model for visual memory tasks. *Hum. Factors* 5, 19–31.
- Tallon-Baudry, C., Bertrand, O., Delpeuch, C., and Pernier, J. (1996). Stimulus specificity of phase-locked and non-phase-locked 40 Hz visual responses in human. *J. Neurosci.* 16, 4240–4249.
- Thorpe, S., Fize, D., and Marlot, C. (1996). Speed of processing in the human visual system. *Nature* 381, 520–522. doi: 10.1038/381520a0
- Thut, G., Miniussi, C., and Gross, J. (2012). The functional importance of rhythmic activity in the brain. *Curr. Biol.* 22, R658–R663. doi: 10.1016/j.cub.2012.06.061
- Thut, G., Veniero, D., Romei, V., Miniussi, C., Schyns, P., and Gross, J. (2011). Rhythmic TMS causes local entrainment of natural oscillatory signatures. *Curr. Biol.* 21, 1176–1185. doi: 10.1016/j.cub.2011.05.049
- Tipper, S. P., Brehaut, J. C., and Driver, J. (1990). Selection of moving and static objects for the control of spatially directed action. *J. Exp. Psychol. Hum. Percept. Perform.* 16:492. doi: 10.1037/0096-1523.16.3.492
- Todd, J. J., and Marois, R. (2004). Capacity limit of visual short-term memory in human posterior parietal cortex. *Nature* 428, 751–754. doi: 10.1038/nature02466
- Treisman, A., and Gelade, G. (1980). A feature integration theory of attention. *Cogn. Psychol.* 12, 97–136. doi: 10.1016/0010-0285(80)90005-5
- Trick, L. M., and Pylyshyn, Z. W. (1994). Why are small and large numbers enumerated differently? A limited-capacity preattentive stage in vision. *Psychol. Rev.* 101:80. doi: 10.1037/0033-295X.101.1.80
- Ullman, S. (1984). Visual routines. *Cognition* 18, 97–159. doi: 10.1016/0010-0277(84)90023-4
- van Koningsbruggen, M. G., Gabay, S., Sapir, A., Henik, A., and Rafal, R. D. (2010). Hemispheric asymmetry in the remapping and maintenance of visual saliency maps: a TMS study. *J. Cogn. Neurosci.* 22, 1730–1738. doi: 10.1162/jocn.2009.21356
- VanRullen, R., Carlson, T., and Cavanagh, P. (2007). The blinking spotlight of attention. *Proc. Natl. Acad. Sci. U.S.A.* 104, 19204–19209. doi: 10.1073/pnas.0707316104
- VanRullen, R., and Koch, C. (2003). Is perception discrete or continuous? *Trends Cogn. Sci. (Regul. Ed.)* 7, 207–213.
- VanRullen, R., Reddy, L., and Koch, C. (2005). Attention-driven discrete sampling of motion perception. *Proc. Natl. Acad. Sci. U.S.A.* 102, 5291–5296. doi: 10.1073/pnas.0409172102
- VanRullen, R., Reddy, L., and Koch, C. (2006). The continuous wagon wheel illusion is associated with changes in electroencephalogram power at ~13 Hz. *J. Neurosci.* 26, 502–507. doi: 10.1523/JNEUROSCI.4654-05.2006
- Van Rullen, R., and Thorpe, S. J. (2001). Rate coding versus temporal order coding: what the retinal ganglion cells tell the visual cortex. *Neural Comput.* 13, 1255–1283. doi: 10.1162/08997660152002852
- Van Veen, B. D., Van Drongelen, W., Yuchtman, M., and Suzuki, A. (1997). Localization of brain electrical activity via linearly constrained minimum variance spatial filtering. *IEEE Trans. Biomed. Eng.* 44, 867–880. doi: 10.1109/10.623056
- Varela, F. J., Toro, A., Roy John, E., and Schwartz, E. L. (1981). Perceptual framing and cortical alpha rhythm. *Neuropsychologia* 19, 675–686. doi: 10.1016/0028-3932(81)90005-1
- Vö, M., and Wolfe, J. M. (2012). Semantic and syntactic inconsistencies in scenes elicit differential ERP signatures. *J. Vis.* 12, 120–120. doi: 10.1167/12.9.120
- Vö, M. L. H., and Wolfe, J. M. (2013). Differential electrophysiological signatures of semantic and syntactic scene processing. *Psychol. Sci.* 24, 1816–1823. doi: 10.1177/0956797613476955
- Vogel, E. K., Woodman, G. F., and Luck, S. J. (2006). The time course of consolidation in visual working memory. *J. Exp. Psychol. Hum. Percept. Perform.* 32, 1436–1451. doi: 10.1037/0096-1523.32.6.1436
- von der Malsburg, C. (1981). *The Correlation Theory of Brain Function*. Internal Report 81-2. Göttingen: Max-Planck-Institute for Bio-physical Chemistry.
- von der Malsburg, C., and Schneider, W. (1986). A neural cocktail-party processor. *Biol. Cybern.* 54, 29–40. doi: 10.1007/BF00337113
- Womelsdorf, T., and Fries, P. (2007). The role of neuronal synchronization in selective attention. *Curr. Opin. Neurobiol.* 17, 154–160. doi: 10.1016/j.conb.2007.02.002
- Wundt, W. (1899). Zur Kritik tachistoskopischer Versuche. *Philos. Stud.* 15, 287–317.
- Wutz, A., Caramazza, A., and Melcher, D. (2012). Rapid enumeration within a fraction of a single glance: the role of visible persistence in object individuation capacity. *Vis. Cogn.* 20, 717–732. doi: 10.1080/13506285.2012.686460
- Wutz, A., and Melcher, D. (2013). Temporal buffering and visual capacity: the time course of object formation underlies capacity limits in visual cognition. *Attent. Percept. Psychophys.* 75, 921–933. doi: 10.3758/s13414-013-0454-9
- Wutz, A., Weisz, N., Braun, C., and Melcher, D. (2014). Temporal windows in visual processing: ‘Pre-stimulus brain state’ and ‘post-stimulus phase reset’ segregate visual transients on different temporal scales. *J. Neurosci.* 34, 1554–1565. doi: 10.1523/JNEUROSCI.3187-13.2014
- Xu, Y. (2007). The role of the superior intra-parietal sulcus in supporting visual short-term memory for multi-feature objects. *J. Neurosci.* 27, 11676–11686. doi: 10.1523/JNEUROSCI.3545-07.2007
- Xu, Y., and Chun, M. M. (2006). Dissociable neural mechanism supporting visual short-term memory for objects. *Nature* 440, 91–95. doi: 10.1038/nature04262
- Xu, Y., and Chun, M. M. (2009). Selecting and perceiving multiple visual objects. *Trends Cogn. Sci. (Regul. Ed.)* 13, 167–174. doi: 10.1016/j.tics.2009.01.008
- Zacks, J. M., and Magliano, J. (2011). “Film, narrative, and cognitive neuroscience,” in *Art and the Senses*, eds D. Melcher and F. Bacci (New York: Oxford University Press).
- Zhang, X., Zhao, L., Zhou, T., and Fang F. (2012). Neural activities in V1 create a bottom-up saliency map. *Neuron* 73, 183–192. doi: 10.1016/j.neuron.2011.10.035
- Zimmermann, E., Morrone, M., and Burr, D. (2013a). Spatial position information accumulates steadily over time. *J. Neurosci.* 33, 18396–18401. doi: 10.1523/JNEUROSCI.1864-13.2013
- Zimmermann, E., Morrone, M. C., Fink, G. R., and Burr, D. (2013b). Spatiotopic neural representations develop slowly across saccades. *Curr. Biol.* 23, R193–R194. doi: 10.1016/j.cub.2013.01.065

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.

Received: 30 May 2014; accepted: 10 August 2014; published online: 27 August 2014.
Citation: Wutz A and Melcher D (2014) The temporal window of individuation limits visual capacity. *Front. Psychol.* 5:952. doi: 10.3389/fpsyg.2014.00952
This article was submitted to Perception Science, a section of the journal Frontiers in Psychology.

Copyright © 2014 Wutz and Melcher. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Neural dynamics of feedforward and feedback processing in figure-ground segregation

Oliver W. Layton^{1,2}, Ennio Mingolla³ and Arash Yazdanbakhsh^{2*}

¹ The Perception and Action Lab, Department of Cognitive Science, Rensselaer Polytechnic Institute, Troy, NY, USA

² Vision Lab, Center for Computational Neuroscience and Neural Technology, Boston University, Boston, MA, USA

³ Computational Vision Laboratory, Department of Speech-Language Pathology and Audiology, Northeastern University, Boston, MA, USA

Edited by:

Haluk Ogmen, University of Houston, USA

Reviewed by:

Ko Sakai, University of Tsukuba, Japan

Saumil Surendra Patel, Baylor College of Medicine, USA

*Correspondence:

Arash Yazdanbakhsh, Vision Lab, Center for Computational Neuroscience and Neural Technology, Boston University, 677 Beacon Street, Boston, MA 02215, USA
e-mail: yazdan@bu.edu

Determining whether a region belongs to the interior or exterior of a shape (figure-ground segregation) is a core competency of the primate brain, yet the underlying mechanisms are not well understood. Many models assume that figure-ground segregation occurs by assembling progressively more complex representations through feedforward connections, with feedback playing only a modulatory role. We present a dynamical model of figure-ground segregation in the primate ventral stream wherein feedback plays a crucial role in disambiguating a figure's interior and exterior. We introduce a processing strategy whereby jitter in RF center locations and variation in RF sizes is exploited to enhance and suppress neural activity inside and outside of figures, respectively. Feedforward projections emanate from units that model cells in V4 known to respond to the curvature of boundary contours (*curved contour cells*), and feedback projections from units predicted to exist in IT that strategically group neurons with different RF sizes and RF center locations (*teardrop cells*). Neurons (*convex cells*) that preferentially respond when centered on a figure dynamically balance feedforward (bottom-up) information and feedback from higher visual areas. The activation is enhanced when an interior portion of a figure is in the RF via feedback from units that detect closure in the boundary contours of a figure. Our model produces maximal activity along the medial axis of well-known figures with and without concavities, and inside algorithmically generated shapes. Our results suggest that the dynamic balancing of feedforward signals with the specific feedback mechanisms proposed by the model is crucial for figure-ground segregation.

Keywords: V4, figure-ground segregation, medial axis transform, ventral stream, feedforward, feedback

INTRODUCTION

Figure-ground segregation refers to the process by which the visual system parses the complex array of luminance that appears on the retina into perceptually grouped foreground objects (figures) and backgrounds (ground). To distinguish between figures and their background, the visual system must perform two complementary processes—detecting defining borders and integrating parts into wholes. How the visual system represents visual figures with respect to these two processes, and the underlying mechanisms, are largely unknown. Emerging neurophysiological and psychophysical evidence suggests that the visual system may rely on multiple parallel “solutions” to segment the visual scene into figures and backgrounds.

One solution likely involves the border-ownership assignment of local edge representations. Figures necessarily share a visual border of an adjacent background region, and border-ownership refers to the association of the border with the figure rather than the ground. Populations of edge-sensitive neurons in primate visual areas V1, V2, and V4 have been shown to exhibit sensitivity to border-ownership: neurons respond with a higher firing rate when the figure to which the edge in the receptive field (RF) is attached appears on the preferred side (Zhou et al., 2000). If the

figure is on the other side of the edge, then the firing rate of the neuron will decrease and another neuron will exhibit enhanced activity. Neural models have suggested that border-ownership selectivity may arise through feedback from neurons with larger RFs in higher visual areas (Kelly and Grossberg, 2000; Craft et al., 2007; Jehee et al., 2007; Layton et al., 2012), through feedforward processing alone (Supér et al., 2010), or through horizontal connections within V2 (Zhaoping, 2005). Border-ownership signals require no more than 25 ms from the presentation of the figure to emerge (Zhou et al., 2000), which constrains the set of possible mechanisms. In early visual areas, feedback connections have the fastest conduction velocities (~ 3.8 m/s) that are considerably faster than those of horizontal connections (~ 0.3 m/s; Girard et al., 2001). Feedback connections are likely involved in border-ownership because they span large cortical areas with minimal delay, unlike horizontal connections.

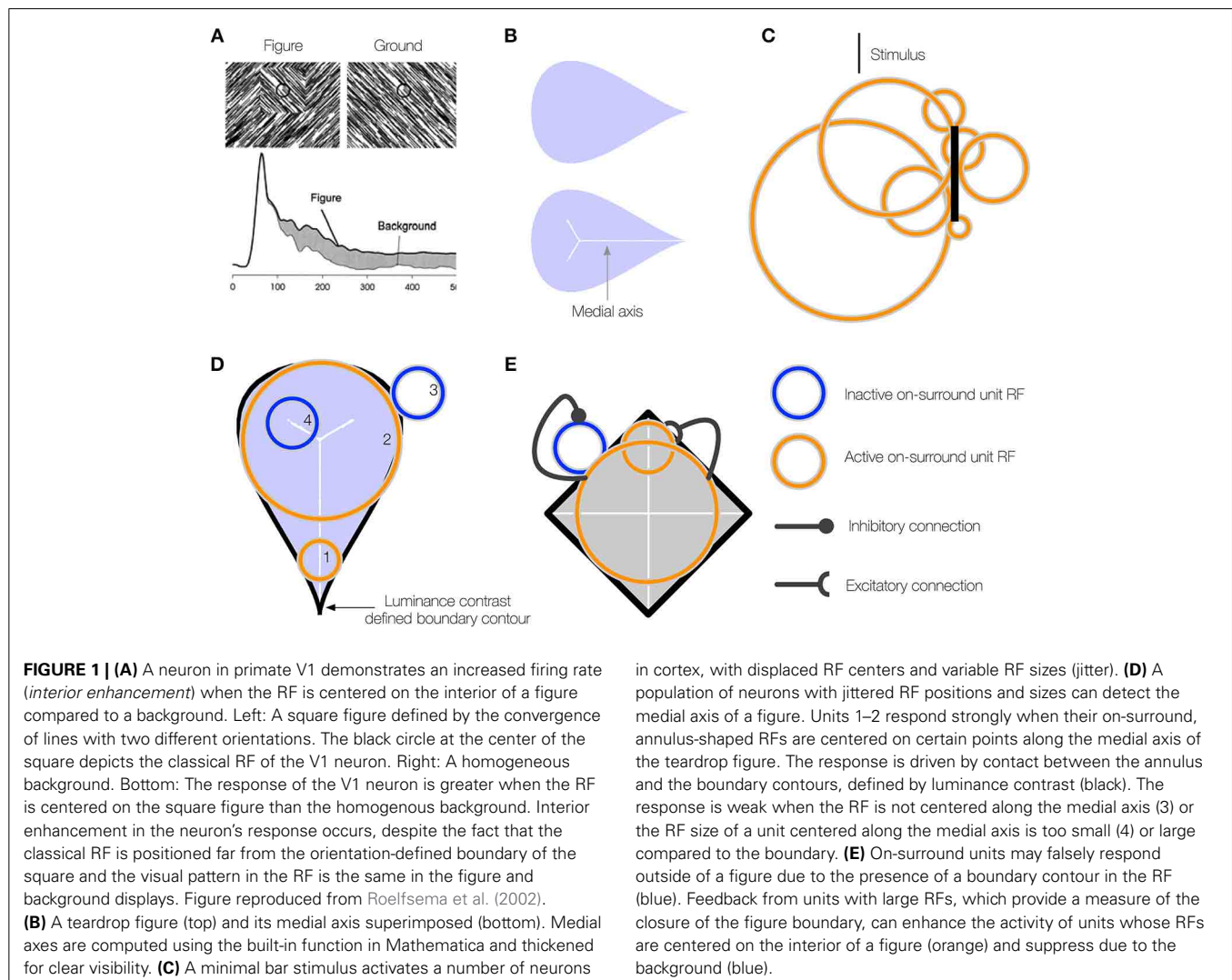
Another solution likely involves an enhancement of neural activity to the interior surface of the figure compared to the exterior (*interior enhancement*). When Lamme and colleagues centered the interior of a texture-defined square within the RF of neurons in early visual areas of monkey, the neurons exhibited an enhanced firing rate compared to when the monkeys were

presented a uniform texture (**Figure 1A**; Lamme, 1995; Zipser et al., 1996). The interior enhancement effect persists when the edges of the square are 8–10° and the modulation occurs after an 80–100 ms latency from the onset of the stimulus, which suggests feedback from neurons with larger RFs may be involved. A temporal analysis indicates that neural activity relating to the edges of the figure emerge first, following a short latency, then interior enhancement occurs in the “late component” of the response (Lamme et al., 1999; Lamme and Roelfsema, 2000). A neuron that shows interior enhancement continues to fire at an elevated rate when the RF is centered at different positions within the texture-defined figure, and the firing rate drops precipitously when the RF is centered on the background (Lee et al., 1998; Friedman et al., 2003). Neurons in V2 demonstrate a greater degree of interior enhancement compared to those in V1 (50% vs. 30%; Marcus and Van Essen, 2002), and the magnitude of interior enhancement response is greatest in V4 (50% greater than in V1; Poort et al., 2012).

Our understanding of the mechanisms underlying interior enhancement of figures is poor. Given that interior enhancement

has only been demonstrated in primary visual cortex, occurs with figures many times larger than the classical RF, and is associated with the late component of the neural response, we wondered if higher visual areas may underlie the effect. That is, we hypothesize that interior enhancement first occurs in higher visual areas and propagates via feedback to early visual areas. Neurons in higher visual areas have larger RF sizes and are ideally suited to determine whether a region belongs to the interior or exterior of a figure. Recurrent connections and multiple feedback loops with early visual areas may explain the late onset latency of interior enhancement.

If higher visual areas mediate the effect, what are neurons with limited RF sizes in early visual cortex that demonstrate interior enhancement signaling about the interior of a figure? We propose that interior enhancement is a means to code the figure with respect to its medial axis (Burbeck and Pizer, 1995; Kovács et al., 1998; Pizer et al., 1998). The medial axis (“skeleton”) of a figure defines the set of points along the interior that run equidistant to points along the boundaries (**Figure 1B**). It is a compact representation of the shape. The “late component” response of



neurons in the primate ventral stream that is characteristic of interior enhancement (Lee et al., 1998) has also been associated with a response to the medial axis of shapes, particularly in inferotemporal cortex (IT; Hung et al., 2012). In humans, fMRI BOLD signals related to the medial axis first emerge in areas V3 and beyond in the ventral stream (Lescroart and Biederman, 2013), which indicates that higher visual areas are important for detecting the medial axis. Medial selectivity in higher visual areas and the late onset of the modulation in early cortical areas suggest that interior enhancement is not a solely feedforward phenomenon. Psychophysical evidence demonstrates that humans exhibit a heightened sensitivity to the medial axis of shapes (Wang and Burbeck, 1998). Julesz and colleagues presented humans with an array of randomly oriented Gabor patches, except for those that collectively composed the boundary of shapes, such as ellipses, cardioids, and triangles (Kovács et al., 1998). Subjects performed a differential contrast detection task of a Gabor pattern that lay some distance on the interior of the shape boundary, and threshold performance was mapped out. The contrast sensitivity of subjects was greatest along the medial axis and the spatial profile of thresholds matched the medial axis representations at different spatial scales. These results indicate that the visual system is particularly sensitive to a figure's medial axis. The medial axis plays an important role in the Core theory of Pizer and colleagues that posits that the visual system represents a figure with respect to its boundary, middle, and width at multiple spatial scales (Pizer et al., 1998). As explained below, the central innovation of the present work is to show that medial representations at multiple spatial scales hold a key to figure-ground segregation, when combined with RF jitter and cooperative-competitive dynamics across neurons in multiple areas of the primate visual system.

If neurons that exhibit interior enhancement code the medial axis of a figure, how do these neurons integrate information about the boundary, given that the classical RF size of a single neuron is fixed and the distance between the medial axis and the boundary may vary? Not many models address the variability in RF sizes in areas of cortex. Contrary to the classical view that a minimal stimulus, such as a small bar, activates neurons with small non-overlapping RFs early in cortex, the neurons that respond to the stimulus occupy a small patch of cortex known as the cortical "point spread" (Das and Gilbert, 1995). Neurons within the "point spread" tend to be spatially close in cortex, but possess a diverse range of RF centers and sizes (Figure 1C; Gilbert et al., 1996). We use the term *jitter* to refer to the displacement of RF centers and variation in RF sizes among nearby neurons in cortex. Within and across visual areas along the ventral stream, RF size and jitter grows proportionately with eccentricity (Gattass et al., 1981, 1988; Bakin et al., 2000). Our model proposes that one of the functions of the naturally occurring jitter in the visual system is to locally "probe" for the medial axis of figures. The activation of some, but not all, neurons with displaced RF centers and sizes within a small patch of cortex provides detailed information about where the medial axis is likely positioned and its spatial extent (Figure 1D). Neurons with a single RF size may not be able to signal the presence of the medial axis of a figure in general.

Our model solves a crucial problem through feedback and the recruitment of neurons with multiple RF sizes that compute a

scale-sensitive estimate of the medial axis of a figure. Although a pair of equidistant contours may locally appear within the RF, the contours may not belong to a figure (Figure 1E). The contours may be incomplete fragments or lie outside of a perceived figure, in which case neurons that demonstrate interior enhancement do not fire (Lee et al., 1998). The visual system appears particularly sensitive to the Gestalt *closure* of a figure's boundary contours, whether they are continuous or fragmented (Elder and Zucker, 1993; Kovács and Julesz, 1993; Gerhardstein et al., 2004; Mathes and Fahle, 2007). We propose that neurons in IT cortex that respond to configurations of contours provide a measure of a figure's closure (Brincat and Connor, 2004, 2006). In our model, signals that emerge from units that collect evidence about a figure's closure send feedback to suppress the activity of units that codes the medial axis when their RFs are centered outside of figures (see blue unit, Figure 1E).

Here we introduce a neural model, called *the teardrop model*, to investigate the hypothesis that interior enhancement occurs in higher visual areas and underlies the effect observed in the primary visual cortex. The model is a multi-level network, consisting of cooperative/competitive interactions at each stage. In the context of figure-ground segregation, models that implement cooperative and competitive dynamics identify a global solution in the large space of possible interpretations (Edelman, 1987; Grossberg, 1994). Units in the teardrop model capitalize on jitter to reinforce the representation of the global figure and suppress other interpretations. We use areas (e.g., V1, V4, etc.) when referring to model layers, analogous to the areas in primate cortex that we believe carry out similar functions and dynamics. To focus on fundamental figure-ground mechanisms, the retina, LGN, and V1 are simplified and lumped together in a preliminary model stage that generates an edge map of figures in the visual display, as is thought to occur in early visual areas. The model has stages corresponding to areas V4, posterior IT (PIT), and anterior IT (AIT). Our model is consistent with physiological evidence that IT sends extensive feedback projections to V4 (Gattass et al., 1988; Piñon et al., 1998). As will be explained below, neurons analogous to those in IT may combine figure representations at multiple spatial scales and propagate information back to neurons that estimate the position of the medial axis.

The mechanisms in the teardrop model bring together aspects of the visual system to support figure-ground segregation in a method not described before. Our model consists of three main propositions. (1) Neurons that show an enhanced response to the interior of a figure signal the figure's medial axis (Figure 1B). (2) The visual system detects the figure's medial axis by recruiting neurons with jittered RF sizes and positions (Figure 1D). (3) Feedback from higher visual areas is necessary to constrain neural responses to the interior of figures (Figure 1E). Model convex cells (model PIT) exhibit enhanced responses to the interior of a figure after the following sequence of operations:

- Start with a V1 complex cell-like edge representation of the boundary contours.
- Detect curved contours by grouping the edge segment output of the complex cells (Figure 2A).

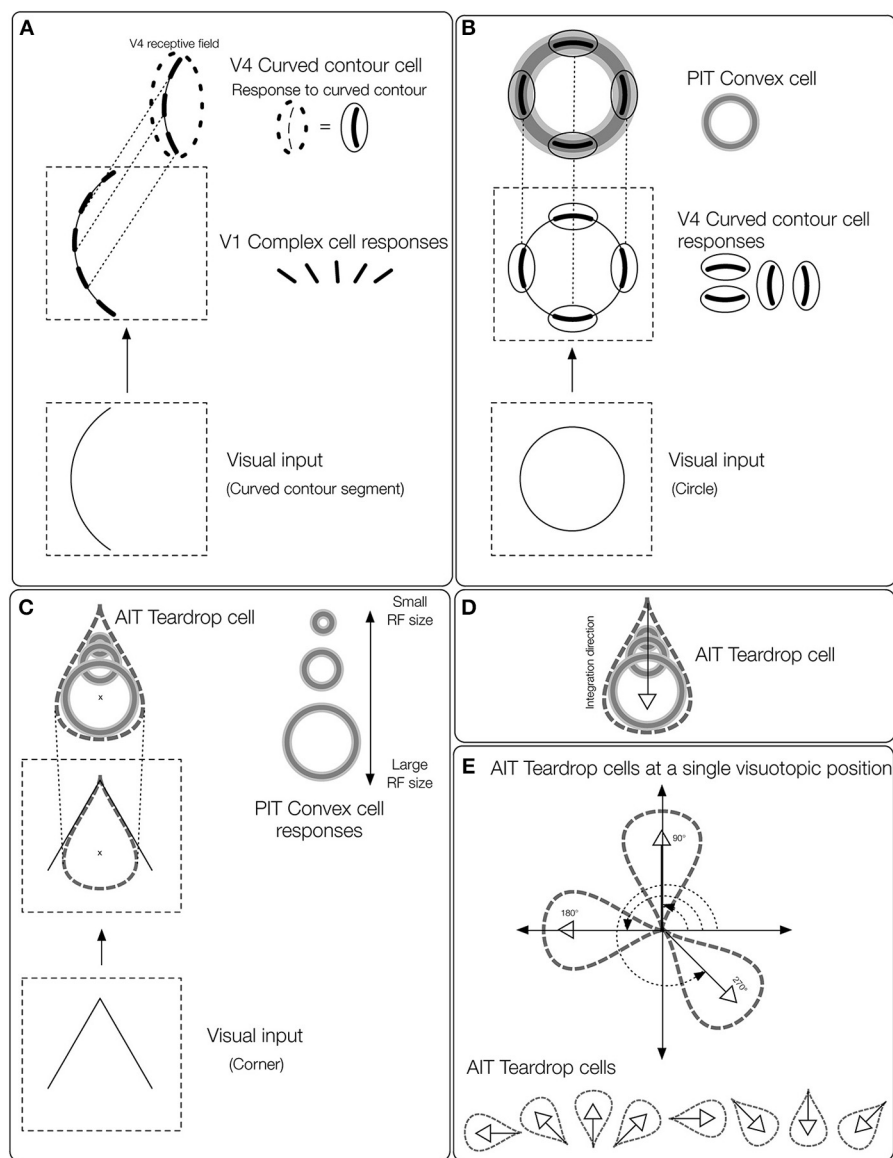


FIGURE 2 | Overview of the teardrop model stages. Network layers are labeled (e.g., V4, PIT, etc.) according to where the computations are proposed to take place in the primate visual system. The input to the model is a preprocessed edge map of the visual display, similar to the output of V1 complex cells. **(A)** The first model stage contains cells selective to curved contours (*curved contour cells*). When a curved segment enters the RF (bottom panel), curved contour cells group the piecewise linear spatial pattern of complex cell outputs (middle panel) to approximate a curved segment (top panel). The dashed ellipse signifies the curved contour cell RF, which hereafter is represented by a curved segment embedded inside a solid ellipse. **(B)** *Convex cells* in model PIT receive input from curved contour cells in an on-surround/annular spatial arrangement. Convex cells respond optimally to circles (bottom panel), because curved contour cell responses to the circular boundary contours

perfectly coincide with the annular receptive field of the convex cell (top panel). Convex cells respond to points along the medial axis of a figure because the units receive input from equidistant curved contour signals about the boundary. **(C)** Model AIT cells are called *teardrop cells* and respond to an ordered (by scale) collection of convex cell outputs along a medial axis segment. The “x” marks the visuotopic position of the teardrop cell RF. Teardrop cells that share the same RF position also receive input from the convex cell whose RF center is marked by the “x.” **(D)** The shown teardrop cell groups convex cells with RF sizes increasing with distance from the base of the arrow and estimates the medial axis of the corner input. Teardrop cells are hereafter depicted by the teardrop outline. **(E)** In our simulations, teardrop cells whose RFs are positioned at a single visuotopic location have one of eight integration directions, indicated by the white outlined arrows.

- Estimate points along the medial axis of the figure using convex cells (**Figure 2B**).
- Detect closure in boundary contour segments by integrating points along the medial axis via teardrop cells (**Figures 2C–E**).

- Teardrop cells are an ordered (by scale) collection of convex cell outputs along a medial axis segment.
- Suppress activity in convex cells to concave regions of the figure (**Figure 4**).

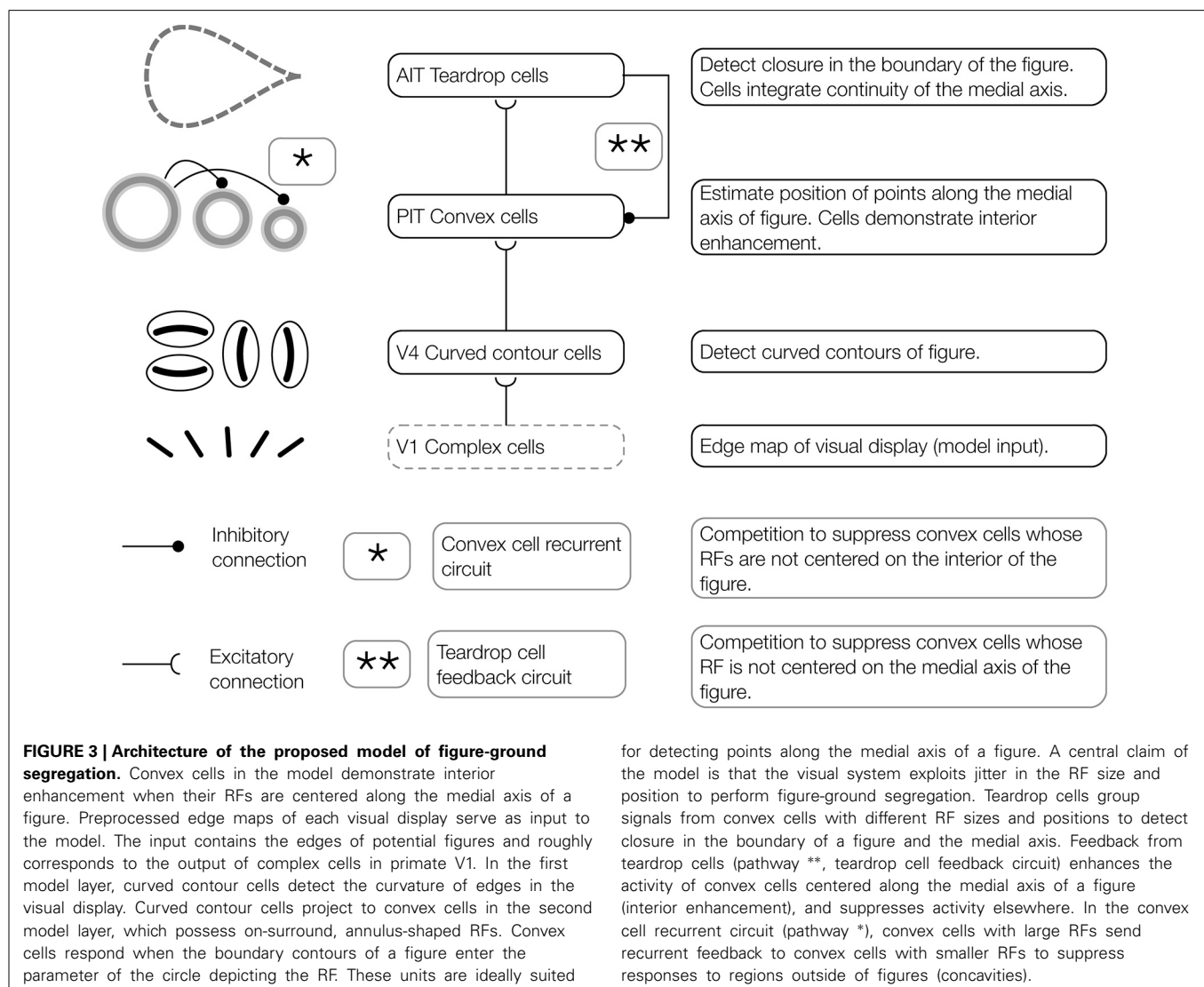
- Suppress activity in convex cells on the exterior of the figure using teardrop cells (**Figure 5**).

To our knowledge, the model created by Roelfsema and colleagues is the only existing investigation of the mechanisms underlying interior enhancement of a figure on a background. The model, however, is restricted to simple texture-defined squares and does not consider more complex shapes and visual scenes (Roelfsema et al., 2002). Our model is capable of performing figure-ground segregation in scenes with any number of figures, whose boundaries form simple closed curves or incomplete fragments thereof. We test our model on images of natural scenes and parametrically generated shapes with varying numbers and degrees of concavities. Our model also addresses response enhancement to a figure's interior in line-drawing or representations of figures whose boundary contours are not continuous. We do not address perceptual grouping that occurs behind occlusion. Several properties emerge through the dynamics of our model that are consistent with physiological data, such as the

size-invariant response properties of IT neurons (Appendix 1 in Supplementary Material; Ito et al., 1995; Logothetis et al., 1995).

MATERIALS AND METHODS

The aim of the present study is to have a better understanding of how interior enhancement occurs in the primate visual system. We use the model to test the hypothesis that dynamical feedforward and feedback interactions with higher visual areas in the ventral stream give rise to interior enhancement. Our model consists of three network layers that we believe correspond to primate visual areas V4, posterior inferotemporal cortex (PIT), and anterior inferotemporal cortex (AIT). We find these areas candidates for the computations carried out by the model based on evidence referenced below. Properties of model curved contour and convex cells are based on known physiology in corresponding areas, and those of teardrop cells are proposed. The proposed model is schematized in **Figure 2** and the model stages are depicted in **Figure 3**.



MODEL V4: CURVED CONTOUR CELLS

The inputs to the model are preprocessed edge maps, which approximate the output of complex cells in primary visual cortex (V1). We refer to the result of complex cells rather than simple cells because the edge maps are contrast polarity insensitive. The first layer of our model corresponds to area V4 in primate cortex (**Figure 2A**). We simulate the dynamics of cells sensitive to the curvature (*curved contour cells*). The behavior of model curved contour cells is similar to that of populations of V4 neurons, which, unlike those in V1 and V2, demonstrate far greater selectivity for curved contours (Pasupathy and Connor, 1999, 2001, 2002) and conjunctions of bars (Hegde and Van Essen, 2006; Yau et al., 2012) at multiple spatial scales (Mineault et al., 2013). Model curved contour cells respond optimally when a contour, such as a curved segment or corner, enters the RF that matches the unit's RF size and preferred curvature sensitivity (**Figure 2A**). At each visuotopic position, we simulate curved contour units tuned to eight arcs about a circle. We construct curved contour units with seven different RF sizes.

MODEL PIT: CONVEX CELLS

Curved contour cells in model V4 project to the second model layer, which corresponds to primate area PIT (**Figure 2B**). The purpose of model units in this network layer is to detect points along the medial axis or "skeleton" of figures. As shown in **Figure 1D**, units that integrate their curved contour inputs in an on-surround fashion, in the shape of an annulus, are ideally suited for detecting the medial axis because they receive bottom-up feedforward signals from the boundary contours when their RFs are centered on the figure. However, units with a single RF size are not sufficient for detecting the medial axis in general. **Figure 1B** shows that in the case of a teardrop shape, the distance changes between points along the medial axis and the boundary. Therefore, units with a single RF size are not sufficient for signaling the location of a figure's medial axis. A subset of units with different RF sizes can detect the medial axis, as indicated in **Figure 1D** by the active units. We call units that detect the medial axis *convex cells* (**Figure 2B**).

Convex cells simulate a number of properties from known neurophysiology and are consistent with findings from psychophysical experiments. Humans demonstrate a bias to judge symmetric, convex regions as figure, and asymmetric, concave regions as the background (Peterson and Salvaggio, 2008; Kim and Feldman, 2009). Two dimensional shapes are more rapidly detected (Elder and Zucker, 1993) with higher accuracy (Kovács and Julesz, 1993; Mathes and Fahle, 2007) in humans, even at a young age (Gerhardstein et al., 2004), when the collection of boundary contours form a continuous closed curve, as opposed to when constituent contours possess different curvatures and orientations that do not align with the overall shape of the figure. These findings are consistent with the possibility that the visual system contains mechanisms that afford sensitivity to convexity and closure (Wagemans et al., 2012). Neurons in PIT appear to integrate multiple curved contour segments when they appear at particular orientations and positions within the RF (Brincat and Connor, 2004, 2006). For example, a neuron in PIT may optimally respond to a crescent shape because a number of curved segments

that form the boundary contours appear together in appropriate positions in the RF (Brincat and Connor, 2004). The annulus has been shown to be an optimal stimulus for many neurons in intermediary areas of the ventral stream (Pollen et al., 2002; Hegde and Van Essen, 2006). An annular RF affords sensitivity to the figure-ground Gestalt properties of convexity and closure.

MODEL AIT: TEARDROP CELLS

Units in the third model layer, model AIT, receive feedforward input from convex cells (**Figure 2C**). The purpose of units in the third network layer is to collect evidence about the presence of a continuous medial axis that spans the interior of a figure. While convex cells detect probable *points* along a figure's medial axis, more is needed to detect its full extent. Units in model AIT spatially integrate signals from convex cells. Recall that the collection of convex cells with a single RF size is in general insufficient for detecting the medial axis of a shape (**Figure 1D**). Therefore, units in model AIT integrate convex cells with different RF positions and sizes (**Figure 2C**). For example, the active units shown in **Figure 1D** collectively signal the medial axis of the teardrop shape.

To integrate signals from convex cells that have different RF sizes and positions, units in model AIT have RFs elongated in a particular spatial direction (*integration direction*). For example, the unit in model AIT that groups the set of convex cell units depicted in **Figure 2D** is elongated in the vertical direction, and therefore has a vertical integration direction. Hence, AIT units respond to the output of convex cells, ordered by scale along a common axis. In our simulations, we used eight integration directions at every location in the visual field (**Figure 2E**). The use of integration directions capitalizes on the jitter in RF size and position found in cortex. We found that units in model AIT that group feedforward signals from convex cells whose RF sizes linearly increase along the integration direction were sufficient for detecting the medial axis in the displays we consider. Therefore, we call units in model AIT *teardrop cells*.

We define the *position* of a teardrop cell's RF to coincide with the RF center of the largest convex cell that sends feedforward input. For example, the "x" marks the position of the teardrop cell depicted in **Figure 2C**. Teardrop cells with different integration directions at the same RF position share a common input from the largest convex cell that falls within the RF.

The behavior of teardrop cells is consistent with properties of cells found in area AIT of primate cortex. Teardrop cells exploit the jitter in RF size and position of neurons in the visual system. Teardrop cells have large RF sizes, by virtue of their integration of convex cell inputs from different sized RFs. Their RF size is at least as large as the largest convex cell unit that provides input. So long as the figure remains within the RF, a teardrop unit yields a response to the medial axis of a figure, irrespective of its retinal size. No attempt was made to quantitatively fit the neurophysiological properties of AIT neurons because we focused on the core figure-ground mechanisms. We selected a set of teardrop cells in our simulations with eight integration directions, corresponding to the horizontal, vertical, and diagonal directions. We found this set was sufficient to yield qualitative matches to the medial axis sensitivity of neurons. Similar to AIT neurons, teardrop cells demonstrate size invariance in their responses (**Figure A1**).

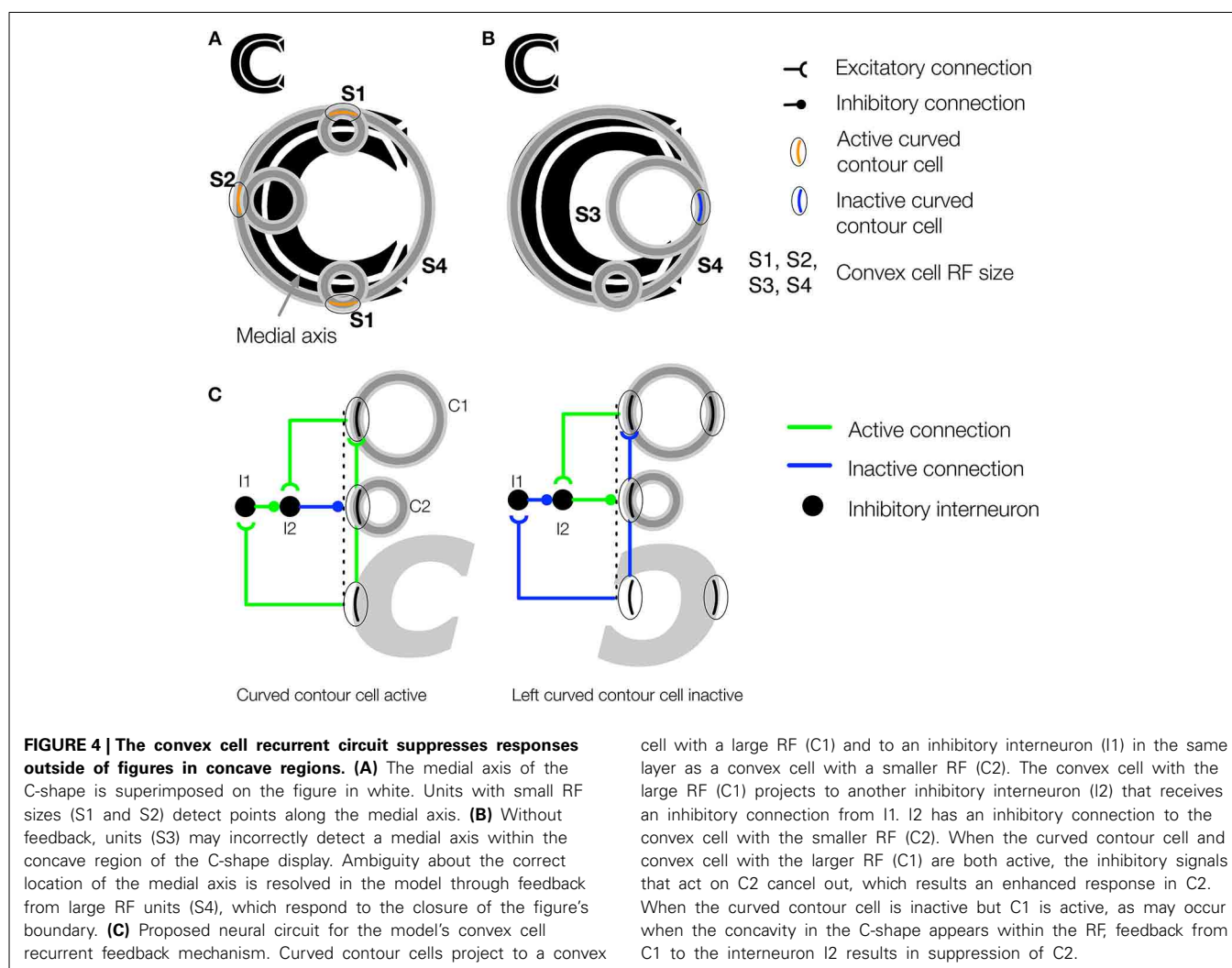
CONVEX CELL RECURRENT CIRCUIT

Although the feedforward model architecture will correctly detect the medial axis of a figure, false positive candidates may emerge in figures with concavities. **Figures 4A,B** shows in black a figure with a concavity called the C-shape. Its medial axis is superimposed in white. The RFs of active convex cell units with three different size RFs are shown. The set of the smallest active convex cell units that are shown (“S1” and “S2”) signals the medial axis of the C-shape to teardrop cells (**Figure 4A**). However, a set of convex cell units with larger RFs (“S3”) signals the presence of a false medial axis that spans the C-shape concavity, outside of the figure (**Figure 4B**). We propose that recurrent feedback connections between convex cell units suppress responses when they are due to a false medial axis outside the figure. The recurrent connections among convex cells are asymmetric: units only receive feedback from others with larger RF sizes. We are not aware of physiological evidence demonstrating asymmetric “coarse to fine” connectivity among cells with different RF sizes, although the idea has been used in existing theory (Grossberg, 1994). **Figure 4C** shows a neural circuit that implements the convex cell recurrent mechanism. An analysis of the convex cell RF organization is shown in (**Figure A2**).

TEARDROP CELL FEEDBACK CIRCUIT

When one or more teardrop cells with different integration directions that share a common RF position is active, it may be because their RFs are positioned on the figure’s medial axis. **Figure 5A** schematically depicts the collection of teardrop cells that have different integration directions whose RFs are positioned on a triangle figure. Because each teardrop cell has the medial axis within the RF (gray line), the units are active (orange). The fact that all three teardrop cells are active provides evidence that they are positioned on the interior of a figure. If sufficiently many teardrop cells are active, they send feedback to enhance the activity of the convex cells from which they received feedforward input (**Figure 5C**). The feedback results in an interior enhancement signal in convex cells centered on the interior of the figure.

Consider the case when few of the teardrop cells that share the same RF position are active. In the example depicted in **Figure 5B**, only one teardrop cell would be active nearby the top-right corner of the triangle because a large segment of the medial axis is in the RF. The other two teardrop cells with the same RF position (blue) are inactive because they are not positioned along the medial axis of the figure. If too few teardrop cells are active, the model sends inhibitory feedback to suppress the activity of



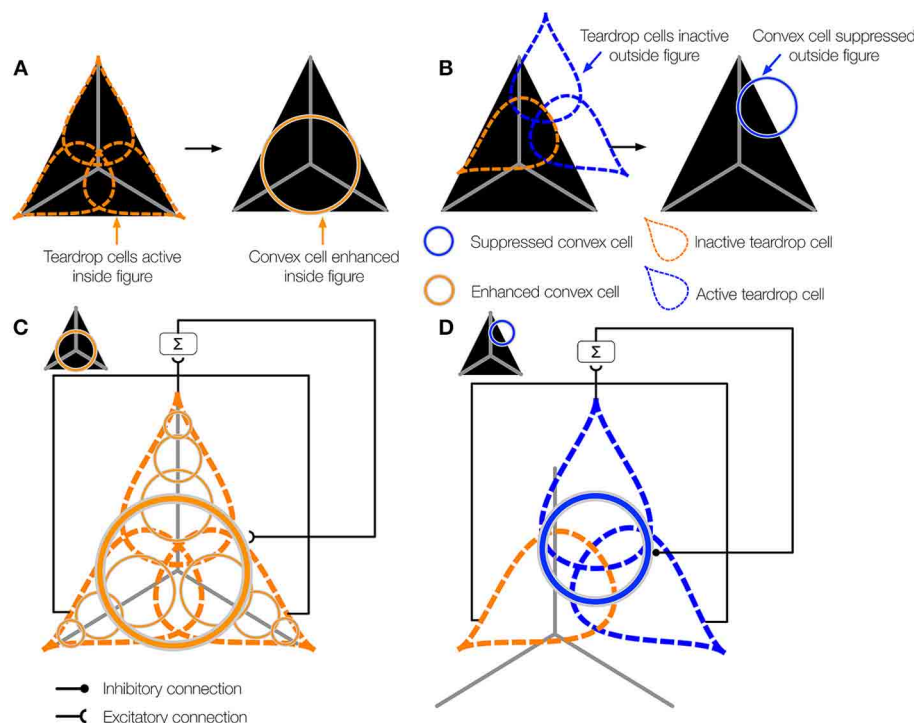


FIGURE 5 | The teardrop cell feedback circuit enhances convex cell activity along the medial axis of a figure. (A) Left panel: Three teardrop cells are shown (outlined by the dashed ellipses) that group convex cells with jittered RF sizes and positions along integration directions that coincide with the medial axis of the triangle figure. Right panel: Feedback from the active teardrop cells (orange) enhances the convex cell centered on the medial axis. **(B)** Left panel: Only the bottom left teardrop cell is active because a large segment of the medial axis does not enter the RF of the other two teardrop cells, so they are inactive. Right panel: The convex cell shown is inhibited because two of the three teardrop cells are inactive. **(C,D)** The teardrop cell feedback circuit mechanism. **(C)**

The teardrop cells in **(A)** share a common RF position, defined by the convex cell with the largest RF size from which they receive feedforward input (thick orange circle in the center). All three teardrop cells are active. Due to the good agreement in the activation of the teardrop cells that share the same RF position, feedback to convex cells in the same position that have the same sized or smaller RFs is enhanced. The feedback results in interior enhancement in convex cells with small RF sizes compared to the figure. **(D)** Only one of three teardrop cells in **(B)** is active, so it is unlikely that there is true medial axis in the RFs. Feedback from teardrop cells to the convex cells in the same position is inhibitory, which suppresses activity away from the medial axis.

the convex cells from which they received feedforward input (**Figure 5D**). This prevents convex cells with RFs centered outside of the figure from demonstrating interior enhancement. In summary, the activity of convex cells on the interior of the figure is enhanced, while the activity of convex cells outside the figure is suppressed.

Convex cells represent the units in our model that demonstrate an enhanced response to the interior of a figure. Our model predicts that these cells that exhibit interior enhancement are aligned with the medial axis. We simulated convex cells with seven different RF sizes. Units with different RF sizes that share a common RF center compete in a contrast-enhancing recurrent network. Cross-scale competition sharpens the network's sensitivity to the position of the medial axis. Activity of units that do not receive input from boundary contours on either side of the RF will be suppressed.

VISUAL DISPLAYS

We sought to test the model's capabilities by simulating parametrically varying versions of figures (**Figure 6**) that

resemble those used in electrophysiological studies of figure-ground segregation (Zipser et al., 1996; Zhou et al., 2000). We tested the model on rectangular (**Figure 6A**), square texture (**Figure 6B**), cross (**Figure 6C**), C-shape (**Figure 6D**), and randomly generated block shapes with varying complexities (**Figures 6E–G**). Concave regions tend to be part of the background rather than the figure and pose a challenge to models of figure-ground segregation. The C-shape and random block displays test the model's ability to avoid these regions when responding to the figure. We produced 500 low (LC), medium (MC), and high (HC) complexity random block displays, and 100 of each type are depicted in **Figures 6E–G**, respectively.

We parametrically varied the aspect ratio of the rectangular displays in the range 1/8 to 8, yielding 64 shapes. The aspect ratio of the C-shape was adjusted in equally spaced increments in the range 1/4 to 4 and the C-shape was 1–6 px thick to yield 96 shapes. We generated 36 crosses (6 thicknesses × 6 sizes) and square-texture displays (6 texture element displacements × 6 element sizes).

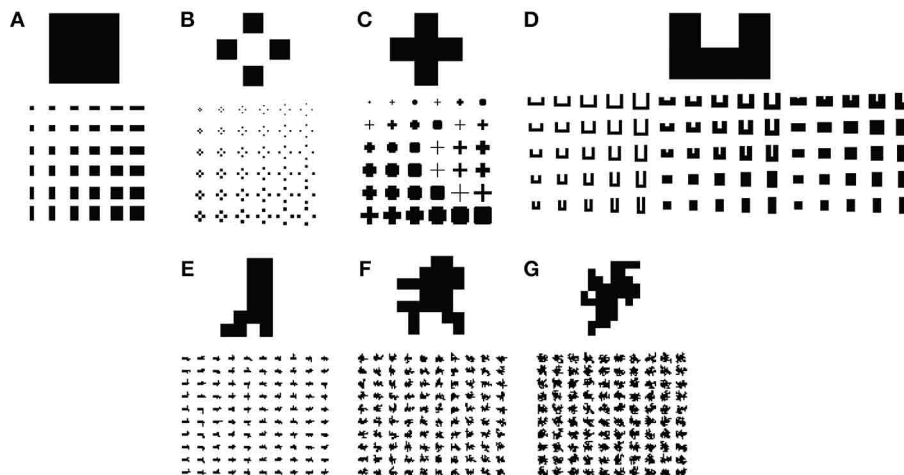


FIGURE 6 | Visual displays used in simulations to test the model. Five hundred of each type of visual display depicted in (E,F) were parametrically generated. Only 100 are shown. (A) Rectangles. (B)

Square textures. (C) Crosses. (D) C-shapes. (E–G) Algorithmically generated random block displays with low (E), medium (F), and high (G) complexity.

The random block displays were generated using a modified version of a random block generation algorithm (Sakai et al., 2012). The block algorithm begins with a base rectangle and iteratively adds an adjacent block to a random location along the rectangle boundary. In the iteration following the addition of a block, locations bordering either the rectangle or newly added block may be randomly selected for the next block addition. We generated LC, MC, and HC random block displays by adding 4, 16, and 32 blocks, respectively. Greater numbers of blocks afford greater complexity due to the increased irregularity in the figure boundaries. We constructed 500 unique blocks of each type in each condition.

FIGURE-GROUND INDICES

To quantify model performance across the visual display sets, we define several indices that assess figure-ground responses in the model. Larger index scores indicate better performance. The In-Out-Index (IOI) provides a measure of how much convex cell activity is distributed on the interior of the figure compared to the background:

$$IOI = \frac{A_{Figure} - A_{Ground}}{A_{Figure} + A_{Ground}} \quad (1)$$

In Equation (1), A_{Figure} and A_{Ground} refer to the mean unit activity inside the figure and ground regions, respectively.

We define two additional indices to assess the spatial distribution of model unit activation in each visual display. Equation (2) defines the medial axis index (MAI), which measures the ratio of unit activity distributed within 1 pixel of the medial axis of the figure (A_{Medial}), as computed Mathematica, to the mean activity on the complementary portion of the interior of the surface ($A_{Interior}$). Greater MAI scores indicate a greater proportion of the model activation due to the figure is distributed along the medial axis.

$$MAI = \frac{A_{Medial} - A_{Interior}}{A_{Medial} + A_{Interior}} \quad (2)$$

Equation (3) defines the boundary index (BI), which measures the ratio between the activity distributed within 1 pixel of the boundary of the figure ($A_{Boundary}$) and the activity garnered to the interior and exterior of the figure ($A_{Elsewhere}$). Greater BI scores indicate that much of the model activation is concentrated around the boundary of the figure.

$$BI = \frac{A_{Boundary} - A_{Elsewhere}}{A_{Boundary} + A_{Elsewhere}} \quad (3)$$

RESULTS

A central focus of the model is to better understand interior enhancement and the signaling of the medial axis of a figure by neurons in the primate visual system. We performed simulations of the model to investigate whether mechanisms in higher visual areas yield interior enhancement, which may underlie the effect observed in primary visual cortex. In Section Interior Enhancement and Medial Axis Sensitivity to Exemplar Figures, we examine medial axis detection and interior enhancement in exemplar visual displays. In Section Spatio-Temporal Dynamics, we focus on the spatio-temporal response of convex cells to show that these units do in fact exhibit interior enhancement, similar to units in primary visual cortex. In Section The Role of Feedback in Interior Enhancement and Figure-Ground Segregation, we describe performance of the model on larger numbers of visual displays, including parametrically generated figures, and analyze the role feedback has on enhanced interior responses. Appendix 3 in Supplementary Material contains the model equations.

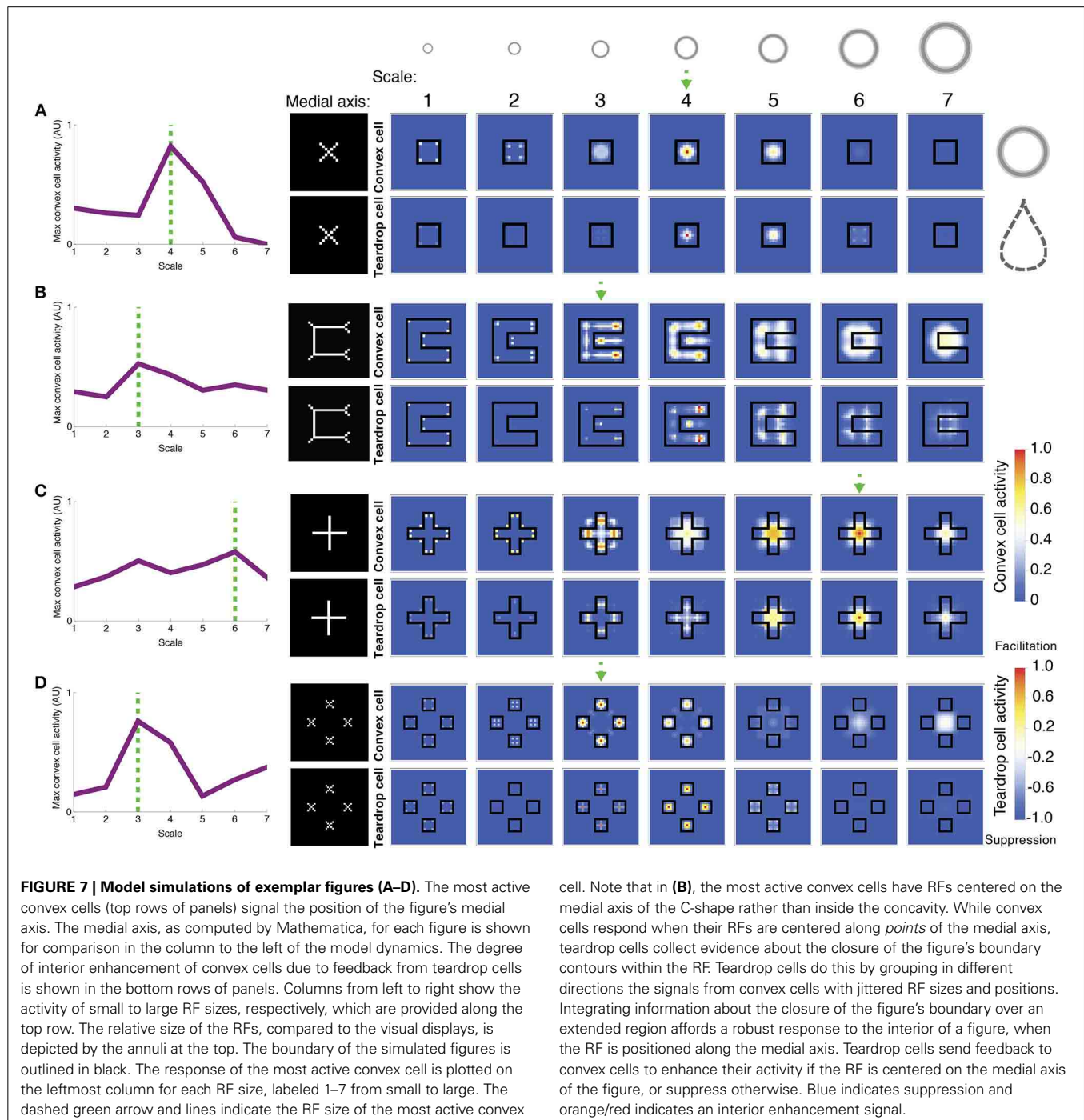
INTERIOR ENHANCEMENT AND MEDIAL AXIS SENSITIVITY TO EXEMPLAR FIGURES

To summarize the model dynamics and behavior, we often plot the activity of convex cells as a measure of the estimated location

of the medial axis (e.g., **Figure 7**). To readout the detected location of the medial axis from the model dynamics, we consider the spatial position of the maximally active convex cell. We do not claim that the brain decodes neural signals to locate the medial axis using maximum likelihood. This approach provides a simple way to readout activity across the network.

Figure 7 depicts the activity of convex cells (top panels), which signal the medial axis, and teardrop cells (bottom panels), which signal interior enhancement. The inputs in each simulation are

exemplar figures from the parametrically generated sets of visual displays shown in **Figure 6**. **Figure 7A** shows the model response to a square. Convex cells with a RF size of 4 yield the greatest activity compared to units with other RF sizes. The activity peak is concentrated at the center of the square. Convex cells with a RF size of 4 yield a MAI score of 0.91, which indicates that a high proportion of the neural activity to the figure interior is distributed along the medial axis. It is also the case that convex cells with smaller RFs yield activity peaks on the medial axis, along the



diagonals of the square. The smaller the RF size, the closer the activity peaks are to the corners.

The activity of convex cells with size 5 RFs and larger is suppressed, due to inhibition from teardrop cells. Recall that the teardrop feedback circuit suppresses convex cell activity when RFs are not positioned on the medial axis of the figure. Convex cells with large RF sizes compared to the square yield broad and weak distributions of activity. The activity is not constrained to the medial axis of the square, and is therefore suppressed. The high concentration of teardrop cell activity at the center of the square in units with size 4 RFs indicates that interior enhancement occurs in convex cells whose RFs are centered on the square. Teardrop cells facilitate an augmented response in convex cells to the figure through feedback.

Figure 7B depicts the model response to a C-shape display. The C-shape represents an important test for models of figure-ground segregation because the concave region is locally similar to the C-shape interior. The greatest convex cell activity is garnered by size 3 units whose RFs are centered along the medial of the C-shape ($MAI = 0.67$). Therefore, the model correctly performs figure-ground segregation because the peak is located inside the C-shape rather than inside the concavity. The teardrop feedback circuit alone does not result in correct figure-ground assignment because both the interior of the C-shape and the concavity are considered in the model as candidates where a medial axis may be located. The convex cell recurrent circuit is an important component of the model that allows it to correctly identify the medial axis of the C-shape. Size 4 teardrop cells yield the maximal activity, which signals interior enhancement to convex cells.

Feedback from teardrop cells does not completely abolish activity due to the concavity, which is consistent with the recent psychophysical finding that figure-ground percepts may reverse when the shape of the concavity is manipulated (Kim and Feldman, 2009). Adjustments to the curvature or junctions of the C-shape may change whether convex cell populations inside the C-shape or the concavity are more active.

Figure 7C depicts the model response to a cross. As shown in the left panel, the distribution of the maximally activity convex cells with different RF sizes is bimodal. The peaks garnered by units with smaller and larger RF sizes correspond to a response to the medial axis along the arms and center of the cross, respectively ($MAI = 0.74$). Units with size 6 RFs produce a strong response to the center of the cross due to feedback signals from teardrop cells (bottom panel), yielding interior enhancement. As shown in the bottom panels, teardrop cell activity is weak outside of the cross, which results in suppression of convex cells whose RFs are centered there. There is facilitation at the interior—particularly in units whose RF sizes are comparable in the length to the arms of the cross (RF sizes 5 and 6). The secondary activity peak produced by convex cells with size 3 RFs occurs due to the convex cell recurrent circuit. Convex cells with large RFs, comparable in size to the cross send feedback signals to enhance the response of units with smaller RFs, comparable in size to the width of the arm of the cross (size 3). The enhancement in the smaller RF units occurs because the small and large RF convex cells share common inputs from curved contour cells that respond to the distal parts of the arms.

Figure 7D shows the model response to the square texture display, which tests performance when there are multiple texture elements with various sizes and displacements. The largest convex cell activity peak occurs in units with size 3 RFs ($MAI = 0.83$). There are four distinct activity peaks that are located at the center of each of the squares. A smaller secondary activity peak occurs in units with size 7 RFs because the RFs are sufficiently large to group the square elements across the center gap. Teardrop cells are most active at the center of the squares, which yields interior enhancement in the convex cells.

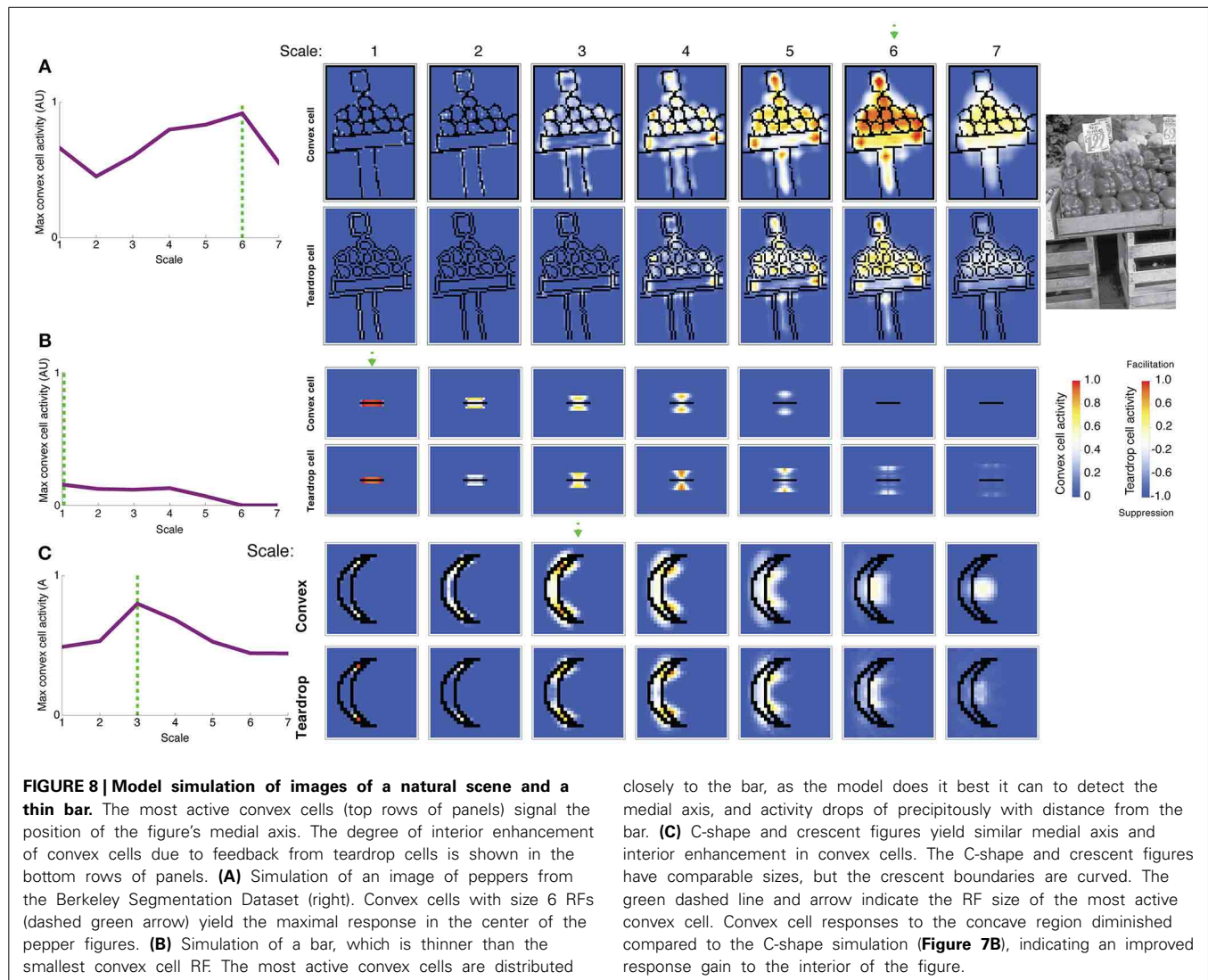
Figure 8A shows the activity of convex cells (top panels) and teardrop cells (bottom panels) with different RF sizes to a natural image of peppers taken from the Berkeley Segmentation Dataset. We wanted to test the model's figure-ground performance and ability to detect the medial axis in a more complex scene. The activity of convex cells with small RF sizes is distributed close to boundary contours. The peak convex cell activity occurs in units with size 5 and 6 RFs, near the center of the peppers. Teardrop cells are mostly quiescent, except for units with size 5 and 6 RFs, and clusters of activity coincide with the medial axis of the peppers. Therefore, feedback from teardrop cells facilitates an enhanced response in convex cells centered along the medial axis of the peppers. Teardrop activity diminishes in units with larger RF sizes, which indicates that the RF size is too large to integrate fine details of the scene.

In **Figure 8B**, we show results of a simulation of a bar that is thinner than the width of the smallest convex cell RF. We wanted to test model performance on a limiting case of when the figure has an infinitesimal width. Only convex cells with small RF sizes centered nearby the bar are active. The activity of units with larger RF sizes centered farther from the bar is greatly reduced. The spatial distribution of convex cell activity remains close to the bar and does not spread far away. The response of teardrop cells follows a similar trend: units with small RFs are active nearby the bar, and the response is lower in units with larger RFs. Convex cells with large RF sizes are not sufficiently active to overcome the suppression from teardrop feedback and are completely inhibited. This indicates that the model does the best job it can to identify a medial axis of an extremely thin figure.

We primarily tested model performance on figures with right angles, such as the C-shape and block visual displays; however, performance remained good on figures with curved contours. Consider the crescent shape shown in **Figure 8C** that approximates the C-shape. Convex cells whose RFs are centered along the medial axis produce the greatest response. Suppression of convex cell responses outside of the figure is greater in the crescent shape simulation, compared to the C-shape (**Figure 7B**). The curvature of the crescent boundary contours more closely matches the preferred sensitivity of the curved contour units than the right angles in the C-shape. This suggests interior enhancement in the C-shape would improve if model V4 included populations of neurons sensitive to conjunctions of bars (Hegde and Van Essen, 2006).

SPATIO-TEMPORAL DYNAMICS

We sought to investigate whether the temporal dynamics of convex cells are similar to those of neurons in V1 that show interior



closely to the bar, as the model does it best it can to detect the medial axis, and activity drops precipitously with distance from the bar. **(C)** C-shape and crescent figures yield similar medial axis and interior enhancement in convex cells. The C-shape and crescent figures have comparable sizes, but the crescent boundaries are curved. The green dashed line and arrow indicate the RF size of the most active convex cell. Convex cell responses to the concave region diminished compared to the C-shape simulation (**Figure 7B**), indicating an improved response gain to the interior of the figure.

enhancement (**Figure 1A**). Following the paradigm of Lamme et al. (1999), we presented the model with a textured scene with (**Figure 9A**; left panel) or without (**Figure 9A**; center panel) a square figure. The right panel of **Figure 9A** shows the temporal dynamics of a convex cell with a RF size of 4 whose RF was centered on the square when it was present. Similar to the V1 neuron responses, the convex cell demonstrates an enhanced response when the figure was present. Similar to the single-cell data, most of the modulation occurs later, following the peak response. The convex cell also demonstrates some interior enhancement prior to the peak response, unlike the neural data. We suspect that this is due to the lack of conduction delays in our model. Feedback from teardrop cells arrives instantaneously, yet *in vivo* there would be a delay for the signal to propagate and act on the target population of neurons. This could shift the onset of the interior enhancement beyond the peak response.

The results shown in **Figures 7, 8** suggest that feedback from teardrop cells plays an important role in enhanced responses along the medial axis. Signals from teardrop cells often suppress

convex cell activity not along the medial axis, even elsewhere within the interior of the figure. We wanted to better understand the role of feedback in the temporal dynamics of interior enhancement. **Figure 9B** plots the temporal response of convex cells whose RFs are centered on the medial axis (top left panel) and on the concavity (bottom left panel). To investigate the importance of feedback, we selectively lesioned different feedback connections in the model. Consistent with the results shown in **Figure 7B**, the response along the medial axis is larger than within the concavity when feedback is intact (**Figure 9B**; top center panel). When feedback is completely abolished, the ordinal relationship between the concavity and medial axis responses reverses: the convex cell activity is slightly larger in the concavity than on the medial axis (**Figure 9B**; bottom center panel). This would indicate an incorrect figure-ground assignment by the model, and shows that feedback is responsible for enhancing activity within the C-shape interior. The dynamics in the Convex-Only and Teardrop-Only Feedback conditions are shown in the right panels. In both cases, the individual types of feedback

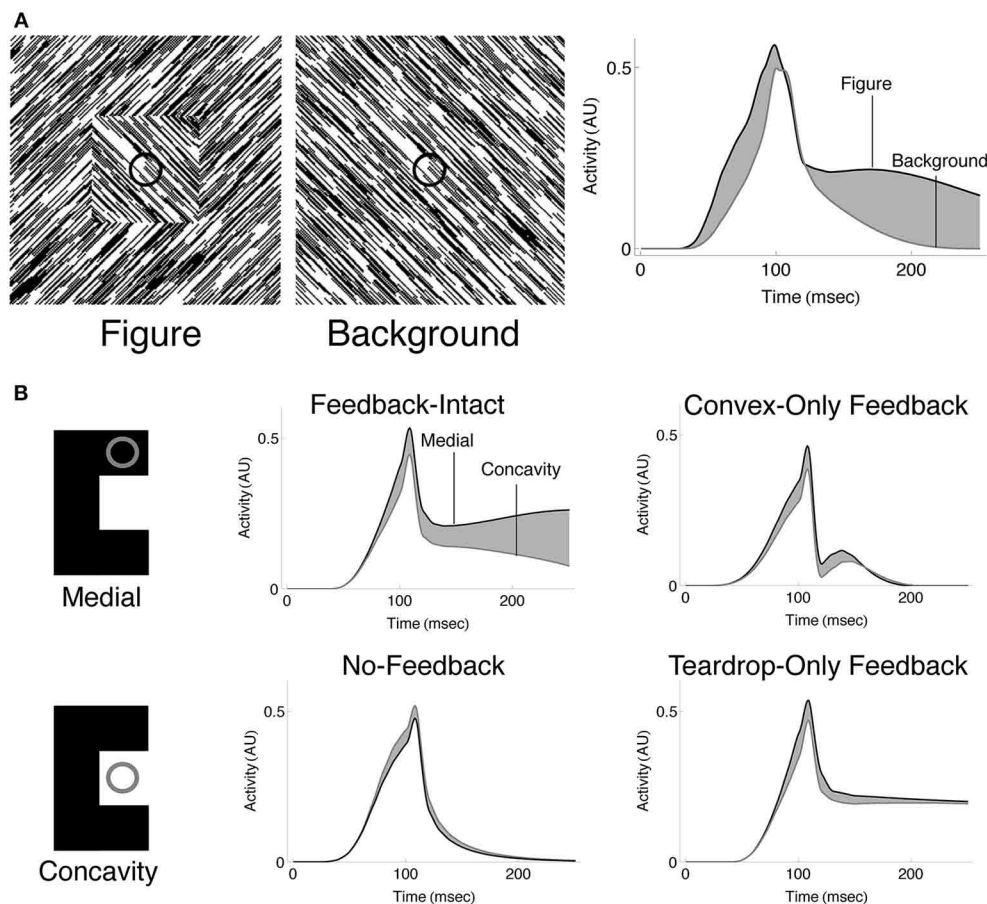


FIGURE 9 | Spatio-temporal dynamics of the model. (A) Convex cells demonstrate interior enhancement similar to single cells in primate area V1 (compare with **Figure 1A**). We presented the model with a textured display that either contained a square figure (left panel) or just the background (middle panel). The square subtended the same size as that simulated in **Figure 7A**. The dynamics of the convex cell whose RF is centered on the square when it was present are shown on the right panel. The response of the convex cell is larger in the presence of the figure (black curve) than when just the background was present (gray curve). **(B)** Interior enhancement along the medial axis of a figure occurs due to feedback in the model. We presented the C-shape used in **Figure 7B** and plotted the dynamics of convex cells whose RFs were

centered on the medial axis (*Medial*; black curves) or on the concavity (*Concavity*; gray curves). The response is enhanced to the medial axis compared to the concavity when feedback connections in the model are not lesioned. This indicates that feedback plays an important role in enhancing convex cell activity to the medial axis and suppressing activation to the background. Our model contains two types of feedback connections: the convex cell recurrent circuit and the teardrop cell feedback circuit. We considered the dynamics of the model when all feedback is intact (*Feedback-Intact* condition) and when all feedback was lesioned (*No-Feedback* condition). We also considered the effect each type of feedback connection had on model behavior through selective lesions (*Convex-Only Feedback* condition and *Teardrop-Only Feedback* condition).

connections yield an increased response to the medial axis relative to the concavity, but the degree of the modulation is less in either case than when both feedback connections are intact.

THE ROLE OF FEEDBACK IN INTERIOR ENHANCEMENT AND FIGURE-GROUND SEGREGATION

The results in **Figure 9** prompted us to quantify the role of feedback on interior enhancement for a broader range of visual displays. **Figure 10** shows model performance as assessed by the IOI, MAI, and BI on the LC, MC, and HC random block displays. The block displays test the model's ability to detect the interior of complicated figures despite the presence of many local concavities along the irregular boundaries. The relative response to the figure compared to the background, as measured by the

IOI, was greatest in Feedback-Intact condition (red), and lowest in the No-Feedback condition (green). The action of the convex cell recurrent circuit alone (Convex-Only Feedback condition, yellow) only slightly improved performance compared to the No-Feedback condition. However, the teardrop cell feedback circuit (Teardrop-Only Feedback condition) alone resulted in substantially improved selectivity to the interior of the figure compared to the background (blue). The absence of lesions (red) improved the relative response to the figure by a margin that often exceeded the combined individual gains obtained from single lesions. The Convex-Only Feedback condition scored the highest BI on the HC random block displays, which indicates a shift in the convex cell activity toward the boundary of the figure compared to the Feedback-Intact condition. The Teardrop-Only Feedback

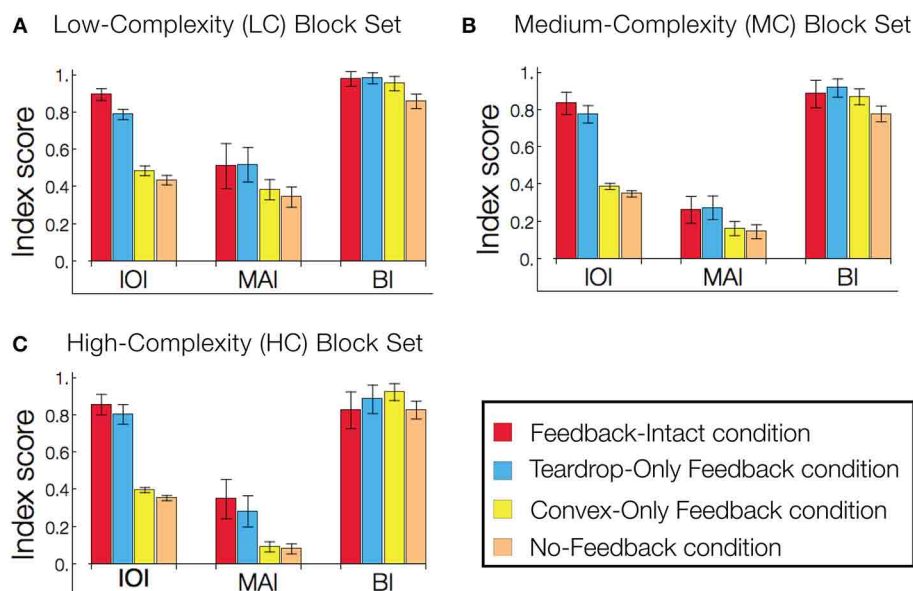


FIGURE 10 | Feedback greatly improves figure-ground segregation.

Performance is expressed with respect to the IOI, MAI, and BI. The in-out index (IOI) assesses the proportion of convex cell responses are due to the figure rather than the background. The medial axis index (MAI) measures how concentrated convex cell responses are on the medial axis of the figure. The boundary index (BI) assesses how proportions of convex cell responses occur near the boundary of the figure, compared to other regions in the display. Performance in the low-complexity (A), medium-complexity (B), and high-complexity (C) block sets is expressed with respect to the IOI, MAI, and BI. All performance indices are normalized such that 1 indicates the best performance and 0 the worst. Each bar represents the mean index score across the entire visual display set, and the error bars correspond to ± 1 standard deviation. The indices are computed based on convex cell activity.

Simulations are performed on the low (LC), medium (MC), and high (HC) complexity random block displays (Figures 6E–G). We assessed model performance when feedback from the convex cell recurrent circuit was lesioned (blue, *Teardrop-Only Feedback* condition), when feedback from teardrop cells was lesioned (yellow, *Convex-Only Feedback* condition), when both types of feedback were lesioned (orange, *No-Feedback* condition), and when all feedback was intact (red, *Feedback-Intact* condition). Convex cell responses were most concentrated on the interior of the figure (IOI) in the Feedback-Intact condition, and performance was the worst in the No-Feedback condition. Lesioning feedback from teardrop cells resulted in the greatest decrease in performance. For the HC display set, which contains figures with irregular boundaries and concavities, the Convex-Only Feedback condition yielded the best figure-ground performance.

condition yielded a high MAI score, which indicates that the feedback mechanism contributes to an increased sensitivity of convex cells to the medial axis of the figure.

For the LC random block displays, the Feedback-Intact condition garnered the largest BI scores (red). However, for the HC random block displays, the Convex-Only Feedback condition (yellow) yielded the highest BI score. Given that the IOI scores for the Convex-Only Feedback condition remained roughly constant irrespective of the input complexity, the increased BI scores indicate that the convex cell recurrent feedback circuit distributed activity closer to the boundary contours, yet still within the interior of the figure.

To quantify how feedback affects the sensitivity of convex cells to the medial axis of the figure, we computed the kurtosis for the distribution of the convex cells yielding the maximal activity with different RF sizes (e.g., Figure 7, left column). Often used in statistics, the kurtosis assesses how modal or “peaked” a distribution appears. For the distribution of maximal convex cell activity, the measure provides a diagnostic to assess how effectively feedback enhances the units with an appropriate RF size to code the medial axis. A large kurtosis indicates that most of convex cells that are active have a common RF size (Figure 11, lower-right panel). A high concentration of activity in units with a single RF size indicates a high degree of confidence in the medial

axis response. A low kurtosis indicates that the energy in convex cell responses is more evenly distributed among units with different RF sizes (Figure 11, top-right panel). A broad distribution indicates a lack of confidence in the medial axis response.

In the majority of the visual display sets we tested (6/7), the Feedback-Intact condition yielded the greatest kurtosis, which suggests that feedback increases the confidence and selectivity of convex cell responses to the medial axis of the figure. The Teardrop-Only Feedback condition generally yielded the next greatest kurtosis. The Convex-Only Feedback condition alone often did not yield a much larger kurtosis than in the No-Feedback condition. This indicates the convex cell recurrent circuit, as presently configured, did not increase the concentration of activity along the medial axis. From Figure 10, as the complexity of the block sets increased, the MAI decreased while the BI increased in the Convex-Only Feedback condition. Given the low kurtosis values, this suggests that convex cell feedback disperses activity more evenly within the figure surface.

In summary, the considerably greater kurtosis in the Feedback-Intact and Teardrop-Only Feedback condition conditions compared to the No-Feedback condition suggest that feedback plays a crucial role in increasing the response gain to the medial axis of a figure. Feedback also increased the confidence of model responses about the location of the medial axis.

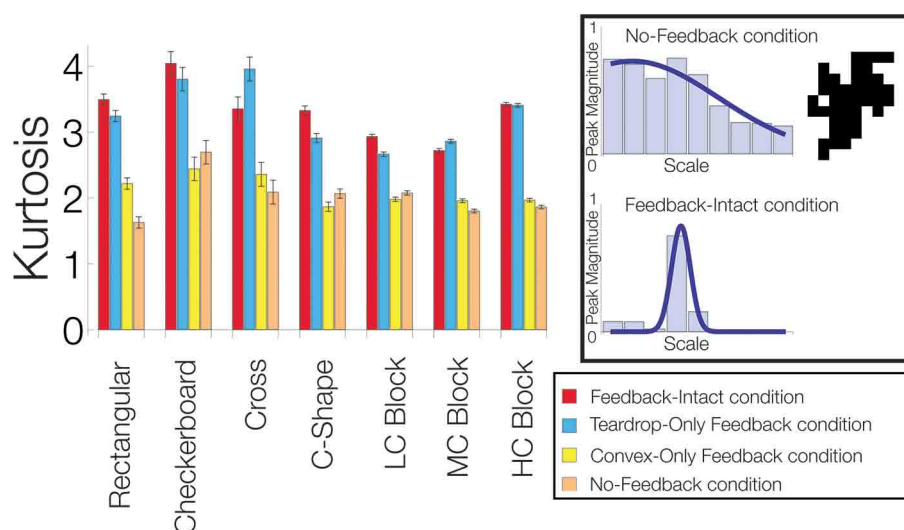


FIGURE 11 | Feedback improves the model's sensitivity to the medial axis of the figure. For each visual display set, the kurtosis of the distribution of the most active convex cells with different RF sizes is plotted. An example of a low kurtosis distribution of most active convex cells with different RF sizes is shown on the top-right panel. The bottom-right panel shows a distribution with a high kurtosis. We compared performance when feedback from the convex cell recurrent circuit was lesioned (blue, *Teardrop-Only Feedback* condition), when feedback from

teardrop cells was lesioned (yellow, *Convex-Only Feedback* condition), when both types of feedback were lesioned (orange, *No-Feedback* condition), and when all feedback was intact (red, *Feedback-Intact* condition). A high concentration of activity in units with a single RF size indicates a high degree of confidence in the medial axis response. A low kurtosis indicates that the energy in convex cell responses is more evenly distributed among units with different RF sizes. Performance was best in the Feedback-Intact and Teardrop-Only Feedback conditions.

DISCUSSION

We presented the teardrop model of figure-ground segregation in the primate ventral stream that explains why neurons demonstrate enhanced activity when their RFs are centered on the interior of a figure compared to the background. Our results support the possibility that interior enhancement arises as the result of dynamical interactions between higher visual areas. The proposed model makes the major theoretical prediction that interior enhancement originates in convex cells and the effect propagates via feedback to cells in earlier visual areas (Lamme, 1995; Lee et al., 1998). More specifically, we predict that cells in area PIT demonstrate interior enhancement prior to those in V1 due to recurrent interactions and feedback from teardrop cells. We also predict that teardrop cells that exploit jitter in convex cell RFs play an important role in modulating the interior enhancement effect. Our model is based on the following three propositions.

Proposition 1: Neurons that demonstrate an enhanced response to the interior of a figure signal the presence of the medial axis. Indeed, the responses of IT neurons support a representation of figures using its medial axis representation (Hung et al., 2012). The “late component” response of neuron in ventral areas not only is associated with interior enhancement, but also to the medial axis of shapes (Hung et al., 2012). Neurons in V1 that demonstrate interior enhancement show elevated responses at the center of texture-defined figures during the “late component” stage, which is consistent with medial axis coding (Lee et al., 1998). Our simulation results show that model convex cells demonstrate enhanced activity along the medial axis of figures due to dynamical cooperative/competitive interactions between higher visual areas. We propose that feedback

signals from convex cells to earlier visual areas may form the basis of the interior enhancement effect observed in V1 neurons. The mechanisms in the present model explain how responses in small RF units are constrained to the medial axis of a figure, which affords a parsimonious and efficient representation of a figure.

Proposition 2: There is a purpose for neurons with different RF sizes in areas in the visual system, aside from potentially detecting and representing figures of different sizes. The existence of neurons with multiple RF sizes in areas throughout the visual system is well known, yet their role is not clear. We claim that jitter in RF size and position serves a crucial role in figure-ground segregation. In our model, teardrop cells demonstrate the advantages that the visual system may garner by exploiting jitter. Grouping of signals from units with different RF sizes and positions by teardrop cells not only leads to a robust detection of a figure's medial axis, but it affords sensitivity to the closure of the figure's boundary contours. The closure of a figure's boundary contours facilitates its detection and the visual system more rapidly detects closed rather than open figures (Elder and Zucker, 1993; Mathes and Fahle, 2007; Wagemans et al., 2012). Sensitivity to closure underlies how the model successfully performs figure-ground segregation in the case of partially concave figures, such as the C-shape.

Proposition 3: Feedback plays a crucial role in yielding enhanced responses to the interior of figures. Our simulations show how interior enhancement occurs in convex cells due to feedback signals from teardrop cells. When we lesioned feedback connections in the model, convex cell activity was less concentrated along the medial axis, and across the population,

there was more “false positive” activation outside the interior of the figure. The action of the teardrop feedback circuit in the model is consistent with existing models (Supèr and Romeo, 2011) and single-cell data (Supèr and Lamme, 2007) that indicate that feedback enhances the response to the figure and suppresses responses to the background.

TEARDROP CELL RFs

In the model teardrop cells group signals from convex cells with jittered RF sizes and positions. In simulations, we assume for simplicity that teardrop cells integrate the signals from convex cells in equally spaced positions along each integration direction. The integration directions extended equally in all radial directions (i.e., isotropic). It is unclear how the visual system would perform the grouping, but that neurons analogous to teardrop cells likely group signals in irregular directions with variable spacing. A consequence of only considering isotropic teardrop cells is that they yield optimal responses to shapes with certain aspect ratios. The response of a teardrop cell to a square (**Figure 7A**) is more concentrated at the center of a square than it would be to an elongated rectangle. We found that varying the aspect ratio of figures did not qualitatively impact figure-ground segregation performance, but it yielded broader, less punctate teardrop activation along the medial axis. Note that this is simply an artifact of making simplifying assumptions for the purposes of simulation. The variability of RF configurations in cortex would be expected to yield comparable responses to figures, irrespective of the aspect ratio. Cortical magnification likely impacts the distribution of RF sizes of convex cells grouped by teardrop cells. An extension of the present model could investigate how these factors impact interior enhancement signals.

REPRESENTATION OF CONCAVITIES

When interpreting the model results for the C-shape, we assume that the concavity is part of the background rather than the foreground. That is, boundaries separating the C-shape and the concavity are grouped with the C-shape rather than the concavity. However, Kim and Feldman found that manipulations to the salience and shape of the concave region might locally reverse border-ownership along different parts of the C-shape boundary, which is inconsistent a globally concave percept of the “negative part” (Kim and Feldman, 2009). Consistent with the possibility that local border-ownership signals may differ from the global interpretation, a substantial number of model convex cells with small RF sizes were active within the “negative part.” “Votes” for the presence of a medial axis from the population of convex cells whose RF is centered on the interior of the C-shape are at odds with those from the competing population whose RFs are centered on the concavity. In the model, the convex cell recurrent circuit enhanced the response of convex cells whose RFs are centered on the C-shape medial axis and suppressed those centered on the concavity. Perhaps the local reversals in border-ownership stem from reversals in the winning populations of convex cells with small RFs. The size and shape of the C-shape “negative part” may modulate the strength of the convex cell recurrent circuit and impact the likelihood that one of the populations win out.

MEDIAL AXIS CODING

Populations of neurons in IT maintain a selective response when 3D rotations of the same figure are presented, which has led to the hypothesis that IT neurons may code shape with respect to a 3D interpretation rather than a set of 2D image features (Janssen et al., 2000; Yamane et al., 2008; Hung et al., 2012). Tuning to the medial axis in IT may similarly occur in 3D (Hung et al., 2012), though presently available evidence is limited. While 3D shape and medial axis tuning makes ecological sense, present data also support coding of 2D figures. IT neurons are well known to exhibit selectivity to line drawing displays and 2D projections of 3D shapes (Logothetis et al., 1995), as well as invariance to planar transformations of planar figures (Ito et al., 1995). Together, these data support the joint coding of 2D and 3D shape in IT cortex, though the primacy of one representation over the other is unclear. For example, Yamane and colleagues found robust tuning in IT neurons to shapes over a range of low-level image manipulations, such as shading, but tuning specificity declined when depth cues were removed (Yamane et al., 2008). On the other hand, Kovacs and colleagues found consistent tuning to 2D caricatures of 3D shapes (Kovács et al., 2003). The mechanisms in our model are agnostic to the issue of 2D vs. 3D coding in IT. The aim of the present paper was to test the core mechanisms of the model, so we focused on simple 2D figures. Modules for binocular disparity, shading, and other depth cues may be integrated into the model to test for 3D selectivity. However, this is outside the scope of the present paper.

MODEL LIMITATIONS

The manner in which teardrop cells combine their inputs likely differs from that of neurons in cortex that respond to the medial axis of figures. In particular, the distribution of RF sizes in each integration direction is unknown, and additional physiological work is required to determine whether regularity exists in how neurons integrate their inputs. Cortical magnification and eccentricity further complicate the picture. An on-surround RF organization also likely represents a significant simplification of the great diversity of RF shapes in cortex.

We did not directly model inhibitory feedforward inputs, although recurrent competition and feedback may afford functionally similar behavior. Others have proposed that inhibitory RF surrounds emerge through feedback, rather than feedforward processes (Hupé et al., 1998). Our model only employs units with convex on-surround RF organizations. Yet, V4 and PIT are functionally diverse areas (Brincat and Connor, 2006; Hegde and Van Essen, 2006), and neurons analogous to teardrop cells in cortex may group their inputs in both convex and concave configurations, which could increase the specificity of figure-ground responses. Our simulations demonstrate that the on-surround RFs are sufficient to detect the medial axis and obtain enhanced responses to the interior of figures.

COMPARISON WITH OTHER MODELS

Our model is not the first to propose that the medial axis of a figure provides a means for the visual system to represent surfaces. The medial axis serves as an attractor in the Bayesian model of Froyen and colleagues for border-ownership signals such that they

are directed toward the interior of a figure (Froyen et al., 2010). Pizer and colleagues developed the Core theory in which a figure is decomposed by the visual system in terms of the boundary, width, and medial axis (Pizer et al., 1998). Unlike existing models, ours is the first to propose that interior enhancement represents the mechanism by which the visual system codes a figural surface with respect to its medial axis.

Another crucial difference is that our model presupposes the recruitment of units in higher visual areas (e.g., PIT, AIT) to determine the medial axis of a figure. A recent computational study has proposed that the visual system only requires areas V1 and V2 to determine the medial axis of a figure (Hatori and Sakai, 2014). Border-ownership signals are first determined among units in model V2, and then the medial axis is resolved through synchronous feedback to area V1. That is, the medial axis computation depends on border-ownership signals, unlike in our model. While populations of neurons involved in coding border-ownership may interact with those in our model in higher visual areas, we demonstrated that the medial axis computation need not depend on border-ownership. An elevated response occurs along the medial axis in the model of Hatori and Sakai (2014) because feedback signals from border-ownership units at boundary contours to either side arrive simultaneously and constructively interfere. It is unclear how the border-ownership signals would be synchronized across cortex, though oscillation is one potential mechanism. However, the existence of coherent oscillations in the context of image feature representations has not been proven and remains controversial (for a discussion see Craft et al., 2007). Moreover, the model of Hatori and Sakai (2014) uses units with a single spatial scale, so it is unclear how the model would determine the medial axis for figures much larger than the RF sizes of V1 and V2 units. By contrast, cooperative and competitive dynamics between units with multiple jittered RF sizes are fundamental in our model for estimating the medial axis. Because units with multiple RFs are at the crux of our model, the model results are robust to figures that have a range of sizes (see Appendix 1 in Supplementary Material).

Whereas the balancing of feedforward and feedback signals is critical in the present model, other models have exclusively used feedforward connections. Supér and colleagues have presented a spiking three level network that uses a combination of excitatory inputs and surround inhibition between model layer connections to determine border-ownership and perform figure-ground segregation (Supér et al., 2010). The model of Sakai and Nishimura also performs border-ownership assignment using asymmetric feedforward surround modulation: units signal a preferred side-of-figure response when the figure falls within facilitatory rather than the inhibitory subfield of the RF (Sakai and Nishimura, 2006). While surround modulation likely plays a crucial role in figure-ground segregation (Walker et al., 1999), we believe feedforward processing alone is too rapid to account for the delayed interior enhancement latency. The effect does not occur in primary visual cortex until ~80–100 ms following the onset of the figure (Lee et al., 1998), yet surround modulation only requires ~7 ms, an order of magnitude faster (Knierim and Van Essen, 1992). In early cortical areas, feedforward signals propagate at ~2.24 m/s (Girard et al., 2001). For example, this means that a

feedforward signal only requires ~9 ms to travel from V2 to V4. Recurrent processing and feedback loops with higher visual areas require additional time and may account for the difference in the latency.

CONCLUSIONS

The enhanced response to the interior of a figure by neurons in primary visual cortex may provide insight into how the visual system performs figure-ground segregation. We presented a model that tests the possibility that interior enhancement arises through dynamical feedforward and feedback interactions between higher visual areas. Our results support the idea that interior enhancement arises in higher visual areas along the medial axis of a figure, and the resulting signals may modulate the activity of neurons in primary visual cortex through feedback. We showed that jitter in RF size and position provides an efficient means for the visual system to determine the medial axis of a figure.

ACKNOWLEDGMENT

This work was supported in part by the Office of Naval Research (ONR N00014-11-1-0535 and ONR N000141310092), Center of Excellence for Learning in Education, Science and Technology (CELEST, NSF/SBE-0354378), and the Air Force Office of Scientific Research (AFOSR 000464-001).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://www.frontiersin.org/journal/10.3389/fpsyg.2014.00972/abstract>

REFERENCES

- Bakin, J. S., Nakayama, K., and Gilbert, C. D. (2000). Visual responses in monkey areas V1 and V2 to three-dimensional surface configurations. *J. Neurosci.* 20, 8188–8198.
- Brincat, S. L., and Connor, C. E. (2004). Underlying principles of visual shape selectivity in posterior inferotemporal cortex. *Nat. Neurosci.* 7, 880–886. doi: 10.1038/nn1278
- Brincat, S. L., and Connor, C. E. (2006). Dynamic shape synthesis in posterior inferotemporal cortex. *Neuron* 49, 17–24. doi: 10.1016/j.neuron.2005.11.026
- Burbeck, C. A., and Pizer, S. M. (1995). Object representation by cores: identifying and representing primitive spatial regions. *Vision Res.* 35, 1917–1930. doi: 10.1016/0042-6989(94)00286-U
- Craft, E., Schutze, H., Niebur, E., and von der Heydt, R. (2007). A neural model of figure-ground organization. *J. Neurophysiol.* 97, 4310–4326. doi: 10.1152/jn.00203.2007
- Das, A., and Gilbert, C. D. (1995). Long-range horizontal connections and their role in cortical reorganization revealed by optical-recording of cat primary visual cortex. *Nature* 375, 780–784. doi: 10.1038/375780a0
- Edelman, G. M. (1987). *Neural Darwinism: The Theory of Neuronal Group Selection*. New York, NY: Basic Books.
- Elder, J., and Zucker, S. (1993). The effect of contour closure on the rapid discrimination of two-dimensional shapes. *Vision Res.* 7, 981–991. doi: 10.1016/0042-6989(93)90080-G
- Friedman, H. S., Zhou, H., and von der Heydt, R. (2003). The coding of uniform colour figures in monkey visual cortex. *J. Physiol.* 548, 593–613. doi: 10.1113/jphysiol.2002.033555
- Froyen, V., Feldman, J., and Singh, M. (2010). A Bayesian framework for figure-ground interpretation. *Adv. Neural Inf.* 631–639.
- Gattass, R., Gross, C. G., and Sandell, J. H. (1981). Visual topography of V2 in the macaque. *J. Comp. Neurol.* 201, 519–539. doi: 10.1002/cne.902010405
- Gattass, R., Sousa, A. P., and Gross, C. G. (1988). Visuotopic organization and extent of V3 and V4 of the macaque. *J. Neurosci.* 8, 1831–1845.

- Gerhardstein, P., Kovacs, I., Ditte, J., and Feher, A. (2004). Detection of contour continuity and closure in three-month-olds. *Vision Res.* 44, 2981–2988. doi: 10.1016/j.visres.2004.06.023
- Gilbert, C. D., Das, A., Ito, M., Kapadia, M., and Westheimer, G. (1996). Spatial integration and cortical dynamics. *Proc. Natl. Acad. Sci. U.S.A.* 93, 615–622. doi: 10.2307/38514
- Girard, P., Hupé, J. M., and Bullier, J. (2001). Feedforward and feedback connections between areas V1 and V2 of the monkey have similar rapid conduction velocities. *J. Neurophysiol.* 85, 1328–1331.
- Grossberg, S. (1973). Contour enhancement, short-term memory, and constancies in reverberating neural networks. *Stud. Appl. Math.* 52, 213–257.
- Grossberg, S. (1994). 3-D vision and figure-ground separation by visual cortex. *Percept. Psychophys.* 55, 48–121. doi: 10.3758/BF03206880
- Hatori, Y., and Sakai, K. (2014). Early representation of shape by onset synchronization of border-ownership-selective cells in the V1-V2 network. *J. Opt. Soc. Am. A Opt. Image Sci. Vis.* 31, 716. doi: 10.1364/JOSAA.31.000716
- Hegde, J., and Van Essen, D. C. (2006). A comparative study of shape representation in macaque visual areas V2 and V4. *Cereb. Cortex* 17, 1100–1116. doi: 10.1093/cercor/bhl020
- Hung, C.-C., Carlson, E. T., and Connor, C. E. (2012). Medial axis shape coding in macaque inferotemporal cortex. *Neuron* 74, 1099–1113. doi: 10.1016/j.neuron.2012.04.029
- Hupé, J. M., James, A. C., Payne, B. R., Lomber, S. G., Girard, P., and Bullier, J. (1998). Cortical feedback improves discrimination between figure and background by V1, V2 and V3 neurons. *Nature* 394, 784–787. doi: 10.1038/29537
- Ito, M., Tamura, H., Fujita, I., and Tanaka, K. (1995). Size and position invariance of neuronal responses in monkey inferotemporal cortex. *J. Neurophysiol.* 73, 218–226.
- Janssen, P., Vogels, R., and Orban, G. A. (2000). Selectivity for 3D shape that reveals distinct areas within macaque inferior temporal cortex. *Science* 288, 2054–2056. doi: 10.1126/science.288.5473.2054
- Jehee, J. F. M., Lamme, V. A. F., and Roelfsema, P. R. (2007). Boundary assignment in a recurrent network architecture. *Vision Res.* 47, 1153–1165. doi: 10.1016/j.visres.2006.12.018
- Kelly, F., and Grossberg, S. (2000). Neural dynamics of 3-D surface perception: figure-ground separation and lightness perception. *Percept. Psychophys.* 62, 1596–1618. doi: 10.3758/BF03212158
- Kim, S.-H., and Feldman, J. (2009). Globally inconsistent figure/ground relations induced by a negative part. *J. Vis.* 9, 8–13. doi: 10.1167/9.10.8
- Knierim, J. J., and Van Essen, D. C. (1992). Neuronal responses to static texture patterns in area V1 of the alert macaque monkey. *J. Neurophysiol.* 67, 961–980.
- Kovács, G., Sári, G., Köteles, K., Chadaide, Z., Tompa, T., Vogels, R., et al. (2003). Effects of surface cues on macaque inferior temporal cortical responses. *Cereb. Cortex* 13, 178–188. doi: 10.1093/cercor/13.2.178
- Kovács, I., Fehér, A., and Julesz, B. (1998). Medial-point description of shape: a representation for action coding and its psychophysical correlates. *Vision Res.* 38, 2323–2333. doi: 10.1016/S0042-6989(97)00321-0
- Kovács, I., and Julesz, B. (1993). A closed curve is much more than an incomplete one: effect of closure in figure-ground segmentation. *Proc. Natl. Acad. Sci. U.S.A.* 90, 7495–7497.
- Lamme, V. (1995). The neurophysiology of figure ground segregation in primary visual-cortex. *J. Neurosci.* 15, 1605–1615.
- Lamme, V., Rodriguez-Rodriguez, V., and Spekreijse, H. (1999). Separate processing dynamics for texture elements, boundaries and surfaces in primary visual cortex of the macaque monkey. *Cereb. Cortex* 9, 406–413. doi: 10.1093/cercor/9.4.406
- Lamme, V., and Roelfsema, P. R. (2000). The distinct modes of vision offered by feedforward and recurrent processing. *Trends Neurosci.* 23, 571–579. doi: 10.1016/S0166-2236(00)01657-X
- Layton, O. W., Mingolla, E., and Yazdanbakhsh, A. (2012). Dynamic coding of border-ownership in visual cortex. *J. Vis.* 12, 8. doi: 10.1167/12.13.8
- Lee, T. S., Mumford, D., Romero, R., and Lamme, V. (1998). The role of the primary visual cortex in higher level vision. *Vision Res.* 38, 2429–2454.
- Lescroart, M. D., and Biederman, I. (2013). Cortical representation of medial axis structure. *Cereb. Cortex* 23, 629–637. doi: 10.1093/cercor/bhs046
- Logothetis, N. K., Pauls, J., and Poggio, T. (1995). Shape representation in the inferior temporal cortex of monkeys. *Curr. Biol.* 5, 552–563. doi: 10.1016/S0960-9822(95)00108-4
- Marcus, D. S., and Van Essen, D. C. (2002). Scene segmentation and attention in primate cortical areas V1 and V2. *J. Neurophysiol.* 88, 2648–2658. doi: 10.1152/jn.00916.2001
- Mathes, B., and Fahle, M. (2007). Closure facilitates contour integration. *Vision Res.* 47, 818–827. doi: 10.1016/j.visres.2006.11.014
- Mineault, P. J., Zanos, T. P., and Pack, C. C. (2013). Local field potentials reflect multiple spatial scales in V4. *Front. Comput. Neurosci.* 7:21. doi: 10.3389/fncom.2013.00021
- Pasupathy, A., and Connor, C. E. (1999). Responses to contour features in macaque area V4. *J. Neurophysiol.* 82, 2490–2502.
- Pasupathy, A., and Connor, C. E. (2001). Shape representation in area V4: position-specific tuning for boundary conformation. *J. Neurophysiol.* 86, 2505–2519.
- Pasupathy, A., and Connor, C. E. (2002). Population coding of shape in area V4. *Nat. Neurosci.* 5, 1332–1338. doi: 10.1038/nn972
- Peterson, M. A., and Salvagio, E. (2008). Inhibitory competition in figure-ground perception: context and convexity. *J. Vis.* 8, 4. doi: 10.1167/8.16.4
- Piñon, M. C., Gattass, R., and Sousa, A. P. (1998). Area V4 in Cebus monkey: extent and visuotopic organization. *Cereb. Cortex* 8, 685–701. doi: 10.1093/cercor/8.8.685
- Pizer, S. M., Eberly, D., Fritsch, D. S., and Morse, B. S. (1998). Zoom-invariant vision of figural shape: the mathematics of cores. *Comput. Vis. Image Underst.* 69, 55–71. doi: 10.1006/cviu.1997.0563
- Pollen, D. A., Przybyszewski, A. W., Rubin, M. A., and Foote, W. (2002). Spatial receptive field organization of macaque V4 neurons. *Cereb. Cortex* 12, 601–616. doi: 10.1093/cercor/12.6.601
- Poort, J., Raudies, F., Wannig, A., Lamme, V. A., Neumann, H., and Roelfsema, P. R. (2012). The role of attention in figure-ground segregation in areas V1 and V4 of the visual cortex. *Neuron* 75, 143–156. doi: 10.1016/j.neuron.2012.04.032
- Roelfsema, P. R., Lamme, V., Spekreijse, H., and Bosch, H. (2002). Figure-ground segregation in a recurrent network architecture. *J. Cogn. Neurosci.* 14, 525–537. doi: 10.1162/08989290260045756
- Sakai, K., and Nishimura, H. (2006). Surrounding suppression and facilitation in the determination of border ownership. *J. Cogn. Neurosci.* 18, 562–579. doi: 10.1162/jocn.2006.18.4.562
- Sakai, K., Nishimura, H., Shimizu, R., and Kondo, K. (2012). Neural networks. *Neural Netw.* 33, 257–274. doi: 10.1016/j.neunet.2012.05.006
- Supér, H., and Lamme, V. A. F. (2007). Altered figure-ground perception in monkeys with an extra-striate lesion. *Neuropsychologia* 45, 3329–3334. doi: 10.1016/j.neuropsychologia.2007.07.001
- Supér, H., and Romeo, A. (2011). Feedback enhances feedforward figure-ground segmentation by changing firing mode. *PLoS ONE* 6:e21641. doi: 10.1371/journal.pone.0021641
- Supér, H., Romeo, A., and Keil, M. (2010). Feed-forward segmentation of figure-ground and assignment of border-ownership. *PLoS ONE* 5:e10705. doi: 10.1371/journal.pone.0010705
- Vogels, R., and Orban, G. A. (1996). Coding of stimulus invariances by inferior temporal neurons. *Prog. Brain Res.* 112, 195–211.
- Wagemans, J., Elder, J. H., Kubovy, M., Palmer, S. E., Peterson, M. A., Singh, M., and von der Heydt, R. (2012). A century of Gestalt psychology in visual perception: I. Perceptual grouping and figure-ground organization. *Psychol. Bull.* 138, 1172–1217. doi: 10.1037/a0029333
- Walker, G. A., Ohzawa, I., and Freeman, R. D. (1999). Asymmetric suppression outside the classical receptive field of the visual cortex. *J. Neurosci.* 19, 10536–10553.
- Wang, X., and Burbeck, C. A. (1998). Scaled medial axis representation: evidence from position discrimination task. *Vision Res.* 38, 1947–1959. doi: 10.1016/S0042-6989(97)00299-X
- Yamane, Y., Carlson, E. T., Bowman, K. C., Wang, Z., and Connor, C. E. (2008). A neural code for three-dimensional object shape in macaque inferotemporal cortex. *Nat. Neurosci.* 11, 1352–1360. doi: 10.1038/nn.2202
- Yau, J. M., Pasupathy, A., Brincat, S. L., and Connor, C. E. (2012). Curvature processing dynamics in macaque area V4. *Cereb. Cortex* 23, 198–209. doi: 10.1093/cercor/bhs004
- Zhaoping, L. (2005). Border ownership from intracortical interactions in visual area V2. *Neuron* 47, 143–153. doi: 10.1016/j.neuron.2005.04.005

- Zhou, H., Friedman, H. S., and von der Heydt, R. (2000). Coding of border ownership in monkey visual cortex. *J. Neurosci.* 20, 6594–6611.
- Zipser, K., Lamme, V., and Schiller, P. H. (1996). Contextual modulation in primary visual cortex. *J. Neurosci.* 16, 7376–7389.

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.

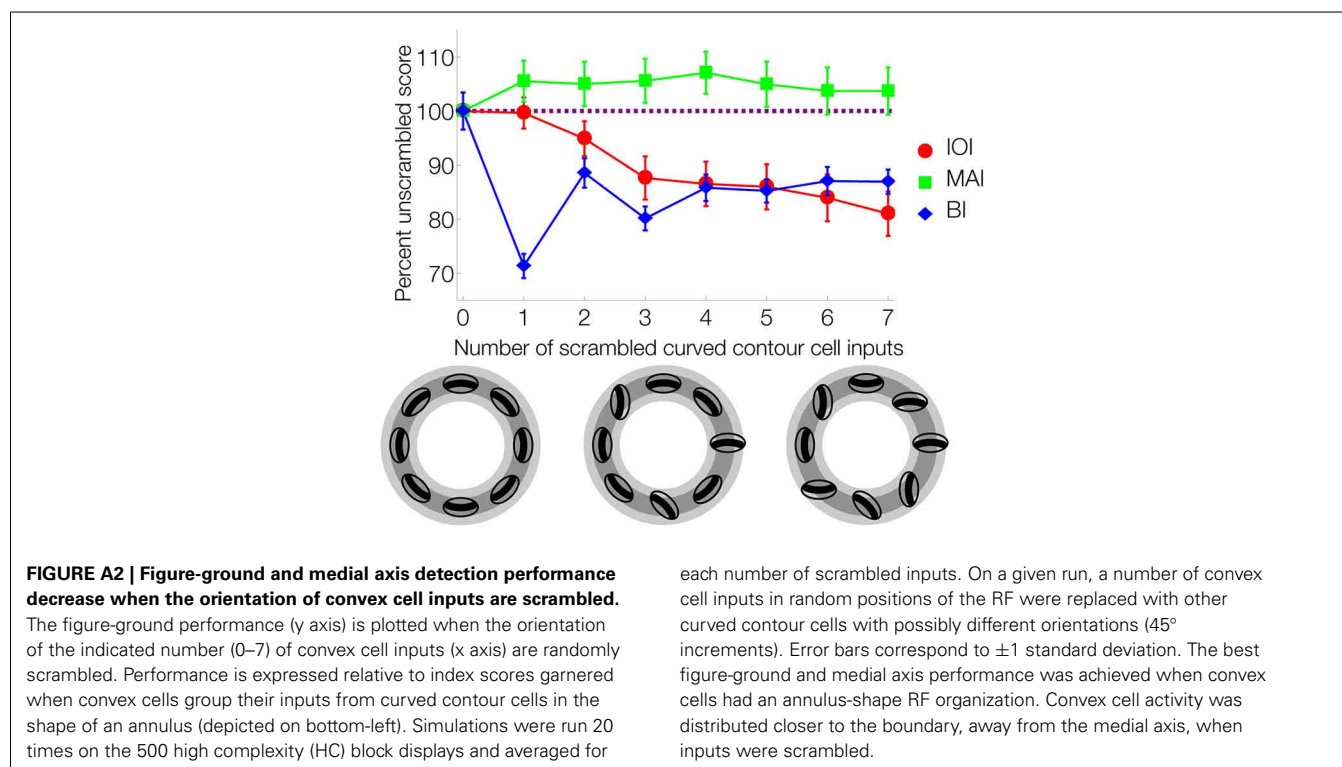
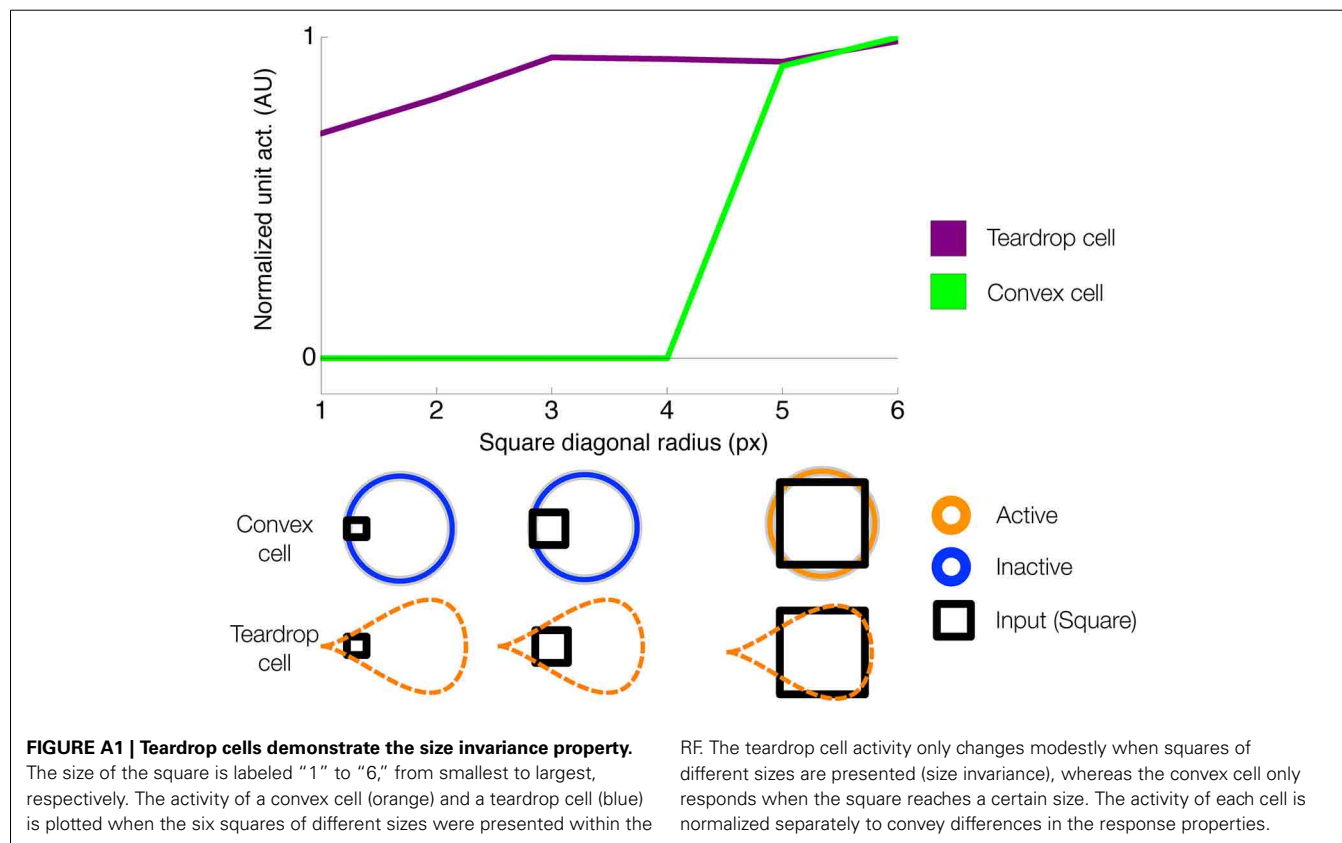
Received: 30 May 2014; accepted: 15 August 2014; published online: 10 September 2014.

Citation: Layton OW, Mingolla E and Yazdanbakhsh A (2014) Neural dynamics of feedforward and feedback processing in figure-ground segregation. *Front. Psychol.* 5:972. doi: 10.3389/fpsyg.2014.00972

This article was submitted to Perception Science, a section of the journal *Frontiers in Psychology*.

Copyright © 2014 Layton, Mingolla and Yazdanbakhsh. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

APPENDIX





Adaptive learning in a compartmental model of visual cortex—how feedback enables stable category learning and refinement

Georg Layher¹, Fabian Schrodtt², Martin V. Butz² and Heiko Neumann^{1*}

¹ Institute of Neural Information Processing, Ulm University, Ulm, Germany

² Department of Computer Science, University of Tübingen, Tübingen, Germany

Edited by:

Haluk Ogmen, University of Houston, USA

Reviewed by:

Oliver Obst, Commonwealth Scientific and Industrial Research Organisation, Australia
Ennio Mingolla, Northeastern University, USA

*Correspondence:

Heiko Neumann, Institute of Neural Information Processing, Ulm University, James-Frank-Ring, Ulm 89069, Germany
e-mail: heiko.neumann@uni-ulm.de

The categorization of real world objects is often reflected in the similarity of their visual appearances. Such categories of objects do not necessarily form disjunct sets of objects, neither semantically nor visually. The relationship between categories can often be described in terms of a hierarchical structure. For instance, tigers and leopards build two separate mammalian categories, both of which are subcategories of the category Felidae. In the last decades, the unsupervised learning of categories of visual input stimuli has been addressed by numerous approaches in machine learning as well as in computational neuroscience. However, the question of what kind of mechanisms might be involved in the process of subcategory learning, or category refinement, remains a topic of active investigation. We propose a recurrent computational network architecture for the unsupervised learning of categorial and subcategorial visual input representations. During learning, the connection strengths of bottom-up weights from input to higher-level category representations are adapted according to the input activity distribution. In a similar manner, top-down weights learn to encode the characteristics of a specific stimulus category. Feedforward and feedback learning in combination realize an associative memory mechanism, enabling the selective top-down propagation of a category's feedback weight distribution. We suggest that the difference between the expected input encoded in the projective field of a category node and the current input pattern controls the amplification of feedforward-driven representations. Large enough differences trigger the recruitment of new representational resources and the establishment of additional (sub-) category representations. We demonstrate the temporal evolution of such learning and show how the proposed combination of an associative memory with a modulatory feedback integration successfully establishes category and subcategory representations.

Keywords: neural model, category learning, subcategory learning, unsupervised learning, feedforward and feedback processing

1. INTRODUCTION

Stimuli presented in isolation cause cortical responses by feeding a representation defined by the feature arrangement that is contained in the current scene. The strength of the response depends on its contrast but is influenced by the local context in which it is embedded. Such (local) context information is integrated and thus made available at a neural site via lateral intra-cortical interactions, preferentially through long-range associative interactions in the superficial layers of cortex (Self et al., 2012). Larger context is integrated through the hierarchical processing of inputs over several stages of the cortical hierarchy where feature specificity of the neurons becomes more and more specific, integrating over an increasingly more widespread space-feature domain (Markov and Kennedy, 2013). At earlier stages, the result of such feature integration is made available via top-down feedback to merge feature representations of higher levels with spatially more

localized responses from initial filtering. Such convergence of feedforward and feedback streams of activation has recently been demonstrated to occur at the level of individual cortical columns (Mountcastle, 1997; Larkum, 2013).

Feedback signals tend to modulate the responses of activations at the earlier representations of raw feature presence (Larkum et al., 2004; Self et al., 2013). Modulating interactions are a common principle of neuronal interaction, which have been observed at different levels of cortical processing, subserving different cognitive computational functions, such as attention, figure-ground segregation, or grouping (Roelfsema et al., 2007; Poort et al., 2012). However, the precise functional role of feedback signals along downstream pathways is largely unclear and a topic of intense research investigation. Specific theoretical frameworks have been proposed that receive support by recent experimental investigations (Markov and Kennedy, 2013). One

such theoretical framework proposes that feedforward sensory activations are amplified by matching feedback such that those cells yield enhanced activations in a competition of cells, that have received a competitive advantage via modulating feedback (biased competition; Girard and Bullier, 1989; Desimone, 1998). Another framework considers the role of feedback as a predictive signal in which a template is activated that predicts the expected input given the evidence derived from current bottom-up input signals. The interaction of feedforward and feedback signals reduces the residual discrepancy between the different signal streams (Ullman, 1995; Rao and Ballard, 1999; Bastos et al., 2012). Overall, the literal difference between these model frameworks lies in the different roles feedback exerts on the bottom-up driven representations, although under certain conditions the two frameworks yield two variants of the same generic principles (Spratling, 2008, 2014).

In this work, we investigate learning and adaptation mechanisms in hierarchical cortical systems to develop a functional account for the role of feedback mechanisms. More specifically, we address the role hierarchical feedback may play in the online learning of visual representations. The study builds upon our previous modeling of a generic cortical architecture at the level of cortical columns. Model areas are defined by regular grids of interconnected columns, which are combined to define cortical subsystems, each composed of distributed networks of interconnected areas. Each model column is described at a mesoscopic level considering a compartmental structure that subdivides a cortical site into an input stage of specific signal filters, as well as superficial and deep layers as columnar compartments. Within this framework, feeding input signals drive the activity of columns and their lateral interactions. Feedback signals are thought to act in a modulating fashion so that responses at higher level cortical stages alone cannot generate activations in earlier representations (thus implementing a no-strong-loops principle; Crick and Koch, 1998). However, we demonstrate that interaction between different groups of cells allows to segregate the feedback signal strength that modulates the feedforward input activation such that the strength of feedback could be traced to serve as a signature how the expectations or predictions converge to the activation distribution of the driving input. The feature specificity of neurons in a cortical column is established through a learning mechanism that evaluates correlative activation in a scheme of modified Hebbian weight adaptation (Grossberg, 1988). During learning the connection strengths of bottom-up weights (to propagate converging driving input signals) are adapted. The applied learning scheme imposes a constraint such that the weights conserve their total energies so that variable input that is distributed over a population of neurons in columns does not lead to any bias in the incremental input segmentation. Thus, segmentations are allowed to build different and partly overlapping categorical patterns in which the total energy of the bottom-up input weights is normalized. The recurrent feedback from higher level representations generates a prediction, which consists of a pattern of the expected input activation, that drives the receiving representation of a column best. For that reason, the modulatory top-down feedback connections are here learned by using a slightly different

weight adaptation mechanism. The feedback weights define a top-down projective field, which represents the expected average input activity distribution of the cell. Taken together, feedforward learning enables the generation of prototypical form pattern representations, whereas feedback weights encode the characteristics of the category a stimulus is currently assigned to by the visual system. Thus, feedback and feedforward learning in combination realize an online associative memory mechanism, allowing the separation of an input stimulus and an according prototypical representation (see Carpenter, 1989). Using a modulation mechanism, the differences between an input pattern and an internal category representation are amplified in the input signal, yielding category building, consolidation, or refinement. The framework thus defines an important building block for the automatic incremental learning of visual categories (at different stages in the visual hierarchy). The compartmental structure and the neuronal interactions allow to stabilize the learning to prevent oscillatory learning as well as effects of overshadowing existing representations, connoted as the plasticity-stability dilemma (Grossberg, 1988). Using simple form patterns as input stimuli, we demonstrate that the model allows to automatically distinguish and refine the encoding of overlapping patterns and to trigger the learning of new categories when the input patterns differ significantly.

2. GENERIC MODEL ARCHITECTURE

2.1. OVERVIEW OF THE MODEL COMPONENTS AND FUNCTIONAL ARCHITECTURE

The function of the proposed network architecture has been discussed in the previous section in order to motivate key aspects of automatic acquisition of shape and object representations and how underlying cortical structural principles and mechanisms might contribute to its realization. In this section we present formal model mechanisms as a sketch of how the processing might be implemented dynamically. The basic structure of the generic model architecture is defined by three layers, each of which consisting of sheets of mutually interconnected computational elements (see **Figure 1**). These layers in the model roughly correspond to areas in cortex. Henceforth, we will address these stages by calling them layers or areas, given the particular context in the text. In the three layer architecture, the input layer is sketched like a simple replica of the input field fed by the current stimulus. The inclusion of such an explicit layer implicitly states that it may represent the result of some complex preprocessing that transforms the raw input into activity distributions referring to certain feature dimensions represented in a distributed fashion in (visual) cortex. As the same structure and composition of abstract columns can be replicated and more fine-tuned at different levels of cortex-like processing, we suggest that the outlined model architecture is generic in its structure and function. The computational elements in layers two and three both consist of an abstract model representation of cortical columns. Each of such columnar units itself is organized in a cascade of three processing stages: (I) input filtering, (II) activity modulation, and (III) pool normalization (details of the functional properties are discussed in, e.g., Neumann and Sepp, 1999; Bouecke et al., 2011; Brosch and Neumann, 2014a,b). These cascade stages roughly

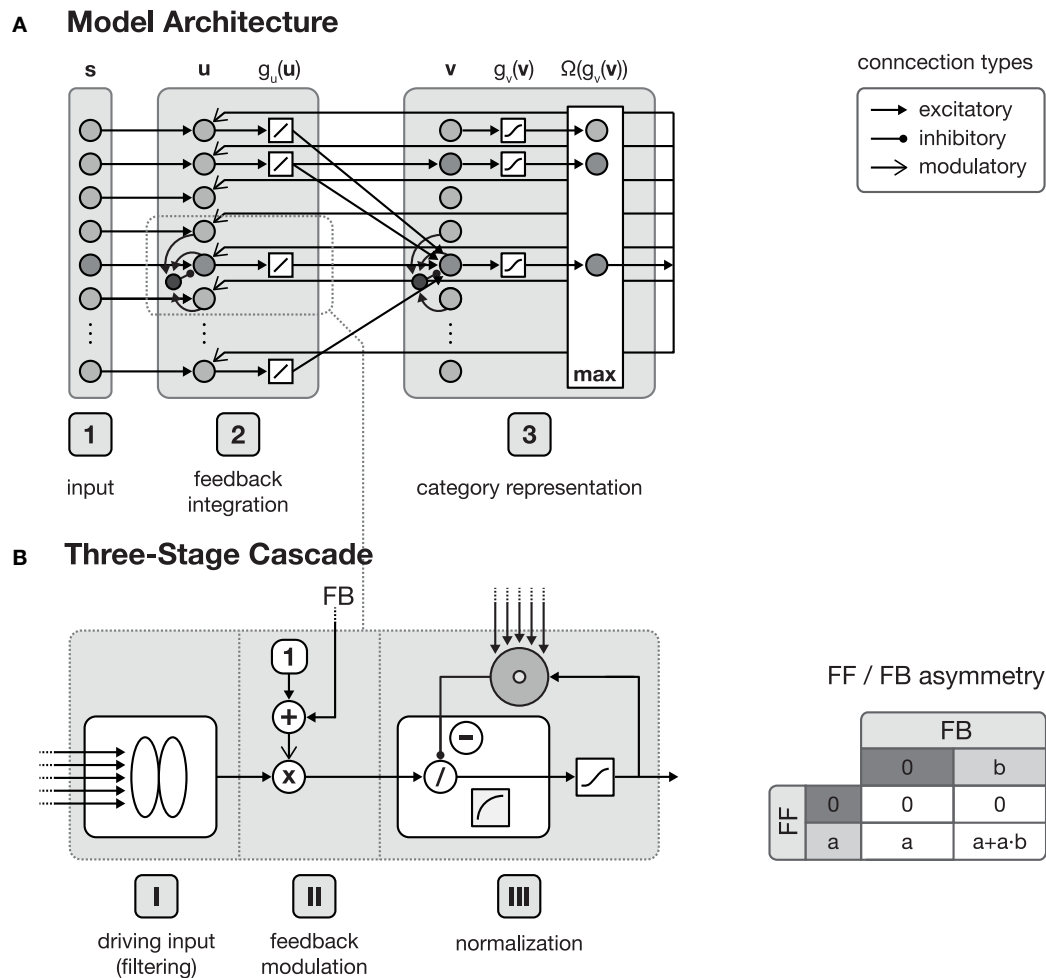


FIGURE 1 | Model architecture. (A) shows the overall structure of the proposed model, which is composed of interconnected cortical columns, subdivided into a compartmental structure of three processing layers. The first layer propagates input activities s to the second layer, where they are combined with a residual signal derived from feedback activities emitted from layer 3 and the current activity $g_u(u)$. After normalization, layer 3 category cells perform a correlation of the current layer 2 activities and their respective synaptic input weights. The cell with the strongest activation $g_v(v)$ is then selected by a winner-take-all (WTA) mechanism for weight adaptation and

activity propagation. In principle all of the model layers consist of a three-stage processing cascade as illustrated in (B). The cascade comprises an initial input filtering (stage I), the modulation of the activity (stage II) and a final pool normalization (stage III). Re-entrant feedback from higher level areas is incorporated in stage II where the current activity is modulated by $(1 + net_{FB})$. This kind of feedback integration is essential, since it results in an asymmetry of the roles the feedforward and the feedback signal play in the signal processing. As illustrated in the table on the right-hand side of (B), without the presence of a feedforward signal, a feedback signal cannot evoke any activity.

correspond to the division of cortical areas, with their six layers (Lui et al., 2011), considering the layer of terminating bottom-up input, as well as the superficial and the deep layers of cortex (Self et al., 2013). Each of these stages is represented by a model neuron that itself is a single-compartment dynamic element with gradual activation dynamics representing the average potential of a group of mutually coupled neurons. A firing-rate function $g(\cdot)$ converts the potentials into an output activation. Feedforward and feedback signal streams are combined at the level of individual columns (Larkum, 2013; see Brosch and Neumann, 2014a for a model implementation). In the proposed architecture, the second layer combines the input multiplicatively with a residual signal that is derived from the current input pattern and a feedback signal emitted from the successive layer 3 which is biased by a tonic

activity level (Eckhorn, 1999; Neumann and Sepp, 1999). Thus, the feedforward signal gates the re-entrant top-down signal so that the gain of existing activity can be increased by matching feedback signals. Feedback signals alone, however, cannot generate any activation for void bottom-up signal input. The feedback signal is generated here by a residual template, which contains the difference between the expected input (of the winning category node) and the current bottom-up input signal. As long as the difference does not vanish, the feedback mechanism leads to an increase in the activity gain of the current input. This mechanism deviates from the scheme described in e.g., (Bouecke et al., 2011), where the top-down signal is used instead of the residual signal. However, the dynamic properties of the non-linear circuit are retained.

Apart from the rather detailed network structure for generating an activation dynamic, the bidirectionally coupled network architecture is capable to adapt its connection weights, and is thus able to learn new category and subcategory representations as well as the expected average input distributions that have been established to drive a specific target category representation. In layer 3 of the generic architecture, category and subcategory representations are established using Hebbian learning mechanisms. Here, two complementary synaptic weight distributions are learned, each serving a different purpose within the proposed network. The feedforward synaptic weights are intended to build the category and subcategory representations during training, whereas the feedback weights are used to propagate an internal representation of the currently best matching category back to layer 2. This allows the estimation of the difference between the current input and the category assigned to the input after the feedforward sweep. Thus, layer 2 cells are able to combine the input with the derived difference signal and potentially evoke the activation of a different category/subcategory cell at the level of layer 3.

We split our presentation of the detailed model components into two major parts. First, we describe the activation dynamics, i.e., the formal definition of the generation of activities in each model computational element along the structure outlined in the previous paragraph. The activations are dependent on the input, the weightings of the spatial couplings for the input, and the current state, or activation of a model neuron. We emphasize how the incorporation of top-down feedback signal pathways can achieve rich and stable computations in such a network architecture. Second, in order to automatically acquire behaviorally relevant feature and category representations, the system can learn by adapting the weightings of the connection patterns between the model areas. We describe the weight, or learning, dynamics separately by focusing on the formal description of the weight adaptation and their key functionality. We finally link activation and learning dynamics to emphasize the capability

of such building blocks for autonomous learning in cortical architectures.

In essence, category and subcategory learning is enabled using two complementary core mechanisms. First, an associative memory is realized through the combination of an instar with an outstar learning scheme (compare Carpenter, 1989; see **Figure 2**). This allows the assignment of a given input to the currently best matching internal representation, as well as the propagation of the corresponding feedback pattern to re-enter at an earlier processing layer. Second, the differences between an input signal and the pattern associated with the best matching internal representation of the input define the modulatory signal to enhance the gain of the bottom-up feedforward signal.

In the following, we first describe the overall properties of the three-stage processing cascade, which forms the generic building block for all of the model layers.

2.2. ACTIVATION DYNAMICS

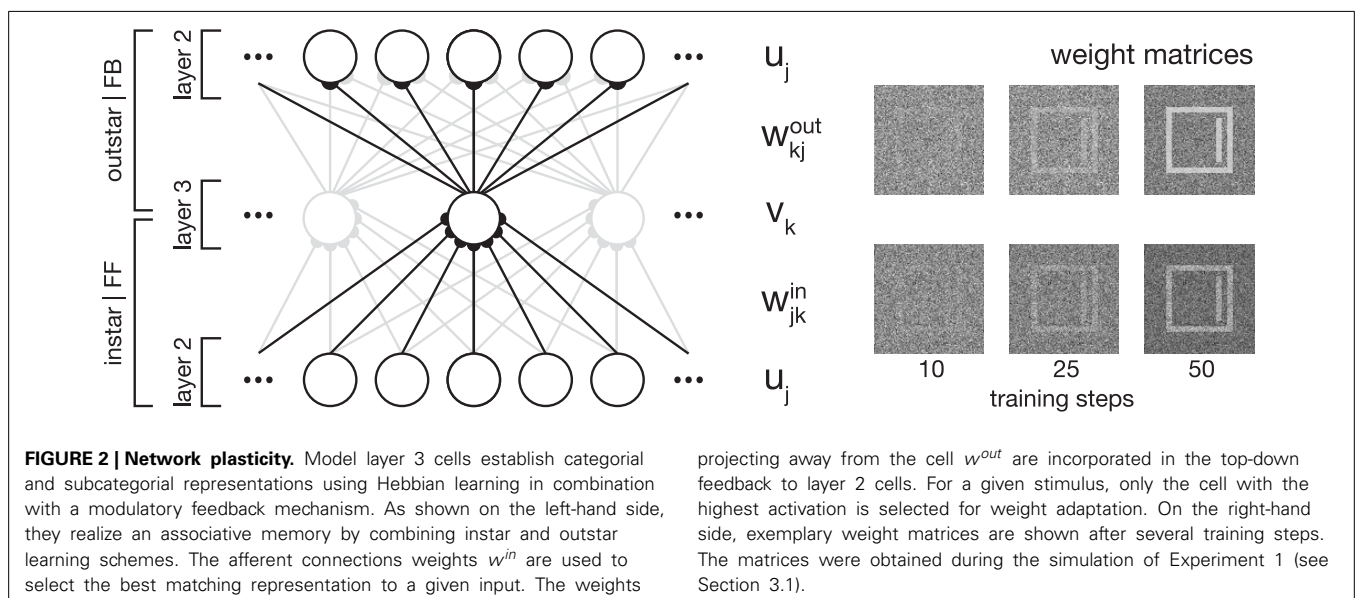
2.2.1. Three-stage processing cascade

The first stage of the model cascade performs a linear filtering of the input. To model the response r of a cell, we calculate the weighted sum on the input to a cell, as defined by

$$r = \sum_{j=1}^N K_j \cdot s_j, \quad (1)$$

with N the number of input cells with activities \mathbf{s} , which are modulated by the weight distribution \mathbf{K} . Within the proposed model, the filtering step either results in the propagation of the impulse response to a given input (for layer 2 cells) or \mathbf{K} corresponds to a weight distribution derived from the input statistics (for layer 3 cells, see Section 2.3.1).

At the second stage of the cascade, responses from the previous filtering are modulated by re-entrant input from higher-level model areas. Modulation is thereby performed in a way, such that



only existing activities in an input signal can be amplified (and thus activities cannot emerge solely provoked by a feedback signal). With r being the unmodulated driving signal and net_{FB} being the strength of the feedback signal, the modulated response of a cell is given by

$$r_{FB} \propto r \cdot (1 + net_{FB}). \quad (2)$$

This kind of feedback incorporation assures that if $r = 0$ no signal is generated as output, independent of the strength of the feedback net_{FB} . On the other hand, the input signal r is left unchanged in the absence of any feedback signal (i.e., $net_{FB} = 0$, see **Figure 1B**).

Prior to the final stage of the processing cascade, we apply a transfer function to convert the responses into a cell activation level. For simplicity we employ a linear transfer function at layer 2 of the proposed model, whereas at layer 3, a non-linear sigmoidal transfer function is used.

At the final stage of the processing cascade, activity normalization through divisive mutual inhibition within a pool of neurons (shunting inhibition) is applied. In its dynamic formulation, the rate change of the a signal r_j^{norm} depends on the current activation level r_j and the amount of inhibitory input activation in the pool q_j

$$\dot{r}_j^{norm} = -\alpha_r \cdot r_j^{norm} + \beta_r \cdot r_j - r_j^{norm} \cdot q_j \quad (3)$$

$$\dot{q}_j = -q_j + \sum_{k=1}^M r_k \cdot \Lambda_{jk}^{pool}, \quad (4)$$

with M denoting the size of the incorporated population in the neighborhood of location j and the weighting function Λ_{jk}^{pool} . The constant β_r controls the scale of the normalized signal, α_r denotes the passive decay rate.

In the following, we first describe the forward sweep throughout the proposed model layers. After the functional differences between the different model layers have been described in detail, we will emphasize the feedback connections and their role for the task of category and in particular subcategory learning.

2.2.2. Model layer 1/2

Layer 1 and layer 2 follow a pairwise connection scheme, such that each input cell in layer 1 is only connected to exactly one cell in layer 2 (see **Figure 1**). At the level of layer 2, the linear filtering step described in Equation (1) is equal to an identity function. Thus, the response of a layer 2 cell is defined by the following equation:

$$\dot{u}_j = -\alpha_u \cdot u_j + \beta_u \cdot s_j - u_j \cdot q_j, \quad (5)$$

where s_j denotes the output of a layer 1 cell, u_j describes the layer 2 cell response which relates to the membrane potential of real cells (j denoting the cell position). The constant α_u denotes the passive decay rate, whereas β_u describes the input scaling factor. The potentials are converted into an activation level, or firing rate, by the transfer function $g_u(u_j)$ (see Brosch and Neumann, 2014a for a formal specification and analysis). Here, we employ a linear transfer function with rectification such that no negative responses occur,

$$g_u(u_j) = [u_j]^+, \quad (6)$$

with $[u]^+ = \max(u, 0)$. The competitive interaction against a pool of cells to accomplish activity normalization is defined as

$$\dot{q}_j = -q_j + \sum_{k=1}^N g_u(u_k) \cdot \Lambda_{jk}^{pool}, \quad (7)$$

with N denoting the size of the incorporated population in the neighborhood of location j , weighted by Λ_{jk}^{pool} . Without the incorporation of any feedback signals, layer 2 cells solely perform an activity normalization on the output activities s of layer 1 and propagate the result to layer 3.

2.2.3. Model layer 3

Layer 2 and layer 3 cells form a complete bipartite connection graph with connections in both directions (see **Figure 1**), with corresponding synaptic coupling strengths w^{in} for feedforward and w^{out} for feedback connections. The output of layer 2 $g_u(\mathbf{u})$ is filtered by the feedforward weights w_{ji}^{in} to generate the strength of the response v_i of a layer 3 cell, which finally enters a competition with the surrounding pool activation (\mathbf{u} denoting the field of input activities represented as a vector), as defined by:

$$\dot{v}_i = -\alpha_v \cdot v_i + \beta_v \cdot \sum_{j=1}^N g_u(u_j) \cdot w_{ji}^{in} - v_i \cdot q_i, \quad (8)$$

with the passive decay rate α_v and the input scaling factor β_v . The response is then converted into an activity level using the non-linear sigmoidal transfer function g_v with the parameters κ_{log} (steepness) and μ_{log} (mean response level),

$$g_v(v_i) = \frac{1}{1 + e^{\kappa_{log} \cdot (\mu_{log} - v_i)}}. \quad (9)$$

As in layer 2, the final competition for activity normalization is defined by a non-linear competition of target activity and the integrated activation over a pool of neurons, which is determined by

$$\dot{q}_i = -q_i + \sum_{k=1}^M g_v(v_k) \cdot \Lambda_{ik}^{pool}, \quad (10)$$

with M denoting the number of cells in layer 3 and the weighting function Λ_{ik}^{pool} .

2.3. NETWORK PLASTICITY

In the previous part we have briefly introduced the formal description that covers the activation dynamics of the model mechanisms in the suggested generic architecture. As already mentioned, the architecture consists of three layers that roughly correspond to model areas of visual cortex. As outlined in **Figure 1**, the first area represents the input, that can be the raw responses of preprocessing the input directly (like in the early stages of the visual hierarchy, e.g., V1 and V2) or the output responses from a cascade of already more sophisticated processing to build intermediate level representations (like in the

higher stages of the visual hierarchy, e.g., V3 and V4). The second and third model areas in the model layout are connected bidirectionally representing feedforward and feedback sweeps of signal propagation in cortex (Lamme and Roelfsema, 2000). We have already explained how the two counterstream signal flows converge to build representations of integrated bottom-up evidences (from signal processing) and top-down predictions or expectations (generated by higher level stages of category representations). In this part we equip the network architecture with mechanisms of adapting the connections to learn representations in specific input weights. We suggest here that learning occurs along the feedforward as well as the feedback pathways (an outline of the learning architecture is shown in **Figure 2**). The functionality behind such a, again generic, principle is that feedforward connections learn weighting profiles that increase the probability for an input activation pattern to generate amplified responses in the recipient unit. Likewise, learning of feedback connections is intended to build up a representation in which source node activations (at the higher-level stage of the architecture) will generate a distribution of (pre-) activations as the expected average activity at the input stage that drives the node. The expectation is thus represented in the top-down connection weights (see Layher et al., 2014 for a model learning architecture that follows the same generic principles). Here, we develop a mechanism with a slightly different emphasis. The network aims to develop categories and also (later) to advance the automatic establishment of subcategories driven by significant local deviations of the already existing category representation. Therefore, the signal that is carried by the top-down feedback connections needs to be transformed into a residual signal such that the difference from the expected activation pattern is registered. We suggest that such residual patterns are generated at the neuronal *activation* pattern, instead of the weighting pattern.

In the following, we present the formal descriptions of the mechanisms used for the weight adaptation. We also briefly sketch how these relate to achieve the target representations for the desired bottom-up and top-down processing. The adaptation of the connection weights, for both feedforward and feedback, can be considered for individual neuronal sites in layer 3: The *receptive* field, or fan-in structure, is defined for connections along the bottom-up signal transmission that converge on a target neuron, $u \rightarrow v$. The *projective* field, or fan-out structure, on the other hand, is defined for connections along the reverse direction that spread out from the target neuron back to the previous stage, $v \rightarrow u$ (compare Carpenter and Grossberg, 1987b; Leaky and Sejnowski, 1988 for discussions of the underlying function of such connection principles). The activity dependent adaptation rules of such connection weights, namely feedforward, w^{in} and feedback w^{out} weights, are governed by modified versions of Hebbian correlation learning principles (Hebb, 1949). These modifications lead to stability and proven convergence properties and it can be shown that the learning rules optimize some target functionals.

The target neurons at layer 3 (with the adaptable fan-in and fan-out connections) are considered here to represent categories in a classification or recognition mechanism. For simplicity, we consider learning by weight adaptation that is allowed only for the category node that is maximally activated, as in many other

related learning paradigms (e.g., Kohonen, 1982; Carpenter and Grossberg, 2003). Such a model neuron is selected by a simple maximum selection operation, or winner-take-all (WTA) mechanism (Grossberg, 1973) and the weight adaptation is triggered subsequently,

$$\Omega(g_v(v_k)) = \begin{cases} 1 & \text{if } k = \arg \max_{i=1 \dots M} g_v(v_i) \\ 0 & \text{otherwise.} \end{cases} \quad (11)$$

It should be noted that the WTA selection is chosen here for simplicity. As an alternative, one could use a softmax mechanism as well (e.g., Roelfsema and van Ooyen, 2005), without changing the overall functionality of the approach. The specific learning rules for feedforward and feedback connections are presented below.

The learning of the feedforward weights w^{in} , as well as the feedback weights w^{out} is realized using Hebbian learning principles, which are described in the following.

2.3.1. Learning of feedforward connections

We utilize a variant of Hebbian correlation learning which prevents the changes of connection weights to grow without bounds. The stabilization is here achieved by a forgetting term that reduces the weight proportionally to the postsynaptic cell activation. The weight change for the receptive fields is formally defined by

$$\dot{w}_{jk}^{in} = \Omega(g_v(v_k)) \cdot \eta_{in} \cdot g_v(v_k) \cdot (g_u(u_j) - g_v(v_k) \cdot w_{jk}^{in}). \quad (12)$$

The r.h.s. of the equation is defined by the switch $\Omega(\cdot)$ to enable/disable neurons for adaptation of their weights and a learning rate η_{in} . The extended Hebbian correlation term is defined by $g_v(v_k) \cdot (g_u(u_j) - g_v(v_k) \cdot w_{jk}^{in})$. In other words, the learning is gated by the activation of the postsynaptic neuron. Here, the Hebbian term $g_u(u_j) \cdot g_v(v_k)$ is combined with the forgetting term $g_v(v_k)^2 \cdot w_{jk}^{in}$ to balance the temporal change and bound the growth of the cell's synaptic input weights. It has been demonstrated that such a learning mechanism extracts the first Eigenvector of the input distribution (Oja, 1982, 1992). Another property of the Oja learning rule is of even more interest here: The learning of the bottom-up feedforward weights approaches a fan-in connection pattern in which the weight energy is conserved (Dayan and Abbott, 2005). The fan-in weight vector \mathbf{w}_k^{in} is adapted over time to reach equilibrium, such that $\lim_{t \rightarrow \infty} \dot{\mathbf{w}}_k^{in} = \mathbf{v}_k \cdot \mathbf{u} - \gamma \mathbf{v}_k^2 \cdot \mathbf{w}_k^{in} = 0$ (with γ as a positive constant value that scales the balancing component). The equilibrium weight energy is then

$$\|\mathbf{w}_k^{in}\|^2 = \frac{1}{\gamma}. \quad (13)$$

Assuming $\gamma = 1$ we get a unit length for the input weights to single category nodes. This, in turn, prevents input activation distributions to bias the output activity at the category representation, given that the input activity distribution is normalized as well. The latter property is achieved by the normalization stage of the pool interaction defined in the activations dynamics of the network stages above.

2.3.2. Learning feedback connections

Again, we utilize a stabilized Hebbian weight adaptation formalism. In its dynamic formulation, the weight changes for projective fields is formally defined by

$$\dot{w}_{kj}^{out} = \Omega(g_v(v_k)) \cdot \eta_{out} \cdot g_v(v_k) \cdot (g_u(u_j) - w_{jk}^{out}). \quad (14)$$

As for the adaptation of the receptive field, or fan-in, weights (Equation 12) we utilize the switch $\Omega(\cdot)$ to enable/disable weight adaptation and a learning rate η_{out} for the projective, or fan-out, weights. The extended Hebbian term is here defined by $g_v(v_k) \cdot (g_u(u_j) - w_{jk}^{out})$. The learning is gated by the activation of the neuron that represents the category, which is presynaptic to the projective field considering the representation generated for the top-down feedback connections. Unlike the learning rule discussed in Equation (12), the forgetting term to balance the temporal change is controlled by the weight only. Such a weight adaptation mechanism defined in Equation (14) has been suggested for gated steepest descent learning in long-term memory formation, e.g., in Adaptive Resonance, or ART networks (Grossberg, 2013b). The adaptation of the fan-out weight vector \mathbf{w}_k^{out} over time reaches equilibrium, such that $\lim_{t \rightarrow \infty} \dot{\mathbf{w}}_k^{out} = \mathbf{v}_k \cdot \mathbf{u} - \gamma \mathbf{v}_k \cdot \mathbf{w}_k^{out} = 0$ (with γ as a positive constant value that scales the balancing component). The equilibrium weight energy is then

$$\mathbf{w}_k^{out} = \frac{1}{\gamma} \mathbf{u}. \quad (15)$$

Assuming $\gamma = 1$ we achieve a projective field, or fan-out, pattern for the connection weights corresponding to the (average) expected input activation represented in \mathbf{u} . Activation of a category node, thus, biases the receiving postsynaptic model neurons according to the predicted pattern the category expects to receive for its best tuning input. Feedback learning may also utilize the learning rule of Oja as for learning the feedforward connections described above. In this case the weight distribution of the projective field would converge to the first Eigenvalue of the expected input, instead of its mean. We have tested this and observed similar network performance. The latter implementation argues in favor of symmetric learning mechanisms for bottom-up and top-down connection weights. We decided to use a version in which the feedback projections approach the expected average input activation that represents the tuning of the individual categories, as in Equation (15).

2.4. FEEDBACK FOR SUBCATEGORY LEARNING

The mechanisms presented so far contributed to the feedforward as well as a generic feedback sweep of the model. The feedback sketched so far generically considered the modulatory influence a feedback signal has on any feedforward input representation. The mechanism emphasized the symmetry breaking property in which bottom-up signals gate the activity generation (at stage 2 of the processing cascade described in Section 2.2.1) which can be selectively amplified by the presence of matching feedback signals. Here, without incorporating the feedback from layer 3, the learning rules defined in Section 2.3 would successfully learn representations of input categories, but without the

potential of further refining them on a subcategorical level. As stated earlier, the feedback allows the estimation of the difference between the current input and the category assigned to the input after the feedforward sweep. Thus, layer 2 cells are able to combine the input with the derived difference signal. If the difference and the modulation strength after the feedback sweep is large enough, learning is potentially triggered such that an associated new subcategory is built using a so far unused layer 3 cell. The enhancement of the layer 2 responses by modulating feedback changes (Equation 5) to

$$\dot{u}_j = -\alpha_u \cdot u_j + \beta_u \cdot s_j \cdot (1 + \lambda \cdot res_j^{templ}) - u_j \cdot q_j, \quad (16)$$

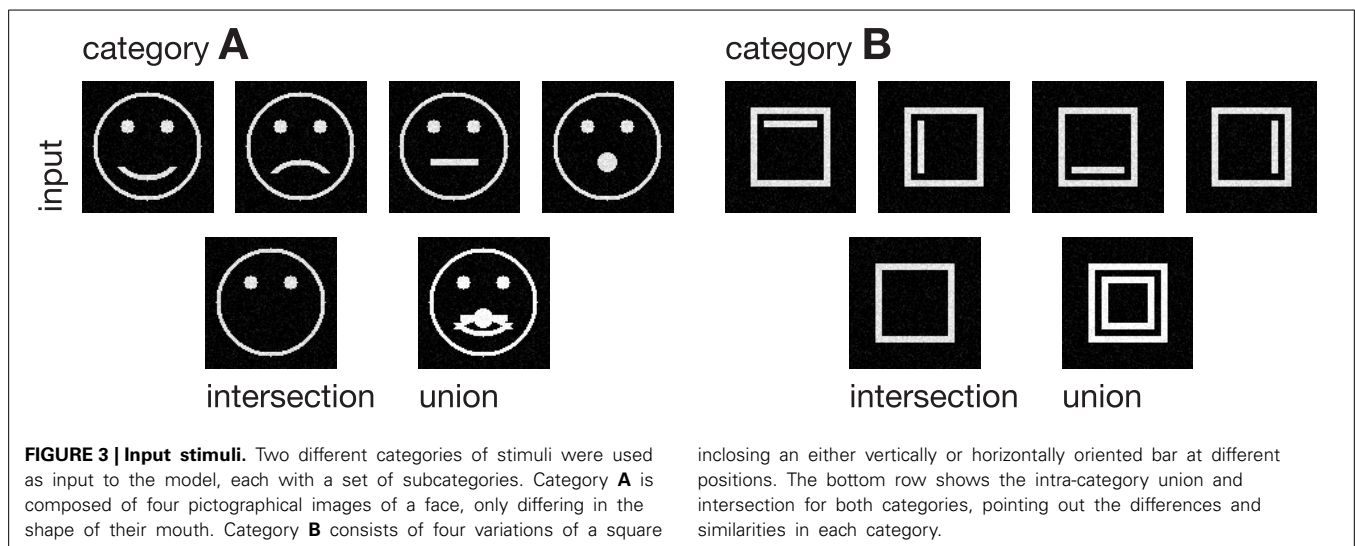
where res_j^{templ} denotes the residual signal derived from the feedback net_j^{FB} of the best matching category cell [selected by $\Omega(g_v(v_k))$] and the current activity $g_u(u_j)$. λ is controlling the influence of res_j^{templ} on u_j and thus is crucial for the extent of the difference between a modulated input and a category assigned in the feedforward sweep. The residual signal res_j^{templ} is defined by

$$res_j^{templ} = [g_u(u_j) - net_j^{FB}]^+ \\ = [g_u(u_j) - \Omega(g_v(v_k)) \cdot w_{kj}^{out}]^+, \quad (17)$$

with $[x]^+ = \max(0, x)$ denoting a rectification operation limiting res_j^{templ} to positive values. A closer look at the presented model dynamics may help us to reveal the potential roles that feedback plays in the context of category learning. According to Equation (17), the feedback signal acts as a predictive coding scheme, since net_j^{FB} expresses what the model expects how an input of a given category looks like on average. On the other hand, the expression $s_j \cdot (1 + \lambda \cdot res_j^{templ})$ in Equation (16) realizes a biased competition mechanism, favoring input components, which are in accordance with the residual signal res_j^{templ} . In essence, this kind of feedback incorporation results in an amplification of the differences between the currently best matching internal representation and the input. During learning, the difference between a category representation and individual instances of the category increases with the number of stimuli of the same category. If the effect of this difference on the input is large enough, a new subcategory representation is established.

3. RESULTS

In the following, we demonstrate the capabilities of the proposed model in learning category and subcategory representations using two categories of artificial input stimuli. As shown in **Figure 3**, category **A** contains four variations of a pictographical face. Category **B** is composed of four squares inclosing an either vertically or horizontally oriented bar at different positions. Without the loss of generality, we used very simplified stimuli to keep the computational complexity and in particular the necessary preprocessing steps as simple as possible. This allows us to keep the focus strictly on the role which feedback might play in the task



of category and subcategory learning. The stimuli were generated with the dimensions of 100×100 px with intensity values ranging from 0 to 1. The number of input units in layer 1 thus is always $100 \cdot 100 = 10000$ units. As mentioned in Section 2, the cells in layer 1 and those in layer 2 follow a pairwise connection scheme, so that layer 2 consists of the equivalent number of 10000 units. The number of layer 3 cells differs from experiment to experiment. Note that in all experiments, there remained at least one unused layer 3 cell after training, which was never selected for weight adaptation. Thus, the number of units in layer 3 never was a limitation to the establishment of a new category or subcategory representation. During training, Gaussian noise with mean $\mu = 0$ and a standard deviation of $\sigma = 0.05$ was added on each of the input stimuli, with values clipped to the range of $[0, 1]$. If not stated otherwise, we used learning rates of $\mu_{out} = \mu_{in} = 2^{-4}$ and a feedback gain factor of $\lambda = 2^5$. These values were found to be a suitable balance between the learning speed and the influence of the feedback. The parameters of the logistic function as defined in Equation (9) were set to $\mu_{log} = 700$ and $\kappa_{log} = 0.0075$, such that the transfer function results in a mean activation level of $g_v(700) = 0.5$ when roughly half of the input energy of one of the used stimuli is present in the input signal. The weights w^{in} and w^{out} of the category cells at model layer 3 were initialized with random values drawn from a normal distribution with mean $\mu = 0.75$ and a standard deviation of $\sigma = 0.1$, allowing empty category cells at layer 3 to be activated by just a small number of active input cells.

For the ease of computational complexity, we simulate the dynamics described in Section 2 using the corresponding steady-state equations. An in depth analysis of the activation dynamics can be found in Brosch and Neumann (2014a). Within the simulations, one (training) step—or iteration—corresponds to the presentation of one input stimulus, consisting of one feedforward and one feedback sweep through the model. Activities of the layer 3 cells are evaluated after the feedforward and after the feedback sweep and both trigger the adaptation of a categorial and/or subcategorial representation.

inclosing an either vertically or horizontally oriented bar at different positions. The bottom row shows the intra-category union and intersection for both categories, pointing out the differences and similarities in each category.

In total, we performed four experiments, each highlighting on a different aspect of the proposed model and learning mechanisms. In the first experiment, we show in principle how the model successfully learns a representation of a category of visual input stimuli and decomposes the category into subcategories. The second experiment is intended to demonstrate the invariance of the proposed learning mechanism to the order in which the stimuli are presented. Experiment 3 focuses on the importance of the feedback signal for the task of subcategory learning by contrasting Experiment 2 with a nearly identical experimental setup. The sole difference to Experiment 2 is that the incorporation of feedback is suppressed by setting the feedback gain parameter λ to $\lambda = 0$. In the last experiment we demonstrate how the model generalizes across the number of categories present in the input data and show how it successfully establishes representations for two categories of visual input and their subcategories.

All simulations were carried out using Mathworks Matlab R2014a.

3.1. EXPERIMENT 1

We trained the proposed model using the rectangular stimuli of category **B** as shown in Figure 3. The stimuli were presented in epochs of four blocks of sorted stimuli, each block containing 100 instances of one of the four rectangle variations. At model layer 3, six cells were used during the training. To slow down the weight adaptation process and highlight on the establishment of new subcategory representations, we used a learning rate of $\mu_{out} = \mu_{in} = 2^{-5}$, set μ_{log} to 800 and initialized w^{in} and w^{out} with random values drawn from a normal distribution with $\mu = 0.5$ and $\sigma = 0.1$. The activities of the layer 3 cells after the feedforward and the feedback sweep are shown in Figure 4 along with the corresponding weights w^{in} and w^{out} after several training steps. Over the first training steps, the model develops a combined representation of the first and the second rectangular shape containing information about the surrounding rectangle, as well as portions of information about the interior of the two shapes. After 200 training steps, the effect of the learning mechanism starts to be twofold. After the feedforward sweep, the overall category

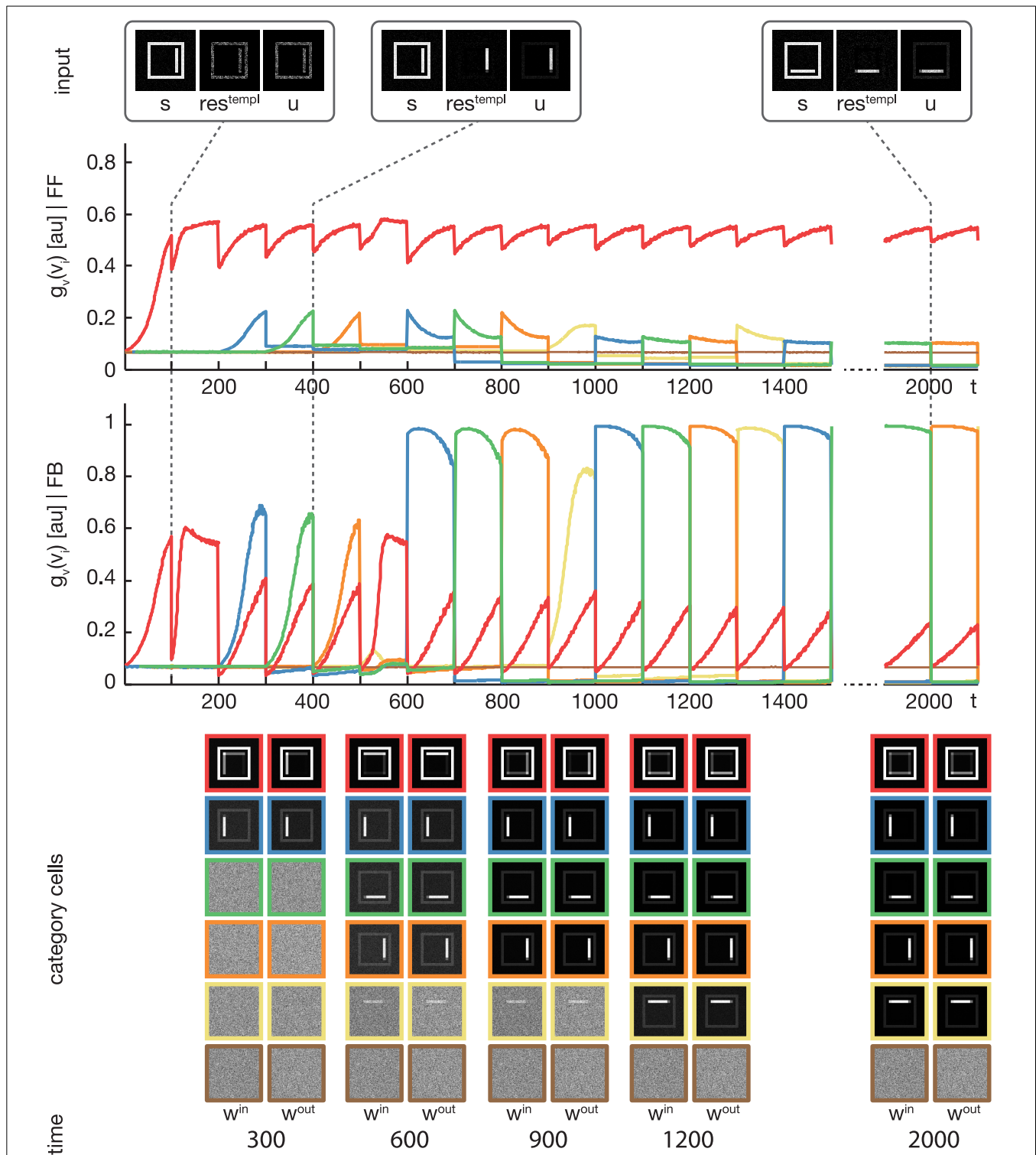


FIGURE 4 | Experiment 1. The model was trained using four rectangular shapes (see **Figure 3**, category **B**) as input stimuli. Stimuli were presented in sorted blocks. Six category cells (color-coded) were initialized with random weights. The first row shows exemplary input configurations s , along with the corresponding residual signal res^{templ} and the input signal u after feedback modulation. In the second row, the activities of the six category cells before the feedback sweep are shown. As can be seen, before the feedback is effective on the input, only one cell (encoded in red) responds to

all input configurations. This cell represents the overall category cell. The second row shows the activities after the feedback sweep. In the last row the corresponding category cell weights are displayed framed by colors according to the activity plots. It can be seen, that in the beginning all inputs are learned into one category cell. After about 200 training steps, the effect of the feedback is high enough to trigger the learning of a new subcategory representation. This process repeats several times, until each subcategory is represented by an own category cell.

representation is adapted to the current input stimulus. On the contrary, after the feedback sweep a subcategorical representation is learned by recruiting an additional layer 3 cell. The effect of the feedback signal now is large enough to suppress the outer rectangular shape and highlight on the differences between the overall category representation and the current input stimulus. This process continues until all of the four input variations are represented in an own subcategory cell. After learning, the feedforward sweep always results in a high activation level $g_v(v_i)$ of the overall category cell that represents the generic shape (refer to the second row of **Figure 4**). After the feedback sweep, however, the subcategory cell representing the specifics of the particular input stimulus is the one with the highest activation level.

3.2. EXPERIMENT 2

In the second experiment, the proposed model was trained using the pictographical faces of category **A** (see **Figure 3**) as input. Stimuli now were presented in random order. As in Experiment 1, six category cells at model layer 3 were used. All training parameters were set to their default values (see Section 3). **Figure 5** shows how category and subcategory cell representations are learned during the simulation. Again, the residual signal res^{temp} increases with the distinctiveness of the already established category representation and thus the effect of the feedback signal increases. Already after 21 training examples, the difference between the current input and the existing category cell is high enough to yield a modulation of the input effective-enough to evoke the establishment of a new subcategory. This process repeats several times, since after 127 learning iterations all of the variations of category **A** are represented in an own subcategory cell. Altogether, the model successfully learns category and subcategory representations, even though the stimuli are presented in random order.

3.3. EXPERIMENT 3

In a third experiment we conducted a simulation equivalent to the one in the second experiment, but now with disabling the feedback signal by setting $\lambda = 0$ (see **Figure 6**). As expected, without the feedback signal no subcategory representations are established and just one overall category representation is learned.

3.4. EXPERIMENT 4

For the last experiment we used both categories **A** and **B** shown in **Figure 3** as input stimuli. The parameters were equivalent to those described in Experiment 1 but now twelve category cells at layer 3 were initialized. Since the differences between the two types of stimuli (circular and rectangular) are already large enough before the feedback takes place, the model establishes two overall category representations and successively builds subcategories to these two categories. **Figure 7** shows the weights of the established two category, as well as the respective four subcategory cells after 1000 learning steps.

4. DISCUSSION

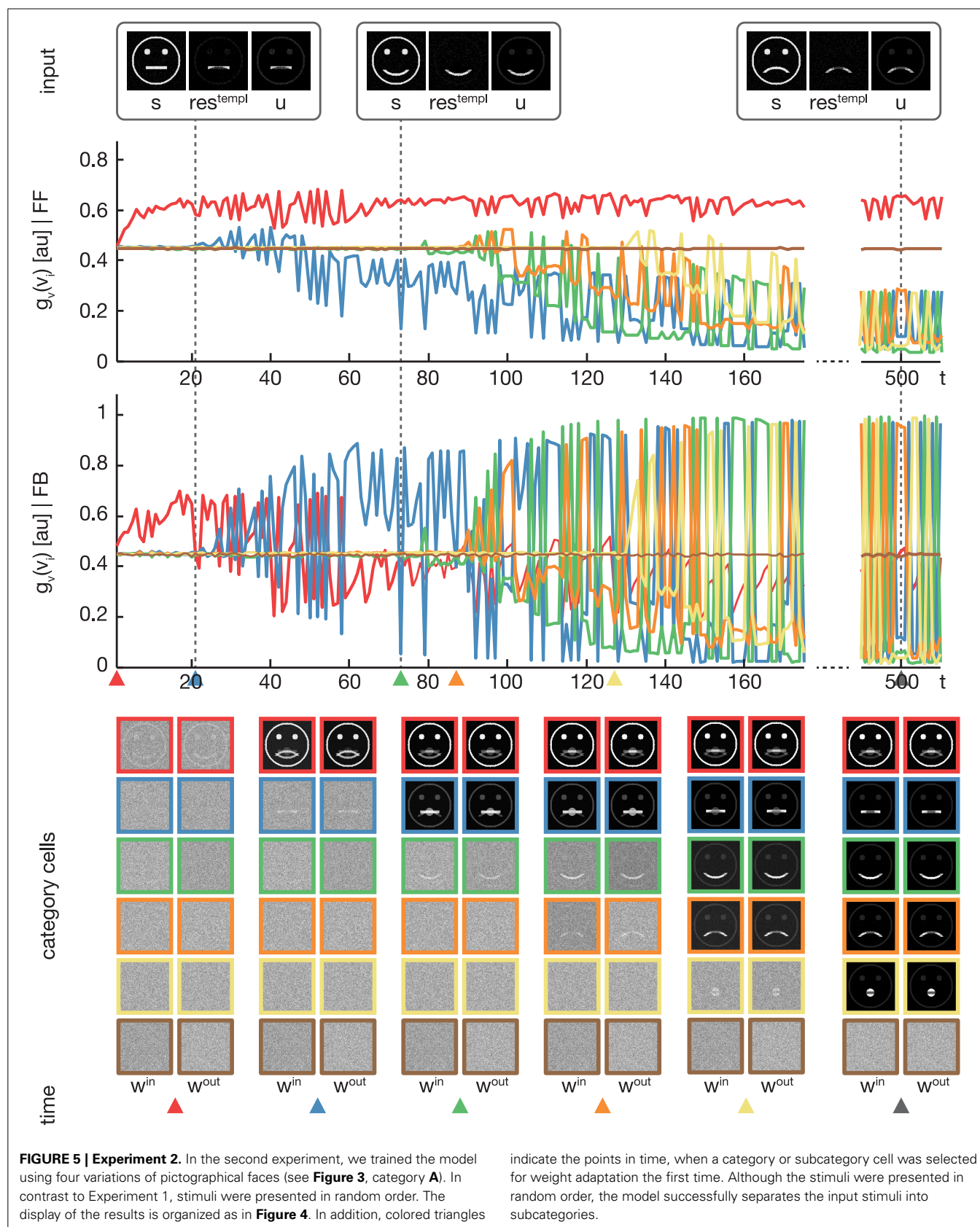
In this work we proposed a hierarchical architecture of cortical feedforward and feedback processing that builds upon previous work on the modeling of recurrent cortical dynamics (Neumann et al., 2007; Brosch and Neumann, 2014a). Here, we particularly

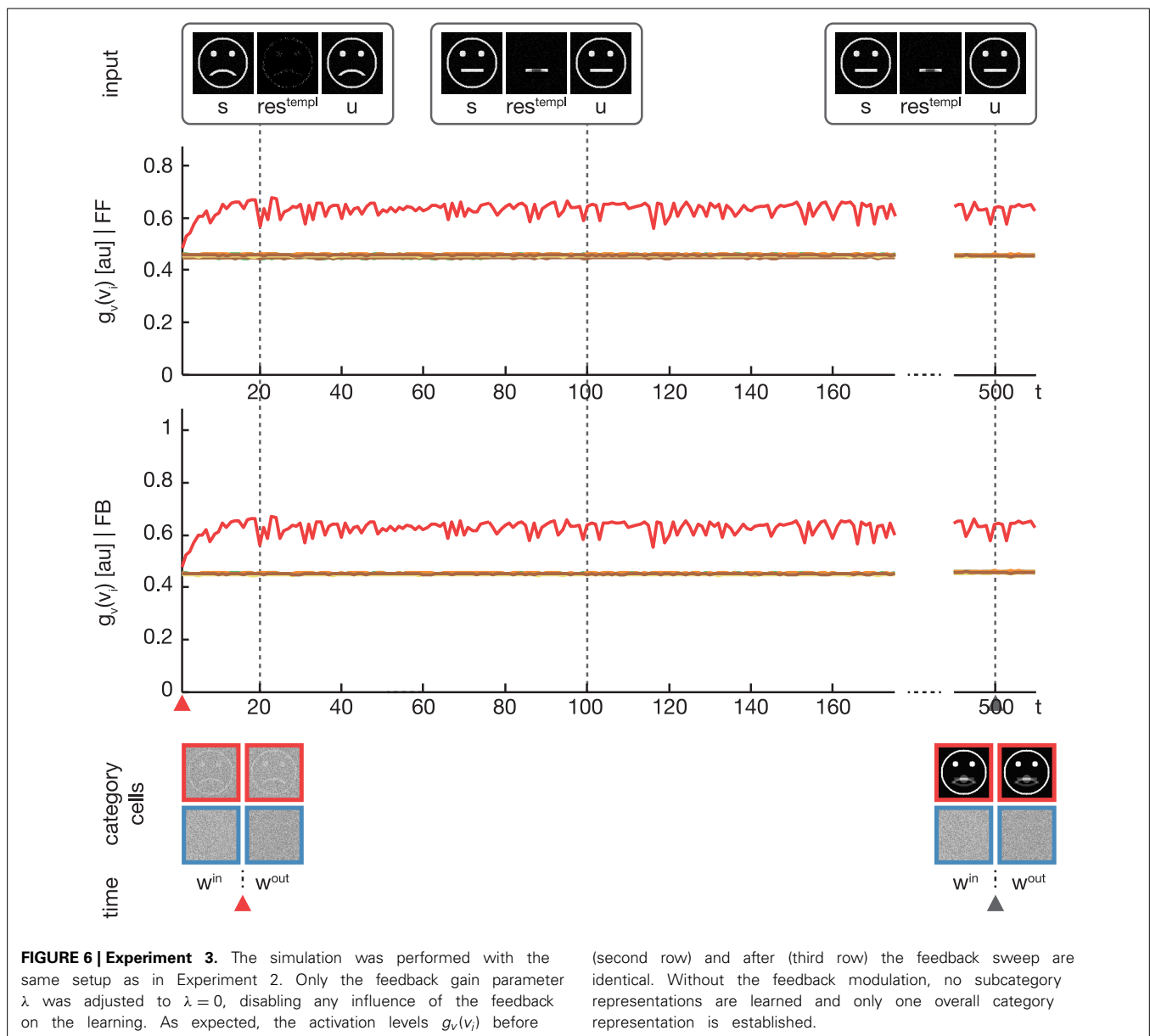
focused on the issue how in such networks feature or category representations could be automatically acquired by unsupervised learning mechanisms, which are seamlessly integrated in the recurrent architecture. The core computational elements assumed are cortical model columns that are abstractly described by a three-stage cascade of processing steps. The same elements have been utilized as generic mechanisms in models of form and motion processing, figure-ground segregation, as well as modeling biological motion perception that fuses segregated form and motion pathways (Neumann and Sepp, 1999; Bayerl and Neumann, 2004; Raudies et al., 2011; Layher et al., 2014). As a specific model feature, we have emphasized the role of feedback that modulates feedforward driving inputs such that their gain is increased dependent on the degree of correlation between feedforward and feedback signal activation. In conjunction with subsequent pool normalization the modulatory feedback sweeps realize a way of biased competition (Girard and Bullier, 1989; Desimone, 1998; Roelfsema et al., 2002; Reynolds and Heeger, 2009). The model now incorporates learning mechanisms to automatically build feature/category representations that are generated by the connection weights through adaptation.¹ Such learning allows to build representations that adapt their specificity to the statistics of the sensory input patterns.

4.1. SUMMARY OF CONTRIBUTIONS

The main contributions of the work presented in the paper are twofold. First, the investigated learning mechanisms occur in the feedforward as well as in the feedback connections. These are driven by bottom-up sensory input and top-down feedback signals to re-enter processing at earlier stages. The latter contain context information that allows to embed local sensory input signals into a larger behavioral context and predictions generated thereof. All this is in the spirit of multi-layer learning networks as discussed in Hinton (2007). In that sense feedforward connections will learn the specific configuration of an (average) appearance of an input feature pattern that the learned category is selectively tuned to. Considering static shape and form input the underlying structural principles are based on the cortical architecture of the ventral pathway with mutual interactions between such distributed representations in different cortical areas (Markov et al., 2014). The feedback connections, on the other hand, also learn by adjusting their weights in order to improve the predicted input pattern that maximally excites the feature/category representation. Second, the top-down feedback learning mechanism combines the modulatory feedback (Girard and Bullier, 1989; De Pasquale and Sherman, 2013) with the concept of top-down predictors that tend to minimize the residual

¹We make the distinction here between feature and category representation in order to emphasize the different locality of representations that are established at different layers in a hierarchical network architecture. With increasing integration sizes of cells at different levels more information from previous stages is integrated. The zones of lateral integration are more localized at earlier stages, thus, we refer to the learning of feature representations. At later stages the convergence zones may range over the full spatial input domain and, therefore, the representations already cover categories that could be shape or motion related.



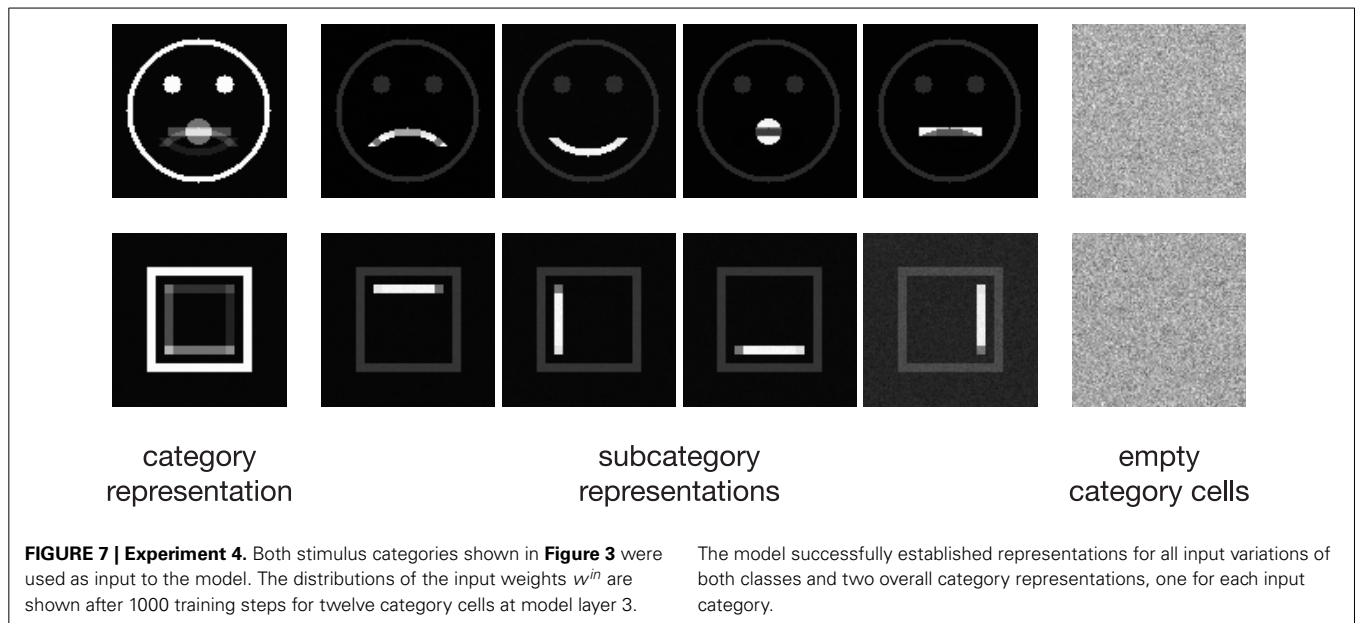


error between feedforward sensory signals and the top-down pattern (Rao and Ballard, 1999; Bastos et al., 2012). The idea behind this concept is that weights will be increased when the predicted pattern and the current input differ. The amount of this gain increase depends on the residual difference between these two patterns. The model defines the basis for more principled investigations how cortical sub-networks that are involved in different tasks might be established. In own previous work (Layher et al., 2014), distributed representations of spatio-temporal patterns in the cortical form and motion pathway were learned for articulated or biological motion perception (Johansson, 1973; Giese and Poggio, 2003). Here, sequence-selective representations were established by learning representations of convergent feedforward responses from form and motion representations. Also top-down weights are learned in which the projective field reaches the two

separate pathways of form and motion. The principles proposed in this work now allow to further develop the understanding of how such complex distributed representations can be learned and how average categories are learned together with subcategories for components that deviated significantly from the average category representation.

4.2. RELATION TO PREVIOUS MODELS OF CORTICAL LEARNING OF REPRESENTATIONS

Learning of feedforward networks has been investigated intensively before. Most importantly, the connection weights in multi-layer networks have been trained by using backpropagation to minimize the residual error of expected output given a specific input pattern (Lehky and Sejnowski, 1988, 1990; LeCun et al., 1989). Such approaches require a teacher signal that



determines the desired target output. The assumption of a supervisor involved in each teaching trial is biologically unrealistic in general. For that reason, a mechanism that is based on reinforcement learning (Sutton and Barto, 1981; Doya, 2007) has been suggested that combines an unspecific global reward-based reinforcement signal with an attentional signal that is backpropagated from the output layer to allow weight adaptation at those units that have been involved in the stimulus-response mapping in the previous processing of the input signals (Roelfsema and van Ooyen, 2005). Also, learning in hierarchical multi-stage architectures for object recognition has been investigated. Approaches range from random sampling of the input pattern space (Riesenhuber and Poggio, 1999; Serre et al., 2007; Mutch and Lowe, 2008; Serre and Poggio, 2010) to clustering techniques to arrive at sparse representations of the input via additional constraints on the connection weight patterns (Aharon et al., 2006) or auto-encoding that minimizes the reconstruction error of the input (LeCun et al., 1998). Recently, learning in multi-layer networks, so-called deep hierarchical networks (Bengio, 2009), has received renewed interest to build networks with high classification rate performance (LeCun et al., 1990; Hinton et al., 2006). Representations in such networks are learned in a sequential manner by learning the connection weight between pairs of layers, starting from the initial sensory-related level. Once learning converges, the next level connection weights are learned. This procedure is recurrently applied until all connections have been determined. The learning mechanisms are based on gradient descent type, for example, realizing stage-by-stage backpropagation learning. Unlike these proposals, the network mechanism here incorporates bidirectional learning of weights along the feedforward as well as the feedback path. The weight adaptation is based on variants of Hebbian correlation learning. These variants stabilize the growth properties of the input and output weight vectors to the computational elements (model columns) in the architecture. As a consequence, the representations built in the

connection patterns have specific interpretations: Along the feedforward path we assume an Oja learning scheme (Oja, 1982, 1992). As a result, the fan-in (or receptive field) weight energy of the total input connections from the previous layer neurons tends to be normalized for feedforward signal filtering. This ensures that different input patterns balance their input weights such they enter any subsequent competition or selection step in an unbiased fashion. Concerning feedback learning, connection weight patterns along the recurrent projection (corresponding to the projective field of a feature or category, Lehky and Sejnowski, 1990) approach the average expected input. In other words, the driving category representation generates a prediction pattern that covers the expected input activation that tends to match the tuning of the representation (Grossberg, 1980).

The proposed architecture is influenced by the conception of adaptive resonance theory (or ART; Grossberg, 1980, 1987; Carpenter, 1989). In a nutshell, learning in ART is organized in stages of feedforward and a feedback sweep processing. During feedforward processing the input signal is weighted by the connection pattern, or filter, between nodes in the feature representation and the category layer. These weightings are initialized by some random values. One category will gain a maximal input from the feature representation activated through the input signal, similar to the feedforward sweep in other networks (Rumelhart and Zipser, 1985), and also in the model proposed in this paper. Similarly, the self-organization of feature maps has also been approached by means of connection weight adaptation in hierarchically organized networks, establishing competitive processes for automatic map formation (von der Malsburg, 1973; Kohonen, 1982). The category that is maximally activated will subsequently suppress all other category representations by recurrent lateral center-surround competition. With supra-linear firing-rate functions such a competitive stage leads to a winner-take-all strategy (Grossberg, 1973). The weightings along the feedforward path can be adjusted to approach the (average) signal features. The

feedback connections fed by the winning category node (the projective field) are then allowed to adapt their weights as well so that they approach the input activation distribution. In other words, the feedback connections learn the input that maximally drives the currently activated category node to maintain a match between the input and the expectation the category has about its input patterns it is tuned to (resonance condition). If, instead, any momentary input feature pattern maximally drives a category with a top-down expectation pattern that does *not* match the input, then a mismatch occurs and the combined bottom-up and top-down expectation patterns annihilate. In order to now select another existing category or recruit a new category item, a reset wave is triggered that instantaneously shuts off the winning category that was activated maximally but has a mismatching representation in its projection field. This allows the top-down weights of a newly selected category to adjust in order to now better match the input that is coherent with the expected pattern represented by the active category representation (for recent comprehensive summaries and overviews of the ART principle, see Grossberg, 2013a,b). Discrete implementations of ART networks for pattern recognition have been described for binary as well as continuous input pattern representations (Carpenter and Grossberg, 1987a,b). A more specific reference to possible biophysical mechanisms underlying the recurrent interaction and learning has been described in Carpenter and Grossberg (1990), while Molenaar and Raijmakers (1997) presented a continuous time network implementation.

Several other network architectures use feedback connections that can be adapted through a learning process, e.g., (Elman, 1990; Hinton et al., 2006; Hinton, 2007; Lazar et al., 2009; Rolfe and LeCun, 2013). While Elman (1990) maps temporal feature history into an explicit representation through recurrences, a more recent approach by Lazar et al. (2009) utilizes a reservoir of connected neurons in a large pool to learn representations of temporal patterns. A read-out mechanism maps the internal state trajectories onto units through reduction of state-space dimension and clustering of activities. This recurrent network architecture with spiking model neurons emphasizes different mechanisms in the learning of connections weights, namely a simplified version of spike-timing dependent plasticity (STDP; Gerstner et al., 1996; Bi and Poo, 2001; Caporale and Dan, 2008) as unsupervised weight adaptation mechanism connecting excitatory cells in the pool, a synaptic scaling mechanisms through weight normalization, and an intrinsic plasticity mechanism for firing threshold adaptation. Our approach makes use of similar mechanism in the learning procedure. Here, we are concerned with networks of gradual activation dynamics, which motivates utilizing standard Hebbian correlation learning instead of the STDP rule. Weight normalization occurs implicitly in our adaptation mechanisms by utilizing modified Hebbian learning. In particular, as discussed in Section 2.3, the bottom-up learning of receptive field weights for individual category nodes approaches a weight energy (Equations 12 and 13). The intrinsic plasticity in our scheme is accomplished through the normalization activations, or firing rates, by the pool of cells in a neighborhood defined in the space-feature domain (compare Equations 5 and 6, Brosch and Neumann, 2014a). The model of Rolfe and LeCun

(2013) stresses the importance of acquisition of representations of categories and subcategories, like in our model. Their network realizes properties of deep networks establishing sparse representations of subcategories, like auto-encoder networks using binary state neuronal elements (Hinton and Salakhutdinov, 2006), and recurrently combine (hidden) representations and their predictions (Hinton et al., 2006) (see Hinton, 2007 for a review). Synaptic scaling (see a recent review in Tetzlaff et al., 2012) is addressed here from the perspective of how the receptive and projective fields learn a particular target activity distribution. In the architecture proposed by Rolfe and LeCun (2013) two types of units emerge that define parts and categories. The time course of the serial learning mechanism suggests that the network first establishes component representations mainly driven by the input. Later and with a slow learning efficacy, categories emerge that combine those units that belong to the category (while those they do not belong to are inhibited). Our proposed network architecture shares the idea of building hierarchical object representations. The acquisition of categories and subcategory, or part, representations operates oppositely: Categories are established as new representations recruiting free capacities from the long-term memory node reservoir in model layer 3 when the current input is significantly dissimilar in comparison to already existing categories. The deviations from a larger category then lead to learning subcategories and these are linked to their category representation by the temporal signature of the activation. Thus, the proposed model may start with only coarse-grained category knowledge, which is subsequently refined when more detailed information is available during the course of interacting with the environment.

While in these approaches the feedback connections serve to incorporate activations over time, feedback in ART architectures is intended to solve the stability-plasticity dilemma. The latter summarizes the necessity that an adaptive system needs to acquire or adapt to new evidence (or knowledge) and, at the same time, to keep those previously acquired representations stable (to prevent catastrophic forgetting). Our proposal differs from these previous model developments in several respects. In our architecture we build upon an abstract though biophysically plausible model of processing in cortical columns. The interaction between signal activations in bottom-up and top-down sweeps is based on modulatory feedback that enhances those sensory signal activation patterns which match the top-down template of activation that is re-entered at earlier stages of processing along the hierarchy. Thus, instead of a similarity calculation between signal patterns, a biologically plausible gain adjustment is assumed (Sherman and Guillery, 1998). The modulation signal we use for the amplification of the input signals is calculated by the difference between the current input signal and the top-down expectation pattern. This effectively combines the key mechanisms underlying the two current main theories of the role of feedback in cortex: top-down modulation and biased competition is assumed for the enhancement of the input gain. Here, the modulation strength is controlled by the difference between bottom-up and top-down signal, or the residual between these two activation patterns. Steering the amount of weight adaptation by the difference between signal and expectation template incorporates the flavor of predictive coding approaches (Rao and Ballard, 1999; Rauss

et al., 2011; Bastos et al., 2012). The logic behind this strategy is that the relative enhancement is reduced monotonically the more the top-down prediction signal approaches the bottom-up signal. As a consequence, the update of the weights will more quickly converge since both, feedforward and feedback, signal remain approximately constant and the weighting pattern approaches the prediction template. Consequently, no external reset mechanism is required that explicitly detects a mismatch discrepancies by a threshold vigilance parameter, as in ART models. In our proposal, the feedback modulatory dynamics and the learning mechanisms automatically tune the average matching activation of the responding category and also select the category or feature representation. Furthermore, and potentially of even more interest is the automatic establishment of categorical representations that capture the average of the input patterns that can drive the corresponding nodes in the columnar architecture. At the same time, subcategory representations are established that represent the significant differences in the detailed feature configurations that differ from the average case. This has been demonstrated in example cases (Section 3) in which, for example, faces are distinguished from non-faces at the categorical level. Smiling facial appearances or faces where the eyes are closed are then also automatically assigned to the average category by learning. However, to distinguish the appearance differences new subcategories are automatically established and learned. This selectivity is realized by two core mechanisms. First, the realization of an associative memory through the combination of an instar with an outstar learning scheme (see Carpenter, 1989), which allows the assignment of a given input to the currently best matching internal representation, as well as the corresponding feedback pattern. Second, the modulatory amplification of the differences between an input signal and the feedback pattern associated with the best matching internal representation of the input. If the amplification after the feedback sweep is effective enough, the correlation between the modulated input and an empty category cell will be higher than to the category representation the input was assigned to in the feedforward sweep. Thus, learning will be triggered for the so far unused category cell and a new subcategory will be built.

The computational mechanisms of activation and weight dynamics support principles that have been predicted to minimize the computational efforts of visual systems to successfully deal with the complexity problem of perception (Tsotsos, 1988, 2005). The hierarchical organization of representations in model areas, the receptive field properties of model columns, the hierarchical pooling of spatially separated input representations, and the top-down feedback together with unsupervised learning are structural principles that enable the visual system to successfully cope with complex input stimuli that are behaviorally relevant. The presented model is able to build the underlying distributed representations at low, intermediate, and higher levels in the cortical hierarchy by means of key cortical principles.

4.3. FEEDBACK—MODULATORS AND PREDICTORS

The hierarchical model architecture proposed here is composed of multiple model areas each of which is represented by a

three-stage columnar cascade model. The cascade consists of input filtering, activity modulation of filter outputs by re-entrant signals, and competitive center-surround interaction of target cells against a pool of cells. The latter stage yields an activity normalization for generating net output responses. Together with the gain enhancement generated by input modulation via re-entrant signals the network interactions achieve a biased competition response characteristics (Desimone, 1998; Reynolds and Heeger, 2009; Carandini and Heeger, 2012). The proposed architecture can be interpreted as an abstracted compartment representation of the layered architecture of cortical areas (Self et al., 2012). The interplay between the normalization of activities and the selective enhancement of activities via feedback establishes the dynamics of cortical processing. Activity normalization at the output stage is computed by a mechanism of shunting inhibition, like the non-linear divisive mechanisms proposed in Carandini and Heeger (1994); Carandini et al. (1999); Kouh and Poggio (2008); Carandini and Heeger (2012) (see Brosch and Neumann, 2014a for a formal analysis of the computational properties). Feedback signals generated at higher-level cortical stages or parallel processing pathways provide context information that is re-entered at the current stage of the processing hierarchy (Grossberg, 1980; Edelman, 1993). While the presence of feedback connections is a well-established principle of cortical signal processing and integration, the exact role of how such feedback signals are re-entered at the earlier stages is a controversial topic of ongoing investigation. We adopt here two principles from the two major frameworks of the functionality of feedback, namely modulatory feedback to bias subsequent competitive mechanisms and predictive coding.

How feedback signals interact and combine with signals delivered in the driving feedforward stream is yet unresolved. Two major conceptual ideas have been developed, each receiving support by experimental evidence (Markov and Kennedy, 2013). In a nutshell, *biased competition* suggests that signals in the feedforward pathway are enhanced by top-down templates (represented by activity distributions) such that they receive a competitive advantage in subsequent mutually competitive processes. As a result, feature responses that receive feedback have a higher gain which, in turn, leads to stronger suppression of activities that were not enhanced (Girard and Bullier, 1989; Desimone, 1998; Roelfsema et al., 2002; Reynolds and Heeger, 2009). In *predictive coding* the goal of computation is to reduce the residual error between the feedforward signal and the (top-down) templates generated at a stage that generates an expectation about the most compatible input. This idea is based upon predictor-corrector mechanisms in optimization (Ullman, 1995; Rao and Ballard, 1999; Bastos et al., 2012). As a consequence the state trajectory of such systems and their activations are different: While in biased competition the activations of the representations that match the predictions will increase, they will decrease in the predictive coding framework. Interestingly, Spratling (2008) has shown that these two approaches are functionally equivalent when the feedback in the biased competition is additive. Here, we utilize multiplicative feedback based on the linking mechanism suggested by Eckhorn et al. (1990); Eckhorn (1999) to account for activity synchronization in networks of spiking neurons and

further evidence that signal amplifications occur at the level of cortical pyramidal cells (Larkum, 2013) (see a model description in Brosch and Neumann, 2014b that accounts for these findings). An influential paper by Crick and Koch (1998) provided strong support for modulatory top-down connections based on theoretical grounds. In the model framework proposed here we adopt the framework of modulatory feedback (thus, biased competition). The feedback signals represent context-sensitive templates and are gated by feedforward driving input signals. In such a modulating feedback driven gain control mechanism spatial detail is generated by feature-driven low-level processes and representations and subsequently associated with coarse-grained context information which is provided by intermediate and higher-levels of cortical computation (Lamme and Roelfsema, 2000; Roelfsema et al., 2002; Roelfsema, 2006). In order to control the weight adaptation for learning, the strength of feedback is calculated by the difference between the feedforward signal and the predictive template that is delivered along the top-down connections. Such a difference represents the residual between the two counter stream representations (Ullman, 1995). In a nutshell, the idea is that the amount of feedback is regulated by the deviation between the two convergent streams (like in predictive coding). The re-entrant combination is, however, based on multiplicative gain enhancement. The strength of the excitatory feedback will vanish when the input is perfectly predicted by the top-down template. In that case, the feedforward signal representation will not be further enhanced. In Bastos et al. (2012) the cortical circuits are present in different compartments of a cortical area (compare Self et al., 2013 for a discussion of the possible roles of input layer and superficial and deep layer compartments in cortical area V1). Our suggested mechanism can be realized assuming subtractive interaction between driving feedforward cells and feedback signals, potentially in the superficial layer compartment. The resulting residual activations can then activate cells in columns via the apical dendrites of pyramidal cells (located either in the superficial or deep layer compartments; Larkum et al., 2004). In Brosch and Neumann (2014b) a firing-rate model of pyramidal cell interaction has been developed that explains such interactions at the level of the columnar architecture adopted here. All these feedforward and feedback interactions combine with learning mechanisms for the feedforward and the feedback connections. The equations supposed to define the weight changes lead to stable convergent weight changes. In the feedforward connection pattern the fan-in, or receptive field, weights to a unit approach a defined weight energy, or length, of the connection coefficients. This is desirable since after a representation accomplished in the weights has been settled, the activation level is not biased by the weights but is determined by the signal input and its changed gain through feedback interaction. In the feedback connection pattern the fan-out, or projective field, weights from a unit approach the (average) activity the representation is tuned to. Thus, the expected input is represented which can be activated as top-down template to instantiate the expected input signal or feature configuration. This leads to resonances in cases where the top-down expectation is retrieved from already established knowledge. In cases of mismatches new feature/category representations can

be automatically recruited to establish new knowledge in the learning cortical architecture.

4.4. MODEL LIMITATIONS AND EXTENSIONS

The proposed model architecture emphasized the computational role of feedforward and feedback mechanisms in order to generate interactive states, or resonances, in a hierarchically organized model system. The re-entrant feedback is assumed to be modulatory such that bottom-up feedforward signals gate the recurrent feedback activations. The interactive processing is combined with a learning mechanism that allows to adjust connection weights along the feedforward as well as the feedback pathways. We have demonstrated the general functionality by using simple shapes that are kept under full control during the design process. Also we employed only a pair of interacting cortical model areas, each composed as a sheet of columnar units with lateral interactions. In addition, a separate input layer that represents the stimulus was incorporated. The proposed model architecture may be investigated along several lines of questions.

In its current form, the proposed model architecture separately evaluates the activities of layer 3 category and subcategory cells before and after the modulation of the residual feedback on the input signal. This results in an activity pattern in which only an overall category cell *or* a subcategory cell can be active at a time. It would be interesting to integrate an additional mechanism which prevents such fluctuations and keeps both the overall category and the subcategory cell active in parallel.

Deep hierarchies have been proposed to accomplish the build-up of rich composite feature representations at different stages of hierarchically organized networks for solving detection and recognition tasks (LeCun et al., 1998; Hinton, 2007; Bengio, 2009). A natural extension of the simplified architecture studied in this paper is to add further model cortical areas and train the feedforward and feedback connection weights at each level. We expect that such an extended architecture allows the construction of multi-level representations of pattern compositions over several stages in a hierarchy. Such an approach should provide the generic structure to automatically build representations of fragments of input stimuli in which recognition is combined with segmenting inputs using the learned top-down templates (Ullman et al., 2002; Ullman, 2007).

The proposed scheme currently utilizes simple input patterns to build categories and associated subcategories to make explicit the variations that deviate from the average category representations. It would be interesting to study the responses for more realistic shape patterns presented as gray level inputs that provide the input to the network architecture. Also in this case, it would be interesting to study the multi-level steps necessary for the proposed model cortical architecture to accomplish the category learning under even more realistic input representations. In a technical instance of processing Borenstein and Ullman (2008) proposed an image segmentation scheme based on bottom-up signal driven processing that is combined with top-down processing to utilize knowledge for improved segmentation. Although the focus there is mainly on the improvement of image processing, the approach might serve as an inspiration for modeling as well. We suggest that the potential power of the network

architecture proposed in this work lies in the automatic learning of templates for feedback expectation (at low and intermediate levels of representation; Hinton, 2007) that could be evaluated in terms of their information content for visual classification tasks (Ullman et al., 2002).

AUTHOR CONTRIBUTIONS

Conceived and designed the experiments: Georg Layher, Heiko Neumann; Performed the experiments: Georg Layher; Analyzed the data: Georg Layher, Fabian Schrodt, Martin V. Butz, Heiko Neumann; Model implementation: Georg Layher; Wrote the paper: Georg Layher, Fabian Schrodt, Martin V. Butz, Heiko Neumann.

ACKNOWLEDGMENTS

Georg Layher and Heiko Neumann have been supported by the Transregional Collaborative Research Centre SFB/TRR 62 “A Companion Technology for Cognitive Technical Systems” funded by the German Research Foundation (DFG).

REFERENCES

- Aharon, M., Elad, M., and Bruckstein, A. (2006). K-SVD: an algorithm for designing overcomplete dictionaries for sparse representation. *IEEE Trans. Signal Process.* 54, 4311–4322. doi: 10.1109/TSP.2006.881199
- Bastos, A. M., Usrey, W. M., Adams, R. A., Mangun, G. R., Fries, P., and Friston, K. J. (2012). Canonical microcircuits for predictive coding. *Neuron* 76, 695–711. doi: 10.1016/j.neuron.2012.10.038
- Bayerl, P., and Neumann, H. (2004). Disambiguating visual motion through contextual feedback modulation. *Neural Comput.* 16, 2041–2066. doi: 10.1162/0899766041732404
- Bengio, Y. (2009). Learning deep architectures for ai. *Found. Trends Mach. Learn.* 2, 1–127. doi: 10.1561/2200000006
- Bi, G.-Q., and Poo, M.-M. (2001). Synaptic modification by correlated activity: Hebb's postulate revisited. *Annu. Rev. Neurosci.* 24, 139–166. doi: 10.1146/annurev.neuro.24.1.139
- Borenstein, E., and Ullman, S. (2008). Combined top-down/bottom-up segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* 30, 2109–2125. doi: 10.1109/TPAMI.2007.70840
- Bouecke, J. D., Tlapale, E., Kornprobst, P., and Neumann, H. (2011). Neural mechanisms of motion detection, integration, and segregation: from biology to artificial image processing systems. *EURASIP J. Adv. Signal Process.* 2011, 6:1–6:22. doi: 10.1155/2011/781561
- Brosch, T., and Neumann, H. (2014a). Computing with a canonical neural circuits model with pool normalization and modulating feedback. *Neural Comput.* 26, 2735–2789. doi: 10.1162/NECO_a_00682
- Brosch, T., and Neumann, H. (2014b). Interaction of feedforward and feedback streams in visual cortex in a firing-rate model of columnar computations. *Neural Netw.* 54, 11–16. doi: 10.1016/j.neunet.2014.02.005
- Caporale, N., and Dan, Y. (2008). Spike timing-dependent plasticity: a hebbian learning rule. *Annu. Rev. Neurosci.* 31, 25–46. doi: 10.1146/annurev.neuro.31.060407.125639
- Carandini, M., and Heeger, D. J. (1994). Summation and division by neurons in primate visual cortex. *Science* 264, 1333–1336. doi: 10.1126/science.8191289
- Carandini, M., and Heeger, D. J. (2012). Normalization as a canonical neural computation. *Nat. Rev. Neurosci.* 13, 51–62. doi: 10.1038/nrn3136
- Carandini, M., Heeger, D. J., and Anthony Movshon, J. (1999). “Linearity and gain control in v1 simple cells,” in *Models of Cortical Circuits*, Volume 13 of *Cerebral Cortex*, eds P. S. Ullinski, E. G. Jones, and A. Peters (New York, NY: Springer US, Kluwer Academic/ Plenum Publishers), 401–443. doi: 10.1007/978-1-4615-4903-1_7
- Carpenter, G. A. (1989). Neural network models for pattern recognition and associative memory. *Neural Netw.* 2, 243–257. doi: 10.1016/0893-6080(89)90035-X
- Carpenter, G. A., and Grossberg, S. (1987a). ART 2: self-organization of stable category recognition codes for analog input patterns. *Appl. Opt.* 26, 4919–4930. doi: 10.1364/AO.26.004919
- Carpenter, G. A., and Grossberg, S. (1987b). A massively parallel architecture for a self-organizing neural pattern recognition machine. *Comput. Vision Graph. Image Process.* 37, 54–115. doi: 10.1016/S0734-189X(87)80014-2
- Carpenter, G. A., and Grossberg, S. (1990). {ART} 3: hierarchical search using chemical transmitters in self-organizing pattern recognition architectures. *Neural Netw.* 3, 129–152. doi: 10.1016/0893-6080(90)90085-Y
- Carpenter, G. A., and Grossberg, S. (2003). “Adaptive resonance theory,” in *The Handbook of Brain Theory and Neural Networks, Second Edition*, ed M. A. Arbib (Cambridge, MA: MIT Press), 87–90.
- Crick, F., and Koch, C. (1998). Constraints on cortical and thalamic projections: the no-strong-loops hypothesis. *Nature* 391, 245–250. doi: 10.1038/34584
- Dayan, P., and Abbott, L. F. (2005). *Theoretical Neuroscience: Computational and Mathematical Modeling of Neural Systems*. Cambridge, MA: The MIT Press.
- De Pasquale, R., and Sherman, S. M. (2013). A modulatory effect of the feedback from higher visual areas to V1 in the mouse. *J. Neurophysiol.* 109, 2618–2631. doi: 10.1152/jn.01083.2012
- Desimone, R. (1998). Visual attention mediated by biased competition in extrastriate visual cortex. *Philos. Trans. R. Soc. Lond. Ser. B Biol. Sci.* 353, 1245–1255. doi: 10.1098/rstb.1998.0280
- Doya, K. (2007). Reinforcement learning: computational theory and biological mechanisms. *HFSP J.* 1, 30–40. doi: 10.2976/1.2732246/10.2976/1
- Eckhorn, R. (1999). Neural mechanisms of visual feature binding investigated with microelectrodes and models. *Visual Cogn.* 6, 231–265. doi: 10.1080/135062899394975
- Eckhorn, R., Reitboeck, H. J., Arndt, M., and Dicke, P. (1990). Feature linking via synchronization among distributed assemblies: simulations of results from cat visual cortex. *Neural Comput.* 2, 293–307. doi: 10.1162/neco.1990.2.3.293
- Edelman, G. M. (1993). Neural darwinism: selection and reentrant signaling in higher brain function. *Neuron* 10, 115–125. doi: 10.1016/0896-6273(93)90304-A
- Elman, J. L. (1990). Finding structure in time. *Cogn. Sci.* 14, 179–211. doi: 10.1207/s15516709cog1402_1
- Gerstner, W., Kempter, R., van Hemmen, J. L., and Wagner, H. (1996). A neuronal learning rule for sub-millisecond temporal coding. *Nature* 383, 76–78. doi: 10.1038/383076a0
- Giese, M., and Poggio, T. (2003). Neural mechanisms for the recognition of biological movements. *Nat. Rev. Neurosci.* 4, 179–192. doi: 10.1038/nrn1057
- Girard, P., and Bullier, J. (1989). Visual activity in area V2 during reversible inactivation of area 17 in the macaque monkey. *J. Neurophysiol.* 62, 1287–1302.
- Grossberg, S. (1973). Contour enhancement, short-term memory, and constancies in reverberating neural networks. *Stud. Appl. Math.* 52, 213–257.
- Grossberg, S. (1980). How does a brain build a cognitive code. *Psychol. Rev.* 87, 1–51. doi: 10.1037/0033-295X.87.1.1
- Grossberg, S. (1987). Competitive learning: from interactive activation to adaptive resonance. *Cogn. Sci.* 11, 23–63. doi: 10.1111/j.1551-6708.1987.tb00862.x
- Grossberg, S. (1988). Nonlinear neural networks: principles, mechanisms, and architectures. *Neural Netw.* 1, 17–61. doi: 10.1016/0893-6080(88)90021-4
- Grossberg, S. (2013a). Adaptive resonance theory. *Scholarpedia* 8, 1569. doi: 10.4249/scholarpedia.1569
- Grossberg, S. (2013b). Adaptive resonance theory: how a brain learns to consciously attend, learn, and recognize a changing world. *Neural Netw.* 37, 1–47. doi: 10.1016/j.neunet.2012.09.017
- Hebb, D. O. (1949). *The Organization of Behavior*. New York, NY: Wiley.
- Hinton, G., Osindero, S., and Teh, Y.-W. (2006). A fast learning algorithm for deep belief nets. *Neural Comput.* 18, 1527–1554. doi: 10.1162/neco.2006.18.7.1527
- Hinton, G. E. (2007). Learning multiple layers of representation. *Trends Cogn. Sci.* 11, 428–434. doi: 10.1016/j.tics.2007.09.004
- Hinton, G. E., and Salakhutdinov, R. R. (2006). Reducing the dimensionality of data with neural networks. *Science* 313, 504–507. doi: 10.1126/science.1127647

- Johansson, G. (1973). Visual perception of biological motion and a model for its analysis. *Percept. Psychophys.* 14, 201–211. doi: 10.3758/BF03212378
- Kohonen, T. (1982). Self-organized formation of topologically correct feature maps. *Biol. Cybernet.* 43, 59–69. doi: 10.1007/BF00337288
- Kouh, M., and Poggio, T. (2008). A canonical neural circuit for cortical non-linear operations. *Neural Comput.* 20, 1427–1451. doi: 10.1162/neco.2008.02-07-466
- Lamme, V. A., and Roelfsema, P. R. (2000). The distinct modes of vision offered by feedforward and recurrent processing. *Trends Neurosci.* 23, 571–579. doi: 10.1016/S0166-2236(00)01657-X
- Larkum, M. (2013). A cellular mechanism for cortical associations: an organizing principle for the cerebral cortex. *Trends Neurosci.* 36, 141–151. doi: 10.1016/j.tins.2012.11.006
- Larkum, M. E., Senn, W., and Lüscher, H. R. (2004). Top-down dendritic input increases the gain of layer 5 pyramidal neurons. *Cereb. Cortex* 14, 1059–1070. doi: 10.1093/cercor/bhh065
- Layher, G., Giese, M. A., and Neumann, H. (2014). Learning representations of animated motion sequences a neural model. *Top. Cogn. Sci.* 6, 170–182. doi: 10.1111/tops.12075
- Lazar, A., Pipa, G., and Triesch, J. (2009). Sorn: a self-organizing recurrent neural network. *Front. Comput. Neurosci.* 3:23. doi: 10.3389/neuro.10.023.2009
- LeCun, Y., Boser, B., Denker, J. S., Henderson, D., Howard, R. E., Hubbard, W., et al. (1989). Backpropagation applied to handwritten zip code recognition. *Neural Comput.* 1, 541–551. doi: 10.1162/neco.1989.1.4.541
- LeCun, Y., Boser, B., Denker, J. S., Howard, R. E., Habbard, W., Jackel, L. D., et al. (1990). “Advances in neural information processing systems 2,” in *Chapter Handwritten Digit Recognition with a Back-propagation Network*, ed D. S. Touretzky (San Francisco, CA: Morgan Kaufmann Publishers Inc.), 396–404.
- LeCun, Y., Bottou, L., Bengio, Y., and Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proc. IEEE* 86, 2278–2324. doi: 10.1109/5.726791
- Lehky, S. R., and Sejnowski, T. J. (1988). Network model of shape-from-shading: neural function arises from both receptive and projective fields. *Nature* 333, 452–454. doi: 10.1038/333452a0
- Lehky, S. R., and Sejnowski, T. J. (1990). Neural network model of visual cortex for determining surface curvature from images of shaded surfaces. *Proc. R. Soc. Lond. B Biol. Sci.* 240, 251–278. doi: 10.1098/rspb.1990.0037
- Lui, J. H., Hansen, D. V., and Kriegstein, A. R. (2011). Development and evolution of the human neocortex. *Cell* 146, 18–36. doi: 10.1016/j.cell.2011.06.030
- Markov, N. T., and Kennedy, H. (2013). The importance of being hierarchical. *Curr. Opin. Neurobiol.* 23, 187–194. doi: 10.1016/j.conb.2012.12.008
- Markov, N. T., Vezoli, J., Chameau, P., Falchier, A., Quilodran, R., Huissoud, C., et al. (2014). Anatomy of hierarchy: feedforward and feedback pathways in macaque visual cortex. *J. Comp. Neurol.* 522, 225–259. doi: 10.1002/cne.23458
- Molenaar, P. C., and Raijmakers, M. E. (1997). Exact ART: a complete implementation of an ART network. *Neural Netw.* 10, 649–669. doi: 10.1016/S0893-6080(96)00111-6
- Mountcastle, V. B. (1997). The columnar organization of the neocortex. *Brain* 120(Pt 4), 701–722. doi: 10.1093/brain/120.4.701
- Mutch, J., and Lowe, D. G. (2008). Object class recognition and localization using sparse features with limited receptive fields. *Int. J. Comput. Vision* 80, 45–57. doi: 10.1007/s11263-007-0118-0
- Neumann, H., and Sepp, W. (1999). Recurrent V1-V2 interaction in early visual boundary processing. *Biol. Cybern.* 81, 425–444. doi: 10.1007/s004220050573
- Neumann, H., Yazdanbakhsh, A., and Mingolla, E. (2007). Seeing surfaces: the brain's vision of the world. *Phys. Life Rev.* 4, 189–222. doi: 10.1016/j.plrev.2007.09.001
- Oja, E. (1982). Simplified neuron model as a principal component analyzer. *J. Math. Biol.* 15, 267–273. doi: 10.1007/BF00275687
- Oja, E. (1992). Principal components, minor components, and linear neural networks. *Neural Netw.* 5, 927–935. doi: 10.1016/S0893-6080(05)80089-9
- Poort, J., Raudies, F., Wannig, A., Lamme, V. A., Neumann, H., and Roelfsema, P. R. (2012). The role of attention in figure-ground segregation in areas V1 and V4 of the visual cortex. *Neuron* 75, 143–156. doi: 10.1016/j.neuron.2012.04.032
- 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. doi: 10.1038/4580
- Raudies, F., Mingolla, E., and Neumann, H. (2011). A model of motion transparency processing with local center-surround interactions and feedback. *Neural Comput.* 23, 2868–2914. doi: 10.1162/NECO_a_00193
- Rauss, K., Schwartz, S., and Pourtois, G. (2011). Top-down effects on early visual processing in humans: a predictive coding framework. *Neurosci. Biobehav. Rev.* 35, 1237–1253. doi: 10.1016/j.neubiorev.2010.12.011
- Reynolds, J. H., and Heeger, D. J. (2009). The normalization model of attention. *Neuron* 61, 168–185. doi: 10.1016/j.neuron.2009.01.002
- Riesenhuber, M., and Poggio, T. (1999). Hierarchical models of object recognition in cortex. *Nat. Neurosci.* 2, 1019–1025. doi: 10.1038/14819
- Roelfsema, P. R. (2006). Cortical algorithms for perceptual grouping. *Annu. Rev. Neurosci.* 29, 203–227. doi: 10.1146/annurev.neuro.29.051605.112939
- Roelfsema, P. R., Lamme, V. A. F., Spekrijse, H., and Bosch, H. (2002). Figure-ground segregation in a recurrent network architecture. *J. Cogn. Neurosci.* 14, 525–537. doi: 10.1162/08999290260045756
- Roelfsema, P. R., Tolboom, M., and Khayat, P. S. (2007). Different processing phases for features, figures, and selective attention in the primary visual cortex. *Neuron* 56, 785–792. doi: 10.1016/j.neuron.2007.10.006
- Roelfsema, P. R., and van Ooyen, A. (2005). Attention-gated reinforcement learning of internal representations for classification. *Neural Comput.* 17, 2176–2214. doi: 10.1162/0899766054615699
- Rolfe, J. T., and LeCun, Y. (2013). Discriminative recurrent sparse auto-encoders. arXiv:1301.3775. Available online at: <http://adsabs.harvard.edu/abs/2013arXiv1301.3775R>
- Rumelhart, D. E., and Zipser, D. (1985). Feature discovery by competitive learning*. *Cogn. Sci.* 9, 75–112. doi: 10.1207/s15516709cog0901_5
- Self, M. W., Kooijmans, R. N., Supér, H., Lamme, V. A., and Roelfsema, P. R. (2012). Different glutamate receptors convey feedforward and recurrent processing in macaque v1. *Proc. Natl. Acad. Sci. U.S.A.* 109, 11031–11036. doi: 10.1073/pnas.1119527109
- Self, M. W., van Kerkoerle, T., Super, H., and Roelfsema, P. R. (2013). Distinct roles of the cortical layers of area V1 in figure-ground segregation. *Curr. Biol.* 23, 2121–2129. doi: 10.1016/j.cub.2013.09.013
- Serre, T., and Poggio, T. (2010). A neuromorphic approach to computer vision. *Commun. ACM* 53, 54–61. doi: 10.1145/1831407.1831425
- 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. doi: 10.1109/TPAMI.2007.56
- Sherman, S. M., and Guillery, R. W. (1998). On the actions that one nerve cell can have on another: distinguishing drivers from modulators. *Proc. Natl. Acad. Sci. U.S.A.* 95, 7121–7126. doi: 10.1073/pnas.95.12.7121
- Spratling, M. W. (2008). Reconciling predictive coding and biased competition models of cortical function. *Front. Comput. Neurosci.* 2:4. doi: 10.3389/neuro.10.004.2008
- Spratling, M. W. (2014). A single functional model of drivers and modulators in cortex. *J. Comput. Neurosci.* 36, 97–118. doi: 10.1007/s10827-013-0471-7
- Sutton, R. S., and Barto, A. G. (1981). Toward a modern theory of adaptive networks: expectation and prediction. *Psychol. Rev.* 88, 135–170. doi: 10.1037/0033-295X.88.2.135
- Tetzlaff, C., Kolodziejski, C., Timme, M., and Wrgtger, F. (2012). Analysis of synaptic scaling in combination with hebbian plasticity in several simple networks. *Front. Comput. Neurosci.* 6:36. doi: 10.3389/fncom.2012.00036
- Tsotsos, J. K. (1988). “How does human vision beat the computational complexity of visual perception?,” in *Computational Processes in Human Vision: An Interdisciplinary Perspective*, ed Z. W. Pylyshyn (Norwood, NJ: Ablex), 286–338.
- Tsotsos, J. K. (2005). “Computational foundations for attentive processes,” in *Neurobiology of Attention*, eds L. Itti, G. Rees, and J. Tsotsos (Amsterdam, NL: Elsevier), 3–7.
- Ullman, S. (1995). Sequence seeking and counter streams: a computational model for bidirectional information flow in the visual cortex. *Cereb. Cortex* 5, 1–11. doi: 10.1093/cercor/5.1.1
- Ullman, S. (2007). Object recognition and segmentation by a fragment-based hierarchy. *Trends Cogn. Sci.* 11, 58–64. doi: 10.1016/j.tics.2006.11.009

- Ullman, S., Vidal-Naquet, M., and Sali, E. (2002). Visual features of intermediate complexity and their use in classification. *Nat. Neurosci.* 5, 682–687. doi: 10.1038/nn870
- von der Malsburg, C. (1973). Self-organization of orientation sensitive cells in the striate cortex. *Kybernetik* 14, 85–100. doi: 10.1007/BF00288907

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.

Received: 01 July 2014; accepted: 23 October 2014; published online: 05 December 2014.

Citation: Layher G, Schrodt F, Butz MV and Neumann H (2014) Adaptive learning in a compartmental model of visual cortex—how feedback enables stable category learning and refinement. *Front. Psychol.* 5:1287. doi: 10.3389/fpsyg.2014.01287

This article was submitted to Perception Science, a section of the journal *Frontiers in Psychology*.

Copyright © 2014 Layher, Schrodt, Butz and Neumann. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Inter-element orientation and distance influence the duration of persistent contour integration

Lars Strother^{1,2} * and Danila Alferov¹

¹ Brain and Mind Institute, University of Western Ontario, London, ON, Canada

² Cognitive and Brain Sciences Program, Department of Psychology, University of Nevada Reno, Reno, NV, USA

Edited by:

Bruno Breitmeyer, University of Houston, USA

Reviewed by:

Udo Ernst, University of Bremen, Germany

Georgios A. Keliris, Max Planck Institute for Biological Cybernetics, Germany

*Correspondence:

Lars Strother, Cognitive and Brain Sciences Program, Department of Psychology, University of Nevada Reno, 1664 North Virginia Street, Mail Stop 296, Reno, NV 89557-0296, USA
e-mail: lars@unr.edu

Contour integration is a fundamental form of perceptual organization. We introduce a new method of studying the mechanisms responsible for contour integration. This method capitalizes on the perceptual persistence of contours under conditions of impending camouflage. Observers viewed arrays of randomly arranged line segments upon which circular contours comprised of similar line segments were superimposed via abrupt onset. Crucially, these contours remained visible for up to a few seconds following onset, but eventually disappeared due to the camouflaging effects of surrounding background line segments. Our main finding was that the duration of contour visibility depended on the distance and degree of co-alignment between adjacent contour segments such that relatively dense smooth contours persisted longest. The stimulus-related effects reported here parallel similar results from contour detection studies, and complement previous reported top-down influences on contour persistence (Strother et al., 2011). We propose that persistent contour visibility reflects the sustained activity of recurrent processing loops within and between visual cortical areas involved in contour integration and other important stages of visual object recognition.

Keywords: contour integration, form perception, perceptual organization, perceptual grouping, association field, collinearity, perceptual hysteresis, visual persistence

INTRODUCTION

The perceptual binding of spatially local edge information into global contours, or *contour integration*, is a crucial stage of visual object recognition. Contour integration is subject to both bottom-up and top-down influences that depend on stimulus regularities, expectations, task demands, and other factors (Hess and Field, 1999; Gilbert and Li, 2013). Much of the psychophysical work on contour integration in human vision involves measuring the detectability of contours embedded in highly camouflage backgrounds. Here we introduce a new method of studying contour integration and its underlying mechanisms. This method is complementary to traditional contour detection methods, and relies on the perceptual decay of a contour under conditions of impending camouflage.

Regan (1986) noted that a highly camouflaged shape (e.g., the outline of a bird) made visible by motion does not disappear immediately after it stops moving¹. Indeed, several studies of this phenomenon have since shown that outlines of recognizable objects and simple shapes persist perceptually for up to several seconds, and furthermore, this persistence of global form is accompanied by persistent neural activity in V1 and higher-tier visual cortical areas (Ferber et al., 2003; Large et al., 2005; Strother et al., 2011, 2012). These findings demonstrate a unique type of perceptual hysteresis, which we refer to here as *contour persistence*. Contour persistence is distinct from other varieties of visual persistence, both in terms of the stimulus conditions

under which it occurs and also its duration. Contour persistence occurs following the offset of a perceptual segmentation cue (e.g., onset or motion) rather than the physical removal of the contour itself or any of the elements comprising the contour. This makes contour persistence distinct from “visible persistence” phenomena (Coltheart, 1980a,b) in which a stimulus continues to be perceived following its physical offset. Contour persistence also differs from other types of visual persistence in terms of its relatively long duration—contour persistence typically lasts >1 s, whereas other visual persistence phenomena typically last <1 s.

Here we measured the duration of contour persistence using a contour fading paradigm in which a circular contour comprised of line segments abruptly onset against a background of randomly oriented line segments². We found that such contours did not disappear immediately following onset, but instead became camouflaged over the course of a few seconds, as in the earlier demonstration (Regan’s bird). Our main goal was to measure the duration of contour persistence as a function of known determinants of contour binding strength (contour smoothness, density, and closure). Complementary to psychophysical studies of contour integration (e.g., Field et al., 1993; Pettet, 1999; Bex et al., 2001; Ledgeway et al., 2005; May and Hess, 2007, 2008; Dakin and Baruch, 2009; Marotti et al., 2012), we used contours comprised of elements that were either co-aligned and tangent to the contour (*snake* contours), co-radial (co-parallel and perpendicular to the contour; *ladder* contours), or randomly

¹<https://sites.google.com/site/visualformpersists/>

²<https://sites.google.com/site/strothertoronto/>

oriented (*jagged* contours). The first question of interest in our study was whether or not local orientation influences the duration of contour persistence. There is substantial evidence of an *association field* (Field et al., 1993) mechanism in visual cortex that enables contour detection of visual elements (e.g., line segments, wavelets) camouflaged within an array of similar elements. The association field consists of neural units tuned to specific orientation, which facilitate the activity of other neural units tuned to similar orientations but at different locations within the visual field, and thus facilitate contour binding and detection. We wondered whether or not local orientation might play a similar role in persistent contour integration. If so, it is possible that the association field maintains a persistent representation of a contour, and thus exhibits visual memory (Magnussen, 2000), either due to the reverberation of feedforward and feedback signals within the neural association field itself, or by virtue of feedback from higher tier visual cortical areas.

We performed additional experiments to examine the effects of relative density and closure contour persistence, both of which have been studied using detection paradigms (Smits et al., 1985; Kovács and Julesz, 1993; Tversky et al., 2004; Mathes and Fahle, 2007), but have not previously been manipulated in studies of contour persistence. We reasoned that if more strongly bound contours persisted longer than less strongly bound contours, this would demonstrate a stimulus-driven influence on the duration of contour persistence. Finally, we performed a control experiment to determine whether or not our results could be accounted for by eye movements. We discuss our results in terms of a recurrent process of feedforward contour integration in primary visual cortex (V1) and shape-related feedback from higher-tier visual cortical areas.

MATERIALS AND METHODS

SUBJECTS

All observers had normal or corrected-to-normal vision. Ten observers participated in Experiment 1. Six new observers participated in Experiment 2 and Experiment 3a. Two new observers (and one observer from the previous experiments) participated in a final control experiment (Experiment 3b) which employed eye-tracking. All observers provided informed consent and were recruited in accordance to University of Western Ontario ethics guidelines.

STIMULI AND PROCEDURE

All experiments employed stimuli comprised of short line segments (**Figure 1**²). Trials began with the appearance of a 'background' array of randomly oriented dark line segments ($\sim 0.3^\circ \times 0.03^\circ$) positioned randomly (overlap allowed) within a lighter $10^\circ \times 10^\circ$ square aperture on an otherwise dark display. A blue fixation cross ($\sim 0.3^\circ \times 0.3^\circ$) was always present in the center of the aperture during the experiments. Shortly (2 s) after the appearance of the background array, a circle or semi-circle comprised of line segments identical to those comprising the background appeared against the background and remained until the end of the trial (total trial duration was always 8 s).

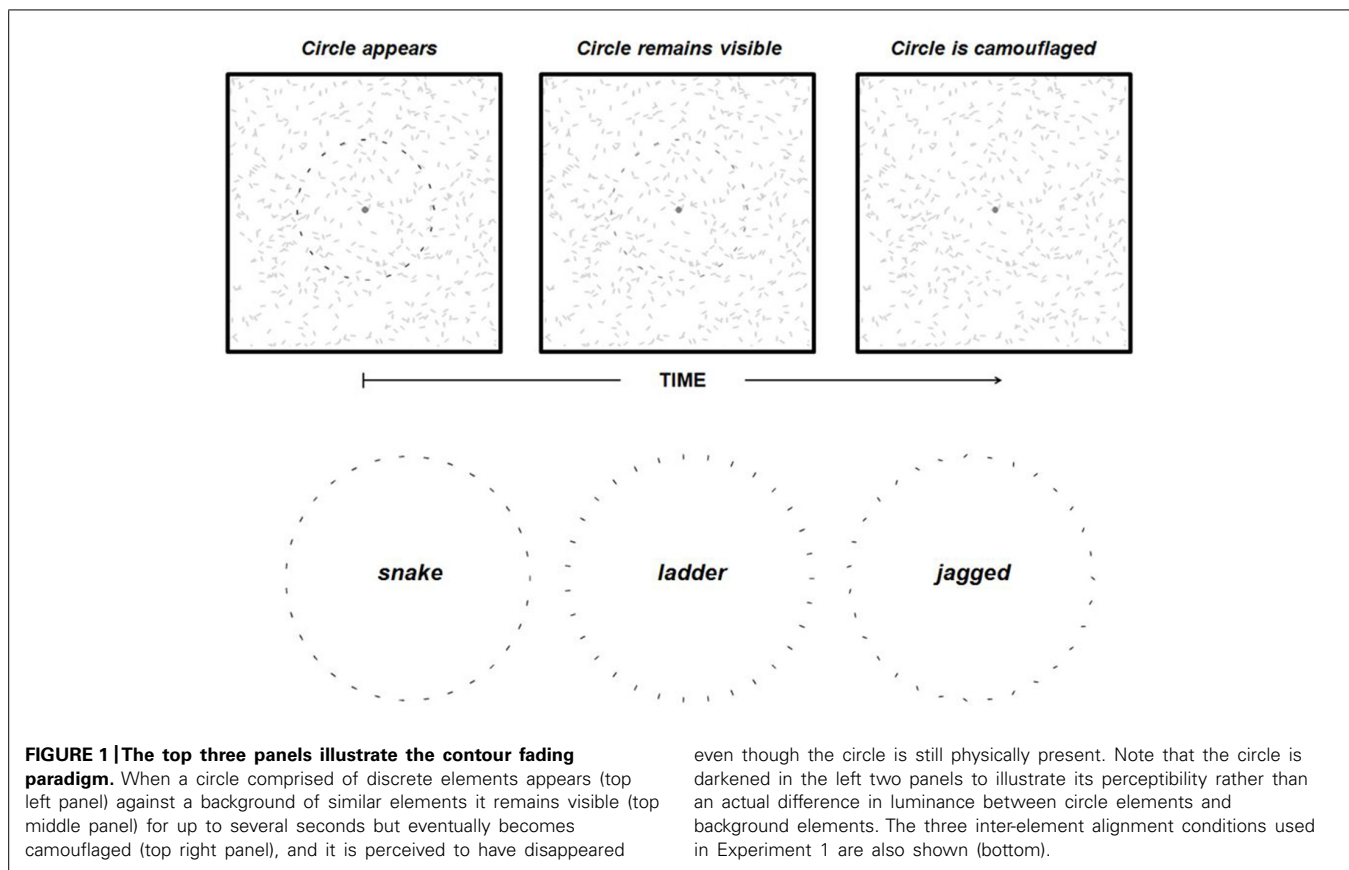
We used *snake* circles comprised of co-circular elements (i.e., smooth contours), *ladder* circles comprised of co-radial (rotated 90° from co-circular), and *jagged* circles comprised of randomly oriented elements (random orientations were generated trial to trial). The absolute positions of each of the elements along a circle or semi-circle were equivalent across all three stimulus types and conditions.

Observers were instructed to maintain fixation throughout the experiment and, on each trial, to press a button when the circle (or semi-circle) was no longer visible; this served as a measure of response time (RT). At the end of each trial a new background array appeared and the sequence was repeated. The appearance of a new background on each trial completely replaced that of the previous trial (novel background elements were generated on every trial). Individual trials ended either with the button press or after 6 s if no button was pressed. Observers were told not to press the button if they never saw the target or if any portion of it never became fully camouflaged (i.e., did not disappear), which occurred on less than 2% of trials for all observers. Observers were always given at least 25 practice trials, the results of which were not included in our analyses. Individual observers completed at least 100 trials in each experiment. Pilot studies for each observer confirmed that circular contours that disappeared were never detectable were it not for the onset cue (i.e., observers could not see the contours when the contour was superimposed against the background in the absence of an onset cue).

In Experiment 1 we were primarily interested in whether or not the duration of continued contour visibility depended on inter-element alignment: *smooth* ("snake") co-radial ("ladder") or *jagged* circles. We also varied the size of the circles and the proximity of the elements making up each circle (*circle density*), and also the density of the background elements (*background density*). We used three circle densities: line segments covered ~ 33 , 25, or 20% of a given circle's circumference. Backgrounds consisted of 2250, 3000, or 3750 elements per $10^\circ \times 10^\circ$ area. Examples of maximally dense or sparse contour-background pairings are shown in **Figure 2** (contour and background elements are shown in different colors to make the contour elements visible; all were the same color in the experiments).

The purpose of varying circle and background density in this experiment was mainly to reduce the predictability of the location of the circle elements (we explore these variables further in Experiment 2). By increasing the variability in the density of the circles relative to that of the background we hoped to increase the overall variability in the latencies of individuals' button presses. Individual observers thus completed at least 33 trials per alignment condition.

In Experiment 2 we further investigated prospective effects of circle and background densities and circle size using *smooth* (snake) circles and *jagged* circles. We were particularly interested in comparing the magnitude of durations for these two alignment conditions to that of the relative densities of the circles and backgrounds. We again used three circles sizes; for each circle size (radii of 1.5, 3, and 4.5°), *dense* circles were comprised of 10, 20, and 30 line segments (respectively), and for *sparse* circles the number of line segments was halved. We also used sparse (500 elements)



and dense (1500 elements) backgrounds (per $10^\circ \times 10^\circ$ area, as in Experiment 1); the condition pairings are illustrated in **Figure 2**.

In Experiment 3a we tested whether or not the effects observed in Experiment 2 could be observed for non-closed contours (semi-circles created from $0.5 \times$ the circumference of the circles used in Experiment 2) by re-testing the same subjects (from Experiment 2) and manipulating a subset of the parameters used in Experiment 2. The location of the arcs in Experiment 3a varied between the upper and lower visual hemifield, and $\pm 2^\circ$ from fixation (thus resulting in greater position uncertainty than in the previous experiments). We ran fewer observers in Experiments 2 and 3a than we did in Experiment 1, but we collected at least twice the amount of data per subject. The primary motivation for this experiment was to test whether or not closure would act as a cue above and beyond inter-element alignment. Finally, in Experiment 3b, we used an eye-tracker (Eyelink; SR Research Ltd., Toronto, ON, Canada) to monitor the eye movements of three observers in a partial replication of Experiment 2. We allowed circle size to vary from 2.7 to 4.5° visual angle; the circles were either smooth (snakes) or jagged (100 trials of each condition), and relative density was held constant (sparse circles on sparse background, as described earlier in this section).

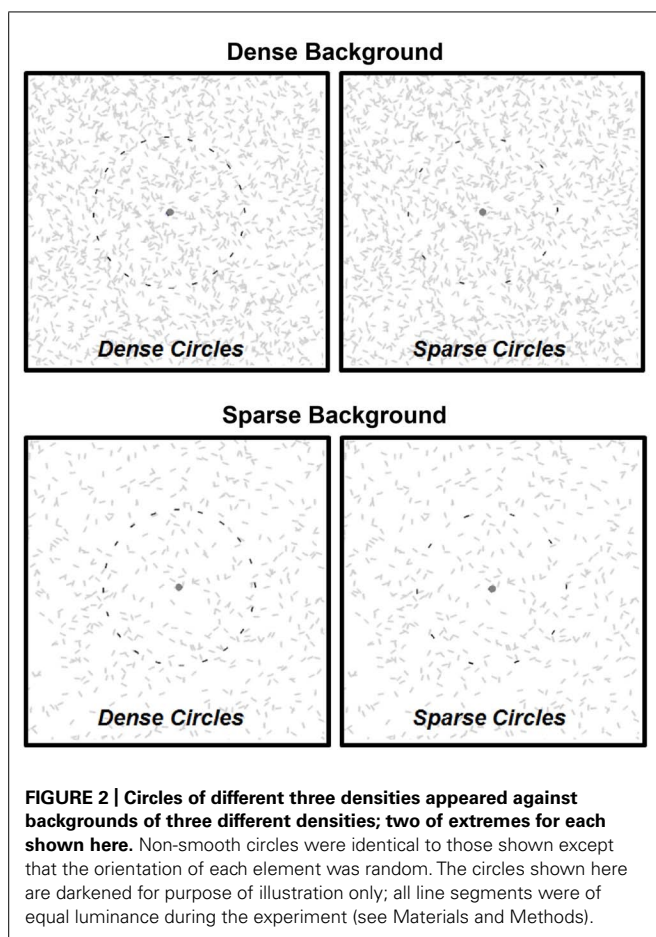
RESULTS

EXPERIMENT 1

The goal of the first experiment was to test for an effect of inter-element alignment on the duration of persistent contour

visibility. Mean RTs were greatest for *smooth* contours (2724 ms), followed by *co-radial* contours (2424 ms) and *jagged* contours (2407 ms). For all remaining statistical analyses RTs were log-transformed to reduce positive skew. Log RTs for the three alignment types are shown in **Figure 3**.

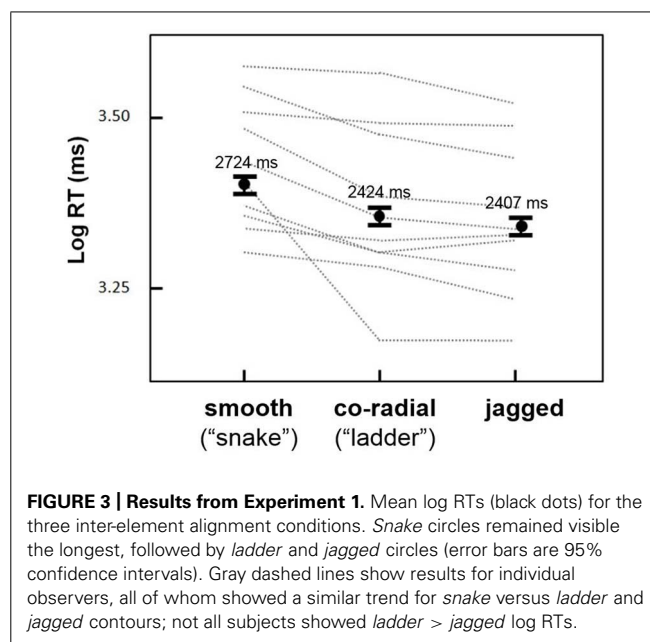
Preliminary repeated measures analyses of variance (ANOVAs) showed no significant main effects or interactions of circle size, circle density, or background density (possibly due to the low number of trials for each condition; we explore these variables further in the next experiment) on log RT, although the interaction of alignment type and circle density approached statistical significance [$F(1,9) = 3.5, p = 0.09$]. A subsequent repeated measures ANOVA based on the three alignment conditions (*smooth*, *co-radial*, *jagged*) showed a highly significant effect of alignment [$F(1,9) = 13.9, p < 0.005$]. *Post hoc* analyses (paired samples *t*-tests, one-tailed) showed significant differences between all three conditions: *smooth* > *co-radial* [$t(9) = 3.38, p < 0.01$]; *smooth* > *jagged* [$t(9) = 3.73, p < 0.01$]; *co-radial* > *jagged* [$t(9) = 2.06, p < .05$]. The *smooth* contours thus evinced a ~ 300 ms increase in RT relative to *co-radial* and *jagged* contours. Although the difference between *co-radial* and *jagged* contours was also statistically significant (*co-radial* > *jagged*), this difference was relatively small (~ 15 ms) compared to the *smooth* > *co-radial* and *smooth* > *jagged* differences, and less consistent across subjects (**Figure 3**; two subjects showed either no difference or greater RTs for *jagged* versus *co-radial* contours).



EXPERIMENT 2

As in Experiment 1, all analyses were conducted on log RTs. **Figure 4** shows that *smooth* circles remained visible longer than *jagged* circles, and the *smooth* > *jagged* log RT trend is apparent across all pairings of circle size, circle density, and background density (except possibly for *sparse* circles paired with *dense* backgrounds, shown in the lower right of **Figure 4**). A repeated measures ANOVA yielded statistically significant main effects for all factors except circle size: *alignment* [$F(1,5) = 13.2, p < 0.05$]; *circle density* [$F(1,5) = 18.4, p < 0.01$]; and *background density* [$F(1,5) = 22.2, p < 0.01$]; although there appears to be a trend in **Figure 4** of decreased log RT with increasing circle size (for *smooth* circles), this was not significant [$F(1,5) = 2.0, p = 0.19$], which means that the effect of alignment was largely scale invariant within 4.5° from fixation.

Two-way interactions between alignment and density were also statistically significant: *alignment* \times *circle density* [$F(1,5) = 11.7, p < 0.05$]; and *alignment* \times *background density* [$F(1,5) = 9.5, p < 0.05$]. These interactions are apparent in **Figure 4** in that the effect of alignment (*smooth* > *jagged*) on log RT was greatest for dense circles and sparse backgrounds. No significant density interaction (*circle density* \times *background density*) was observed [$F(1,5) = 0.1, p = 0.73$]. A three-way interaction between these variables (*alignment* \times *circle density* \times *background density*) was also significant [$F(1,5) = 18.2, p < 0.01$]. Paired-samples *t*-tests



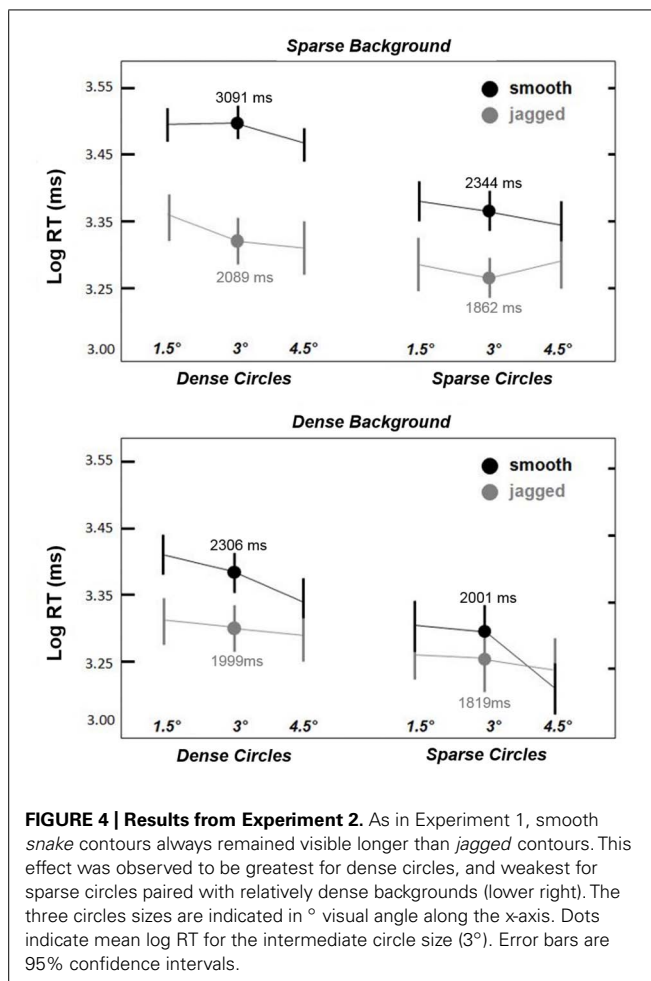
confirmed that *smooth* > *jagged* log RTs across all combinations of circle and background density [$t(5) = 2.8$ – 4.1 , always $p < 0.05$, two-tailed], except for *sparse* circles and *dense* backgrounds [$t(5) = 2.3, p = 0.07$], which approached statistical significance. Thus, while effect of alignment (*smooth* versus *jagged*) varied with the relative density of the circle and background elements, *jagged* circles always tended to disappear more quickly than *smooth* circles. In short, the results shown in **Figure 4** indicate the greatest log RTs for dense *smooth* circles superimposed on sparse backgrounds.

EXPERIMENT 3a

This experiment was a partial replication of Experiment 2 (same observers) in which we sought to replicate the *smooth* > *jagged* log RT result for non-closed contours (semi-circles). **Figure 5A** shows mean log RTs corresponding to circles (solid bars) and semi-circles (dots with error bars) obtained in Experiment 3a. The same *smooth* > *jagged* log RT trend was observed in all three cases. A repeated measures ANOVA showed a main effect of inter-element alignment [*smooth* > *jagged*; $F(1,5) = 7.7, p < 0.01$]; a main effect of closure (with circles persisting longer than semi-circles) approached significance [$F(1,5) = 2.9, p = 0.09$], and there were no significant interactions. This means that the *smooth* > *jagged* effect shown in **Figures 3** and **4** is not limited to closed contours.

EXPERIMENT 3b

The purpose of this experiment was to determine whether or not the greater persistence of smooth (snake) contours versus jagged contours could be explained by differences in eye movements. Our logic was as follows: if differences in eye movements explains our results, then equating for eye movements between our two conditions—*smooth* and *jagged*—should result in equivalent contour persistence durations. **Figure 5B** shows similar contour persistence durations (*smooth* duration > *jagged* duration) for the data based on all 200 trials (100 *smooth*, 100 *jagged*).

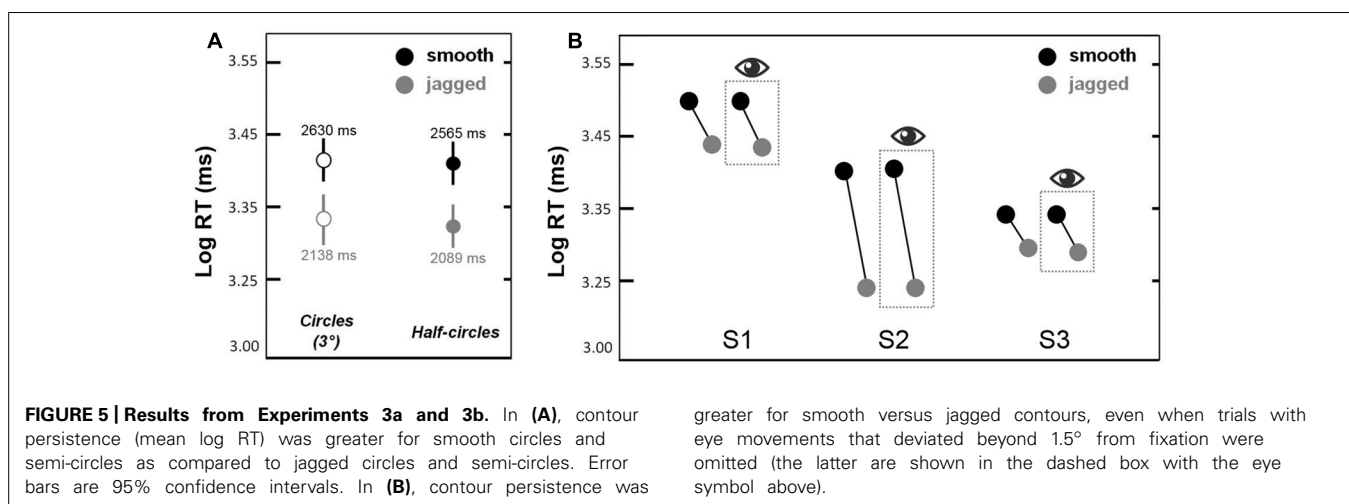


For all three subjects, discarding trials during which gaze shifted beyond 1.5° from fixation, resulted in the exclusion of >25% of the original data. Nevertheless, as is clear in **Figure 5B**, this filtering of the original trials had no effect on the pattern of results, namely that smooth contours consistently persisted

longer than jagged contours. The results in the dashed box in **Figure 5B** are based on the filtered data, and results from the original data are shown to the left of each box. The slopes of the lines connecting the black dots (smooth) and gray dots (jagged) are similar for all within-subject pairings. This means that the *smooth* > *jagged* duration result is not due to the effects of eye movements toward the circular contours used in each condition. Even when eye movements were restricted to within 1.5° from the fixation cross, and did not impinge on even the smallest circle used in the experiment (radius = 2.7°), the influence of inter-element alignment on contour persistence was the same (**Figure 5B**). Additional eye movement results reported in Supplementary Material (**Figure S1**).

DISCUSSION

We used a new perceptual fading paradigm to study persistent contour integration under conditions of impending camouflage. In Experiments 1 and 2, we found that the duration of contour persistence was influenced by stimulus properties known to influence contour salience in traditional contour detection paradigms, namely inter-element distance and co-alignment (Smits et al., 1985; Field et al., 1993; Geisler et al., 2001; Elder and Goldberg, 2002). Given major differences between our contour fading paradigm and detection paradigms typically used to study contour integration, this was not an inevitable result. For instance, the contours used in our study were undetectable were it not for the onset cue, and thus synchronous onset alone could have resulted in persistence (Wong et al., 2009), without the additional influence of inter-element alignment. Furthermore, the results observed here cannot be fully accounted for by differences in eye movements for smooth versus non-smooth contours (Experiment 3b). The persistent visibility of highly camouflaged contours observed in our study is consistent with a recurrent processing loop in which high-tier neural representations global form interact with low-level neural mechanisms that bind local edges into global contours. Given the absence of contour persistence when contour elements are physically removed (Ferber et al., 2003; Wong et al., 2009; Strother et al., 2012), it is highly plausible that feedforward responses in primary visual



cortex provide a neural basis upon which all feedback effects are exerted.

CONTOUR PERSISTENCE AND THE 'ASSOCIATION FIELD'

It has long been recognized that the sensation produced by a visual stimulus can persist after its offset, and the term "visual persistence" has been used to denote many different examples of short-term perceptual memory. Coltheart (1980a,b) used "visible persistence" to refer to cases when a visual stimulus continues to be perceived after its offset, and to distinguish these cases from "iconic memory" (Sperling, 1960), which is not accompanied by persistent perception of a physically absent stimulus. In contrast to both visible persistence and iconic memory, contour persistence occurs in the absence of physical removal (offset) of the contour. Indeed, several previous studies showed that global contours do not persist when the elements comprising the contour are removed (Ferber et al., 2003; Large et al., 2005; Wong et al., 2009; Strother et al., 2011, 2012). That is, contour persistence reflects the sustained perceptual organization of elements after an initial binding cue (onset in this case) has ended, rather than the sustained perceptual representation of a visual stimulus that has physically disappeared. Furthermore, contour persistence typically lasts considerably longer than iconic memory and other types of short-term visual memory (which are usually <1 s).

The results of the present study showed clear influences of physical properties of a contour on the duration of its perceptual persistence under camouflaging conditions. Experiment 1 showed that smooth contours showed the greatest degree of persistent contour visibility. When elements were equally co-aligned but perpendicular to the tangent of the circular global contour (the *co-radial* condition), the facilitative effect of inter-element alignment on contour persistence was reduced (**Figure 3**). Randomizing the orientations of contour elements in the *jagged* condition had a similar (but slightly greater) effect, and showed the weakest degree of persistence of the three contour conditions used in Experiment 1. The difference in duration of contour persistence for *snake* and *jagged* contours (*smooth* > *jagged*) was replicated in Experiment 2, and shown to be modulated by inter-element distance, such that decreasing the inter-element distances of the contour elements relative to the background elements decreased or eliminated the *smooth* > *jagged* effect (**Figure 4**). Experiment 3a showed that similar effects of co-alignment and density are not limited to closed circular contours (**Figure 5**), although an additional facilitative effect of closure is nevertheless a possibility—circular arcs (semi-circles) did not persist as long in general, but this trend was not statistically significant. While it is well-known that for curved contours comprised of discrete oriented elements, smooth contours are easier to detect than jagged contours, this study is the first to show that increasing contour density and smoothness facilitate contour persistence under conditions of extreme camouflage. Previous studies have recognition-related effects on the duration of visual persistence (Ferber et al., 2005; Ferber and Emrich, 2007; Emrich et al., 2008; Strother et al., 2011), but none of these systematically manipulated contour properties in a manner consistent with detection studies of contour integration. The results of the present study are thus an

important step toward identifying common mechanisms involved in contour integration and contour persistence, and the relationship of these mechanisms to feedforward and feedback processes in human vision.

Hebb (1949) proposed that short-term memory consists in the persistent reverberation of activity in neuronal assemblies. A plausible explanation of the results of the present study is that contour persistence reflects persistent reverberation of an *association field* mechanism in visual cortex (Field et al., 1993). The association field is a neuronal assembly consisting of cells with similar orientation preferences and receptive fields at different retinal locations. These cells exhibit mutually facilitative interactions, and the more similar adjacent cells are in receptive field location and orientation preference, the stronger the facilitation. This mechanism thus shows greater mutual facilitation with increasing edge co-alignment. It is plausible that an association field mechanism is responsible, at least in part, for both the initial perception and persistence of global contours in the present study. This would be consistent with the effects of inter-element distance and co-alignment on the duration of persistence reported here, which parallel similar effects on the detectability of contours (e.g., Field et al., 1993; Bex et al., 2001; Ledgeway et al., 2005; Marotti et al., 2012).

FEEDFORWARD AND FEEDBACK INFLUENCES

Our findings are consistent with the view that neural mechanisms in higher-tier visual cortical areas represent hypotheses about low-level visual input, and in doing so, reinforce inferences (e.g., about shape) via feedback to lower level visual cortical mechanisms to facilitate efficient extraction and encoding of visual features (Engel et al., 2001; Murray et al., 2002; Cardin et al., 2011). There is growing consensus that top-down feedback plays an integral role in contour integration (Gilbert and Li, 2013), but the precise nature of the effects of this feedback is not known. One possibility is that feedforward contour integration processes are accompanied by feedback processes that serve to disambiguate and enhance the salience of global contour by suppressing background noise (Strother et al., 2012; Chen et al., 2014). In this framework, extrastriate feedback could serve to modulate the responses of neurons in primary visual cortex (V1). More specifically, the responses of neurons stimulated by background elements would be suppressed and the responses of neurons stimulated by contour elements would be facilitated by extrastriate feedback in addition to facilitation by an association field within V1. The crucial result of this feedback would be the facilitation of inter-element binding within the contour and the suppression of background noise, and ultimately, the perceptual segmentation of the contour from its surroundings.

In addition to the facilitation of contour binding by an association field in V1, extrastriate feedback may also play an important role, not only in contextually modulating the responses of individual V1 neurons (Zipser et al., 1996; Lamme et al., 1998), but also in temporarily sustaining the joint activity of neurons in the association field. It is worth noting that the duration of contour persistence in the *jagged* condition (**Figure 4**) was surprisingly long, even when the density of these jagged contours was similar to that of the background. This surprising effect highlights the

importance of synchronous onset in the persistence of global form (Wong et al., 2009), and role of temporal synchrony as a powerful binding cue in contour integration (Usher and Donnelly, 1998; Beaudot, 2002), and even higher stages of the visual object recognition process (Singer and Kreiman, 2014). The fact that synchronous neuronal firing is a common feature of neural network models of contour integration and other types of perceptual organization (Sporns et al., 1991; Yen and Finkel, 1998), it is conceivable that the synchronous onset of elements comprising a contour could result in the persistent activity of visual cortical neurons irrespective of contour smoothness. This prediction is consistent with findings that global contours are represented in shape-selective cortex irrespective of the local features (Altmann et al., 2003; Kourtzi et al., 2003).

Taken together with the results of previous studies (Strother et al., 2011, 2012), the results reported here lead us to propose that contour persistence reflects sustained feedforward and feedback visual processing. Some of this processing involves the binding of local visual elements into global form, which involves feedforward processing in visual cortex as well as feedback processing, both within and between visual cortical areas (Chen et al., 2014). Our results show that this complex circuit exhibits short-term memory, as evidenced by the persistence of a contour under conditions of impending camouflage. It is not clear whether the persistent contour integration reported here is due to hysteresis intrinsic to mechanisms in visual cortex alone—for example, sustained neural reverberation within the association field—or involves neural reverberation at a larger cortical scale, such as a recurrent processing loop between shape-selective neural mechanisms in extrastriate visual cortex (Kourtzi and Connor, 2011), and those in earlier visual cortical areas (e.g., V1). For instance, feedforward activity in V1 could be subsequently modulated by interactions within the association field, which may serve to enhance the perceptual salience of a contour relative to its background (Gilbert and Li, 2013). Additionally, shape-related feedback from V4 and higher-tier areas could exert an additional influence on the responses of V1 neurons, by facilitating the responses of those corresponding to contour elements, and by inhibiting the responses of neurons responding to background elements (Chen et al., 2014). Persistent contour integration could therefore reflect hysteresis in both types of mechanisms. Persistent contour integration could also involve sustained neural activity in more anterior cortical areas that play a top-down role in visual memory (Curtis and D'Esposito, 2003).

It is worth noting that our proposed feedforward-feedback account is not the only possible explanation for our results. An alternative account could predict greater persistence for smooth contours without the need for recurrent processing loops. For instance, contour onset could elicit transient activity in orientation-selective neurons, and during this initial surge of neural activity (which could occur within <100 ms), contour integration mechanisms (e.g., the association field) could enhance the representation of contour elements and their configuration, which could be transferred to a higher tier cortical area. Once transferred, it is possible that this high-tier representation no longer receives input from earlier visual areas (e.g., after the initial ~100 ms surge

of activity), and thus decays. While this is possible, it is not clear why this initial higher-tier representation should be stronger for smooth versus jagged contours since local orientation information is thought to be less important than global form in high-tier cortical representations of contour shape (Altmann et al., 2003; Kourtzi et al., 2003). Moreover, results from fMRI studies of contour persistence show an effect of familiarity on persistent neural activity in early visual areas, including V1 (Strother et al., 2011). Additionally, persistent neural activity in V1 was subsequently shown to be limited to the retinal location of the contour elements, and also to correspond to the duration of contour visibility (Strother et al., 2012). It should nevertheless be acknowledged that persistent neural activity in early visual areas, such as V1, could be epiphenomenal rather than evidence of a recurrent feedback loop between visual cortical areas. Further studies are necessary to test whether perceptual decay of a camouflage contour corresponds to persistent shape representation in high-tier visual cortical areas, earlier visual areas such as V1, or the persistent activation of a recurrent processing loop between several areas.

To conclude, the results reported here were obtained using a novel psychophysical method, and show that the neural mechanisms responsible for contour integration exhibit short-term memory, the duration of which is sensitive to the spatial properties of visual elements comprising the contour. Future studies could employ a more continuous range of element orientations and test for a possible within-observer correlation between contour detection performance and contour persistence. If observed, a correlation would strengthen the link between contour persistence and its neural basis in the association field. Additional studies could also employ neurophysiological measures to assess the concurrent operation of feedforward and feedback processes during persistent contour integration.

ACKNOWLEDGMENTS

We thank T. Vilis for discussing some of the results and ideas presented here, and Z. Zhou for his help with eye-tracking.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://www.frontiersin.org/journal/10.3389/fpsyg.2014.01273/abstract>

Figure S1 | Gaze position results from a control experiments with three subjects (S1, S2, S3). To obtain the results of **Figure 5B**, trials were excluded when gaze position deviated >1.5° from the fixation cross at the center of the screen. This means that the resultant data did not include trials for which the observer shifted his or her gaze to the actual contour (which occurred for <5% of trials for each observer). The smallest dashed circle indicates a distance of 1.5° from the fixation cross (distance used for filtering trials); the largest dashed circle indicates size of largest circle (4.5°); and the intermediate-sized dashed circle above indicates the size of the smallest contour circle used in the experiment.

REFERENCES

- Altmann, C. F., Bühlhoff, H. H., and Kourtzi, Z. (2003). Perceptual organization of local elements into global shapes in the human visual cortex. *Curr. Biol.* 13, 342–349. doi: 10.1016/S0960-9822(03)00052-6
- Beaudot, W. H. (2002). Role of onset asynchrony in contour integration. *Vision Res.* 42, 1–9. doi: 10.1016/S0042-6989(01)00259-0

- Bex, P. J., Simmers, A. J., and Dakin, S. C. (2001). Snakes and ladders: the role of temporal modulation in visual contour integration. *Vision Res.* 41, 3775–3782. doi: 10.1016/S0042-6989(01)00222-X
- Cardin, V., Friston, K. J., and Zeki, S. (2011). Top-down modulations in the visual form pathway revealed with dynamic causal modeling. *Cereb. Cortex* 21, 550–562. doi: 10.1093/cercor/bhq122
- Chen, M., Yan, Y., Gong, X., Gilbert, C. D., Liang, H., and Li, W. (2014). Incremental integration of global contours through interplay between visual cortical areas. *Neuron* 82, 682–694. doi: 10.1016/j.neuron.2014.03.023
- Coltheart, M. (1980a). Iconic memory and visible persistence. *Percept. Psychophys.* 27, 183–228. doi: 10.3758/BF03204258
- Coltheart, M. (1980b). The persistences of vision. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 290, 57–69. doi: 10.1098/rstb.1980.0082
- Curtis, C. E., and D'Esposito, M. (2003). Persistent activity in the prefrontal cortex during working memory. *Trends Cogn. Sci.* 7, 415–423. doi: 10.1016/S1364-6613(03)00197-9
- Dakin, S. C., and Baruch, N. J. (2009). Context influences contour integration. *J. Vis.* 9, 13.1–13.13.
- Elder, J. H., and Goldberg, R. M. (2002). Ecological statistics of Gestalt laws for the perceptual organization of contours. *J. Vis.* 2, 324–353. doi: 10.1167/2.4.5
- Emrich, S. M., Ruppel, J. D., and Ferber, S. (2008). The role of elaboration in the persistence of awareness for degraded objects. *Conscious. Cogn.* 17, 319–329. doi: 10.1016/j.concog.2006.12.001
- Engel, A. K., Fries, P., and Singer, W. (2001). Dynamic predictions: oscillations and synchrony in top-down processing. *Nat. Rev. Neurosci.* 2, 704–716. doi: 10.1038/35094565
- Ferber, S., and Emrich, S. M. (2007). Maintaining the ties that bind: the role of an intermediate visual memory store in the persistence of awareness. *Cogn. Neuropsychol.* 24, 187–210. doi: 10.1080/02643290601046598
- Ferber, S., Humphrey, G. K., and Vilis, T. (2003). The lateral occipital complex subserves the perceptual persistence of motion-defined groupings. *Cereb. Cortex* 13, 716–721. doi: 10.1093/cercor/13.7.716
- Ferber, S., Humphrey, G. K., and Vilis, T. (2005). Segregation and persistence of form in the lateral occipital complex. *Neuropsychologia* 43, 41–51. doi: 10.1016/j.neuropsychologia.2004.06.020
- Field, D. J., Hayes, A., and Hess, R. F. (1993). Contour integration by the human visual system: evidence for a local “association field.” *Vision Res.* 33, 173–193. doi: 10.1016/0042-6989(93)90156-Q
- Geisler, W. S., Perry, J. S., Super, B. J., and Gallogly, D. P. (2001). Edge co-occurrence in natural images predicts contour grouping performance. *Vision Res.* 41, 711–724. doi: 10.1016/S0042-6989(00)00277-7
- Gilbert, C. D., and Li, W. (2013). Top-down influences on visual processing. *Nat. Rev. Neurosci.* 14, 350–363. doi: 10.1038/nrn3476
- Hebb, D. O. (1949). *The Organization of Behavior: A Neuropsychological Theory*. New York: Wiley.
- Hess, R., and Field, D. (1999). Integration of contours: new insights. *Trends Cogn. Sci.* 3, 480–486. doi: 10.1016/S1364-6613(99)01410-2
- Kourtzi, Z., and Connor, C. E. (2011). Neural representations for object perception: structure, category, and adaptive coding. *Annu. Rev. Neurosci.* 34, 45–67. doi: 10.1146/annurev-neuro-060909-153218
- Kourtzi, Z., Tolias, A. S., Altmann, C. F., Augath, M., and Logothetis, N. K. (2003). Integration of local features into global shapes: monkey and human fMRI studies. *Neuron* 37, 333–346. doi: 10.1016/S0896-6273(02)01174-1
- Kovács, I., and Julesz, B. (1993). A closed curve is much more than an incomplete one: effect of closure in figure-ground segmentation. *Proc. Natl. Acad. Sci. U.S.A.* 90, 7495–7497. doi: 10.1073/pnas.90.16.7495
- Lamme, V. A., Supér, H., and Spekreijse, H. (1998). Feedforward, horizontal, and feedback processing in the visual cortex. *Curr. Opin. Neurobiol.* 8, 529–535. doi: 10.1016/S0959-4388(98)80042-1
- Large, M. E., Aldcroft, A., and Vilis, T. (2005). Perceptual continuity and the emergence of perceptual persistence in the ventral visual pathway. *J. Neurophysiol.* 93, 3453–3462. doi: 10.1152/jn.00934.2004
- Ledgeway, T., Hess, R. F., and Geisler, W. S. (2005). Grouping local orientation and direction signals to extract spatial contours: empirical tests of “association field” models of contour integration. *Vision Res.* 45, 2511–2522. doi: 10.1016/j.visres.2005.04.002
- Magnussen, S. (2000). Low-level memory processes in vision. *Trends Neurosci.* 23, 247–251. doi: 10.1016/S0166-2236(00)01569-1
- Marotti, R. B., Pavan, A., and Casco, C. (2012). The integration of straight contours (snakes and ladders): the role of spatial arrangement, spatial frequency and spatial phase. *Vision Res.* 71, 44–52. doi: 10.1016/j.visres.2012.07.021
- Mathes, B., and Fahle, M. (2007). Closure facilitates contour integration. *Vision Res.* 47, 818–827. doi: 10.1016/j.visres.2006.11.014
- May, K. A., and Hess, R. F. (2007). Ladder contours are undetectable in the periphery: a crowding effect? *J. Vis.* 7, 9.1–9.15. doi: 10.1167/7.13.9
- May, K. A., and Hess, R. F. (2008). Effects of element separation and carrier wavelength on detection of snakes and ladders: implications for models of contour integration. *J. Vis.* 8, 4.1–4.23. doi: 10.1167/8.13.4
- Murray, S. O., Kersten, D., Olshausen, B. A., Schrater, P., and Woods, D. L. (2002). Shape perception reduces activity in human primary visual cortex. *Proc. Natl. Acad. Sci. U.S.A.* 99, 15164–15169. doi: 10.1073/pnas.192579399
- Pettet, M. W. (1999). Shape and contour detection. *Vision Res.* 39, 551–557. doi: 10.1016/S0042-6989(98)00130-8
- Regan, D. (1986). Form from motion parallax and form from luminance contrast: vernier discrimination. *Spat. Vis.* 1, 305–318. doi: 10.1163/156856886X00106
- Singer, J. M., and Kreiman, G. (2014). Short temporal asynchrony disrupts visual object recognition. *J. Vis.* 14:7. doi: 10.1167/14.5.7
- Smits, J. T., Vos, P. G., and van Oeffelen, M. P. (1985). The perception of a dotted line in noise: a model of good continuation and some experimental results. *Spat. Vis.* 1, 163–177. doi: 10.1163/156856885X00170
- Sperling, G. (1960). The information available in brief visual presentations. *Psychol. Monogr.* 74, 1–29. doi: 10.1037/h0093759
- Sporns, O., Tononi, G., and Edelman, G. M. (1991). Modeling perceptual grouping and figure-ground segregation by means of active reentrant connections. *Proc. Natl. Acad. Sci. U.S.A.* 88, 129–133. doi: 10.1073/pnas.88.1.129
- Strother, L., Lavell, C., and Vilis, T. (2012). Figure-ground representation and its decay in primary visual cortex. *J. Cogn. Neurosci.* 24, 905–914. doi: 10.1162/jocn_a_00190
- Strother, L., Mathuranath, P. S., Aldcroft, A., Lavell, C., Goodale, M. A., and Vilis, T. (2011). Face inversion reduces the persistence of global form and its neural correlates. *PLoS ONE* 6:e18705. doi: 10.1371/journal.pone.0018705
- Tversky, T., Geisler, W. S., and Perry, J. S. (2004). Contour grouping: closure effects are explained by good continuation and proximity. *Vision Res.* 44, 2769–2777. doi: 10.1016/j.visres.2004.06.011
- Usher, M., and Donnelly, N. (1998). Visual synchrony affects binding and segmentation in perception. *Nature* 394, 179–182. doi: 10.1038/28166
- Wong, Y. J., Aldcroft, A. J., Large, M. E., Culham, J. C., and Vilis, T. (2009). The role of temporal synchrony as a binding cue for visual persistence in early visual areas: an fMRI study. *J. Neurophysiol.* 102, 3461–3468. doi: 10.1152/jn.00243.2009
- Yen, S. C., and Finkel, L. H. (1998). Extraction of perceptually salient contours by striate cortical networks. *Vision Res.* 38, 719–741. doi: 10.1016/S0042-6989(97)00197-1
- Zipser, K., Lamme, V. A., and Schiller, P. H. (1996). Contextual modulation in primary visual cortex. *J. Neurosci.* 16, 7376–7389.

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.

Received: 30 June 2014; accepted: 20 October 2014; published online: 06 November 2014.

Citation: Strother L and Alferov D (2014) Inter-element orientation and distance influence the duration of persistent contour integration. *Front. Psychol.* 5:1273. doi: 10.3389/fpsyg.2014.01273

This article was submitted to Perception Science, a section of the journal *Frontiers in Psychology*.

Copyright © 2014 Strother and Alferov. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Combined contributions of feedforward and feedback inputs to bottom-up attention

Peyman Khorsand¹, Tirin Moore^{2,3} and Alireza Soltani^{4*}

¹ Jefferies International Limited, London, UK

² Department of Neurobiology, Stanford University School of Medicine, Stanford, CA, USA

³ Howard Hughes Medical Institute, Stanford, CA, USA

⁴ Department of Psychological and Brain Sciences, Dartmouth College, Hanover, NH, USA

Edited by:

Bruno Breitmeyer, University of Houston, USA

Reviewed by:

Peter König, University of

Osnabrück, Germany

Simon P. Kelly, City College of

New York, USA

Milica Mormann, University of

Miami, USA

*Correspondence:

Alireza Soltani, Department of Psychological and Brain Sciences,

Dartmouth College,

Moore Hall 6207, Hanover,

NH 03784, USA

e-mail: alireza.soltani@

dartmouth.edu

In order to deal with a large amount of information carried by visual inputs entering the brain at any given point in time, the brain swiftly uses the same inputs to enhance processing in one part of visual field at the expense of the others. These processes, collectively called bottom-up attentional selection, are assumed to solely rely on feedforward processing of the external inputs, as it is implied by the nomenclature. Nevertheless, evidence from recent experimental and modeling studies points to the role of feedback in bottom-up attention. Here, we review behavioral and neural evidence that feedback inputs are important for the formation of signals that could guide attentional selection based on exogenous inputs. Moreover, we review results from a modeling study elucidating mechanisms underlying the emergence of these signals in successive layers of neural populations and how they depend on feedback from higher visual areas. We use these results to interpret and discuss more recent findings that can further unravel feedforward and feedback neural mechanisms underlying bottom-up attention. We argue that while it is descriptively useful to separate feedforward and feedback processes underlying bottom-up attention, these processes cannot be mechanistically separated into two successive stages as they occur at almost the same time and affect neural activity within the same brain areas using similar neural mechanisms. Therefore, understanding the interaction and integration of feedforward and feedback inputs is crucial for better understanding of bottom-up attention.

Keywords: saliency map, saliency computation, top-down attention, computational modeling, feedforward, feedback, lateral interaction, NMDA

INTRODUCTION

Bottom-up, saliency-driven attentional selection is the mechanism through which the brain uses exogenous signals to allocate its limited computational resources to further process a part of visual space or an object. Early investigations into bottom-up attention showed that this form of attention is fast and involuntary, and purely relies on external inputs that impinge on the retina at a given time (Treisman, 1985; Braun and Julesz, 1998). Therefore, early on, vision scientists hypothesized that bottom-up attention should rely only on parallel, feedforward processes (Treisman and Gelade, 1980; Treisman and Gormican, 1988; Nakayama and Mackeben, 1989). Accordingly, various computational models of attention adopted a similar architecture for bottom-up visual processing (Koch and Ullman, 1985; Wolfe, 1994; Itti and Koch, 2001). More specifically, these models assume that bottom-up attention relies on feedforward processes and computations that terminates in the formation of the saliency (or priority) map, a feature-independent topographical map that represents the visual salience of the entire visual field and can guide covert attention. Nonetheless, all of these models also assume that feedback is involved at some point in visual processing, but this

occurs late in processing and only due to top-down signals in tasks which involve top-down attention (e.g., conjunction search, or the search for a target distinguished from other stimuli by more than one feature).

There are a few aspects of bottom-up attentional processes that explain how the hypothesis for the purely feedforward nature of bottom-up attention was originated and why it is still influencing the field, despite more recent contradictory evidence. Specifically, in comparison to top-down attention, bottom-up attention is fast and is relatively unaffected by aspects of the visual stimulus, such as the number of targets on the screen (Treisman and Sato, 1990) or the presence or absence of visual cues (Nakayama and Mackeben, 1989). The relative independence of bottom-up attention from the number of targets is taken as evidence that during bottom-up selection, exogenous signals should be processed in a parallel instead of a serial fashion. Combining this behavioral evidence with the presumption that feedback and recurrent processes are slower than feedforward processes, and that parallel processing excludes feedback, made it appear less likely that bottom-up attention relies on feedback.

However, a number of recent experimental and modeling studies have challenged most of the rather intuitive reasoning mentioned above. On the one hand, there is recent experimental evidence that top-down signals (via inputs to higher cortical areas representing saliency or to lower-level visual areas) can not only alter the previously established behavioral signatures of bottom-up attention (Joseph et al., 1997; Krummenacher et al., 2001; Einhäuser et al., 2008) but also its neural signature (Burrows and Moore, 2009). On the other hand, more recent models of visions have tried to incorporate top-down effects into bottom-up attention in order to design more efficient models of vision that can match human performance in different visual tasks (Oliva et al., 2003; Navalpakkam and Itti, 2005, 2006; see Borji and Itti, 2013 for a review). Importantly, results from a recent biophysically plausible computational model of bottom-up attention, which is mainly concerned with underlying neurophysiological mechanisms, demonstrate that recurrent and feedback inputs do not slow down the saliency computations necessary for bottom-up attention, and instead enhance them (Soltani and Koch, 2010).

Here, we review recent studies that challenge the idea that bottom-up attention solely relies on feedforward processes. Moreover, findings in these studies suggest that mechanistically one cannot separate the feedforward and feedback processes into two successive stages as they occur concurrently and within the same brain areas by using similar neural mechanisms. Therefore, we propose that while thinking in terms of separate feedforward and feedback processes was or maybe is still useful for explaining some behavioral observations, this approach is neither fruitful nor constructive for interpreting the neural data and revealing the neural mechanisms underlying bottom-up attention. Instead, we suggest that understanding the interaction and integration of feedforward and feedback inputs is crucial for understanding bottom-up attention.

EXPERIMENTAL EVIDENCE FOR THE ROLE OF FEEDBACK

Despite its intuitive appeal, even early studies of attention yielded behavioral evidence against the hypothesis that bottom-up attention relies solely on feedforward processes. This evidence includes, but is not limited to, asymmetries between the search time when targets and distracters are switched (Treisman, 1985), and the impairment of visual search in the presence of a concurrent visual task for the least salient (but not the most salient) target (Braun, 1994). However, these findings were used to argue for parallel versus serial attentional processes and to separate visual processes to “preattentive” (i.e., processes that precede top-down attention and so do not require it) and attentive processes (i.e., processes that require top-down attention; Treisman, 1985; Braun, 1994). That is, instead of assuming a function for feedback in bottom-up attention, they equated feedback processes with the involvement of top-down attention.

The first clear evidence for the role of top-down signals (and therefore feedback) in bottom-up attention comes from a study by Joseph et al. (1997) where they showed that even a visual search for popout targets (which is traditionally considered as a preattentive process) can be impaired in the presence of a demanding central task. Specifically, the authors showed that the performance for detection of an oddball target (defined by a

simple feature such as orientation) was greatly impaired when the subjects were simultaneously engaged in reporting a white letter in a stream of black letters. This impairment in performance was alleviated as the lag between the demanding central task and oddball detection was increased, indicating that the impairment was not due to interference between responses in the two tasks. Interestingly, the subjects did not become slower in oddball detection as the number of distracters was increased, a hallmark of parallel processing in visual search tasks. These behavioral results demonstrate that top-down signals are important even for the oddball detection task, which was considered to only rely on preattentive processes, as the shift of such signals to other part of space changes the bottom-up characteristics of performance in the task.

There is other experimental evidence that indicates bottom-up saliency computations are strongly modulated by top-down signals. Some of this evidence is based on inter-trial effects in visual search tasks where the reaction time (RT) for detection of popout targets is influenced by the feature that defined the target on the preceding trial (Maljkovic and Nakayama, 1994; Found and Müller, 1996; Krummenacher et al., 2001, 2010; Mortier et al., 2005). For example, Krummenacher et al. (2001) showed that RT for popout targets was shorter when the feature defining the target on trial “*n*” was the same as the feature defining the target on trial “*n*-1.” Because these effects are task-dependent and can survive an inter-trial time interval of a few seconds, it is unlikely that they are caused by activity-dependent changes in the feedforward pathway such as short-term synaptic plasticity which are mostly dominated by depression rather than facilitation (which itself is only prominent on a timescale of a few hundreds milliseconds; Zucker and Regehr, 2002). Overall, these inter-trial effects indicate that not only feedback but also memory can influence bottom-up saliency computations (Krummenacher et al., 2010).

One of the most successful models of bottom-up attention, the saliency model of Itti et al. (1998), assumes the existence of a unique saliency map that represents the visual salience of the entire visual field by integrating saliency across individual features. In order to calculate the most salient locations, the model relies on series of successive computations that separately enhance contrast between neighboring locations for different features of the stimulus such as intensity, orientation, color, motion, etc. This gives rise to the formation of the so-called conspicuity maps for each visual feature which are then further processed and combined to form a single saliency map that has no feature selectivity. This saliency map is proposed to be instantiated in superior colliculus (Kustov and Robinson, 1996), pulvinar (Shipp, 2004), V4 (Mazer and Gallant, 2003), lateral intraparietal cortex (LIP; Gottlieb et al., 1998), or the frontal eye field (FEF; Thompson and Bichot, 2005). Finally, this model assumes that top-down effects could happen via changes at different stages of saliency computations (Itti and Koch, 2001; Navalpakkam and Itti, 2005). Alternatively but not exclusively, top-down effects could directly influence bottom-up attention after the completion of saliency computations (Ahissar and Hochstein, 1997).

There is evidence from viewing (eye movement) behavior that top-down signals can interact with bottom-up saliency signals. In one study, Einhäuser et al. (2008) used a visual search task

(using images with manipulated saliency, e.g., by imposing a gradient in contrast across them) to show that task demands can override saliency-driven signals which otherwise bias eye movements. These top-down effects on eye movements could be due to adjustments of weighting of different features involved in saliency computations or direct influence of task demands after bottom-up saliency computations are performed (the so-called weak versus strong top-down effects), or a combination of the two. For example, measuring webpage viewing during different tasks, Betz et al. (2010) argued that the influence of task on viewing behavior could not be merely explained by reweighing of features. Whereas effects of task demands on viewing behavior are present in both of these studies, the exact locus and relative contribution of top-down signals to bottom-up processes could depend on the task (e.g., visual search versus information gathering from texts). Moreover, it is more biophysically plausible (in terms of existing feedback connections and neural circuitry) that top-down signals and task demands exerts their effects on bottom-up attention via modulating the saliency computations as they progress, rather than overriding the final computations.

Overall, these behavioral results demonstrate that bottom-up saliency computations (e.g., detecting an oddball) are strongly modulated by feedback signals and processes that include working memory. Moreover, they provide an alternative way to interpret the aforementioned asymmetries in the detection of a salient object, or, the dichotomy between preattentive and attentive processes. That is, the detection of any target (salient on non-salient) requires some amount of feedback from higher visual areas; however, the necessary amount of feedback depends on the configuration of targets and distracters (see below).

Despite earlier behavioral evidence for the role of top-down signals in bottom-up attention, the corresponding neural evidence has been demonstrated only recently (Burrows and Moore, 2009). More specifically, Burrows and Moore (2009) examined the representation of salience in area V4, as previous attempts at finding these signals in lower visual areas were equivocal (Hegdé and Felleman, 2003), and moreover, examined the effects of top-down signals on this representation. In order to distinguish pure salience signals from signals that merely reflect a contrast between the center and surround (such as orientation contrast reported by Knierim and van Essen, 1992), the authors measured the response of V4 neurons to different types of stimuli (singleton, color and orientation popout, combined popout, and conjunction), for which the target has different levels of saliency. Interestingly, they found that V4 neurons carry pure saliency signals reflected in their differential firing responses to popout and conjunction stimuli. Next, they measured the response to the same stimuli while a monkey prepared a saccade to a location far from a neuron's receptive field. Interestingly, they found that saccade preparation eliminated the saliency signals observed in V4. Later, our computational modeling showed that these observations can be explained by alterations of feedback from neurons in a putative saliency map due to saccade preparation (Soltani and Koch, 2010). Overall, these results demonstrate that the most basic computations underlying bottom-up attention, which enable the brain to discriminate between salient and non-salient objects, are strongly modulated by top-down signals.

MODELING EVIDENCE FOR THE ROLE OF FEEDBACK

Most computational models of bottom-up attention rely on feedforward processes as the main source of computations during visual search tasks (Koch and Ullman, 1985; Treisman and Gormican, 1988; Wolfe, 1994; Itti and Koch, 2001). Whereas some of these models were constructed keeping neural substrates in mind, they lack enough detail to be able to elucidate biophysical mechanisms or constraints underlying bottom-up attention. As described below, some of these biophysical constraints are the main reasons why feedforward processes are not sufficient to adequately account for the behavioral and neural signatures of bottom-up attention.

In a recent study, Soltani and Koch (2010) constructed a detailed, biophysically plausible computational model to examine neural mechanisms and constraints underlying the formation of saliency signals. The model network consisted of populations of spiking model neurons representing primary visual areas (V1, V2, and V4) and a higher visual area representing the saliency (or priority) map, a topographical map that represents the visual salience of the entire visual field. Similar to the saliency model of Itti et al. (1998), Soltani and Koch (2010) assumed that the neural population in the saliency map integrates the output of neural populations in V4 with different features selectivity. Therefore, the saliency signals in visual areas V1–V4 were feature-dependent whereas this signal was feature-independent in the saliency map. The input to the model was generated by filtering stimuli used in Burrows and Moore's (2009) study based on response properties of neurons in LGN and V1. Using this model, the authors studied both the formation of saliency signals in successive populations of neurons (which mimic visual areas V1–V4) and how these signals are modulated by the feedback from a putative saliency map (assumed to be instantiated in the LIP or FEF).

The results from this computational study challenge the idea that bottom-up, exogenous attention solely relies on feedforward processing at various levels. Firstly, this study provides evidence that saliency processing relies heavily on recurrent connections (so it is not solely feedforward) with slow synaptic dynamics operating via NMDA receptors. However, the involvement of NMDA-mediated currents does not slow down the emergence of saliency signals. More specifically, the onsets of saliency signals in successive layers of the network were delayed by only a few milliseconds (and were advanced for some stimuli), while the strength of signals greatly increased. Secondly, as shown experimentally and computationally, recurrent reverberation through NMDA is crucial for working memory (Wang, 1999; Tsukada et al., 2005; Wang et al., 2013) and decision making (Wang, 2002). Therefore, an equally important role for reverberation through NMDA in saliency computations makes these computations more similar to cognitive processes that are not considered feedforward, such as working memory and decision making. Thirdly, this study demonstrates that whereas saliency signals do increase across successive layers of neurons, they could be significantly improved by feedback from higher visual areas that represent the saliency map.

But how is it that recurrent and feedback inputs do not slow down saliency computations in the model? The formation of saliency signals relies heavily on slow recurrent inputs (dominated

by NMDA receptors), but at the same time these signals propagate through successive layers of the network via fast AMPA currents. Computation at successive layers with slow synapses reduces noise and enhances signals such that higher visual areas carry the saliency signals earlier than the lower visual areas. Consequently, feedback from the higher visual areas via fast AMPA synapses can enhance the saliency signals in the lower visual areas. Importantly, all these results depend on the presence of cortical noise. In the absence of noise, saliency computations could be accomplished merely by AMPA currents and do not require successive layers of neural populations (as in the saliency model of Itti et al., 1998).

Another important aspect of modeling results is that due to noise and basic mechanisms for saliency computations (i.e., center-surround computations via lateral interaction) the optimal architecture for these computations is for them to process visual inputs in separate populations of neurons selective for individual features. This feature could explain the inter-trial effects similarly to the parallel coactivation model of Krummenacher et al. (2001). In that model, the feature of the target on the preceding trial could deploy top-down attention to enhance processing in population selective to that feature, therefore, decreasing RT for the “same” versus “different” trials. In our model, the saliency signals in a population selective to the repeated feature could be enhanced due to feedback signals (caused by working memory of previously selected target), while the same feedback increases noise in the non-repeated population and results in a slower RT. Despite this advantage for separate processing of various features, future studies are required to explore the role of neural populations with mixed selectivity in saliency computations.

In the aforementioned model, only feedback from neurons in the saliency map to those in early visual areas was considered. However, we propose a more general form of feedback that also includes feedback between visual areas (from the next layer/population) as well as top-down signals from other cortical areas to the saliency map(s) (Figure 1). Moreover, because the projections that mediate feedback are active whenever the presynaptic neurons are active, independently of the task demands, the feedback is always present (unless top-down signals suppress these activities at their origin) and exerts their effects on visual processes. Considering the short delays in transmission of visual signals across brain areas, separating bottom-up attentional processes into feedforward and feedback components could be mechanistically impossible.

There are high-level models of bottom-up attention that address the influence of top-down signals on bottom-up attention in general (see Borji and Itti, 2013 for a review) or for improving object recognition (Oliva et al., 2003; Navalpakkam and Itti, 2005, 2006). In some of these models, top-down effects are simulated via multiplicative gain modulations of bottom-up computations (Navalpakkam and Itti, 2006) or as an abstract term (contextual priors) in computing the posterior probability of an object being present (Oliva et al., 2003). However, in most computational models of visual attention that strive to predict the pattern of eye movements in real time, the distinction between top-down and bottom-up processes are not clear (Borji and Itti, 2013). Importantly, the main result of those modeling works is that top-down signals are crucial to achieve performance that matches

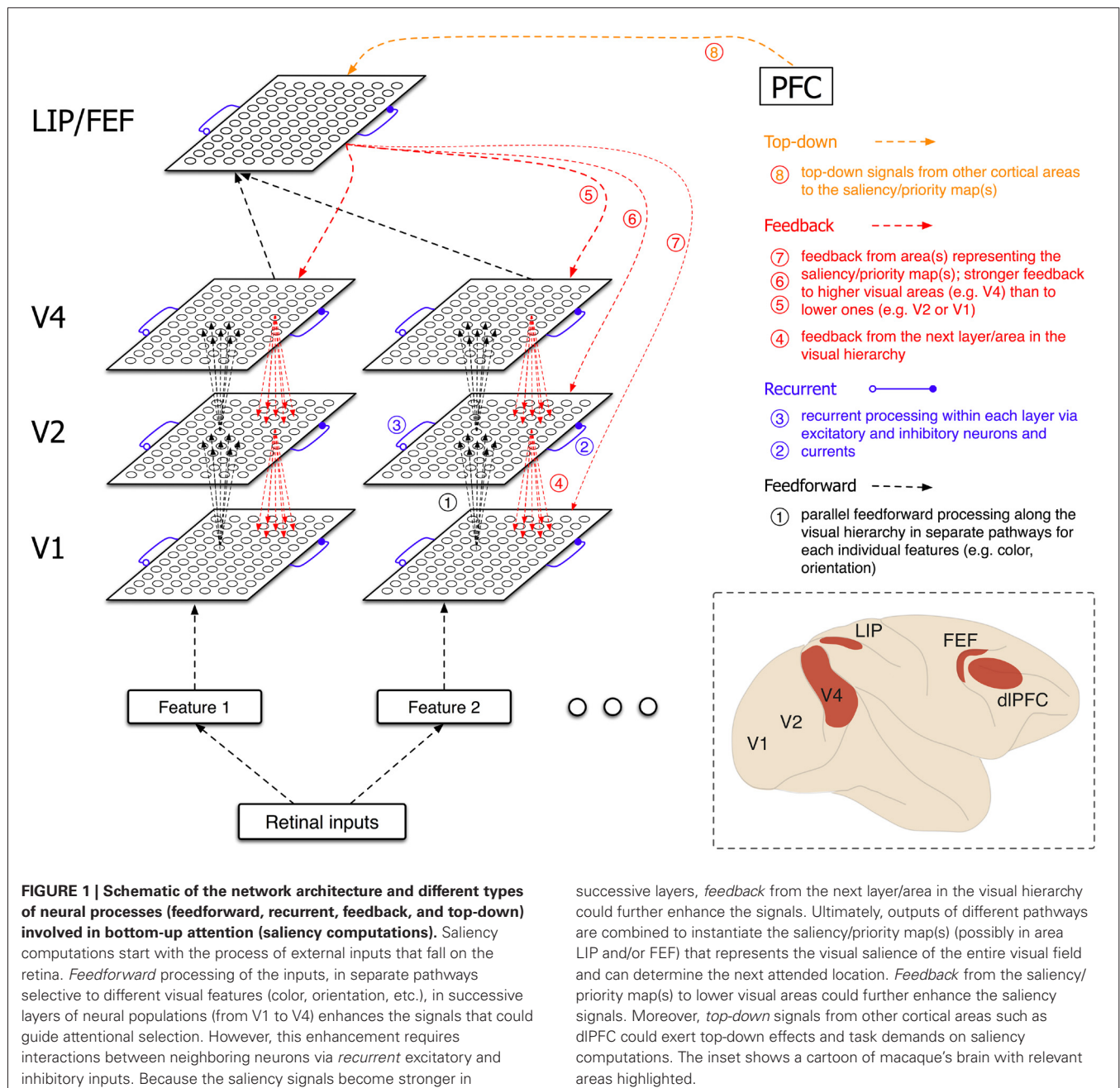
human visual performance and can accurately predict eye movements. However, because of the high-level nature of these models, computations performed by these models are not constrained and so are not biophysically plausible. Therefore, these models do not elucidate biophysical constraints underlying bottom-up attention that could reveal the role of feedback on bottom-up attention. Perhaps, the lack of distinction between bottom-up and top-down processes in more advanced models of visual attention is an indication that one cannot separate these processes based on behavior alone.

MORE EVIDENCE FOR THE ROLE OF FEEDBACK: SPEED OF FEEDFORWARD, RECURRENT, AND FEEDBACK PROCESSES

As described above, one of the reasons for assuming that bottom-up attention relies on feedforward processes is because feedback and recurrent inputs are not fast enough, e.g., by considering the time it takes for the visual signals to travel from lower to higher visual areas and back. Nevertheless, as shown by the recent modeling work, even recurrent inputs through slow NMDA synapses do not impede the emergence of saliency signals in successive layers of neural populations, and feedback can enhance those signals (Soltani and Koch, 2010). Interestingly, compatible with the model's assumptions and predictions, there is growing neurophysiological evidence that feedback and recurrent inputs actually do contribute to bottom-up attention (see below).

As shown by computational models with different levels of detail, saliency computations heavily rely on center-surround computations (Itti and Koch, 2001). One prevalent form of center-surround computations recorded neurophysiologically is the surround suppression (i.e., suppression of response by stimuli outside the classical RF (CRF); Cavanaugh et al., 2002). Surround suppression is observed even in the primary visual cortex as well as retinal ganglion (Kruger et al., 1975) and LGN cells (Levick et al., 1972), and has been assumed to be instantiated by horizontal connections from neighboring neurons with similar selectivity. However, by analyzing the timing of surround suppression and how it depends on the distance of the stimuli outside the CRF, Bair et al. (2003) found not only that the latency of suppression depends on its strength but also that this suppression could arrive faster than the excitatory CRF response and does not depend on the distance of the surround stimuli. To explain these results, Bair et al. (2003) suggested that in addition to recurrent inputs, surround suppression in V1 might be strongly influenced by feedback from higher visual areas (e.g., V2) with a larger RF. In another experiment, Hupe et al. (1998) found that feedback from higher visual areas (area V5) is crucial for surround suppression within early visual areas (V1, V2, and V3). More specifically, the authors showed that inactivation of V5 greatly reduces surround suppression in V3 neurons. These findings corroborate the idea that even the simplest form of saliency computations depends on feedback which could enhance the speed of computations within the same layer simply because feedback connections are faster than horizontal connections by an order of magnitude (Bringuier et al., 1999; Girard et al., 2001).

While feedforward connections are faster than horizontal connections, it is known that feedforward and feedback connections are equally fast (about 3.5 m/s) and have latencies



as short as 1.5 ms (Girard et al., 2001). Therefore, feedback processing can be as fast as feedforward processing, but with the advantage that higher visual areas carry larger saliency signals as shown experimentally and computationally (Hegdé and Felleman, 2003; Burrows and Moore, 2009; Soltani and Koch, 2010; Bogler et al., 2011; Melloni et al., 2012; see below). Interestingly, the difference in the response latency in different visual areas can be very small, while the represented signals can be very different at different time points. For example, Bisley et al. (2004) showed that the visual response in LIP could emerge as quickly as 40 ms, which matches the latency of the visual response in the primary visual cortex. This could happen by bypassing

successive processing of visual information (Schmolesky et al., 1998), and indicates that signals from salient targets may emerge in higher visual areas very quickly.

CONTRIBUTION OF DIFFERENT BRAIN AREAS TO SALIENCY COMPUTATIONS

Whereas early studies that investigated the neural representation of bottom-up attention found saliency signals in early visual areas such as V1 (Knierim and van Essen, 1992), later studies showed that distinct saliency signals are only present in higher visual areas (Hegdé and Felleman, 2003; Burrows and Moore, 2009; Betz et al., 2013). As mentioned earlier, electrophysiological studies were able

to distinguish pure saliency signals from signals that merely reflect a contrast between the center and surround by measuring the response of neurons to different types of stimuli for which the target has different levels of saliency (Hegd  and Felleman, 2003; Burrows and Moore, 2009).

Using a similar approach, recent fMRI studies indicate that saliency signals emerge gradually over successive brain areas. For example, an attempt to identify how saliency signals progress through the brain demonstrated that while activity in early visual areas is correlated with the graded saliency in natural images, activity in higher visual areas (such as anterior intraparietal sulcus and the FEF) is correlated with the signal associated with the most salient location in the visual field (Bogler et al., 2011). The latter observation supports the idea that a winner-take-all mechanism results in selection of the most salient location only in higher visual areas. Another recent study showed the gradual emergence of bottom-up signals in early visual areas (Melloni et al., 2012). Specifically, considering a “TSO-DSC” stimulus (a stimulus that contains a target that was singleton in orientation but also contains a highly salient distractor in a task-irrelevant dimension) as a conjunction stimulus, the patterns of activation across successive visual areas are similar to the results from the computational model of Soltani and Koch (2010). That is, only in V4, the response to both types of popout is larger than the response to the conjunction stimulus. Finally, compatible with what Burrows and Moore (2009) reported, an fMRI study found that in the presence of a demanding central task, saliency signals (in the form of orientation popout) are only present in higher visual areas (V3 and V4) and not in V1 (Bogler et al., 2013).

Despite strong neural evidence for the instantiation of saliency map in higher cortical area, it has been argued that the saliency map could be represented by V1 neurons (Li, 2002). The support for this proposal has been mainly based on behavioral data, but a recent fMRI study has provided some neural evidence for saliency signals (in the absence of awareness) in V1–V4 and not higher cortical areas (Zhang et al., 2012). However, using stimuli for which saliency and luminance contrast were uncorrelated, a more recent study showed that most BOLD activity in early visual areas (V1–V3) is dominated by contrast-dependent processes and does not comprise contrast invariance which is necessary for saliency representation (Betz et al., 2013).

Moreover, instantiation of saliency map in early visual areas is not very feasible and imposes serious constraints for saliency computations and the observed effects of top-down signals. Firstly, area V1 is not well-equipped for performing saliency computations. For example, V1 neurons lack certain feature selectivity and therefore, saliency computations in V1 imply that those features cannot contribute to saliency and bottom-up attention. Secondly, saliency computations (center-surround computations, pooling of signals over different features) eliminate some of the information presents in V1 and therefore, limiting information processing that higher visual area can perform on the output of V1. Thirdly, feedback projections to V1 are not very strong and this significantly limits the effects of top-down signals on saliency computations. Finally, our computational results show that saliency computations require successive processing of visual information over multiple layers and cannot be replaced by

a stronger interaction within one layer of neural population (Soltani and Koch, 2010). For these reasons, we think that instantiation of a real saliency map in V1 is not plausible.

As mentioned earlier, the observed asymmetries in popout detection reveal the importance of feedback in bottom-up attention. For example, the finding of Schiller and Lee that lesions of V4 differentially affect detection of the most and least salient targets (Schiller and Lee, 1991; which was used by Braun, 1994 as evidence for different attentional strategies) could indicate that detection of any target requires feedback. More specifically, in the case of detecting the least salient target, feedback from higher visual areas is required to suppress the activity in most parts of the visual space, a process that could be easily interrupted by V4 lesions. On the other hand, detection of the most salient target requires only feedback that enhances activity in the target location, a process that could be only mildly disrupted by V4 lesions.

Whereas we mainly discussed the modulation of bottom-up attention by top-down signals via their effects on early visual areas, there is experimental evidence that even activity in the putative saliency map is modulated by top-down signals (Thompson et al., 2005; Ipata et al., 2006; Suzuki and Gottlieb, 2013). In one study Thompson et al. (2005) showed that during a popout search task, where target and distractor colors switched unpredictably, monkeys made more erroneous saccades to distractors on the first trial after the switch. Importantly, presaccadic neural activity in the FEF was informative about the selected stimulus independently of whether the stimulus was a popout target or one of many distractors. Moreover, the signal conveyed by FEF neurons was correlated with the probability that a given target would be selected, indicative of this area to instantiate the saliency map. In another study Ipata et al. (2006) trained a monkey to ignore the presentation of a popout distractor during a visual search task, while they recorded from LIP neurons. They found that on trials where the monkey ignored the distractor, the LIP response to the salient distractor was smaller than the response to a non-salient distractor. Recently, Suzuki and Gottlieb (2013) compared the ability of LIP and dorsolateral prefrontal cortex (dlPFC) in suppressing distractors using a memory saccade task where a salient distractor was flashed at variable delays and locations during the memory delay. Interestingly, they found that not only dlPFC neurons showed stronger distractor suppression than LIP neurons, but also reversible inactivation of dlPFC gave rise to larger increases in distractibility than inactivation of LIP. Overall, these results show that even the activity of neurons in the putative saliency map is modulated by top-down signals and moreover, these signals strongly contribute to performance in attention tasks.

Considering strong projections from areas representing the putative saliency map (LIP/FEF) to lower cortical areas (Blatt et al., 1990; Schall et al., 1995) and the fact that this feedback is present as long as the former areas are active (in both correct and incorrect trials), one can predict specific effects of activity in the saliency map on neural processes in lower visual areas. Interestingly, the modeling results described above indicate that the main reason for a concurrent task (or even saccade planning as in Burrows and Moore) interfering with bottom-up computations

is the influence of the concurrent task on the activity in the saliency map (LIP or FEF). This happens because the bump of activity from planning a saccade suppresses neural activity in most parts of the saliency map except the saccade location, interrupting and altering feedback from neurons in those parts of the map. The behavioral results for detecting two popout targets at various distances show that the RT redundancy gain (shortening of RT when popout is defined by two features compared to when it is defined by one feature) decreases as the distance between the two targets increases (Krummenacher et al., 2002). This may be explained by the fact that two bumps of activity in the saliency map interact weakly if they are too far from each other (or alternatively due to interactions in early visual areas, which is less likely due to weaker interactions between neurons selective to different features in these areas). On the other hand, at short distances these bumps compete (with higher probability of winning for the faster detected (more salient) target) resulting in an increase in feedback based on the most salient location and therefore higher RT gains. Future experiments are needed to study the effects of inter-trial variability of neural responses in higher cortical areas on bottom-up attentional processes in lower visual areas.

SIMILARITIES BETWEEN BOTTOM-UP AND TOP-DOWN ATTENTION

Considering that top-down attention likely involves feedback inputs, examining similarities between bottom-up and top-down attention can further shed light on the role of feedback in bottom-up attention. These include similarities in: the timing of bottom-up and top-down attentional signals in different brain areas; neural substrates of bottom-up and top-down attention; and involved neurotransmitters.

Importantly, a few studies have examined the timing of bottom-up and top-down attentional signals in different brain areas. In one study, Buschman and Miller (2007) found earlier bottom-up signals in LIP than in lateral prefrontal cortex and the FEF whereas FEF neurons detected conjunction targets before LIP neurons. Other studies, however, point to a more complicated formation of attentional signals in prefrontal and parietal cortices. For example, a recent study by Katsuki and Constantinidis (2012) showed that neurons in dlPFC and posterior parietal cortex signal bottom-up attention around the same time. Interestingly, there is evidence that top-down attentional enhancements of activity within visual cortices are larger and earlier in higher areas (V4) compared to lower areas (V1), indicative of a “backward” propagation of modulatory signals (Mehta et al., 2000a,b; Buffalo et al., 2010). Moreover, the laminar source of attentional modulations in primary visual cortices supports the idea that feedback from the next visual area in the hierarchy is the origin of these modulations (Mehta et al., 2000a,b). This is compatible with the finding that during top-down attention, the FEF neurons exhibit attentional modulation about 50 ms before V4 neurons (Gregoriou et al., 2009). These observed trends of neural modulations resemble successive processing of bottom-up attentional signals, and earlier emergence of saliency signals in higher visual areas.

Interestingly, even the timing of top-down attentional signals could be similar between the lower and higher visual areas.

A recent study found that signals related to object-based attention can be detected in primary visual areas and the FEF at the same time (by simultaneous recording from V1 and the FEF), and that the interaction between these areas determines the dynamics of target selection (Pooremaeili et al., 2014). These observations challenge the feedforward assumption behind the formation of bottom-up attentional signals and point to the role of reciprocal interactions within lower and higher visual/cortical areas. An interesting aspect of the observed neural response in the FEF (which was not present in V1) was an increase in the differential response to target and distracter over time, indicative of a winner-take-all process in the FEF. Comparing recordings from V1 and the FEF, which seem to reside on the opposite sides of visual hierarchy for visual attention, shows that while the visual response in V1 occurs earlier than in the FEF, the selection signal occurs at the same time in both of these areas (Khayat et al., 2009). However, the modulation index of neuronal response in area V1 was much smaller than the one in the FEF indicating more enhanced signals in the latter area. Interestingly, Khayat et al. (2009) also found that on error trials FEF activity precedes V1 activity and therefore imposes its erroneous decision. These results show the important role of ever-present feedback from higher cortical areas in object-based attention.

Another piece of evidence supporting similarities between neural substrates underlying bottom-up and top-down attention comes from two separate experiments measuring the effects of FEF microstimulation on information processing in other visual areas. Considering the FEF as a higher visual area that controls top-down attention, one would assume that its microstimulation would enhance visual signals in lower visual areas that show attentional modulations, independently of bottom-up driven signals in the latter areas. However, using different methods for measuring signals (single cell recordings and fMRI), two separate experiments found that induced enhancements of visual signals depended on the already present bottom-up signals. In one study, Moore and Armstrong (2003) found an increase in spiking activity in V4 only when a target was present in the V4 RF, and this enhancement was larger in the presence of a competing distracter. In another study, in which changes in fMRI BOLD responses throughout visual cortex were measured, Ekstrom and Roelfsema (2008) found that the effect of FEF microstimulation on posterior visual areas (such as V4) depends on the stimulus contrast and the presence of distracters. These results demonstrate that even artificially simulated top-down effects are not independent of bottom-up saliency signals, which renders the distinction between feedforward and feedback processes even more unnecessary.

The modeling results also predicted that saliency computations should rely on excitatory and inhibitory recurrent inputs within each layer of neural populations and the excitatory recurrent input should be dominated by NMDA receptors (and not AMPA receptors), in order to integrate saliency signals in the presence of cortical noise (Soltani and Koch, 2010). There is recent experimental evidence supporting this prediction. In one study, Self et al. (2012) used different drugs to measure the contribution of AMPA and NMDA receptors to figure-ground modulations (the increased activity of neurons representing the figure compared with the background) in V1. They found that

AMPA currents mainly contribute to feedforward processing and not to the figure-ground modulations, whereas NMDA blockade reduces figure-ground modulations. Another recent study showed that NMDA, and not AMPA, receptors contribute to the reduction of variance and noise correlation due to attention (Herrero et al., 2013). Both these results corroborate the modeling results that NMDA receptors are crucial for saliency computations and bottom-up attention.

Interestingly, NMDA receptors are modulated by dopamine (Cepeda et al., 1998; Seamans et al., 2001), the main neurotransmitter for signaling reward that also influences working memory (Williams and Goldman-Rakic, 1995) and can alter visual processes on a long timescale (Bao et al., 2001). However, in the absence of strong dopaminergic projections to primary visual cortex (Lewis and Melchitzky, 2001), most dopamine-dependent modulations of visual processing may occur via dopamine effects on prefrontal activity and resulting modulated feedback. For example, recent studies found that dopamine effects on the FEF activity can enhance the visual response in V4 neurons (Noudoost and Moore, 2011) and contribute to adaptive target selection (Soltani et al., 2013) via specific types of receptors. Considering the effects of dopamine on working memory and the fact that top-down attention requires some forms of working memory, one might regard dopamine as the primary neuromodulator for top-down attention. However, one needs to exercise caution because of the aforementioned evidence for the role of feedback in bottom-up attention suggesting that dopamine could have a significant role in bottom-up attention.

In summary, the role of NMDA in saliency computations highlights shared neural substrates for bottom-up attention and cognitive processes that are not considered feedforward (such as working memory and decision making). Moreover, the strong effects of neuromodulators on NMDA receptors indicates how various neuromodulators could affect bottom-up attention via their effects on higher visual areas that provide feedback to early visual areas, or by directly altering saliency computations.

The aforementioned similarities between bottom-up and top-down attentional processes removes a clear distinction between neural substrates of bottom-up and top-down attention based on the location of a given area in the visual hierarchy, and point to a stronger role of feedback in bottom-up attention. Moreover, similarities between bottom-up and top-down attention, which are originally assumed to rely on feedforward and feedback inputs, respectively, indicate that both these inputs are important for both types of attention. These observations signify that the main dichotomy of visual attention should be disregarded in the search of more unified models of attention.

Recently, Awh et al. (2012) have elegantly challenged the bottom-up and top-down dichotomy and instead proposed a framework that relies on a priority map that integrates multiple selection mechanisms and biases including: current goals, selection history, and physical salience. Specifically, they summarized experimental evidence supporting the idea that both recent history of attentional deployment as well as the reward history can bias visual selection independently of the current goals (top-down signals) or stimulus salience (bottom-up signals). Interestingly, the main argument of Awh et al. (2012) for the failure of atten-

tional dichotomy is unexplained selection biases due to lingering effects of past experience, either selection history and reward history. However, the only feasible mechanism for the effects of past experience on attentional selection could be synaptic plasticity, if one wants to truly separate mechanisms underlying these effects from those serving the influence of top-down signals (due to some sustained activity in some higher cortical areas). Because of different timescales of selection history and reward history, their effects should rely on short-term and long-term synaptic plasticity, respectively. This has important implications for the effects of neuromodulators on bottom-up attention.

DISCUSSION AND FUTURE DIRECTIONS

We have reviewed the experimental and modeling evidence for the role of feedback in bottom-up attention. From a behavioral point of view, there is evidence that top-down signals are necessary for observing characteristics of bottom-up attentional processes such as the very fast detection of oddball targets. From a neuronal point of view, signals reflecting the saliency of an object can be diminished when top-down signals are interrupted, for example during saccade preparation. From a computational point of view, bottom-up, saliency computations can be enhanced by feedback from higher visual areas that represent the saliency map. Considering this evidence, it may be logical to replace bottom-up attention with salience-dependent attention, as the former term implies a specific direction for information processing which is not compatible with most experimental or computational results.

As suggested by Awh et al. (2012), some of the experimental findings reviewed here can be considered as the lingering effects of past experience on attentional selection (due to recent history of attentional deployment). This includes inter-trial effects on performance and RT (e.g., work of Krummenacher and colleagues). On the other hand, effects of reward history on attentional selection are not discussed here but are of great importance for understanding attentional processes. Both their and our proposals challenge the bottom-up versus top-down dichotomy, but point to different mechanisms that could account for unexplained observations. More specifically, Awh et al. (2012) point to the role of short-term and long-term synaptic plasticity in the feedforward pathways to explain some of the observed selection biases. In contrast, we assign an important role for feedback between successive stages of saliency computations and from the saliency map(s) to lower visual areas, as well as interaction and integration of top-down and bottom-up signals within the saliency map(s). While these mechanisms are not exclusive, future work is needed to clarify the specific role and relative contribution of them in attentional selection.

Overall, the reviewed findings indicate that in order to reveal the neural substrates of attentional processes, the focus should be shifted toward understanding biophysical mechanisms through which the necessary computations could be performed, and whether a specific brain area has the proper neural type and connectivity to perform those computations. Therefore, even though the results described above reduce the role of unidirectional, hierarchal computations (i.e., from lower to higher visual areas) and minimize the distinction between feedforward

and feedback inputs in bottom-up attention, one should not ignore the anatomical and biophysical constraints underlying these computations. For example, whereas successive processing across visual hierarchy can be bypassed, area V4 still sends more projections to higher visual areas (such as the FEF) than area V1 (Schall et al., 1995). On the other hand, the projections from the FEF to area V4 mostly target pyramidal neurons in primary visual areas (Anderson et al., 2011). The observed lack of projections to inhibitory neurons limits mechanisms through which feedback projections could exert modulatory effects, instead of just driving the recipient areas. Understanding the implications of these and other constraints on feedforward and feedback processing could provide valuable insight into understanding bottom-up attention in particular and vision in general.

There are still many unanswered questions about the role of feedback in bottom-up, exogenous attention. Firstly, while the benefit of feedback from a higher visual area representing the saliency map has been established, there is need for further research that investigates the effects of feedback between each successive layers/areas using detailed computational models. Secondly, the saliency signals are observed in many brain areas (FEF, LIP, superior colliculus, dIPFC), all of which provide feedback to early visual areas. This indicates that there should be interaction between these signals in order to deploy attention to a unique location; understanding this interaction is crucial for understanding bottom-up attention. Interestingly, some of these areas contribute to top-down attention, which requires working memory, and it is important to see how saliency and working memory signals interact and integrate within the saliency/priority maps. Thirdly, all primary visual areas receive feedback from higher visual areas representing the saliency map. Future computational work is needed to elucidate the relative contribution of feedback to a specific brain area (e.g., V1 in comparison to V4). Overall, considering the complexity of behavioral and neural data, more detailed computational models are needed to study interaction and integration of feedforward and feedback inputs in order to provide a more coherent account of bottom-up attention and its underlying neural mechanisms.

As experimental methods for manipulations and simultaneous measurements of neural activity improve, there is a greater need for more extensive and detailed computational models to interpret the outcome data and provide predictions for future experiments. Future experiments with simultaneous recording of neural activity should allow us to study the relationship between feature selectivity (tuning) and saliency signals for individual neurons. Computational models are needed to explain such relationships and how different neural types contribute to bottom-up attention. Similarly, drug manipulations of various brain areas provide another opportunity for computational modeling to contribute, considering the large number of involved receptors and contradictory possible outcomes (e.g., Disney et al., 2007; Herrero et al., 2008).

REFERENCES

- Ahissar, M., and Hochstein, S. (1997). Task difficulty and the specificity of perceptual learning. *Nature* 387, 401–406. doi: 10.1038/387401a0
- Anderson, J. C., Kennedy, H., and Martin, K. A. C. (2011). Pathways of attention: synaptic relationships of frontal eye field to V4, lateral intraparietal cortex, and area 46 in macaque monkey. *J. Neurosci.* 31, 10872–10881. doi: 10.1523/JNEUROSCI.0622-11.2011
- Awh, E., Belopolsky, A. V., and Theeuwes, J. (2012). Top-down versus bottom-up attentional control: a failed theoretical dichotomy. *Trends Cogn. Sci.* 16, 437–443. doi: 10.1016/j.tics.2012.06.010
- Bair, W., Cavanaugh, J. R., and Movshon, J. A. (2003). Time course and time-distance relationships for surround suppression in macaque V1 neurons. *J. Neurosci.* 23, 7690–7701.
- Bao, S., Chan, V. T., and Merzenich, M. M. (2001). Cortical remodelling induced by activity of ventral tegmental dopamine neurons. *Nature* 412, 79–83. doi: 10.1038/35083586
- Betz, T., Kietzmann, T., Wilming, N., and König, P. (2010). Investigating task-dependent top-down effects on overt visual attention. *J. Vis.* 10, 1–14. doi: 10.1167/10.3.15
- Betz, T., Wilming, N., Bogler, C., Haynes, J.-D., and König, P. (2013). Dissociation between saliency signals and activity in early visual cortex. *J. Vis.* 13, 1–12. doi: 10.1167/13.14.6
- Bisley, J. W., Krishna, B. S., and Goldberg, M. E. (2004). A rapid and precise on-response in posterior parietal cortex. *J. Neurosci.* 24, 1833–1838. doi: 10.1523/JNEUROSCI.5007-03.2004
- Blatt, G. J., Andersen, R. A., and Stoner, G. R. (1990). Visual receptive field organization and corticocortical connections of the lateral intraparietal area (area LIP) in the macaque. *J. Comp. Neurol.* 299, 421–445. doi: 10.1002/cne.902990404
- Bogler, C., Bode, S., and Haynes, J.-D. (2011). Decoding successive computational stages of saliency processing. *Curr. Biol.* 21, 1667–1671. doi: 10.1016/j.cub.2011.08.039
- Bogler, C., Bode, S., and Haynes, J.-D. (2013). Orientation pop-out processing in human visual cortex. *Neuroimage* 81, 73–80. doi: 10.1016/j.neuroimage.2013.05.040
- Borji, A., and Itti, L. (2013). State-of-the-art in visual attention modeling. *IEEE Trans. Pattern Anal. Mach. Intell.* 35, 185–207. doi: 10.1109/TPAMI.2012.89
- Braun, J. (1994). Visual search among items of different salience: removal of visual attention mimics a lesion in extrastriate area V4. *J. Neurosci.* 14, 554–567.
- Braun, J., and Julesz, B. (1998). Withdrawing attention at little or no cost: detection and discrimination tasks. *Percept. Psychophys.* 60, 1–23. doi: 10.3758/BF03211915
- Bringuier, V., Chavane, F., Glaeser, L., and Frégnac, Y. (1999). Horizontal propagation of visual activity in the synaptic integration field of area 17 neurons. *Science* 283, 695–699. doi: 10.1126/science.283.5402.695
- Buffalo, E. A., Fries, P., Landman, R., Liang, H., and Desimone, R. (2010). A backward progression of attentional effects in the ventral stream. *PNAS* 107, 361–365. doi: 10.1073/pnas.0907658106
- Burrows, B. E., and Moore, T. (2009). Influence and limitations of popout in the selection of salient visual stimuli by area V4 neurons. *J. Neurosci.* 29, 15169–15177. doi: 10.1523/JNEUROSCI.3710-09.2009
- Buschman, T., and Miller, E. (2007). Top-down versus bottom-up control of attention in the prefrontal and posterior parietal cortices. *Science* 315, 1860–1862. doi: 10.1126/science.1138071
- Cavanaugh, J. R., Bair, W., and Movshon, J. A. (2002). Nature and interaction of signals from the receptive field center and surround in macaque V1 neurons. *J. Neurophysiol.* 88, 2530–2546. doi: 10.1152/jn.00692.2001
- Cepeda, C., Colwell, C. S., Itri, J. N., Chandler, S. H., and Levine, M. S. (1998). Dopaminergic modulation of NMDA-induced whole cell currents in neostriatal neurons in slices: contribution of calcium conductances. *J. Neurophysiol.* 79, 82–94.
- Disney, A. A., Aoki, C., and Hawken, M. J. (2007). Gain modulation by nicotine in macaque v1. *Neuron* 56, 701–713. doi: 10.1016/j.neuron.2007.09.034
- Einhäuser, W., Rutishauser, U., and Koch, C. (2008). Task-demands can immediately reverse the effects of sensory-driven saliency in complex visual stimuli. *J. Vis.* 8, 1–19. doi: 10.1167/8.2.2
- Ekstrom, L., and Roelfsema, P. (2008). Bottom-up dependent gating of frontal signals in early visual cortex. *Science* 321, 414–417. doi: 10.1126/science.1153276
- Found, A., and Müller, H. J. (1996). Searching for unknown feature targets on more than one dimension: investigating a “dimension-weighting” account. *Percept. Psychophys.* 58, 88–101. doi: 10.3758/BF03205479
- Girard, P., Hupé, J. M., and Bullier, J. (2001). Feedforward and feedback connections between areas V1 and V2 of the monkey have similar rapid conduction velocities. *J. Neurophysiol.* 85, 1328–1331.

- Gottlieb, J. P., Kusunoki, M., and Goldberg, M. E. (1998). The representation of visual salience in monkey parietal cortex. *Nature* 391, 481–484. doi: 10.1038/35135
- Gregoriou, G. G., Gotts, S. J., Zhou, H., and Desimone, R. (2009). High-frequency, long-range coupling between prefrontal and visual cortex during attention. *Science* 324, 1207–1210. doi: 10.1126/science.1171402
- Hegd , J., and Felleman, D. (2003). How selective are V1 cells for pop-out stimuli? *J. Neurosci.* 23, 9968–9980.
- Herrero, J. L., Gieselmann, M. A., Sanaye, M., and Thiele, A. (2013). Attention-induced variance and noise correlation reduction in macaque V1 is mediated by NMDA receptors. *Neuron* 78, 729–739. doi: 10.1016/j.neuron.2013.03.029
- 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. doi: 10.1038/nature07141
- Hupe, J., James, A., Payne, B., and Lomber, S. (1998). Cortical feedback improves discrimination between figure and background by V1, V2, and V3 neurons. *Nature* 394, 1–4. doi: 10.1038/27710
- Ipata, A. E., Gee, A. L., Gottlieb, J., Bisley, J. W., and Goldberg, M. E. (2006). LIP responses to a popout stimulus are reduced if it is overtly ignored. *Nat. Neurosci.* 9, 1071–1076. doi: 10.1038/nn1734
- Itti, L., Koch, C., and Niebur, E. (1998). A model of saliency-based visual attention for rapid scene analysis. *IEEE Trans. Pattern Anal. Mach. Intell.* 20, 1254–1259. doi: 10.1109/34.730558
- Itti, L., and Koch, C. (2001). Computational modelling of visual attention. *Nat. Rev. Neurosci.* 2, 1–11. doi: 10.1038/35058500
- Joseph, J. S., Chun, M. M., and Nakayama, K. (1997). Attentional requirements in a “preattentive” feature search task. *Nature* 387, 805–807. doi: 10.1038/42940
- Katsuki, F., and Constantinidis, C. (2012). Early involvement of prefrontal cortex in visual bottom-up attention. *Nat. Neurosci.* 15, 1160–1166. doi: 10.1038/nn.3164
- Khayat, P., Pooremaeli, A., and Roelfsema, P. (2009). Time course of attentional modulation in the frontal eye field during curve tracing. *J. Neurophysiol.* 101, 1813–1822. doi: 10.1152/jn.91050.2008
- Knierim, J. J., and van Essen, D. C. (1992). Neuronal responses to static texture patterns in area V1 of the alert macaque monkey. *J. Neurophysiol.* 67, 961–980.
- Koch, C., and Ullman, S. (1985). Shifts in selective visual attention: towards the underlying neural circuitry. *Hum. Neurobiol.* 4, 219–227.
- Kruger, J., Fischer, B., and Barth, R. (1975). The shift-effect in retinal ganglion cells of the rhesus monkey. *Exp. Brain Res.* 23, 443–446. doi: 10.1007/BF00238025
- Krummenacher, J., Grubert, A., and M ller, H. J. (2010). Inter-trial and redundant-signals effects in visual search and discrimination tasks: separable pre-attentive and post-selective effects. *Vision Res.* 50, 1382–1395. doi: 10.1016/j.visres.2010.04.006
- Krummenacher, J., M ller, H. J., and Heller, D. (2001). Visual search for dimensionally redundant pop-out targets: evidence for parallel-coactive processing of dimensions. *Percept. Psychophys.* 63, 901–917. doi: 10.3758/BF03194446
- Krummenacher, J., M ller, H. J., and Heller, D. (2002). Visual search for dimensionally redundant pop-out targets: parallel-coactive processing of dimensions is location specific. *J. Exp. Psychol. Hum. Percept. Perform.* 28, 1303–1322. doi: 10.1037/0096-1523.28.6.1303
- Kustov, A. A., and Robinson, D. L. (1996). Shared neural control of attentional shifts and eye movements. *Nature* 384, 74–77. doi: 10.1038/384074a0
- Levick, W. R., Cleland, B. G., and Dubin, M. W. (1972). Lateral geniculate neurons of cat: retinal inputs and physiology. *Invest. Ophthalmol. Vis. Sci.* 11, 302–311.
- Lewis, D., and Melchitzky, D. (2001). Dopamine transporter immunoreactivity in monkey cerebral cortex: regional, laminar, and ultrastructural localization. *J. Comp. Neurol.* 432, 119–136. doi: 10.1002/cne.1092
- Li, Z. (2002). A saliency map in primary visual cortex. *Trends Cogn. Sci.* 6, 9–16. doi: 10.1016/S1364-6613(00)01817-9
- Maljkovic, V., and Nakayama, K. (1994). Priming of pop-out: I. Role of features. *Mem. Cogn.* 22, 657–672. doi: 10.3758/BF03209251
- Mazer, J. A., and Gallant, J. L. (2003). Goal-related activity in V4 during free viewing visual search. Evidence for a ventral stream visual salience map. *Neuron* 40, 1241–1250. doi: 10.1016/S0896-6273(03)00764-5
- Mehta, A. D., Ulbert, I., and Schroeder, C. E. (2000a). Intermodal selective attention in monkeys. I: distribution and timing of effects across visual areas. *Cereb. Cortex* 10, 343–358. doi: 10.1093/cercor/10.4.343
- Mehta, A. D., Ulbert, I., and Schroeder, C. E. (2000b). Intermodal selective attention in monkeys. II: physiological mechanisms of modulation. *Cereb. Cortex* 10, 359–370. doi: 10.1093/cercor/10.4.359
- Melloni, L., van Leeuwen, S., Alink, A., and M ller, N. G. (2012). Interaction between bottom-up saliency and top-down control: how saliency maps are created in the human brain. *Cereb. Cortex* 22, 2943–2952. doi: 10.1093/cercor/bhr384
- Moore, T., and Armstrong, K. (2003). Selective gating of visual signals by microstimulation of frontal cortex. *Nature* 421, 370–373. doi: 10.1038/nature01341
- Mortier, K., Theeuwes, J., and Starreveld, P. (2005). Response selection modulates visual search within and across dimensions. *J. Exp. Psychol. Hum. Percept. Perform.* 31, 542–557. doi: 10.1037/0096-1523.31.3.542
- Nakayama, K., and Mackeben, M. (1989). Sustained and transient components of focal visual attention. *Vision Res.* 29, 1631–1647. doi: 10.1016/0042-6989(89)90144-2
- Navalpakkam, V., and Itti, L. (2005). Modeling the influence of task on attention. *Vision Res.* 45, 205–231. doi: 10.1016/j.visres.2004.07.042
- Navalpakkam, V., and Itti, L. (2006). An integrated model of top-down and bottom-up attention for optimizing detection speed. *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.* 2, 2049–2056. doi: 10.1109/CVPR.2006.54
- Noudoost, B., and Moore, T. (2011). Control of visual cortical signals by prefrontal dopamine. *Nature* 474, 372–375. doi: 10.1038/nature09995
- Oliva, A., Torralba, A., Castelano, M. S., and Henderson, J. M. (2003). Top-down control of visual attention in object detection. *IEEE Int. Conf. Image Proc.* 1, 253–256.
- Pooremaeli, A., Poort, J., and Roelfsema, P. R. (2014). Simultaneous selection by object-based attention in visual and frontal cortex. *Proc. Natl. Acad. Sci. U.S.A.* 111, 6467–6472. doi: 10.1073/pnas.1316181111
- Schall, J. D., Morel, A., King, D. J., and Bullier, J. (1995). Topography of visual cortex connections with frontal eye field in macaque: convergence and segregation of processing streams. *J. Neurosci.* 15, 4464–4487.
- Schiller, P., and Lee, K. (1991). The role of the primate extrastriate area V4 in vision. *Science* 106, 1251–1253. doi: 10.1126/science.2006413
- Schmolesky, M. T., Wang, Y., Hanes, D. P., Thompson, K. G., Leutgeb, S., Schall, J. D., et al. (1998). Signal timing across the macaque visual system. *J. Neurophysiol.* 79, 3272–3278.
- Seamans, J. K., Durstewitz, D., Christie, B. R., Stevens, C. F., and Sejnowski, T. J. (2001). Dopamine D1/D5 receptor modulation of excitatory synaptic inputs to layer V prefrontal cortex neurons. *PNAS* 98, 301–306. doi: 10.1073/pnas.98.1.301
- Self, M. W., Kooijmans, R. N., Sup r, H., Lamme, V. A., and Roelfsema, P. R. (2012). Different glutamate receptors convey feedforward and recurrent processing in macaque V1. *PNAS* 109, 11031–11036. doi: 10.1073/pnas.1119527109
- Shipp, S. (2004). The brain circuitry of attention. *Trends Cogn. Sci.* 8, 223–230. doi: 10.1016/j.tics.2004.03.004
- Soltani, A., and Koch, C. (2010). Visual saliency computations: mechanisms, constraints, and the effect of feedback. *J. Neurosci.* 30, 12831–12843. doi: 10.1523/JNEUROSCI.1517-10.2010
- Soltani, A., Noudoost, B., and Moore, T. (2013). Dissociable dopaminergic control of saccadic target selection and its implications for reward modulation. *PNAS* 110, 3579–3584. doi: 10.1073/pnas.1221236110
- Suzuki, M., and Gottlieb, J. (2013). Distinct neural mechanisms of distractor suppression in the frontal and parietal lobe. *Nat. Neurosci.* 16, 98–104. doi: 10.1038/nn.3282
- Thompson, K. G., and Bichot, N. P. (2005). A visual salience map in the primate frontal eye field. *Prog. Brain Res.* 147, 251–262. doi: 10.1016/S0079-6123(04)47019-8
- Thompson, K. G., Bichot, N. P., Sato, T. R., Kirk, G., Bichot, N. P., and Sato, T. R. (2005). Frontal eye field activity before visual search errors reveals the integration of bottom-up and top-down salience. *J. Neurophysiol.* 93, 337–351. doi: 10.1152/jn.00330.2004
- Treisman, A. (1985). Preattentive processing in vision. *Comput. Vis. Graph. Image Process.* 177, 156–177. doi: 10.1016/S0734-189X(85)80004-9
- Treisman, A., and Gelade, G. (1980). A feature-integration theory of attention. *Cogn. Psychol.* 136, 97–136. doi: 10.1016/0010-0285(80)90005-5
- Treisman, A., and Gormican, S. (1988). Feature analysis in early vision: evidence from search asymmetries. *Psychol. Rev.* 95, 15–48. doi: 10.1037/0033-295X.95.1.15
- Treisman, A., and Sato, S. (1990). Conjunction search revisited. *J. Exp. Psychol. Hum. Percept. Perform.* 16, 459–478. doi: 10.1037/0096-1523.16.3.459

- Tsukada, H., Nishiyama, S., Fukumoto, D., Sato, K., Kakiuchi, T., and Domino, E. F. (2005). Chronic NMDA antagonism impairs working memory, decreases extracellular dopamine, and increases D1 receptor binding in prefrontal cortex of conscious monkeys. *Neuropsychopharmacology* 30, 1861–1869. doi: 10.1038/sj.npp.1300732
- Wang, M., Yang, Y., Wang, C. J., Gamo, N. J., Jin, L. E., Mazer, J. A., et al. (2013). NMDA receptors subserve persistent neuronal firing during working memory in dorsolateral prefrontal cortex. *Neuron* 77, 736–749. doi: 10.1016/j.neuron.2012.12.032
- Wang, X. J. (1999). Synaptic basis of cortical persistent activity: the importance of NMDA receptors to working memory. *J. Neurosci.* 19, 9587–9603.
- Wang, X. J. (2002). Probabilistic decision making by slow reverberation in cortical circuits. *Neuron* 36, 955–968. doi: 10.1016/S0896-6273(02)01092-9
- Williams, G., and Goldman-Rakic, P. (1995). Modulation of memory fields by dopamine D1 receptors in prefrontal cortex. *Nature* 376, 572–575. doi: 10.1038/376572a0
- Wolfe, J. M. (1994). Guided Search 2.0 A revised model of visual search. *Psychon. Bull. Rev.* 1, 202–238. doi: 10.3758/BF03200774
- Zhang, X., Zhaoping, L., Zhou, T., and Fang, F. (2012). Neural activities in V1 create a bottom-up saliency map. *Neuron* 73, 183–192. doi: 10.1016/j.neuron.2011.10.035
- Zucker, R. S., and Regehr, W. G. (2002). Short-term synaptic plasticity. *Annu. Rev. Physiol.* 64, 355–405. doi: 10.1146/annurev.physiol.64.092501.114547
- 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.
- Received: 27 July 2014; accepted: 30 January 2015; published online: 02 March 2015.
- Citation: Khorsand P, Moore T and Soltani A (2015) Combined contributions of feedforward and feedback inputs to bottom-up attention. *Front. Psychol.* 6:155. doi: 10.3389/fpsyg.2015.00155
- This article was submitted to Perception Science, a section of the journal *Frontiers in Psychology*.
- Copyright © 2015 Khorsand, Moore and Soltani. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Limits to the usability of iconic memory

Ronald A. Rensink*

Department of Psychology and Department of Computer Science, University of British Columbia, Vancouver, BC, Canada

Edited by:

Bruno Breitmeyer, University of Houston, USA

Reviewed by:

Bruno Breitmeyer, University of Houston, USA

Ijla G. Sligte, University of Amsterdam, Netherlands

*Correspondence:

Ronald A. Rensink, Department of Psychology and Department of Computer Science, University of British Columbia, 2136 West Mall, Vancouver, BC V6T 1Z4, Canada
e-mail: rensink@psych.ubc.ca; rensink@cs.ubc.ca

Human vision briefly retains a trace of a stimulus after it disappears. This trace—iconic memory—is often believed to be a surrogate for the original stimulus, a representational structure that can be used as if the original stimulus were still present. To investigate its nature, a flicker-search paradigm was developed that relied upon a full scan (rather than partial report) of its contents. Results show that for visual search it can indeed act as a surrogate, with little cost for alternating between visible and iconic representations. However, the duration over which it can be used depends on the type of task: some tasks can use iconic memory for at least 240 ms, others for only about 190 ms, while others for no more than about 120 ms. The existence of these different limits suggests that iconic memory may have multiple layers, each corresponding to a particular level of the visual hierarchy. In this view, the inability to use a layer of iconic memory may reflect an inability to maintain feedback connections to the corresponding representation.

Keywords: iconic memory, feedback connections, visual search, visual attention, visual memory

INTRODUCTION

It has long been known that human vision retains a brief trace of any stimulus it encounters (see e.g., Loftus and Irwin, 1998). This trace, often referred to as *iconic memory*, has been a focus of investigation for several decades (e.g., Sperling, 1960; Coltheart, 1980; Ruff et al., 2007; Sligte et al., 2010). It is sometimes considered to be a “visual echo” that can act as a surrogate, i.e., that as long as it lasts, its contents can be used in much the same way as if the stimulus were still visible. But there is little consensus as to what—if any—function iconic memory may have (see e.g., Pashler, 1998). On one hand, it has sometimes been considered a simple side effect, with potentially deleterious effects on perception (Haber, 1983). On the other, it could potentially increase the amount of information that could be extracted from a brief presentation (Haber, 1971).

Iconic memory has most often been studied via *partial report*, in which observers are briefly shown an array of a dozen or so items and then asked to report a subset that is cued after the array disappears (Sperling, 1960; Averbach and Coriell, 1961). Various studies have also examined the extent to which iconic representations can be used in memorization and recognition tasks (e.g., Loftus et al., 1992; Keyser et al., 2005) as well as change detection (e.g., Becker et al., 2000; Sligte et al., 2010). All assume that iconic memory is equally available to any visual process. But is this really so? Or might it be used to different extents by different processes?

To investigate this, a *flicker search* paradigm was developed (Figure 1). This is a variant of visual search, where the observer must determine as quickly as possible the presence or absence of a given *target* among a set of non-target items (or *distractors*) in a display; different visual operations can be tested by different choices of target and distractors (e.g., Treisman and Gormican, 1988; Wolfe and Horowitz, 2004). In flicker search, observers search displays that are visible only intermittently: after a fixed time (the display duration, or *on-time*), the display is blanked for some fixed interval [the interstimulus interval (ISI),

or *off-time*], this cycle then repeated until the observer responds or times out. (To enable maximal use of iconic memory, no masks are present.) For many kinds of search task, the time needed to respond is proportional to the *set size* (the number of items in the display), likely reflecting the application of an attentional mechanism (Treisman and Gormican, 1988; Wolfe and Horowitz, 2004). If this mechanism is sufficiently slow, search will require the scan of the blank intervals¹. The question then is whether the speed of search through a blank interval (i.e., iconic memory) is the same as through the representation that gave rise to it. This can be answered by comparing performance when iconic memory is used for different fractions of the display cycle.

Such a “full scan” technique removes several potential problems of partial report, such as complications due to memory consolidation and transfer; it also reduces the likelihood of observers using different strategies (cf. Estes and Taylor, 1964). Consequently, it may provide a more precise estimate of iconic properties. Importantly for the issue at hand, it also allows a wide variety of tasks to be examined using the same general framework.

¹ If the start of search after display onset is stochastic, and the variance of this is sufficient, random sampling will ensure that the fraction of on- or off-time encountered will on average be that in the display cycle. To help with this, observers were dropped from the analysis if search was over before the first display cycle was complete—i.e., before a full testing of the first iconic representation could be made. The criterion used was that search should be slow enough to allow the complete testing of 10 items (the maximum present) duration a single display cycle at the slowest cadence (320 ms). Note that this does not assume an item-by-item scan of the display; attention could be allocated to the items in parallel. On the basis of this criterion, two observers were removed: one from Experiment 1A, and one from Experiment 3C. More severe criteria did not significantly change the overall pattern of results.

For even the fastest search encountered here (c. 50 ms/item in Experiment 1A), a scan of both visual and iconic representations was essentially complete for displays containing only six items. Importantly, cadence affected only the slopes and not the shapes of the response-time curves (Figures 1 and 4). This provides evidence that the timing assumptions underlying this technique are reasonably accurate for the conditions examined here.

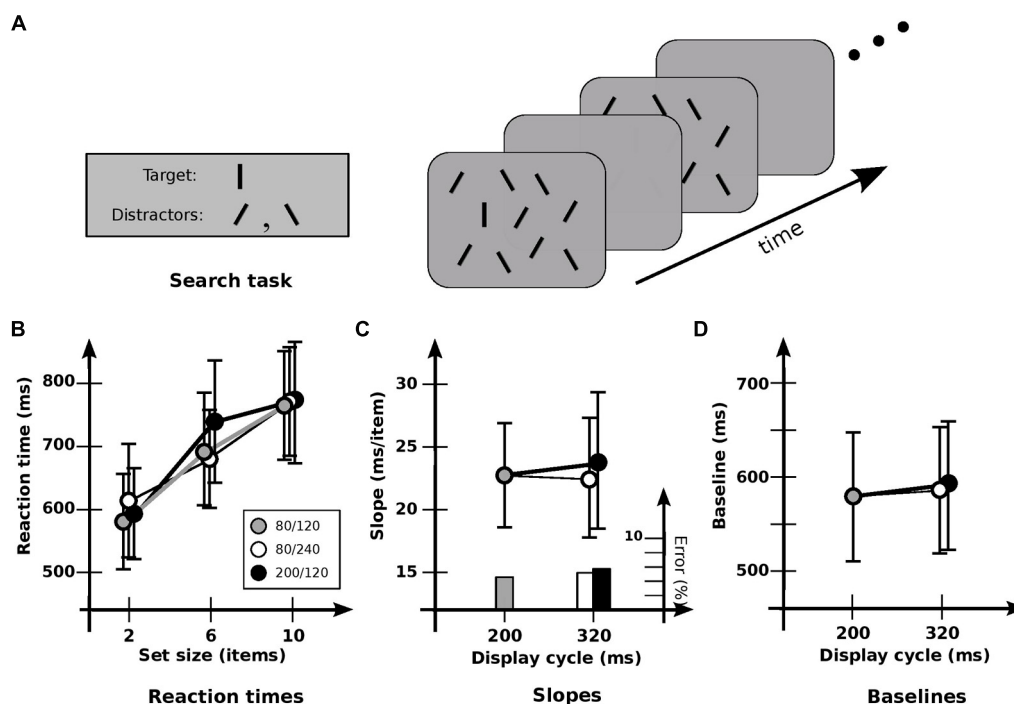


FIGURE 1 | Experiment 1A: detection of orientation. (A) General setup. Target is a vertical line; distractors are lines tilted $\pm 30^\circ$. Displays “flickered” until subject responded, or 5 s elapsed. (B) Response times and error rates as a function of set size for the three cadences. (C) Data recast as slopes. Slope for base cadence (23.0 ms/item) is unaffected by either an increase in off-time (22.1 ms/item) or an increase in on-time (24.4 ms/item). Note that

since these are target-present slopes from a presumably self-terminating search, the search speed itself is obtained by multiplying by a factor of about 2. The resultant speeds are about 50 ms/item, similar to those found elsewhere. (D) Data recast as baselines. Values for the base cadence (564 ms) are not significantly affected by an increase in off-time (576 ms) or on-time (580 ms). Error bars indicate standard error of the mean.

In what follows, it will be shown that this approach can indeed work, and provides converging evidence that iconic memory can act as a surrogate for a stimulus that has suddenly disappeared. But it will also be shown that iconic memory is available to different tasks for different amounts of time, with these limits clustering into a few groups, each likely corresponding to a particular level of the visual hierarchy. As such, it will be argued that this approach can shed considerable light on the nature of the various levels of the visual hierarchy, and on the nature of the feedforward and feedback² connections between them.

GENERAL METHOD

Unless otherwise specified, each experimental condition used three timing patterns, or *cadences*: a base cadence of 80/120 (80 ms on; 120 ms off), and two longer cadences of 80/240 and 200/120, created by increasing the off- and on-times respectively by 120 ms. Each condition tested 12 observers, with order of cadence counter-balanced. Observers were seated 57 cm from the monitor. Displays subtended $11.5^\circ \times 8.5^\circ$ in visual angle, and contained 2, 6, or 10 items, with spacing controlled to keep item density constant. For

detection conditions, the target was present on a randomly selected half of trials; otherwise, the target was always present in each display. Items were $\sim 1^\circ$ in extent, the exact size depending on the condition tested.

Lighting level was sufficient to allow color to be easily seen (i.e., above the mesopic range). A cathode-ray tube (CRT) display was used for all conditions. Blank fields and display backgrounds were both medium gray, resulting in a continual flickering of the items on a static background. All items were black, apart from those in the contrast polarity condition. The appearance of a gray field after the disappearance of an item therefore corresponded to an increase rather than a decrease of phosphor activation, ensuring that phosphor persistence could not significantly affect the results.

All experimental conditions were run on a Macintosh computer using VSearch software (Enns et al., 1990). Observers were instructed to maintain fixation during each trial, to detect the target as quickly as possible, and to keep error rates below 5%. Responses were given via one of two response keys. All observers completed four sets of 60 trials in each condition. Performance was measured in terms of reaction times (RTs) that were averaged for each observer; these were then recast into search speed (average target-present slope³) and baseline (estimated time

²The term “re-entry” generally denotes a particular type of feedback, viz., that in which density of back connections is similar to or exceeds the density of forward connections, and for which the mapping of back connections is not haphazard, but has a mapping similar to that of the feedforward connections. In the context here, “re-entry” and “feedback” will be considered synonymous.

³Slopes for each observer were calculated by determining mean response time for each set size, and calculating a least-squares fit through these points. Analysis used repeated-measures ANOVAs, and paired, two-tailed *t*-tests. Target-present slopes

needed for a single item in the display). A trial timed out—and was considered an incorrect response—if more than 5 s was needed.

EXPERIMENT 1

This experiment examined whether iconic memory can support visual search for a simple feature. The target was a black vertical line 0.8° long; distractors (non-targets) were similar lines oriented $\pm 30^\circ$ to the vertical (**Figure 1A**). Observers were asked to detect the presence or absence of the target.

Condition 1A examined detection for the three cadences of 80/120, 200/120, and 80/240. Search of this kind typically has target-present slopes of 15–30 ms/item in a static display (cf. Treisman and Gormican, 1988). Search here was similar: RTs showed a strong effect of set size [$F(2,10) = 22.8$; $p < 0.0001$], with an average slope of 23.2 ms/item (**Figure 1B**). However, no significant effect of cadence was found [$F(2,10) = 0.711$; $p > 0.5$], nor any significant interaction between set size and cadence [$F(4,10) = 1.29$; $p > 0.3$]. Cadence had no significant effect on either slopes [$F(2,10) = 0.151$; $p > 0.8$; **Figure 1C**], or baselines [$F(2,10) = 0.47$; $p > 0.6$; **Figure 1D**]. Error rates were much the same for all cadences, indicating that no speed-accuracy trade-offs occurred. As such, these results indicate that the information in iconic memory can survive without serious degradation for at least 240 ms, consistent with conclusions obtained elsewhere (e.g., Sperling, 1960; Graziano and Sigman, 2008). And the lack of effect of different cadences—essentially, different switching rates—indicates little cost of switching between visual and iconic representations.

As a test of whether the memory being used actually is iconic memory, Condition 1B compared performance for the 80/240 cadence against two others: a 80/0 cadence (i.e., a display that remained on), and a 80/320 cadence (in which the blank interval was 320 ms). Paired *t*-tests showed that slopes and baselines for 80/240 and 80/0 conditions were virtually identical ($p > 0.9$ and $p > 0.5$, respectively), both with a slope of 20.5 ms, indicating that the flicker had little effect. Extending the blank duration to 320 ms showed a similar lack of effect ($p > 0.2$ and $p > 0.9$, respectively). However, slopes for the 80/240 and 80/320 conditions were 20.5 and 25.2 ms/item respectively, suggesting a slight degradation for the longer blank; indeed, a more detailed analysis⁴ indicates that

were used; target-absent slopes either followed the same pattern or showed no strong effects. Error rates in the target-absent condition were generally low (below 2%) and did not vary much over different conditions. Errors for target-present conditions either followed the pattern of the slopes or showed no strong effects, indicating that speed-accuracy trade-off was not a factor.

⁴Usable memory duration u can be calculated in the following way. The total usable time in each alteration is taken to be the duration of the visible component plus the usable duration of the iconic component. Assuming the usable duration in the 80/120 and 200/120 cadences is 120 ms or more, and that speed is the same for visible and iconic inputs (both assumptions supported by the results of Experiment 1), search speed can be estimated by averaging the slopes of the two short-ISI cadences to get slope s_V , corresponding to search through a visible representation. The usable fraction f over a complete display cycle is s_V/s_L , where s_L is the slope of the long-ISI cadence. For a long-ISI condition with on-time of 80 ms and display cycle (= on-time + off-time) of D ms, f is also $(80+u)/D$; rewriting, $u = Df - 80 = D(s_V/s_L) - 80$. The standard error of the mean of u can be determined from this formula, via the standard errors of the slopes.

performance is a function of on-time plus a *usable duration* (u) of 246 ± 57 ms.

Taken together, these results are consistent with other findings showing that the information in iconic memory can survive without serious degradation for several 100 ms (e.g., Sperling, 1960; Graziano and Sigman, 2008). The speed of search was much the same throughout, not only supporting the proposal that attentional selection and iconic memory involve common representations (Ruff et al., 2007), but indicating that the iconic representation can be used as easily and effectively as the one used in “regular” vision, with the switch between visible and iconic representations requiring little or no time.

EXPERIMENT 2

To examine the extent to which iconic memory can be used for other tasks, Experiment 2 examined its involvement in change detection. Based on the difficulty of detecting change in the absence of attention (i.e., change blindness), it has been proposed that most unattended structure is detailed but volatile, with iconic memory being the quickly dissipating remnant of this representation after the stimulus disappears (Rensink et al., 1997; Rensink, 2000a). Subsequent work (Becker et al., 2000) supported this proposal, indicating that the cueing of iconic memory can guide attention, and thereby facilitate change detection.

Experiment 2 used the same set sizes and much the same items as in Experiment 1A. The same cadences were also used, so that any interference from the flickering displays would be about the same. However, each display now contained approximately equal numbers of vertical lines and lines tilted counterclockwise by 30° . The target was now the item that *changed* its orientation by 30° between displays (**Figure 2A**).

As before, set size had a strong effect on RT [$F(2,11) = 172.1$; $p < 0.0001$; **Figure 2B**]. But there was now a significant effect of cadence [$F(2,11) = 27.4$; $p < 0.0001$] and a significant interaction between set size and cadence [$F(4,11) = 24.5$; $p < 0.0001$]. In particular, cadence had a strong effect on slopes [$F(2,11) = 33.0$; $p < 0.0001$; **Figure 2C**], which were higher with increased off-time ($p < 0.001$). However, there was no effect with increased on-time ($p > 0.2$), again indicating that the different rates of switching between visual and iconic representations had little effect. Baselines (**Figure 2D**) were not reliably affected [$F(2,11) = 1.31$; $p > 0.2$]. (In general, baselines were never reliably different in all the conditions that follow, and so are omitted from subsequent analyses.)

Interpreting slopes in terms of the number of items held across the blank interval (Rensink, 2000b), a strong effect of cadence was again evident [$F(2,11) = 20.1$; $p < 0.0001$]. However, the opposite pattern now occurred: hold did not differ significantly with greater off-time ($p > 0.05$), but increased with greater *on-time* ($p < 0.005$). This is consistent with the proposal that under these conditions the speed of change detection is largely governed by the loading of information into visual short-term memory (vSTM) and its subsequent comparison (Rensink, 2000b). It also suggests that these operations take place largely during on-times alone, being largely unable to use iconic memory. Indeed, a more

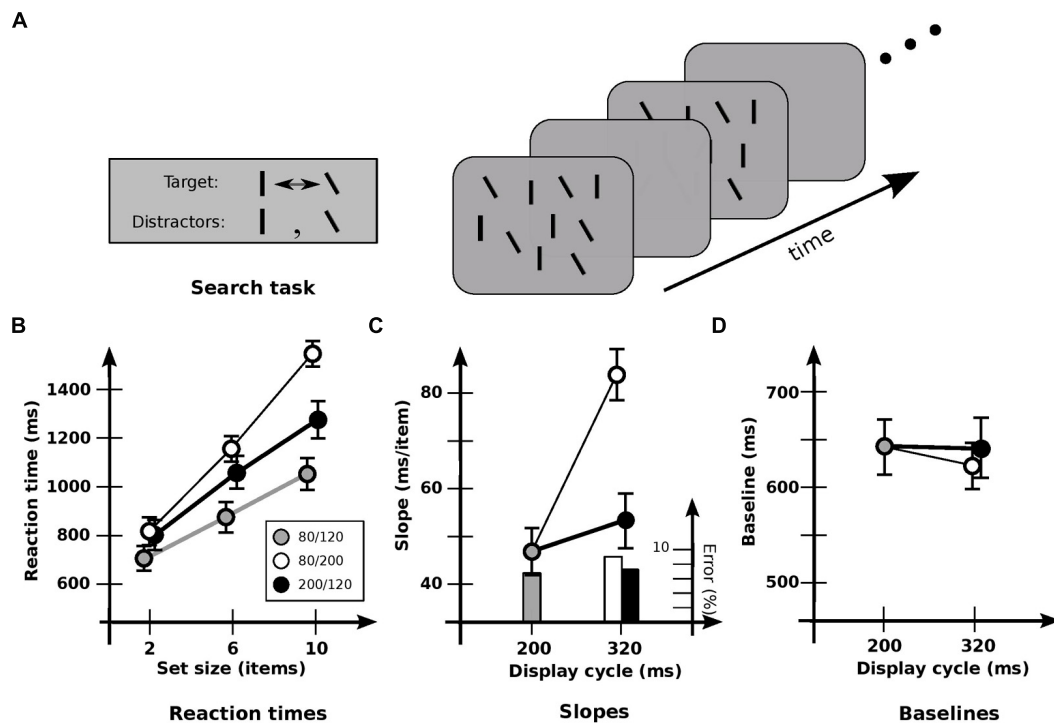


FIGURE 2 | Experiment 2: detection of orientation change. (A) Stimuli used. ~50% of lines in each display are vertical, and 50% are tilted by 30° counterclockwise. Target is the item that changes between vertical and tilted; distractors are those items that maintain a constant orientation. **(B)** Response times and error rates as a function of set size for the three cadences. **(C)** Data recast as slopes. Slope for base cadence (47.3 ms/item) is strongly affected

by an increase in off-time (83.0 ms/item) but not by an increase in on-time (53.7 ms/item). **(D)** Data recast as baselines. Values for the 80/240 and 200/120 cadences have been subtracted by 120 ms to equate the time of first appearance of the changed item. Baseline for base cadence (645 ms) is not significantly affected by an increase in off-time (615 ms) or on-time (641 ms). Error bars indicate standard error of the mean.

detailed analysis of the slopes shows that performance is a function of on-time plus a usable duration of $u = 115 \pm 18$ ms. [Note that if usable duration started from stimulus onset, the similar speeds for the 80/120 and 200/120 cadences would require a value of at least 320 ms. But then there would be similar speeds for the 80/240 and 200/120 cadences, which was not the case ($p < 0.0001$). Thus, usable duration apparently begins at stimulus offset.]

For the detection of both orientation and contrast changes, the loading of information into vSTM is proportional to the duration of the display plus ~110 ms (Rensink, 2000b, Figure 6). Since the ISI in those conditions was 120 ms, this indicates that usable duration u is not the “worth” of iconic memory (Loftus et al., 1992), but an actual time limit. Once this limit is exceeded, iconic memory simply cannot be used for change detection, even though the results of Experiment 1 indicate that it still exists, and contains potentially usable information.

EXPERIMENT 3

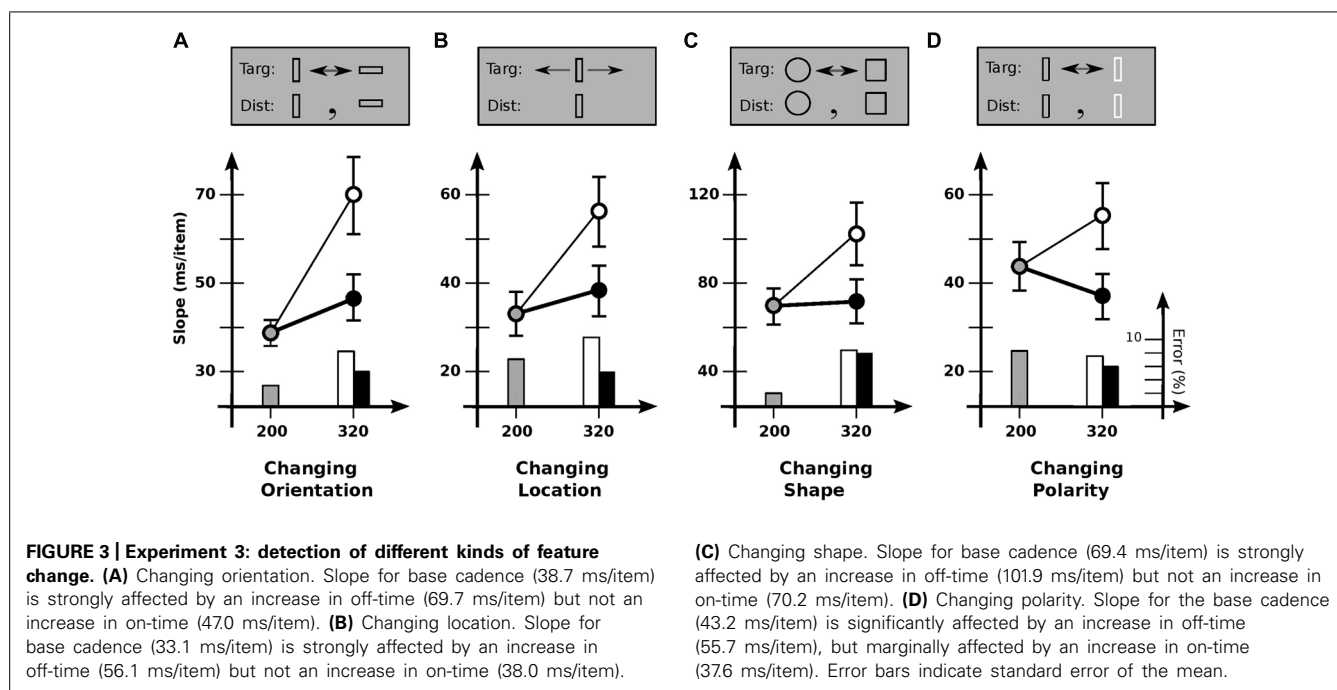
To explore the generality of the limited usability found in Experiment 2, Experiment 3 investigated other kinds of items and kinds of change (Figure 3). Conditions were otherwise much the same. In Condition 3A, items were rectangular outlines $0.4^\circ \times 1.2^\circ$, with targets changing orientation 90°

between vertical and horizontal (Figure 3A). As in Experiment 2, slopes depended strongly on cadence [$F(2,11) = 14.4$; $p < 0.0001$], with search slowing reliably for increased off-time ($p < 0.001$) but not increased on-time ($p > 0.05$). Usable duration u was 117 ± 27 ms, much the same as before.

Condition 3B examined change in location. Here, the target jumped back and forth 1.2° each alternation, with distractors remaining stationary. Slopes again depended on cadence [$F(2,10) = 12.7$; $p < 0.0002$], with search slowing for increased off-time ($p < 0.0002$) but not increased on-time ($p > 0.3$). Usable duration u was 123 ± 34 ms, similar to previous values.

Condition 3C looked at shape change, with the target alternating between a circle and a square. Although more difficult than the other conditions, similar results were found: slope depended on cadence [$F(2,11) = 9.9$; $p < 0.001$], with search slowing for increased off-time ($p < 0.01$) but not increased on-time ($p > 0.9$). Usable duration u was 139 ± 37 ms, comparable to previous values.

Finally, condition 3D examined changes in contrast polarity (black vs. white). Slopes again depended on cadence [$F(2,11) = 8.8$; $p < 0.002$]. Search slowed down with increased off-time ($p < 0.05$), and tended to speed up with increased on-time, although statistical reliability was marginal ($p = 0.06$). [This latter effect has been found elsewhere, where it was taken to indicate a grouping process—based on polarity—that takes place



over several 100 ms (Rensink, 2000b).] Comparing the 80/240 and 200/120 cadences (which equates time per alternation) shows search to be reliably faster with greater on-time ($p < 0.005$); relative speeds yield $u = 137 \pm 34$ ms, similar to the values for other kinds of change. In summary, then, all change detection tasks appeared to show the same kind of behavior, with the same usable duration of about 120 ms.

EXPERIMENT 4

Experiment 4 investigated why the usability of iconic memory might be limited for some tasks but not others. To determine if task difficulty was important, Condition 4A gave observers a simple detection task (as in Experiment 1), with the target defined by a horizontal bar only slightly higher than those of the distractors. Speeds were now comparable to several of those in Experiments 2 and 3 (Figure 4A). However, cadence did not have much of an effect [$F(2,11) = 0.28$; $p > 0.7$], indicating that difficulty *per se* was not the critical factor.

To determine if usable duration might be different if a report is required of the target, Condition 4B used much the same items as in Condition 4A, but with half being black and half white; observers were asked to *identify* the contrast of the target rather than detect it (Figure 4B). Dependence on cadence now reappeared [$F(2,11) = 4.0$; $p < 0.05$], with search slowing for increased off-time ($p < 0.01$) but not increased on-time ($p > 0.3$). Usable duration u was 202 ± 29 ms, less than the 240 ms (or higher) limit of a static detection task, but greater than the values for a change detection task.

To determine if this value might have somehow been due to the mixed polarity of the items, Condition 4C tested report of the orientation of a T-shaped target (left or right) among L-shaped distractors; all items were black (Figure 4C). Search again depended on cadence [$F(2,11) = 10.3$; $p < 0.001$], slowing for increased

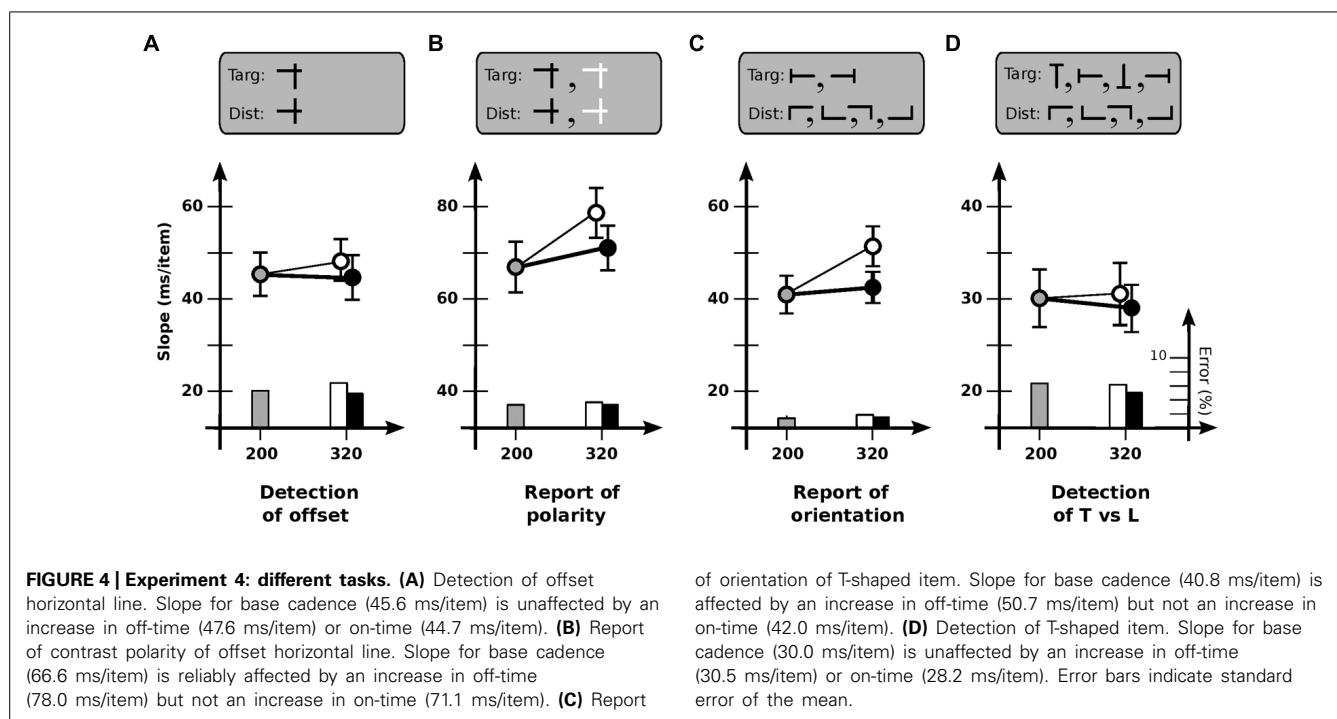
off-time ($p < 0.002$) but not increased on-time ($p > 0.5$). Usable duration u was 181 ± 26 ms, similar to that for Condition 4B.

Finally, to examine whether the key factor in Conditions 4B and 4C might have been the existence of multiple kinds of target, Condition 4D asked observers to detect (but not report on) a T-shaped target among L-shaped distractors, with all items—targets as well as distractors—in any of four orientations (Figure 4D). Dependence on cadence now vanished [$F(2,11) = 0.49$; $p > 0.6$], indicating that multiplicity was not important.

Taken together, then, the results above suggest that the critical factor determining the extent to which iconic memory can be used is not the difficulty of the task or the kinds of items involved, but something about the task itself. A common element of change detection and report—but not static detection—is the need for an item to be *individuated*, i.e., treated as a particular individual at a particular location (Smith, 1998; Pylyshyn, 2003). In change detection, for example, an item that is initially seen (and stored in vSTM) must be re-identified as the same item in the subsequent presentation. Likewise in report, an item detected on the basis of some given feature must be identified as such by whatever process underlies the subsequent report. Such individuated items are believed to play a key role in many visual processes (Ullman, 1984).

GENERAL DISCUSSION

The results above indicate that for all visual search tasks, iconic memory can act as a surrogate for about 120 ms: during this time it can be used as easily and effectively as if the original stimulus were present. Results also show that for some—but not all—tasks, it is available for much longer. The key factor is not the difficulty of the task or the type of feature involved; instead, it appears to be the extent to which the task relies on individuation. Three groups of limits were encountered: for change detection, ~ 120 ms; for



report, 190 ms; for static detection, at least 240 ms. The existence of these groups suggests that iconic memory is not a monolithic structure, but involves several (spatially organized) layers, drawn upon by different tasks to different extents.

Traditionally, iconic memory is taken as having two components: the first a high-density, retinotopic *visible persistence* existing up to 200 ms from stimulus onset (exact value depending on lighting level), and the second a longer-lasting *informational persistence* that is more abstract and mediated more centrally (Coltheart, 1980; Loftus and Irwin, 1998). Since visible persistence can last on the order of a 100 ms under some conditions (Coltheart, 1980), it may be part of the fastest-decaying layer. However, access to the other layers lasts much longer; as such, they would likely involve only informational persistence.

What might these layers correspond to? One possibility involves re-entrant connections from higher level visual areas to lower level ones. Complex static patterns can be detected by neurons in areas such as temporal cortex; cells here have a considerable degree of spatial invariance, responding to much of the visual field (e.g., Felleman and Van Essen, 1991). But to individuate an item—to see it as a particular individual at a particular location—requires linking these spatially invariant representations to lower level retinotopic ones. This can be done, for example, by correlating downward, spatially diffuse signals from higher levels with upward, spatially precise ones from striate cortex (Di Lollo et al., 2000; Tsotsos, 2011).

Results from several lines of research are consistent with this general view. Massive numbers of re-entrant connections exist between the cortical areas involved in visual perception (e.g., Felleman and Van Essen, 1991; Bullier, 2004). Such connections can explain phenomena such as common-onset

masking (Di Lollo et al., 2000) and context effects in recognition (Weinstein and Harris, 1974); indeed, they are believed to be involved in a large variety of visual processes (Fukushima et al., 1991; Tsotsos, 2011). As such, the representation of an item—a *visual object*—is distributed over several levels, with its representation at these levels “knit” together by feedforward and feedback circuits (e.g., Rensink, 2000a, 2002).

Looked at in this way, the different layers of iconic memory could correspond to the memory traces at these different levels (cf. Keyser et al., 2005; Ruff et al., 2007). After a stimulus disappears, representations at the various levels—or at least, their connections—begin to decay, with different time constants at each level. Given that durations are generally longer at higher visual areas (Keyser et al., 2005), the more detailed representations at lower levels would likely be the first to go. If so, the layer accessible for only 120 ms would likely correspond to the lower level representations. (Visible persistence may be part of this.) Given that this layer is needed for change detection, it would likely contain relatively precise spatial information, needed to ensure continuity of representation over time (Rensink, 2000a, 2013). Meanwhile, layers that are usable for longer durations might reflect higher level representations, which are more abstract and have poorer spatial localization.

Such a *multi-layer* theory of iconic memory could explain the usable durations for the different kinds of task as follows:

- (a) *Static detection* (≥ 240 ms). Information carried by the feedforward “wave” created by the appearance of an item reaches high levels relatively quickly. After a brief time (c. 100 ms), access to high-precision spatial information in the low iconic layers begins to degrade. But since detection does not require precise spatial information, it can still be “driven” by the information

at the higher layers of iconic memory for several 100 ms longer. This can explain many classic partial report results, which require only a report of a stimulus (generally, a letter) at some coarsely specified location, but not its precise position. Note that although absolute position is eventually lost at higher levels, precise *relative* positions could still be maintained. For example, the targets in Condition 4A differed from the distractors by only a small shift in the position of a horizontal bar; this information remained available for at least 240 ms. Consistent with this, partial report studies suggest that shape information in iconic memory can remain fairly accurate for over 300 ms (Gegenfurtner and Sperling, 1993; Graziano and Sigman, 2008).

- (b) *Change detection (c. 120 ms)*. The relatively short usable duration (120 ms) for change detection could reflect the need for precise spatial location, which is required for item continuity (Rensink, 2000a, 2013). An important issue is whether this duration reflects the decay of the *contents* of the low-level representation, or just the *connections* to it. Studies based on exogenous cues indicate that positional information does not degrade greatly for at least 300 ms (Graziano and Sigman, 2008). And since exogenous cues can make use of—and transmit—the location of these cues, it would appear that feed-forward connections can be maintained, at least for spatial information of moderate resolution. In contrast, the process of establishing a feedback connection to lower levels needs spatial information that is very precise (Di Lollo et al., 2000); such connections might therefore fail relatively quickly.
- (c) *Report (c. 190 ms)*. For the report tasks of Conditions 4B and 4C, usable duration is greater than that for change detection but less than that for static detection. At least two explanations are possible. First, it may be that detection proceeds as usual, but a subsequent individuation stage is needed to report the associated properties of the detected item; usable duration would then reflect the relative amount of time needed for each of these stages. Consistent with this, slopes of the report tasks were 10–20 ms/item greater than those of their detection equivalents (Figure 4), suggesting the involvement of an additional processing stage. Alternatively, individuation may only need to be partial—i.e., the representation of the target item need only be linked back to a level where its location can be readily distinguished from those of the others. If so, feedback connections may only be established with a mid-level layer, which may endure somewhat longer than those at lower levels.

Relation to other work

Among other things, the proposal here is consistent with results on attentional capture and apparent motion that show a visual continuity for 100 ms after the disappearance of an item (e.g., Yantis and Gibson, 1994). It is also consistent with findings of partial report experiments that (i) when a mask is shown after stimulus disappearance, identification errors arise only if the mask is shown within 150 ms or so of stimulus onset, while localization errors can be induced even if the mask is presented much later, and (ii) if a mask is not used, localization errors begin soon after stimulus disappearance, while identification errors remain low (Mewhort et al., 1981). These patterns can be explained by the

existence of a durable array (or “buffer”) of fairly complex but poorly localized information at higher levels, along with a relatively fast decay of their connections to spatial locations at lower ones.

The proposal of multiple layers of iconic memory is also similar in some ways to the proposal of multiple systems of visual memory (e.g., Sligte et al., 2010). There is general agreement with the idea of detailed, volatile representations at the lower levels, along with a single detailed, longer-lasting representation (corresponding to a visual object) held in vSTM (cf. Rensink, 2000a, 2002). Multiple-systems experiments are based on the use of positional cues with delays of several seconds. Since this is beyond the lifetime of “classic” iconic memory, they are likely concerned with longer-lasting—and likely more limited—representations. The exact nature of this memory is not completely understood; indeed, the existence of a distinct “fragile” vSTM is still controversial (see e.g., Makovski, 2012). But if multiple systems do exist, they could be higher level counterparts of the layers proposed here.

Iconic memory, feedback connections, and visual attention

The theory of iconic memory described here also has implications for the role of feedback processes in human vision. Anatomical and physiological studies indicate that human vision relies upon two main types of feedback connections (e.g., Bullier, 2004). The first are *horizontal* connections of adjacent cells at the same level of the processing stream; these converge quickly and can potentially support rapid local computation of considerable complexity, such as determination of local shape. Given the durability of high-accuracy (local) shape representation in iconic memory (Mewhort et al., 1981), such connections appear to be relatively long-lasting. Longer-range connections can also exist between corresponding locations in representations at the same level of the visual hierarchy (e.g., representations of color and orientation). The second type of connection involves *vertical* links between corresponding cells at different levels. As discussed above, the memory at each of these levels—and in the connections between them—may be the basis of the iconic layers proposed here.

It has been proposed that “the representation of any item in this form of storage [iconic memory] is achieved by creating a temporary file of information about the item” (Coltheart, 1983, p. 291), with relatively complex structure (such as characters) created in parallel across the visual field, but susceptible to overwriting by the subsequent appearance of other structures (Mewhort et al., 1981; Coltheart, 1983). This is similar to the proposal of *proto-objects* (Rensink and Enns, 1998), which are relatively complex structures of limited extent formed rapidly and in parallel in the (near-) absence of attention; these too are temporary, either fading away within a few 100 ms, or being overwritten by the representation of a new item that appears at their location (Rensink et al., 1997). Fast-acting horizontal connections could explain why the within-item binding needed for proto-objects can be achieved using so little time and so little attention. They could also explain why considerable binding exists in iconic memory (Landman et al., 2003), even in its lowest layers (Experiments 1 and 4A).

Meanwhile, vertical connections could be the basis of larger-scale representations. Feedforward and feedback links likely connect corresponding locations at different levels in a fairly dense way (e.g., Di Lollo et al., 2000). Such links could enable retinotopic representations at low-levels (e.g., in striate cortex) to connect to spatiotopic representations at high ones (e.g., in temporal cortex) via a series of stages in which position is increasingly less tied to retinal location (e.g., Tsotsos, 2011). And attention might act by establishing long-range feedforward-feedback loops to represent a coherent visual object, resulting in a representation distributed across the various levels, their contents linked via circuits connecting contents at the same (relative) spatial location (Rensink, 2000a, 2002; Lamme, 2003; Sligte et al., 2010; Tsotsos, 2011).

Characterizing “iconic,” “preattentive,” and “attentive” representations in this way can account for why performance on iconic and visible representations is so similar (Experiment 1), why selective attention and readout from iconic memory involve common neural mechanisms (e.g., Ruff et al., 2007), and why there is little cost for switching between the two (Experiments 1 and 4A). Said simply, there is no separate “iconic” memory system: the layers of iconic memory are just the traces of the representations through which normal visual perception proceeds (see also Keyser et al., 2005; Ruff et al., 2007).

In this view, iconic memory—or at least, informational persistence—has a clear purpose: *to help establish and maintain links between the various spatially organized representations of an item*. Given the decreasing precision of representations with increasing level, processes based on a feedforward sweep of information could continue to use such information even after the contents at the lower levels have faded. However, processes relying on feedback from higher levels would not always have access to the more detailed (but volatile) representations at lower ones; when this happens, the process must wait for the contents of these to be re-instantiated.

The extent to which this proposal adequately captures the operation of the visual system is unclear. But to the degree that it is relevant, the “usability logic” developed here could provide a useful way to investigate the various feedforward and feedback mechanisms involved.

AUTHOR CONTRIBUTIONS

Much of this work was done while the author was with Cambridge Basic Research, Nissan Research & Development, Inc., Cambridge, MA, USA. Portions were presented at the annual conference of the Association for Research in Vision and Ophthalmology in May 1997, and the European Conference on Visual Perception in August 2008. Many thanks to Duncan Bryce, Emily Cramer, Puishan Lam, Kyle Melnick, Nayantara Santhi, and Monica Strauss for their help in running the experiments.

ACKNOWLEDGMENTS

This research was supported by Nissan Motor Co., Ltd., and the Natural Sciences and Engineering Research Council of Canada (NSERC). Thanks to Vince Di Lollo, Dan Simons, and an anonymous reviewer for feedback on earlier versions of this paper.

REFERENCES

- Averbach, E., and Coriell, A. S. (1961). Short-term memory in vision. *Bell Systems Tech. J.* 40, 309–328. doi: 10.1002/j.1538-7305.1961.tb03987.x
- Becker, M. W., Pashler, H., and Anstis, S. M. (2000). The role of iconic memory in change-detection tasks. *Perception* 29, 273–286.
- Bullier, J. (2004). “Communications between cortical areas of the visual system,” in *The Visual Neurosciences*, Vol. 1, eds L. M. Chalupa and J. S. Werner (Cambridge, MA: MIT Press), 522–540.
- Coltheart, M. (1980). Iconic memory and visible persistence. *Percept. Psychophys.* 27, 183–228. doi: 10.3758/BF03204258
- Coltheart, M. (1983). Iconic memory. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 302, 283–294. doi: 10.1098/rstb.1983.0055
- Di Lollo, V., Enns, J. T., and Rensink, R. A. (2000). Competition for consciousness among visual events: the psychophysics of reentrant visual processes. *J. Exp. Psychol. Gen.* 129, 481–507. doi: 10.1037/0096-3445.129.4.481
- Enns, J. T., Ochs, E. P., and Rensink, R. A. (1990). VSearch: macintosh software for experiments in visual search. *Behav. Res. Methods Instrum. Comput.* 22, 118–122. doi: 10.3758/BF03203130
- Estes, W. K., and Taylor, H. A. (1964). A detection method and probabilistic models for assessing information processing from brief visual displays. *Proc. Natl. Acad. Sci. U.S.A.* 52, 446–454. doi: 10.1073/pnas.52.2.446
- Felleman, D. J., and Van Essen, D. C. (1991). Distributed hierarchical processing in primate visual cortex. *Cereb. Cortex* 1, 1–47. doi: 10.1093/cercor/1.1.1
- Fukushima, K., Imagawa, T., and Ashida, E. (1991). “Character recognition with selective attention,” in *Proceedings of the International Joint Conference on Neural Networks*, 1, (Seattle, WA: IJCNN-91-Seattle International Joint Conference on Neural Networks), 593–598.
- Gegenfurtner, K., and Sperling, G. (1993). Information transfer in iconic memory experiments. *J. Exp. Psychol. Hum. Percept. Perform.* 19, 845–866. doi: 10.1037/0096-1523.19.4.845
- Graziano, M., and Sigman, M. (2008). The dynamics of sensory buffers: geometric, spatial, and experience-dependent shaping of iconic memory. *J. Vis.* 8, 9.1–13. doi: 10.1167/8.5.9
- Haber, R. N. (1971). “Where are the visions in visual perception?,” in *The Adaptive Function of Imagery*, ed. S. J. Segal (New York: Academic Press), 33–48.
- Haber, R. N. (1983). The impending demise of the icon: a critique of the concept of iconic storage in visual information processing. *Behav. Brain Sci.* 6, 1–54. doi: 10.1017/S0140525X0001428X
- Keyser, C., Xiao, D.-K., Földiák, P., and Perrett, D. (2005). Out of sight but not out of mind: the neurophysiology of iconic memory in the superior temporal sulcus. *Cogn. Neuropsychol.* 22, 316–332. doi: 10.1080/02643290442000103
- Lamme, V. A. F. (2003). Why visual attention and awareness are different. *Trends Cogn. Sci. (Regul. Ed.)* 7, 12–18. doi: 10.1016/S1364-6613(02)00013-X
- Landman, R., Spekreijse, H., and Lamme, V. A. (2003). Large capacity storage of integrated objects before change blindness. *Vision Res.* 43, 149–164. doi: 10.1016/S0042-6989(02)00402-9
- Loftus, G. R., Duncan, J., and Gehrig, P. (1992). On the time course of perceptual information that results from a brief visual presentation. *J. Exp. Psychol. Hum. Percept. Perform.* 18, 530–549. doi: 10.1037/0096-1523.18.2.530
- Loftus, G. R., and Irwin, D. E. (1998). On the relations among different measures of visible and informational persistence. *Cogn. Psychol.* 35, 135–199. doi: 10.1006/cogp.1998.0678
- Makovski, T. (2012). Are multiple visual short-term memory storages necessary to explain the retro-cue effect? *Psychon. Bull. Rev.* 19, 470–476. doi: 10.3758/s13423-012-0235-9
- Mewhort, D. J. K., Campbell, A. J., Marchetti, F. M., and Campbell, J. I. D. (1981). Identification, localization, and “iconic memory”: an evaluation of the bar-probe task. *Mem. Cognit.* 9, 50–67. doi: 10.3758/BF03196951
- Pashler, H. E. (1998). *The Psychology of Attention*. Cambridge, MA: MIT Press, 107–108.
- Pylyshyn, Z. W. (2003). *Seeing and Visualizing: it's Not What You Think*. Cambridge, MA: MIT Press, 201–279.
- Rensink, R. A. (2000a). The dynamic representation of scenes. *Vis. Cogn.* 7, 17–42. doi: 10.1080/135062800394667
- Rensink, R. A. (2000b). Visual search for change: a probe into the nature of attentional processing. *Vis. Cogn.* 7, 345–376. doi: 10.1080/135062800394847
- Rensink, R. A. (2002). Changes. *Prog. Brain Res.* 140, 197–207. doi: 10.1016/S0079-6123(02)40051-9

- Rensink, R. A. (2013). "Perception and Attention," in *Oxford Handbook of Cognitive Psychology*, ed. D. Reisberg (Oxford: University Press), 97–116.
- Rensink, R. A., and Enns, J. T. (1998). Early completion of occluded objects. *Vision Res.* 38, 2489–2505. doi: 10.1016/S0042-6989(98)00051-0
- Rensink, R. A., O'Regan, J. K., and Clark, J. J. (1997). To see or not to see: the need for attention to perceive changes in scenes. *Psychol. Sci.* 8, 368–373. doi: 10.1111/j.1467-9280.1997.tb00427.x
- Ruff, C. C., Kristjansson, A., and Driver, J. (2007). Readout from iconic memory and selective spatial attention involve similar neural processes. *Psychol. Sci.* 18, 901–909. doi: 10.1111/j.1467-9280.2007.01998.x
- Sligte, I. G., Vandenbroucke, A. R. E., Scholte, H. S., and Lamme, V. A. F. (2010). Detailed sensory memory, sloppy working memory. *Front. Psychol.* 1:175. doi: 10.3389/fpsyg.2010.00175
- Smith, B. C. (1998). *On the Origin of Objects*. Cambridge, MA: MIT Press.
- Sperling, G. (1960). The information available in brief visual presentations. *Psychol. Monogr.* 74, 1–29. doi: 10.1037/h0093759
- Treisman, A., and Gormican, S. (1988). Feature analysis in early vision: evidence from search asymmetries. *Psychol. Rev.* 95, 15–48. doi: 10.1037/0033-295X.95.1.15
- Tsotsos, J. K. (2011). *A Computational Perspective on Visual Attention*. Cambridge, MA: MIT Press. doi: 10.7551/mitpress/9780262015417.001.0001
- Ullman, S. (1984). Visual routines. *Cognition* 18, 97–159. doi: 10.1016/0010-0277(84)90023-4
- Weisstein, N., and Harris, C. S. (1974). Visual detection of line segments: an object-superiority effect. *Science* 186, 752–755. doi: 10.1126/science.186.4165.752
- 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. doi: 10.1038/nrn1411
- Yantis, S., and Gibson, B. S. (1994). Object continuity in motion perception and attention. *Can. J. Exp. Psychol.* 48, 182–204. doi: 10.1037/1196-1961.48.2.182

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.

Received: 21 May 2014; accepted: 15 August 2014; published online: 29 August 2014.
Citation: Rensink RA (2014) Limits to the usability of iconic memory. *Front. Psychol.* 5:971. doi: 10.3389/fpsyg.2014.00971

This article was submitted to Perception Science, a section of the journal *Frontiers in Psychology*.

Copyright © 2014 Rensink. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Reentrant processing mediates object substitution masking: comment on Pöder (2013)

Vincent Di Lollo*

Department of Psychology, Simon Fraser University, Burnaby, BC, Canada

Edited by:

Bruno Breitmeyer, University of Houston, USA

Reviewed by:

Bruno Breitmeyer, University of Houston, USA

Duncan Guest, Nottingham Trent University, UK

Talis Bachmann, University of Tartu, Estonia

*Correspondence:

Vincent Di Lollo, Department of Psychology, Simon Fraser University, 8888 University Drive, Burnaby, BC V5A 1S6, Canada
e-mail: enzo@sfu.ca

Object-substitution masking (OSM) occurs when a target stimulus and a surrounding mask are displayed briefly together, and the display then continues with the mask alone. Target identification is accurate when the stimuli co-terminate but is progressively impaired as the duration of the trailing mask is increased. In reentrant accounts, OSM is said to arise from iterative exchanges between brain regions connected by two-way pathways. In an alternative account, OSM is explained on the basis of exclusively feed-forward processes, without recourse to reentry. Here I show that the feed-forward account runs afoul of the extant phenomenological, behavioral, brain-imaging, and electrophysiological evidence. Further, the feed-forward assumption that masking occurs when attention finds a degraded target is shown to be entirely *ad hoc*. In contrast, the evidence is uniformly consistent with a reentrant-processing account of OSM.

Keywords: visual masking, object substitution masking, feed-forward, reentrant processing, attention

Visual masking refers to an impairment in the perception of a briefly presented object (the target) by the presentation of a second object (the mask) in close spatiotemporal proximity. The present work is concerned with a form of masking known as *object-substitution masking* (OSM) that occurs when a brief simultaneous display of the target and the mask continues with a display of the mask alone (Di Lollo et al., 2000).

Figure 1 illustrates the basic OSM paradigm. The display sequence begins with a brief presentation of a variable number of rings, each with a gap in one of the four cardinal orientations. Observers indicate the orientation of the gap in the target ring, which is singled out by four surrounding small dots that act as both cue and mask. After a brief exposure, all elements in the display are turned off except for the four dots which remain on view for a variable period of up to several hundred ms. When the target and the mask terminate together (i.e., when there is no trailing display of the four dots alone) the target is identified accurately. Masking develops rapidly, however, as the duration of the trailing four-dot mask is increased up to about 200 ms (see Figure 2).

Early theoretical accounts of OSM were couched in terms of reentrant processes that take place after an initial feed-forward sweep (Di Lollo et al., 2000; Lleras and Moore, 2003). More recently, an exclusively feed-forward account has been proposed by Pöder (2013). That account is examined and questioned in the present work.

A REENTRANT ACCOUNT OF OSM

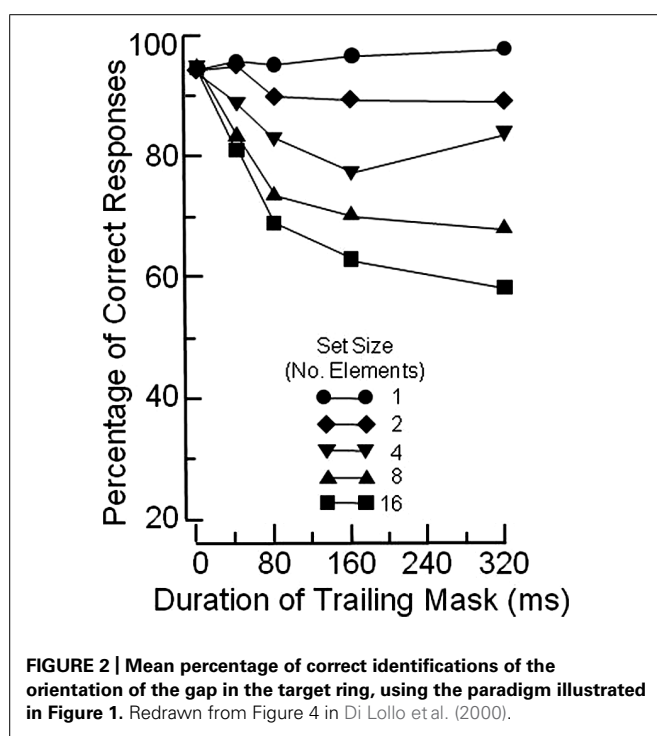
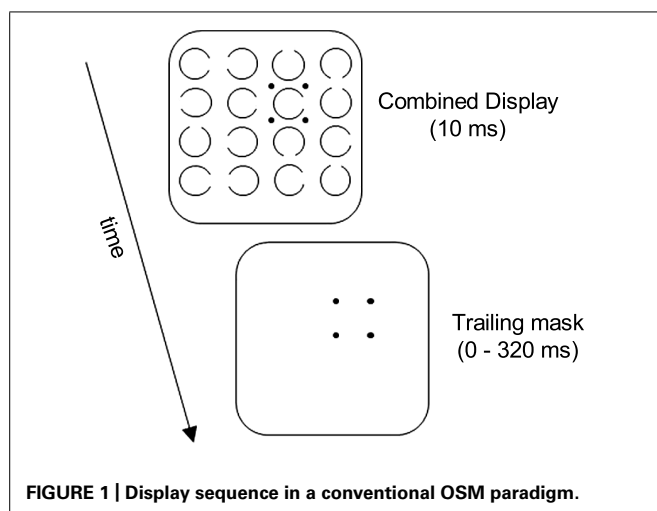
In the conventional OSM paradigm (see Figure 1) the target and the mask have a common onset; therefore, no unique onset transient is generated by the mask. This rules out onset transients as a source of masking (e.g., Breitmeyer and Ganz, 1976; see Di Lollo et al., 2000, for a more detailed account of the role of transient

responses in OSM). Rather, OSM is thought to be mediated by reentrant signals between brain regions connected by two-way pathways.

In the feed-forward sweep, the neural activity triggered by the initial display ascends to higher brain regions, where it activates a large number of perceptual hypotheses that are in some way compatible with the sensory input. The perceptual hypotheses then descend to lower levels, where they attempt to match themselves to the pattern of ongoing activity through a process of correlation. Matches that yield low correlations are discarded, whereas the hypothesis that yields the highest correlation is confirmed and eventually leads to conscious awareness (Mumford, 1991, 1992; Grossberg, 1995; Di Lollo et al., 2000).

Masking occurs when a mismatch arises between the reentrant signals and the ongoing activity at the lower level. At short durations of the trailing mask, the reentrant signals find a pattern of ongoing low-level activity that, although decayed, is of relatively uniform strength. Notably, the brief additional display of the four dots causes the low-level representation of the mask to be only slightly stronger than that of the target. In this case, little or no masking occurs because the similarity between the reentrant hypothesis and the low-level representation allows for an adequate correlation. This leads to confirmation of that perceptual hypothesis, and to relatively accurate target identification, as illustrated by the short-mask-duration points in Figure 2.

In contrast, at long durations of the trailing mask, the reentrant signals find a pattern of ongoing low-level activity of non-uniform strength. To wit, the representation of the target has decayed, but the mask remains at full strength because of the continued external input. This mismatch reduces the correlation with the reentrant hypothesis, which consists of a representation of the target and the mask at uniform strength. The ensuing low correlation causes the current perceptual hypothesis to be discarded,



and a new “mask-alone” hypothesis to be generated, with consequent impairment of target identification, as illustrated by the long-duration points in **Figure 2**.

PÖDER’S FEED-FORWARD ACCOUNT OF OSM

A simpler, strictly feed-forward account has been proposed by Pöder (2013). The account is based on two assumptions. First, the continued presence of the mask after the offset of the initial display is held to add noise at the target’s internal representation, causing its signal-to-noise (S/N) ratio to be reduced. Because of temporal integration, the noise continues to grow while the mask remains on view. For this reason, the reduction in S/N ratio is said to be proportional to the exposure duration of the trailing mask. Second, masking is assumed to occur when attention is deployed to the

target location. Upon its deployment, attention finds a degraded representation of the target due to reduced S/N ratio, and accuracy of target identification is impaired correspondingly.

The two assumptions were embodied in a computational model (Pöder, 2013) that provided an excellent fit for the OSM data reported by Di Lollo et al. (2000; see present **Figure 2**). This buttressed the claim that OSM can be explained in strictly feed-forward terms, without recourse to reentry.

Pöder’s (2013) assumptions are examined in the remainder of this article. The assumption that the trailing mask reduces the target’s S/N ratio is shown to run afoul of the phenomenological, behavioral, brain-imaging, and electrophysiological evidence. Further, the assumption that masking occurs when attention finds a degraded target is shown to be *ad hoc*.

ASSUMPTION OF REDUCED TARGET S/N RATIO

Pöder’s (2013) account of how the noise generated by the extended presentation of the four-dot mask may affect the target’s internal representation does not draw a distinction between sensory noise and non-specific internal “system” noise. In what follows, I endeavor to show that externally generated noise stemming from the prolonged exposure of the four-dot-mask is inadequate as a determinant of OSM. Furthermore, an account based on non-specific internally generated noise is just as inadequate¹.

Non-specific “internal” or “system” noise is often used to introduce an element of variability in models such as the Computer Model of Object Substitution (CMOS); (Di Lollo et al., 2000). It has never been used as a masking agent (either forward, simultaneous, or backward) in any form of masking (metacontrast, pattern, camouflage, conceptual, etc.) in the vast masking literature. Masking by non-specific noise is certainly not listed in Breitmeyer’s (1984) definitive treatise on masking (Breitmeyer and Öğmen, 2006). More important, it is not mentioned explicitly in Pöder’s (2013) Attentional Gating Theory (AGT). To be sure, the claim that internal noise may be a determinant of OSM could be a bold, imaginative step, as long as strong logical and empirical documentation were provided to justify it. As it is, such a claim is *ad hoc* and not part of the AGT as stated in Pöder (2013).

PHENOMENOLOGICAL EVIDENCE

On Pöder’s (2013) assumption that the internal representation of the target is degraded because of reduced S/N ratio, one could reasonably expect some distortive effects of the noise to be evidenced in the appearance of the target. In fact, what is seen is a blank area demarcated by the four-dot mask. A compelling description has been provided by Neill et al. (2002, p. 683) as follows:

... in our own experiments the general notion of object substitution is consistent with the phenomenal experience of the masked target: not only does the space inside the dots appear blank, but there is a strong subjective impression of the contours of a square connecting the dots. Furthermore, there is a subjective impression of enhanced brightness of the area within the square, very similar to the brightness enhancement that occurs within illusory contours or subjective contours resulting

¹I thank an anonymous reviewer for suggesting the possibility that OSM may arise from internal noise.

from long-duration inducing elements (Coren, 1972; Kanizsa, 1976; Petry and Meyer, 1987; Purghé and Coren, 1992).

Such a phenomenological appearance is far from that of a degraded target postulated in Pöder's account. On such an account, additional processes need to be invoked to explain why the reduced S/N ratio causes the target to disappear without a trace instead of appearing merely as degraded. Rather, this phenomenology is precisely what is expected on the basis of OSM: at long-durations of the trailing mask, a mismatch arises between the ongoing pattern of activity at the lower level (four dots alone) and the reentrant perceptual hypothesis (target surrounded by four dots). The mismatch causes that perceptual hypothesis to be discarded and replaced by a new hypothesis consisting of four dots demarcating a blank square area, and that's what is eventually perceived.

BEHAVIORAL EVIDENCE

Results inconsistent with the claim that the four-dot mask degrades the target by adding noise to its internal representation have been reported by Lleras and Moore (2003). They showed that OSM was fully in evidence even when the four dots were not physically present around the target after its offset. Rather, what was necessary was the presence of the trailing mask in a location next to the target, under conditions of apparent motion that supported the perception of the target morphing into the mask. Increased noise at the target location can hardly be regarded as a critical determinant of OSM in Lleras and Moore's study, simply because the target was unobscured by the trailing four-dot mask. Further evidence that OSM occurs when the mask is presented in a location other than that of the target has been reported by Jiang and Chun (2001) and by Guest et al. (2011).

Pöder's assumption that the four dots add noise to the target is also questioned by the results of Bouvier and Treisman (2010) who found that a target's low-level features can be detected accurately even when OSM prevents identification of the target's configuration. If, as Pöder asserts, a critical factor in OSM were the increased visual noise at the target's location, what needs to be asked is why the noise spared the target's low-level features but not its configuration. The likely answer is that OSM interferes with reentrant signaling, leaving the low-level features in the feed-forward sweep largely intact. Evidence consistent with the findings of Bouvier and Treisman has been reported by Guest et al. (2011), and by Binsted et al. (2007) who found that OSM occurs after the physical features of the target have been processed.

More behavioral evidence inconsistent with Pöder's claim that the principal role of the mask is to add noise to the target's representation has been reported by Jannati et al. (2013). In Experiment 1 of that study, the mask was a solid ring surrounding the target. In Experiment 3, the mask consisted of four small dots, as seen in **Figure 1**. On Pöder's hypothesis, the sizeable contours of the ring should have generated substantially more noise than the sparse contours of the four dots. The strength of masking, therefore, should have been greater in Experiment 1 than in Experiment 3. In fact, the results revealed the opposite pattern, at least numerically.

Another aspect of Jannati et al.'s (2013) study is inconsistent with a key assumption in Pöder's account. Namely, that the

amount of noise added to the target is proportional to the mask's exposure duration. In the study of Jannati et al. (2013) the display sequence began with a brief combined presentation of target and mask, continued with a blank inter-stimulus interval (ISI) of variable duration, and ended with a brief re-presentation of the mask alone. The important point is that, because the duration of the trailing mask was fixed, the amount of noise supposedly added to the target should also have been fixed. This should have given rise to a correspondingly fixed level of OSM. Instead, the results revealed a non-monotonic *U*-shaped function of accuracy over ISI, as predicted in Di Lollo et al.'s (2000) reentrant-processing account.

ELECTROPHYSIOLOGICAL AND BRAIN-IMAGING EVIDENCE

The electrophysiological and brain-imaging evidence is uniformly supportive of a reentrant-processing account of OSM. To wit, there is broad agreement that OSM interferes with the reentrant sweep while leaving the feed-forward sweep largely unaffected.

Especially relevant to a comparison of reentrant and feed-forward accounts of OSM are two ERP experiments by Woodman and Luck (2003). Experiment 1 employed a search display in which the target was singled out by four dots that either co-terminated with the target or remained on the screen alone for 600 ms after target offset. Two results are directly relevant to the present purpose. First, accuracy of target identification was impaired when the offset of the four-dot mask was delayed (a conventional OSM effect). Second, the target-elicited N2pc (an ERP component said to index target localization, as distinct from target consolidation) was the same in the delayed as in the co-termination conditions. Namely, unlike identification accuracy, the N2pc was unaffected by OSM. This strongly suggested that OSM interfered with later processes of target consolidation, while leaving earlier processes of target localization essentially unaffected. As pointedly noted by Woodman and Luck (2003, p. 608): "The finding of lateralized response to the target (i.e., the N2pc) indicates that on both trial types, the brain was able to determine which side of the array contained the target, which implies that the target was detected by the visual system even though the observers could not accurately report it."

Pöder's noise-based hypothesis was further disconfirmed in Woodman and Luck's (2003) Experiment 2 in which the four-dot mask always co-terminated with the target. The critical manipulation was whether or not the target was embedded in visual noise. An important procedural detail was that the strength of the noise was adjusted so that it produced the same degree of behavioral impairment as the delayed-offset mask in Experiment 1.

The results were unambiguous: the N2pc was fully in evidence when the target was unencumbered by visual noise, but was totally absent when the target was embedded in noise. This finding rules out the option that in the delayed-mask-offset condition in Experiment 1 target identification was impaired by visual noise. Had visual noise caused that impairment, it should also have eliminated the N2pc, as it did in Experiment 2. Rather, this pattern of results is consistent with the idea that target identification in Experiment 1 was impaired because the extended four-dot mask interfered with the reentrant signaling. From a reentrant perspective, no suitable perceptual hypotheses could be generated in Experiment 2 when

the target was embedded in noise. Whatever perceptual hypotheses were generated consisted largely of visual noise, and that's what was eventually perceived.

From a broader perspective, it is fitting to ask whether, in principle, four small dots displayed outside the spatial confines of the target can produce sufficient noise to prevent target identification. Or, for that matter, whether they can introduce any manifest noise at all. Experiment 2 of Woodman and Luck (2003) offers important evidence in this respect. In order to match the impairment produced by the extended mask in Experiment 1, the noise mask in Experiment 2 required 23 dots placed directly on top of the target. This raises a further question regarding Pöder's noise-based account of OSM. What needs to be asked is by what means four small dots that remain on view around the target can produce an amount of noise equivalent to that produced by 23 dots placed directly on the target itself. This equivalence cannot be accepted uncritically as stated: it is in need of empirical verification. Similarly, the validity of the claim that four small dots placed as much as 40 min arc away from the target (Di Lollo et al., 2000) can produce sufficient noise to prevent target identification cannot merely be assumed: it needs to be empirically verified.

The idea that OSM interferes with the reentrant sweep while leaving the feed-forward sweep essentially intact is supported by a number of other ERP studies (e.g., Reiss and Hoffman, 2007; Harris et al., 2013). That idea is also buttressed by a functional magnetic-resonance adaptation study by Carlson et al. (2007). Contrary to the hypothesis of increased visual noise at the target's location, that study revealed no effect of OSM in early visual areas. In contrast, powerful effects of OSM were in evidence at higher cortical regions. Further brain-imaging evidence supportive of the reentrant account of OSM has been reported in an fMRI study by Weidner et al. (2006).

I hasten to note that the evidence listed in the foregoing is not – nor was it intended to be – exhaustive. Rather, the intent was to cite examples of phenomenological, behavioral, electrophysiological, and brain-imaging evidence inconsistent with Pöder's (2013) claim that a critical factor in OSM is the degradation of the internal representation of the target by visual noise generated by the four-dot mask.

ASSUMPTION OF THE ROLE OF ATTENTION IN OSM

According to Pöder (2013), OSM occurs when attention is deployed to the target's location and finds a representation degraded by visual noise. What is not specified is the mechanism presumed to be involved in the attentional processing.

Attention has been described as a limited resource (Norman and Bobrow, 1975; Lavie and Tsai, 1994), a filter (Broadbent, 1958), a spotlight (Posner et al., 1980), a zoom lens (Eriksen and St. James, 1986), and a glue (Treisman and Gelade, 1980). A major drawback of these metaphors is that they do not specify what underlying mechanisms mediate the purported function. As pointedly noted by Chun et al. (2011, p. 74): "Attention has become a catch-all term for how the brain controls its own information processing...." So, when Pöder (2013) invokes "attention" to explain OSM, one is left wondering just what it is that he means. To be useful, accounts of OSM – or, for that matter, accounts of any other phenomenon – should endeavor to make explicit the mechanisms

underlying such a nebulous and ill-defined concept as "attention." It is time to recognize that the indiscriminate use of attention as an explanatory panacea can be an impediment to communication and understanding.

Come to think of it, the function performed by "attention" in Pöder's account of OSM is a more vague – though in some ways equivalent – incarnation of the function performed by reentry in the OSM account of Di Lollo et al. (2000). In the former account, OSM is said to occur when attention is deployed to the target location and finds an item that has been degraded beyond recognition. In the latter, OSM is said to occur when the reentrant signals arrive on their return and find an item that does not match any of the perceptual hypotheses. From a comparison of Pöder's use of "attention" and Di Lollo et al. (2000) use of "reentry" there appears to be a good deal of commonality in the two accounts of OSM.

CONCLUDING COMMENTS: OF QUANTITATIVE MODELS

Having reviewed the pertinent empirical evidence, we now turn to the quantitative models of OSM: the CMOS proposed by Di Lollo et al. (2000) and the AGT proposed by Pöder (2013). CMOS provides an excellent fit to the empirical data illustrated in Figure 2; AGT provides an even better fit.

Not to cut too fine a point, it can be confidently stated that both models are misguided. This is because the data that they purport to model (see Figure 2) are now known to be vitiated by a confounding. The reasoning is as follows: OSM is defined as the difference in the level of performance observed when the mask co-terminates with the target *minus* the level of performance observed when the mask continues to be on view after target offset. By that criterion, the functions in Figure 2 indicate that the magnitude of OSM varies with the size of the search display: OSM is maximal at set size 16, and absent at set size 1.

What vitiates the data in Figure 2 is a response ceiling imposed by the 100% limit of the response scale. When that response ceiling is removed by making the task more difficult, as was done by Argyropoulos et al. (2013; see also Jannati et al., 2013), the functions turn out to be parallel across set sizes. This means that, although the level of performance varies as a function of set size, the magnitude of OSM does not. The invariance of OSM with set size obviously invalidates both the CMOS and the AGT models. Importantly, however, invariance of OSM across set sizes in no way impugns reentry as the underlying mechanism, witness the experimental evidence adduced in the present article.

ACKNOWLEDGMENTS

This work was supported by a Discovery grant from the Natural Sciences and Engineering Research Council of Canada. I thank Thomas Spalek, Hayley Lagroix, Ali Jannati, and James Patten for commenting on an earlier version of this article.

REFERENCES

- Argyropoulos, I., Gellatly, A., Pilling, M., and Carter, W. (2013). Set size and mask duration do not interact in object-substitution masking. *J. Exp. Psychol. Hum. Percept. Perform.* 39, 646–661. doi: 10.1037/a0030240
- Binsted, G., Brownell, K., Vorontsova, Z., Heath, M., and Saucier, D. (2007). Visuo-motor system uses target features unavailable to conscious awareness. *Proc. Natl. Acad. Sci. U.S.A.* 104, 12669–12672. doi: 10.1073/pnas.0702307104

- Bouvier, S., and Treisman, A. (2010). Visual feature binding requires reentry. *Psychol. Sci.* 21, 200–204. doi: 10.1177/0956797609357858
- Breitmeyer, B. G. (1984). *Visual Masking: An Integrative Approach*. New York: Oxford University Press.
- Breitmeyer, B. G., and Ganz, L. (1976). Implications of sustained and transient channels for theories of visual pattern masking, saccadic suppression, and information processing. *Psychol. Rev.* 83, 1–36. doi: 10.1037/0033-295X.83.1.1
- Breitmeyer, B. G., and Ögmen, H. (2006). *Visual Masking: Time Slices Through Conscious and Unconscious Vision*. New York: Oxford University Press. doi: 10.1093/acprof:oso/9780198530671.001.0001
- Broadbent, D. E. (1958). *Perception and Communication*. New York: Oxford University Press. doi: 10.1037/10037-000
- Carlson, T. A., Rauschenberger, R., and Verstraten, F. A. (2007). No representation without awareness in the lateral occipital cortex. *Psychol. Sci.* 18, 298–302. doi: 10.1111/j.1467-9280.2007.01892.x
- Chun, M. M., Golomb, J. D., and Turk-Browne, N. B. (2011). A taxonomy of external and internal attention. *Annu. Rev. Psychol.* 62, 73–101. doi: 10.1146/annurev.psych.093008.100427
- Coren, S. (1972). Subjective contours and apparent depth. *Psychol. Rev.* 79, 359–367. doi: 10.1037/h0032940
- Di Lollo, V., Enns, J. T., and Rensink, R. A. (2000). Competition for consciousness among visual events: the psychophysics of reentrant visual processes. *J. Exp. Psychol. Gen.* 129, 481–507. doi: 10.1037/0096-3445.129.4.481
- Eriksen, C. W., and St. James, J. D. (1986). Visual attention within and around the field of focal attention: a zoom lens model. *Percept. Psychophys.* 40, 225–240. doi: 10.3758/BF03211502
- Grossberg, S. (1995). The attentive brain. *Am. Sci.* 83, 438–449.
- Guest, D., Gellatly, A., and Pilling, M. (2011). The effect of spatial competition between object-level representations of target and mask on object substitution masking. *Atten. Percept. Psychophys.* 73, 2528–2541. doi: 10.3758/s13414-011-0196-5
- Harris, J. A., Ku, S., and Woldorff, M. G. (2013). Neural processing stages during object-substitution masking, and their relationship to perceptual awareness. *Neuropsychologia* 51, 1907–1917. doi: 10.1016/j.neuropsychologia.2013.05.023
- Jannati, A., Spalek, T. M., and Di Lollo, V. (2013). A novel paradigm reveals the role of reentrant visual processes in object substitution masking. *Atten. Percept. Psychophys.* 75, 1118–1127. doi: 10.3758/s13414-013-0462-9
- Jiang, Y., and Chun, M. M. (2001). Asymmetric object substitution masking. *J. Exp. Psychol. Hum. Percept. Perform.* 27, 895–918. doi: 10.1037/0096-1523.27.4.895
- Kanizsa, G. (1976). Subjective contours. *Sci. Am.* 234, 48–52. doi: 10.1038/scientificamerican0476-48
- Lavie, N., and Tsal, Y. (1994). Perceptual load as a major determinant of the locus of selection in visual attention. *Percept. Psychophys.* 56, 183–197. doi: 10.3758/BF03213897
- Lleras, A., and Moore, C. M. (2003). When the target becomes the mask: using apparent motion to isolate the object-level component of object substitution masking. *J. Exp. Psychol. Hum. Percept. Perform.* 29, 106–120. doi: 10.1037/0096-1523.29.1.106
- Mumford, D. (1991). On the computational architecture of the neocortex I. The role of the thalamo-cortical loop. *Biol. Cybern.* 65, 135–145. doi: 10.1007/BF00202389
- Mumford, D. (1992). On the computational architecture of the neocortex II. The role of cortico-cortical loops. *Biol. Cybern.* 66, 241–251. doi: 10.1007/BF00198477
- Neill, W. T., Hutchison, K. A., and Graves, D. F. (2002). Masking by object substitution: dissociation of masking and cuing effects. *J. Exp. Psychol. Hum. Percept. Perform.* 28, 682–694. doi: 10.1037/0096-1523.28.3.682
- Norman, D. A., and Bobrow, D. G. (1975). On data-limited and resource-limited processes. *Cogn. Psychol.* 7, 44–64. doi: 10.1016/0010-0285(75)90004-3
- Petry, S., and Meyer, G. E. (eds) (1987). *The Perception of Illusory Contours*. New York: Springer-Verlag. doi: 10.1007/978-1-4612-4760-9
- Pöder, E. (2013). Attentional gating models of object substitution masking. *J. Exp. Psychol. Gen.* 142, 1130–1141. doi: 10.1037/a0030575
- Posner, M. I., Snyder, C. R., and Davidson, B. J. (1980). Attention and the detection of signals. *J. Exp. Psychol. Gen.* 109, 160–174. doi: 10.1037/0096-3445.109.2.160
- Purghue, F., and Coren, S. (1992). Subjective contours 1900–1990: research trends and bibliography. *Percept. Psychophys.* 51, 291–304. doi: 10.3758/BF03212255
- Reiss, J. E., and Hoffman, J. E. (2007). Disruption of early face recognition processes by object substitution masking. *Vis. Cogn.* 15, 789–798. doi: 10.1080/13506280701307035
- Treisman, A. M., and Gelade, G. (1980). A feature-integration theory of attention. *Cogn. Psychol.* 12, 97–136. doi: 10.1016/0010-0285(80)90005-5
- Weidner, R., Shah, N. J., and Fink, G. R. (2006). The neural basis of perceptual hypothesis generation and testing. *J. Cogn. Neurosci.* 18, 258–266. doi: 10.1162/jocn.2006.18.2.258
- Woodman, G. F., and Luck, S. J. (2003). Dissociations among attention, perception, and awareness during object-substitution masking. *Psychol. Sci.* 14, 605–611. doi: 10.1046/j.0956-7976.2003.psci_1472.x

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.

Received: 23 April 2014; accepted: 10 July 2014; published online: 04 August 2014.

Citation: Di Lollo V (2014) Reentrant processing mediates object substitution masking: comment on Pöder (2013). *Front. Psychol.* 5:819. doi: 10.3389/fpsyg.2014.00819

This article was submitted to Perception Science, a section of the journal *Frontiers in Psychology*.

Copyright © 2014 Di Lollo. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



The changing picture of object substitution masking: reply to Di Lollo (2014)

Endel Pöder*

Department of Experimental Psychology, Institute of Psychology, University of Tartu, Tartu, Estonia

*Correspondence: endel.poder@ut.ee

Edited by:

Bruno Breitmeyer, University of Houston, USA

Reviewed by:

Greg Francis, Purdue University, USA

Bruce Bridgeman, University of California, Santa Cruz, USA

Keywords: visual masking, object substitution, attention, modeling, reentrant, feed-forward

In his recent comment, Di Lollo (2014) criticizes my proposal (Pöder, 2013) that the attentional gating model (Reeves and Sperling, 1986; Sperling and Weichselgartner, 1995) might be the most simple and reasonable explanation for the results of object substitution masking (OSM) experiments. He argues that OSM cannot be explained without a reentrant hypotheses-testing mechanism (as proposed in Di Lollo et al., 2000). A closer look at his arguments reveals that they are partly based on an inaccurate interpretation of my study, and partly, on some highly problematic assumptions about visual processing.

The goal of my study (Pöder, 2013) was to understand the mechanisms behind the results of a typical OSM experiment with varied duration of masker and set-size. There were two main points in my study that are relevant in the present context. First, I analyzed the computational model (CMOS) proposed by Di Lollo et al. (2000) and found it to be identical with the attentional gating model, which has no direct relationship with any kind of “reentrant processing.” Second, I proposed an improved mechanism of attention to be combined with this model (or other possible masking models). I have never proposed or tested any new model of masking.

Di Lollo (2014) seems to have a rather subjective view of my study. He criticizes something named Pöder’s feed-forward account (or model) of OSM, which is supposedly based on two assumptions: reduction of signal to noise ratio (SNR) as a result of integrated noise from the masker, and delayed deployment of attention. Both

assumptions are declared to be wrong or at least unjustified.

As I mentioned, I have not built a new model of OSM but just reinterpreted Di Lollo et al.’s (2000) CMOS. Therefore, these assumptions can only be the assumptions underlying CMOS, and actually, they are. I believe that CMOS was a reasonably good model for OSM experiments and that its main assumptions cannot be fundamentally wrong. However, Di Lollo (2014) missed some possibly important details. In CMOS, SNR was reduced not only because of accumulating “noise” from the masker but also because of decay of the target signal. Neither Di Lollo et al. (2000) nor Pöder (2013) supposed that SNR is proportional (or inversely proportional) to the duration of the masker.

The model that was tested in my study tried to explain the set-size effects better than the simple attention deployment mechanism used in CMOS. My model assumes that the set-size effect is caused by an initial stage of divided attention. In this model masking *per se* is independent of set-size. A similar idea about the invariance of masking to set-size was independently discovered by Argyropoulos et al. (2013). My model with divided attention goes a bit further and proposes a plausible explanation for the observed set size effects.

Having explained away the set-size effect, a simple masking effect remains. In my study (Pöder, 2013), I did not attempt to reveal its exact mechanisms. I indicated that the combination of the decay of the target signal, integration of the masker signal, and a delayed attention as

implemented in CMOS (Di Lollo et al., 2000) might do the job. However, there are many other possibilities. After the removal of the burden to explain set-size effects, the classic models of masking (Weinstein, 1968; Bridgeman, 1971; Francis, 1997) that were analyzed in Francis and Hermens (2002) become fully applicable for OSM (note that the unsatisfactory modeling of attention/set-size effects was the main problem with these models; Di Lollo et al., 2002). Of course, Di Lollo’s reentrant hypotheses-testing idea can be included in the candidate list too, as well as Bachmann’s (1994) non-specific amplification idea. Hopefully, future studies will be able to discriminate between these models.

Although the set-size effect has been quite convincingly separated from the masking effect, some role of attention in OSM is still not excluded. In a recent study, Pilling et al. (2014) found a modest effect of spatial pre-cueing in one out of the five experiments. Up to now, nobody has explained away the position uncertainty and pre-cueing effects reported in earlier studies (Enns and Di Lollo, 1997; Neill et al., 2002; Tata and Giaschi, 2004; Luiga and Bachmann, 2007). If attention is still important then the models of masking developed by Smith et al. (Smith and Wolfgang, 2004; Smith et al., 2009) or by Bridgeman (2007) may be considered.

Di Lollo (2014) mainly argues against a CMOS-like masking account, but apparently supposes that any kind of essentially feed-forward model cannot explain masking with a sparse trailing masker. The majority of his arguments are based

on a quite strange view on the visual system. Di Lollo (2014) seems to ignore the hierarchical nature of visual processing and assume that all the masking-related processes should occur at some low-level retinotopic layer. Therefore, only retinotopic picture-level noise, integration, and masking are possible. The real visual system consists of at least 4–5 (possibly more) processing levels (e.g., DeYoe and van Essen, 1988; Riesenhuber and Poggio, 1999). The target and masker signals are processed and temporally and spatially integrated throughout this hierarchy. Thus, the “noise” from irrelevant stimuli may interact with relevant signals at any level of processing. The higher levels are increasingly invariant to spatial positions and combine all visual features including motion. It is therefore not surprising at all that the dots far from the target, or a masker that was retinotopically moved away from the target location (Lleras and Moore, 2003), can still interact at higher object recognition levels. Note that the higher-level masking does not need anything like sending perceptual hypotheses back to the lower levels.

A large part of Di Lollo's (2014) critique is directed against using SNR in the models of OSM (although the idea itself was introduced by Di Lollo et al., 2000). Di Lollo (2014) argues that a noisy representation of visual objects is not consistent with the phenomenal experience of not seeing them as noisy pictures; and with clearly different effects of the external pixel noise compared to four-dot masker. His arguments apparently challenge some points of traditional psychophysics. In usual psychophysical models (e.g., Macmillan and Creelman, 2005), noise is a random trial-by-trial variability of internal representations that causes incorrect perceptual decisions. This noise can make a letter A look like a letter B, or like a chicken, or like a blank screen, in some trials. We can manipulate this noise (or SNR) by varying stimulus contrast, size, or exposure duration, presenting distractors, or forward or backward maskers, pre-cueing attention, simultaneous eye movements, etc., besides adding external pixel noise. There is no reason to suppose that the decision-level noise should be visible and look like noise within a single image.

New studies have forced Di Lollo and colleagues to make some changes to their theory. The original account of OSM (Di Lollo et al., 2000) was heavily based on both reentrant hypotheses testing and deployment of attention. The Argyropoulos et al. (2013) results indicated that something is wrong with this theory. The simplest way out was to leave out attention. However, attention had a key role in CMOS and in the predictions related to the Di Lollo et al. (2000) theory. Jannati et al. (2013) found an innovative solution. Nominally, they removed attention but attributed its properties to “reentrant processing.” In the original model (Di Lollo et al., 2000), the reentrant processing was supposed to be perpetual generation and testing perceptual hypotheses with periodicity of about 13 ms. In their new account (Jannati et al., 2013), the reentrance “arrives” at about 80–120 ms after stimulus onset, a typical delay of focusing spatial attention (e.g., Cheal and Lyon, 1991). Overall, their revised theory still follows the attentional gating logic of CMOS. At the same time, they claim that their experiment falsifies the attentional gating account of OSM. A closer look at their arguments reveals that their description of the “attentional gating model” does not contain attention at all. It is not surprising that such a model cannot fit any (old or new) experimental results.

In conclusion, I would describe the present situation as follows. The attentional gating idea effectively explained the effects of attention and simplified the problem of OSM tremendously. Now, one may take a single target stimulus with a common-onset masker and present them at a fixed position of the visual field, with full attention available, and try to observe OSM. There is a chance that Di Lollo (or somebody else) can demonstrate the action of reentrant hypotheses-testing mechanism in that simple experiment. It would be an interesting and surprising (at least for me) finding. But it would not contradict my attentional gating account of OSM experiments with set-size variation.

REFERENCES

Argyropoulos, Y., Gellatly, A., Pilling, M., and Carter, W. (2013). Set size and mask duration do not interact in object-substitution masking. *J. Exp. Psychol. Hum. Percept. Perform.* 39, 646–661. doi: 10.1037/a0030240

- Bachmann, T. (1994). *Psychophysiology of Visual Masking: The Fine Structure of Conscious Experience*. Commack, NY: Nova Science.
- Bridgeman, B. (1971). Metacontrast and lateral inhibition. *Psychol. Rev.* 78, 528–539. doi: 10.1037/h0031782
- Bridgeman, B. (2007). Common-onset masking simulated with a distributed-code model. *Adv. Cogn. Psychol.* 3, 33–40. doi: 10.2478/v10053-008-0012-5
- Cheal, M. L., and Lyon, D. R. (1991). Central and peripheral precueing of forced-choice discrimination. *Q. J. Exp. Psychol.* 43A, 859–880. doi: 10.1080/14640749108400960
- DeYoe, E. A., and van Essen, D. C. (1988). Concurrent processing streams in monkey visual cortex. *Trends Neurosci.* 5, 219–226. doi: 10.1016/0166-2236(88)90130-0
- Di Lollo, V. (2014). Reentrant processing mediates Object Substitution Masking: comment on Pöder (2013). *Front. Psychol.* 5:819. doi: 10.3389/fpsyg.2014.00819
- Di Lollo, V., Enns, J. T., and Rensink, R. A. (2000). Competition for consciousness among visual events: the psychophysics of reentrant visual processes. *J. Exp. Psychol. Gen.* 129, 481–507. doi: 10.1037/0096-3445.129.4.481
- Di Lollo, V., Enns, J. T., and Rensink, R. A. (2002). Object substitution without reentry? *J. Exp. Psychol. Gen.* 131, 594–596. doi: 10.1037/0096-3445.131.4.594
- Enns, J. T., and Di Lollo, V. (1997). Object substitution: a new form of visual masking in unattended visual locations. *Psychol. Sci.* 8, 135–139. doi: 10.1111/j.1467-9280.1997.tb00696.x
- Francis, G. (1997). Cortical dynamics of lateral inhibition: metacontrast masking. *Psychol. Rev.* 104, 572–594. doi: 10.1037/0033-295X.104.3.572
- Francis, G., and Hermens, F. (2002). Comment on “Competition for consciousness among visual events: the psychophysics of reentrant visual processes” (Di Lollo, Enns, and Rensink, 2000). *J. Exp. Psychol. Gen.* 131, 590–593. doi: 10.1037/0096-3445.131.4.590
- Jannati, A., Spalek, T. M., and Di Lollo, V. (2013). A novel paradigm reveals the role of reentrant visual processes in object substitution masking. *Atten. Percept. Psychophys.* 75, 1118–1127. doi: 10.3758/s13414-013-0462-9
- Lleras, A., and Moore, C. M. (2003). When the target becomes the mask: using apparent motion to isolate the object-level component of object substitution masking. *J. Exp. Psychol. Hum. Percept. Perform.* 29, 106–120. doi: 10.1037/0096-1523.29.1.106
- Luiga, I., and Bachmann, T. (2007). Different effects of the two types of spatial pre-cueing: what precisely is “attention” in Di Lollo's and Enns' substitution masking theory? *Psychol. Res.* 71, 634–640. doi: 10.1007/s00426-006-0052-8
- Macmillan, N. A., and Creelman, C. D. (2005). *Detection Theory: A User's Guide, 2nd Edn.* New York, NY: Lawrence Erlbaum Associates.
- Neill, W. T., Hutchison, K. A., and Graves, D. F. (2002). Masking by object substitution: dissociation of masking and cuing effects. *J. Exp. Psychol. Hum. Percept. Perform.* 28, 682–694. doi: 10.1037/0096-1523.28.3.682
- Pilling, M., Gellatly, A., Argyropoulos, Y., and Skarratt, P. (2014). Exogenous spatial precueing

- reliably modulates object processing but not object substitution masking. *Atten. Percept. Psychophys.* doi: 10.3758/s13414-014-0661-z
- Pöder, E. (2013). Attentional gating models of object substitution masking. *J. Exp. Psychol. Gen.* 142, 1130–1141. doi: 10.1037/a0030575
- Reeves, A., and Sperling, G. (1986). Attention gating in short-term visual memory. *Psychol. Rev.* 93, 180–206. doi: 10.1037/0033-295X.93.2.180
- Riesenhuber, M., and Poggio, T. (1999). Hierarchical models of object recognition in cortex. *Nat. Neurosci.* 2, 1019–1025. doi: 10.1038/14819
- Smith, P. L., Lee, Y.-E., Wolfgang, B. J., and Ratcliff, R. (2009). Attention and the detection of masked radial frequency patterns: data and model. *Vision Res.* 49, 1363–1377. doi: 10.1016/j.visres.2008.04.024
- Smith, P. L., and Wolfgang, B. J. (2004). The attentional dynamics of masked detection. *J. Exp. Psychol. Hum. Percept. Perform.* 30, 119–136. doi: 10.1037/0096-1523.30.1.119
- Sperling, G., and Weichselgartner, E. (1995). Episodic theory of the dynamics of spatial attention. *Psychol. Rev.* 102, 503–532. doi: 10.1037/0033-295X.102.3.503
- Tata, M. S., and Giaschi, D. E. (2004). Warning: attending to a mask may be hazardous to your perception. *Psychon. Bull. Rev.* 11, 262–268. doi: 10.3758/BF03196568
- Weisstein, N. (1968). A Rashevsky–Landahl neural net: simulation of metacontrast. *Psychol. Rev.* 75, 494–521. doi: 10.1037/h0026769
- commercial or financial relationships that could be construed as a potential conflict of interest.
- Received: 29 July 2014; accepted: 22 August 2014; published online: 09 September 2014.
- Citation: Pöder E (2014) The changing picture of object substitution masking: reply to Di Lollo (2014). *Front. Psychol.* 5:1004. doi: 10.3389/fpsyg.2014.01004
- This article was submitted to *Perception Science*, a section of the journal *Frontiers in Psychology*.
- Copyright © 2014 Pöder. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.
- Conflict of Interest Statement:** The author declares that the research was conducted in the absence of any



A computational investigation of feedforward and feedback processing in metacontrast backward masking

David N. Silverstein^{1,2,3*}

¹ PDC Center For High Performance Computing, KTH Royal Institute of Technology, Stockholm, Sweden

² Department of Computational Biology, KTH Royal Institute of Technology, Stockholm, Sweden

³ Stockholm Brain Institute, Karolinska Institute, Solna, Sweden

Edited by:

Hulusi Kafaligonul, Bilkent University, Turkey

Reviewed by:

Greg Francis, Purdue University, USA

Frouke Hermens, University of Aberdeen, UK

*Correspondence:

David N. Silverstein, PDC Center for High Performance Computing, KTH Royal Institute of Technology, SE-100 44 Stockholm, Sweden
e-mail: davidsi@kth.se

In human perception studies, visual backward masking has been used to understand the temporal dynamics of subliminal vs. conscious perception. When a brief target stimulus is followed by a masking stimulus after a short interval of <100 ms, performance on the target is impaired when the target and mask are in close spatial proximity. While the psychophysical properties of backward masking have been studied extensively, there is still debate on the underlying cortical dynamics. One prevailing theory suggests that the impairment of target performance due to the mask is the result of lateral inhibition between the target and mask in feedforward processing. Another prevailing theory suggests that this impairment is due to the interruption of feedback processing of the target by the mask. This computational study demonstrates that both aspects of these theories may be correct. Using a biophysical model of V1 and V2, visual processing was modeled as interacting neocortical attractors, which must propagate up the visual stream. If an activating target attractor in V1 is quiesced enough with lateral inhibition from a mask, or not reinforced by recurrent feedback, it is more likely to burn out before becoming fully active and progressing through V2 and beyond. Results are presented which simulate metacontrast backward masking with an increasing stimulus interval and with the presence and absence of feedback activity. This showed that recurrent feedback diminishes backward masking effects and can make conscious perception more likely. One model configuration presented a metacontrast noise mask in the same hypercolumns as the target, and produced type-A masking. A second model configuration presented a target line with two parallel adjacent masking lines, and produced type-B masking. Future work should examine how the model extends to more complex spatial mask configurations.

Keywords: backward masking, visual cortex, feedback projections, conscious processing, neural attractor dynamics

INTRODUCTION

Visual backward masking is a classic technique used to examine differences between conscious and unconscious visual processing (Breitmeyer and Ogmen, 2006). It is employed by presenting a target image followed closely in time by a mask image. The target image exposure is typically very short, often around 20 or 16.7 ms, but may be limited by monitor refresh rates. The mask typically has longer exposure, often at least 50 ms, but sometimes up to hundreds of milliseconds. The time from the start of the target exposure to the time of the start of the mask is experimentally varied, and this is commonly known as the stimulus onset asynchrony (SOA). While there are many experimental variations, target and mask exposure times often remain fixed while the SOA is varied. When the SOA is 20–60 ms, a face target is sometimes not consciously perceived (Rolls, 2004). The measured response from recognizing a masked target has been characterized as type-A and type-B masking. In type-A masking, the masking effect monotonically decreases with increasing SOA. This is often

associated with a stronger masking stimulus. In type-B masking, the masking effect is weaker at low SOAs, becomes stronger at some point with SOAs less than 100 ms and then diminishes again with increasing SOA, with a response curve sometimes referred to as a U-shaped function (Breitmeyer and Ganz, 1976). Different types of masks are possible. Pattern masking occurs when the mask shares some features with the target or is superimposed. Metacontrast masking occurs when the mask features are non-overlapping with the target, but some features may be in close spatial proximity. Masks may also be different forms of noise, and might also be a flash of light (Breitmeyer and Ogmen, 2000).

There are two broad classes of conceptual models for explaining backward masking. One states that visual sensory information is stored in a visual sensory buffer (or iconic memory) for processing, but can be interrupted by a mask (Sperling, 1963; Di Lollo, 1980). The other states that information propagates in dual channels (such as parvocellular and magnocellular pathways), with one faster and more transient and the other slower

and more sustained. When the target and mask are presented to both channels, the fast transient activity of the mask suppresses the slow sustained activity of the target through inter-channel inhibition. The psychophysics of masking have been characterized, although individual differences have been observed in stable masking functions (Albrecht and Mattler, 2012). Less understood are the underlying cortical dynamics, which are still deeper in debate (Macknick and Martinez-Conde, 2007). There are several prevailing theories on the mechanisms of backward masking and visual masking in general. One view states that this is primarily caused by feedforward lateral inhibition (Macknick, 2006). The mask spatiotemporally interferes with the target through inhibition, preventing further processing. Another view asserts that the mask interferes with feedback processing from higher areas, preventing the discrimination between the figure and background which makes visual awareness possible (Lamme and Roelfsema, 2000; Super et al., 2001; Lamme et al., 2002).

Several computational models have been developed over time and at different levels of abstraction, a subset of which will be discussed here. Earlier models focused more on the temporal aspects of the masking function, with later models incorporating some spatial aspects as well (Francis, 1997, 2009; Hermens et al., 2008). The retino-cortical dynamics (RECOD) model (Ogmen, 1993) is a dual-channel approach which incorporates neural representations as well as feedforward dynamics and feedback inhibition. It utilizes transient-on-sustained inhibition to explain some backward masking properties (Breitmeyer and Ogmen, 2000). The Boundary Contour System (BCS) originally developed by Grossberg and Mingolla (1985) and extended by Francis (1997) can reproduce many aspects of metacontrast masking. It uses model neurons, can spatially represent two orientation preferences and includes elements of lateral inhibition and feedback. Bugmann and Taylor (2005) also developed a detailed neural model with feedforward and lateral connections, which was able to produce U-shaped masking functions. Spatial aspects of backward masking have also been explored by modeling the shine-through effect (Herzog et al., 2001). When a vernier target with two adjacent and offset vertical lines is masked by a grating with five straight lines, target perception is impaired. However, if masked with a grating of seven or more straight lines, the vernier target is more easily perceived, and “shines through” the grating. The 3D-LAMINART (Grossberg, 1997; Francis, 2009) and WCTM (Herzog et al., 2003) computational models have been able to reproduce some but not all aspects of these phenomena (Rüter et al., 2011). 3D-LAMINART is a general purpose visual model that utilizes binocular vision to perceive the vernier target. WCTM is a simpler two-layer model which uses lateral inhibition to suppress repeating patterns such as lines.

This study seeks to model the cortical dynamics of metacontrast backward masking at a biophysically detailed level, to investigate the roles of feedforward, feedback and lateral connections, specifically in the context of interacting neural attractor networks (Hopfield, 1982; Amit, 1989; Hertz et al., 1991). This spiking neural attractor model is conceptually related to the sensory store model or iconic memory, because a neural attractor is a recurrent store of activity for associative processing. Among existing neural models (Francis, 1997, 2009; Bugmann and Taylor,

2005) the work presented here is perhaps the most biophysically detailed cortical model to date used to simulate the temporal aspects of backward masking. The spatial aspects are currently limited to abstract metacontrast representations where the target and mask are represented in close proximity in common hypercolumns or as parallel lines, although this could be extended with biophysical feature detectors (Rehn et al., 2011). A neural attractor in this case is considered an activated stored memory pattern, which is a neural assembly of sparse and distributed pyramidal cells recurrently connected with excitatory synapses. When a stored memory pattern (or attractor memory) is partially stimulated, it can become fully active across the distributed representation through recurrent excitation. Over time, it adapts and burns out, due to short-term synaptic plasticity and calcium dynamics, both of which can have near-second time constants. Many attractor memories can co-exist in the same neural population, and may mutually exclude each other when activated, through lateral and di-synaptic inhibition. These neural attractor memories can also activate each other associatively when overlapping and be nested and hierarchical as well. In the case of primary visual cortex, these attractor memories can represent features as grouped orientation preferences. It is hypothesized that targets consist of a set of feature detectors in individual visual areas, each with an associated patch-level attractor memory, containing minicolumns that are themselves small-world networks and mini-attractors. These patch-level (i.e., V1 or V2) attractor memories are interconnected across visual areas, activating regional-level attractors through feedforward and feedback projections. With feedforward activity, attractor activations propagate up the ventral stream (Kravits et al., 2013) as a traveling wave (Sato et al., 2012), while feedback activity provides competitive reinforcement from previous perceptual memories, or resolves ambiguity and expectation partially on the regional level (Wyatte et al., 2014). Eventually, this traveling wave is postulated to reach the pre-frontal cortex for global-level attractor activation or “ignition” for conscious access (Dehaene and Changeux, 2011). The model in this study hypothesizes that regional-level attractor memories exist across V1 and V2 and is limited to those areas. When a patch-level attractor memory is stimulated, it takes time for recurrence to fully activate it, sometimes up to 50 ms. During this time, it can be more vulnerable to interference such as a metacontrast noise mask, which may produce a monotonically decreasing masking function as the attractor builds and becomes more stable. If two competing attractor memories are activated as a target and mask, the interference between them can build as the attractors build, depending on spatial overlap or proximal contours. With spatial overlap, it is hypothesized that activation is more likely to transition to the masking attractor memory, if the target is not reinforced by feedback. In common-offset masking where the target and mask are presented simultaneously, transitions to the masking attractor memory can also occur (Enns and Di Lollo, 2000). Proximal contours during masking may also interfere with target attractor activation via lateral inhibition.

Evidence suggests that the latency of projections between V1 and V2 is about 10 ms in both directions (Nowak and Bullier, 1997; Girard et al., 2001), while horizontal propagation has been

found to be significantly slower (Sugihara et al., 2011). This suggests that, considering the synaptic integration delays in V2, feedback to V1 may arrive before lateral processing is complete. Thus, this feedback may also be a factor in how that lateral processing completes. Both excitation and inhibition have been identified in both feedforward and feedback projections in rat primary visual cortex, although feedback inhibition appears to be less (Shao and Burkhalter, 1996). If feedback recurrently excites currently activated features, the target feature attractors will be enhanced, and be more likely to become fully active and propagate. However, if excitatory feedback were to activate attractor memories for features not present in the target, the target attractor could be inhibited through competition. Alternatively, if feedback inhibits other feature attractors through di-synaptic inhibition, then the target attractor will be enhanced through lower competition or suppressed noise, or at least would not be diminished.

MATERIALS AND METHODS

A biophysical model was constructed of early visual cortex, with two different instantiations. The first instantiation (called model 1) entailed using an abstract target and metacontrast noise mask in close spatial proximity. The second instantiation (called model 2) entailed using a single vertical line for the target and two adjacent parallel lines for the mask, with the intention of a more specific spatial representation. The models represent a subset of the ventral stream of primate visual cortex and includes the lateral geniculate nucleus (LGN), areas V1 and V2 and the projections between them. While projections between the LGN and V1 layer 4 are feedforward only, V1 and V2 are bidirectionally connected. The LGN is represented as a grid of 256 locations in model 1 and 648 locations in model 2, each containing a stack of 10 relay cells, acting as on-center cells. Stimuli presented to the LGN are not actual images, but are abstract representations. Off-center cells were not included. Each LGN location projects to pyramidal cells in one minicolumn of V1 layer 4 and surrounding interneurons (small basket cells), which in turn inhibit pyramidal cells in surrounding minicolumns within the same hypercolumn. The neocortical patches of V1 and V2 represent a square matrix of hypercolumns, each containing internal minicolumns. In model 1, the 4 mm² patch of cortex is composed of 4 × 4 hypercolumns, subsampled with 16 minicolumns each. In model 2, the 20 mm² patch of cortex is composed of 9 × 9 hypercolumns, subsampled with eight minicolumns each. The structure is similar to Silverstein and Lansner (2011), with the addition of a regular spiking non-pyramidal (RSNP) interneuron into the neocortical microcircuit (Lundqvist et al., 2006), a more complete layer 4 and the addition of layer 5. Di-synaptic inhibition and competition from RSNP interneurons occurred when attractor memories had intersecting hypercolumns, which occurred in model 1 but not model 2. The microcircuit of V1 is illustrated in **Figure 1**. The minicolumns are also subsampled, and contain pyramidal cells and interneurons for layers 2/3, 4, and 5. Each layer contains 20 pyramidal cells, two basket cells, and two interneurons allocated per minicolumn, although the basket cells physically reside outside the minicolumn. V1 layer 4 is known to largely contain spiny stellate cells, but pyramidal cells are used in their place for simplicity. While V1 and V2 are known to have

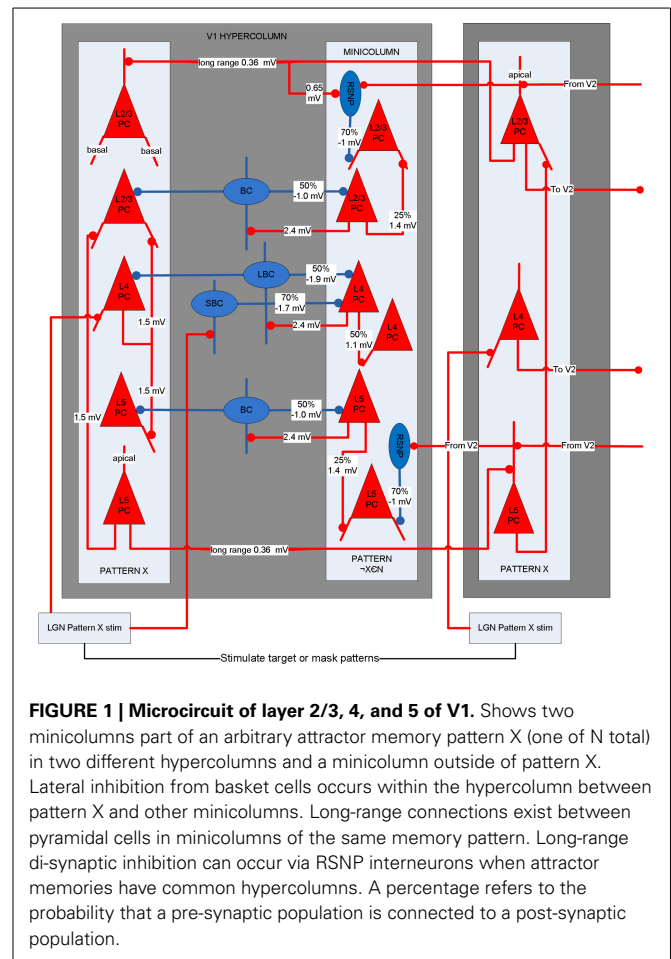
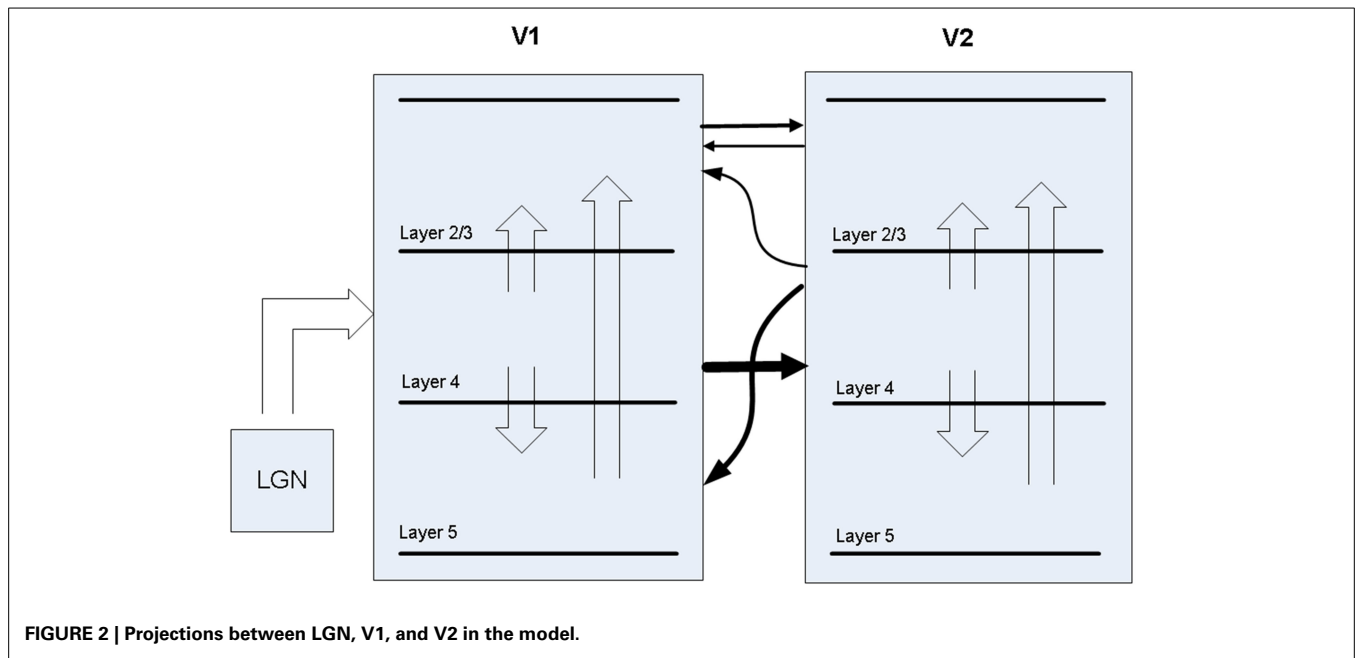


FIGURE 1 | Microcircuit of layer 2/3, 4, and 5 of V1. Shows two minicolumns part of an arbitrary attractor memory pattern X (one of N total) in two different hypercolumns and a minicolumn outside of pattern X. Lateral inhibition from basket cells occurs within the hypercolumn between pattern X and other minicolumns. Long-range connections exist between pyramidal cells in minicolumns of the same memory pattern. Long-range di-synaptic inhibition can occur via RSNP interneurons when attractor memories have common hypercolumns. A percentage refers to the probability that a pre-synaptic population is connected to a post-synaptic population.

different structure, they are both thought to have hypercolumns (Ts'o et al., 2009) and the same structure was used for both here. The V1 model represents interblobs (or interpatches) for orientation as hypercolumns, but does not include blobs (or patches) for color. It is also monocular, so does not include binocular stripes. Orientation preferences are represented in minicolumns. In model 1, these orientations remain abstract and are not tuned to particular feature preferences. However, randomly selected minicolumns in different hypercolumns are connected in stored memory patterns, representing linked orientation preferences for feature detection. While abstract, it is meant to generally represent features. In model 2, minicolumns have vertical orientation preferences for the more specific representation of line detection. V2 is known to have thin, pale and thick stripes, and the model represents the pale stripes only, which are known to also project to V4 and on along the ventral stream. Feed-forward projection streams from V1 interblobs to V2 pale stripes have been identified in Macaque (Sincich and Horton, 2005; Federer et al., 2013). These include projections from layer 2/3 and 4 of V1 interblobs to layer 2/3 and 4 of V2 pale stripes (Federer et al., 2013). Feedback projections from V2 originate from layer 2/3 and 6 and target layers 1, 2/3, and 5 of V1 (Sincich and Horton, 2005). A subset of these projections have been implemented, as can be seen in **Figure 2**.



Between V1 and V2, the model has feed-forward projections from V1 layer 4 to V2 layer 4 in addition to weaker projections from V1 layer 2/3 to V2 layer 2/3. Feed-back projections from V2 are predominantly from layer 4 to V1 layer 5, but also include layer 4 to V1 layer 2/3, which are about 10% of the strength. While anatomical data suggests most V2 feedback originates in layer 3, layer 4 is used for simplicity, considering dendrites from layer 3 pyramidal cells are likely to drop down into layer 4, where early activations are likely to occur after target presentation. The latencies of all projections between V1 and V2 projections are set to 10 ms, based on the findings mentioned earlier.

The model contained four different types of cells, which included spiking pyramidal cells, basket cells, RSNP interneurons and relay cells, all of which utilized the Hodgkin-Huxley formalism. The equations and parameters for these neurons are included in the Appendix. The pyramidal cells contained compartments for the soma, initial segment, basal dendrite, and apical dendrite, while the rest contained compartments for the soma, initial segment, and dendrite. With calcium dynamics, the pyramidal cells were adapting, the RSNP interneurons were weakly adapting, and the rest were not. The pyramidal cells and RSNP interneurons had Kainate/AMPA, NMDA, and GABA_A channels, while the basket cells had Kainate/AMPA and GABA_A channels. All synaptic channels had synaptic depression. However, the relay cells were stimulated only through a time-activated noise source applied to an alpha channel on the soma, and only projected to Kainate/AMPA and NMDA channels on V1 layer 4 pyramidal cells. All but the relay cells received 300 Hz of background Poisson noise and produced a positive bias.

In model 1, each area had a total of 18 stored attractor memories. Each attractor memory was created by randomly choosing one minicolumn from 10 of the 16 hypercolumns, an example of which can be seen in **Figure 3A**. The minicolumn sampling was restricted to prevent a minicolumn from being chosen in

more than one memory pattern, making the memories sparse and orthogonal. In model 2, each area had a total of 72 stored attractor memories, each containing nine minicolumns across the 81 hypercolumns, and organized as vertical lines. Once the minicolumns in an attractor memory were selected, long-range connections were created between them within the patch, which included both excitatory and inhibitory synapses. If a pairwise connection probability determined that two minicolumns in a stored memory pattern are to be connected, a pyramidal cell in the source minicolumn was randomly chosen to originate the axon. In the destination minicolumn, pyramidal cells received synapses with a 25% probability, and di-synaptic interneurons received synapses on surrounding minicolumns. All excitatory synapses had the same conductance, as did all the inhibitory (di-synaptic) synapses. For projections, attractor memories were connected across areas, similar to the descriptions in Szalicszno et al. (2013). To connect two attractor memories in two different areas, the minicolumns of the memory pattern in the source area projected axons to the minicolumns of the memory pattern in the destination area. These pattern projections were not all-to-all since it was assumed that projections are only a cue to activate a remote attractor memory that would necessarily have further local representations. Thus, four minicolumns in the corresponding attractor memory were selected on the destination side to receive the axons of the pattern projection.

BACKWARD MASKING SIMULATION

To present a target or mask stimulus to the model, 4 LGN locations, each with 10 relay cells in the LGN patch were stimulated, activating 40 relay cells in total. The target stimulus appears to the model as four dots in the grid and is sparse, representing 40% of the full target. The length of the target stimulation was always 20 ms and the length of the mask stimulation was 60 ms in model 1 and 50 ms in model 2. It was assumed that LGN relay cells fire

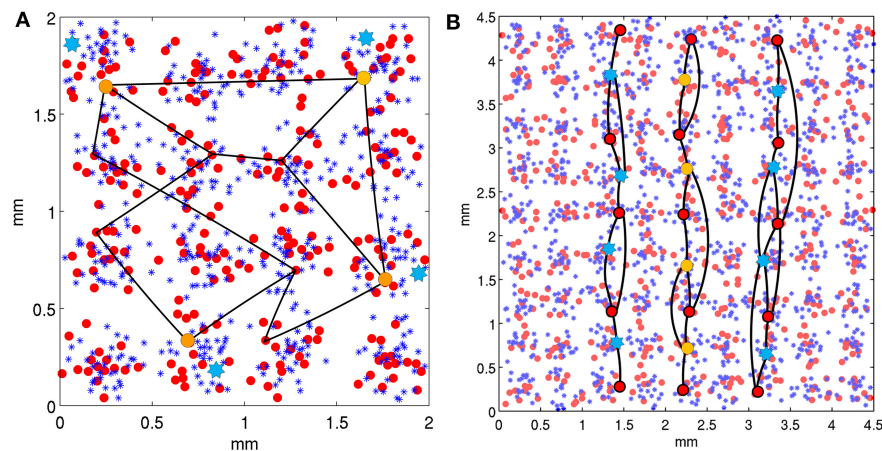


FIGURE 3 | Neocortical patches of V1 for two model configurations.

Within hypercolumns (1/2 mm in diameter) are minicolumns shown as small red circles and basket cells shown as blue asterisks. Example stored attractor memories are illustrated as black lines, which connect single minicolumns (via internal pyramidal cells) in independent hypercolumns, with a uniform connection probability. Only several of many connections of these attractor memories are illustrated. In a backward masking trial, minicolumns at orange circles are stimulated as the target and blue stars are stimulated as the mask, both via the LGN. **(A)** Shows

the model 1 configuration with 16 hypercolumns, each containing 16 minicolumns. Stored attractor memories are 10 random minicolumns in separate hypercolumns across the patch. Mask stimulation occurs in the same hypercolumns as target stimulation. **(B)** Shows the model 2 configuration with 81 hypercolumns, each containing 8 minicolumns. The stored attractor memories contain 9 minicolumns each and are organized as vertical lines across hypercolumns. The target is activated as the middle vertical line and the mask is activated as the two adjacent parallel lines two hypercolumns away.

at about 50 Hz, which meant each relay cell in a presented target stimulation would fire once over 20 ms and each relay cell included in the mask stimulation would fire three times over a 60 ms stimulation. These cell firings were uniformly distributed over the stimulation intervals. The relay cells in turn project to and stimulate minicolumns in V1, as can be seen in **Figure 3**. In the case of a target, the minicolumns are part of a stored memory pattern representing a feature detector, distributed across hypercolumns. In the case of a metacontrast noise mask, they are minicolumns selected from different attractor memories other than the target, which corresponds to parts of uncorrelated features. In the case of competing metacontrast line masks, the selected minicolumns were from a single attractor memories as the target was.

In model 1 as seen in **Figure 3A**, both the target and mask were presented as stimulated minicolumns in common hypercolumns for spatial proximity, which would roughly correspond to a visual angle of within about 10 min. Simulations were performed on model 2 with modifications for additional spatial context, to use lines in one dimension for both the target and mask, similar to stimuli presented in Growney et al. (1977). As seen in **Figure 3B**, the target was presented as a single, straight broken vertical line, and the mask was presented as two broken vertical parallel lines, flanking both sides of the target and equidistant from it. The patch size was changed from model 1 to 9×9 hypercolumns to accommodate the short lines, with eight minicolumns per hypercolumn. The feature detectors, as attractor memories, were modified (from random assembly) to assemble selected minicolumns (as orientation preferences) vertically, along each column of hypercolumns in the 9×9 matrix. Each of the eight minicolumns in every hypercolumn was used in a single independent, vertically

oriented feature detector, creating a total of 72 attractor memory patterns. These feature detectors were spatially redundant, but implemented so that an individual corresponding target or mask feature detector was activated for only one SOA interval during a trial run, which consisted of multiple sliding SOA intervals. This was done because the attractor memories did not completely recover from adaptation between the selective SOAs tested during each cortical second of each trial run, so couldn't be reused during a following SOA interval. Lateral inhibition in model 1 was within the hypercolumn only, but was changed to extend beyond the hypercolumn horizontally in model 2, for competition between the vertical target and mask lines. Lateral inhibition beyond the hypercolumn had a reduced basket-pyramidal synaptic connection probability of 50% one hypercolumn away, 25% two hypercolumns away and 0% outside of this. Simulations were performed with the mask 1, 2, and 3 hypercolumns away, which roughly corresponds to a fovea visual field arc of 10–20, 20–30, and 30–40 min. respectively.

For each model, five different individuals were simulated by generating 5 different neural sets, connection matrices and projections for the LGN, V1, and V2. Each of these individual instantiations was simulated for five trials with different seeds, for a total of 25 trials per trial set. Each trial consisted of presenting the target alone, the mask alone, and both target and mask with a sliding SOA of 20, 40, 60, 80, 100, and 120 ms. Feature attractors can become fully activated in Layer 2/3 and/or 5 of V1 and/or V2.

It is assumed that for the possibility of conscious perception, the linked attractor memory patterns must become fully active in layer 2/3 of both V1 and V2, indicating regional activation. To determine if this occurred, layer 2/3 of V1 and V2 were analyzed on each trial. For the attractor pattern to be considered fully

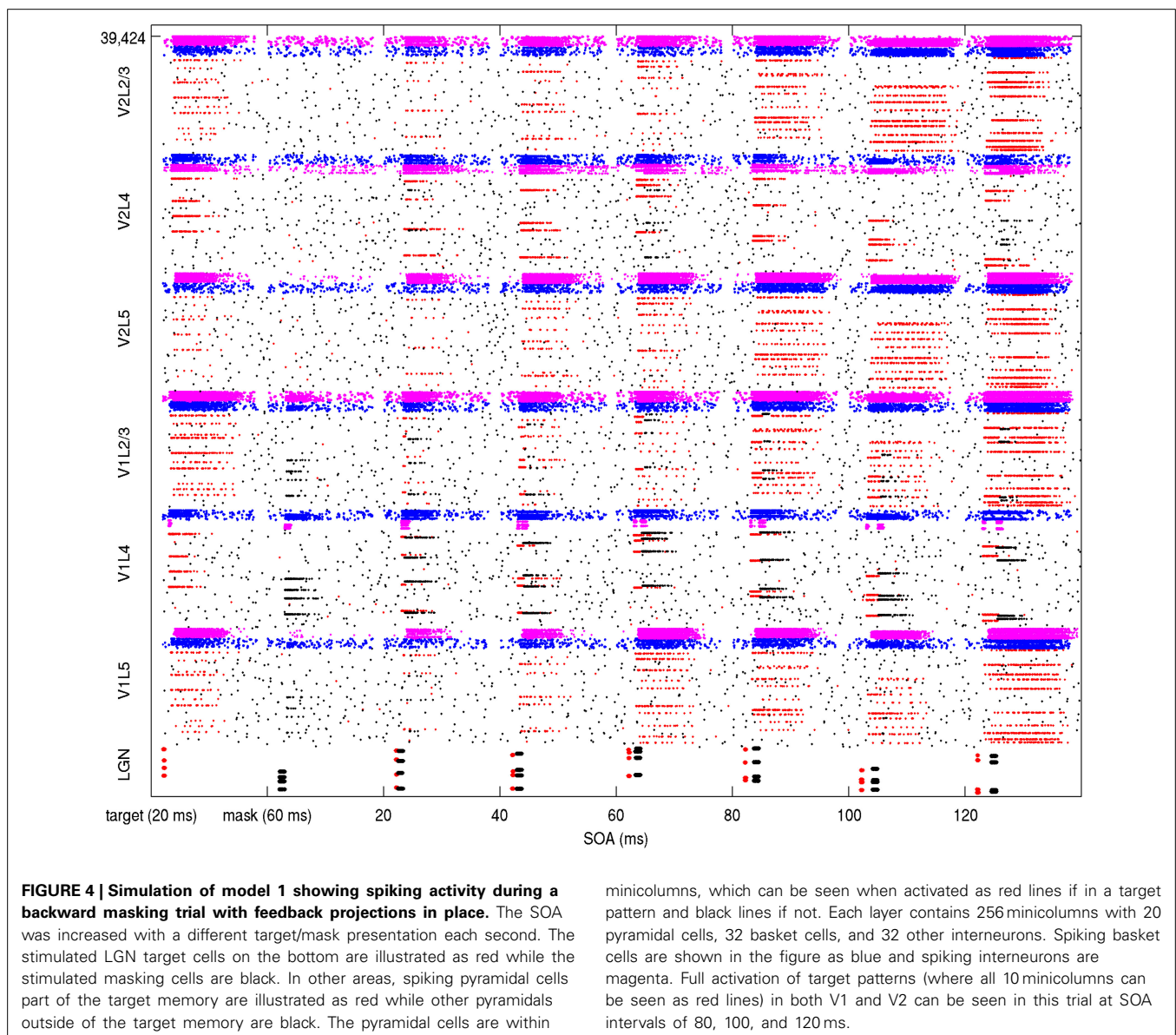
activated or complete in each area, pyramidal cells in 7 of the 10 minicolumns in the memory pattern were assumed to require at least 10 spikes during the SOA trial, indicating substantial recurrent activity within the attractor memory.

The models were implemented using the CORTSIM library (manuscript forthcoming) that was written using the native Hoc and Mod languages of the parallel NEURON simulator, version 7.3 (Carnevale and Hines, 2006) and run on a Cray XC30 system. Construction of the model geometry, synaptic connection matrices and analysis of the spiking output from the NEURON simulation were done in Matlab. There were 25 trials in each trial set, which ran both with and without feedback connections, on both model 1 and model 2. Model 1 had a total of 39,424 neurons and each trial took about 4 min. to run on 256 cores. Model 2 had a total of 99,792 neurons and each trial took about 5 min. to run on 648 cores.

RESULTS

Both lateral inhibition in V1 and V2 and feedback from V2 were factors in the backward masking effects observed in the models. When the target and mask presentations were close in time and space, they mutually inhibited each other, first in V1 layer 4 and later in layer 2/3 and 5. As the SOA increased, the target pattern was more likely to become a fully activated attractor before the mask stimulus could begin to interfere via basket cells and di-synaptic inhibition. Feedback from V2 could reinforce the target attractor and be a factor in achieving full activation locally in V1 and regionally in both V1 and V2, if arrival was early enough, before the mask stimulus arrived to compete.

The round-trip signaling latency of a target attractor in V1 feeding forward to V2 and feeding back to V1 is a minimum of about 25–40 ms, given a 10 ms latency of excitatory projections in each direction and synaptic integration at a single hop in V2.



More robust feedback from V2 to V1 can take longer, once an attractor activates in V2. Other feedback can occur via secondary excitatory activity and pattern completion from other layers, but this can take longer, even more than 50 ms. The reason for this is not just the synaptic integration times of secondary, tertiary and greater hops, but the longer latencies of horizontal connections. From model 1 results, an example backward masking trial with feedback in place is shown in **Figure 4**. Results varied between trials from individual connection matrices and trial seed, but here full activation of the target pattern in layers 2/3 of V1 and V2 can be seen with an SOA of 80 ms and greater, with near activation at an SOA of 60 ms. This activation was due to competition in V1 layer 4 between the target and mask (red and black lines in area V1L4), allowing activity to propagate to V2. Following this, the recurrent feedback from V2 reinforced and sustained the activated target. **Figure 5** shows this behavior as aggregated spiking

activity on a different example, comparing trials with and without feedback connections.

Depending on the level of stimulus response and dynamics, some attractor memories did regionally complete in both V1 layer 2/3 and V2 layer 2/3 without feedback projections, but this activity was less likely than with feedback projections in place. Both excitatory and inhibitory feedback (via di-synaptic inhibition) from V2 contributed to enhancing the target attractor by increasing the likelihood of full activation of the target pattern.

Full target pattern activation usually took 25–50 ms or longer, depending on local connectivity and conduction strengths. Reinforcement of memory attractors from recurrent feedback sometimes needed to occur before a masking stimulus arrived, or the target attractor would be quiesced. **Figure 6** shows aggregate results of two simulations for model 1, each consisting of 25 trials

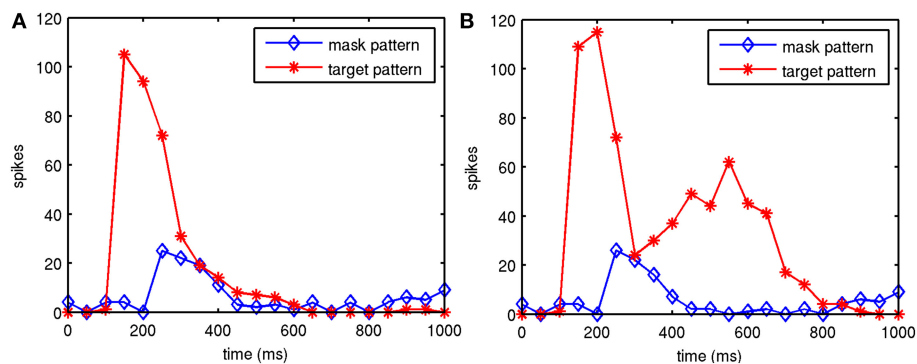


FIGURE 5 | Example of model 1 spiking activity in V1 layer 2/3 during two trials. Shown are the target attractor and noise mask during backward masking trials, one with and one without feedback connections. Feedback activity reinforces and sustains the target

attractor in the presence of the mask. Target stimulation starts at 100 ms, the SOA was 100 ms and the spikes were summed in 50 ms bins. **(A)** Without feedback from V2 to V1. **(B)** With feedback from V2 to V1.

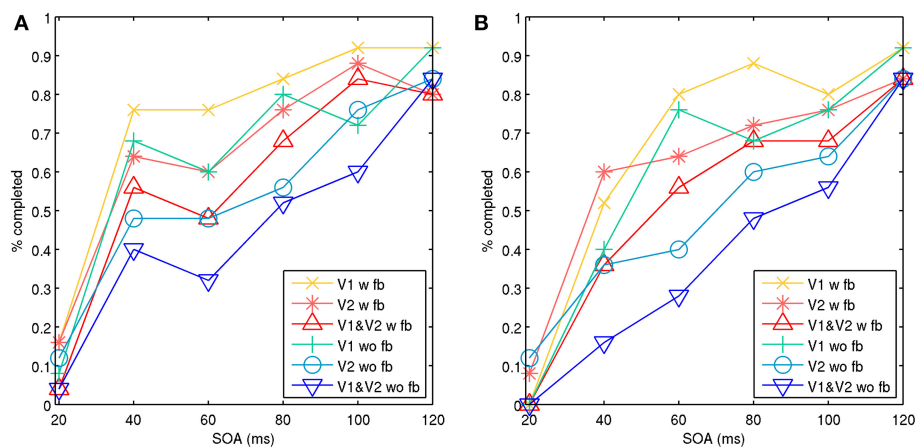


FIGURE 6 | Model 1 with noise masks showing aggregate percentage of targets completed in V1, V2, and both V1 and V2 with an increasing SOA. The target and noise mask stimulation occurred in the same hypercolumn. Activity is shown with feedback connections (w fb) from V2 to V1 and without feedback connections (wo fb). The behavior represents type-A and type-B masking effects. Activation of both V1 and V2 represents

activation across visual areas, which is assumed to be necessary for signal propagation up the visual stream to achieve conscious perception. **(A)** Illustrates aspects of a type-B masking with stimulation of four points for the target and noise mask. **(B)** Illustrates aspects of type-A masking with stimulation of four points for the target and five points for the noise mask, representing a higher masking salience.

with feedback connections and 25 trials without. This demonstrated aspects of a type-B masking, as well as type-A masking at a higher noise salience, achieved by increasing the number of stimulated minicolumns in the noise mask from 4 to 5. With the presence of feedback connections, the masking effect was significantly reduced. The feedback connections also appeared to aid target pattern completion, and made the target attractor more stable. With the model 1 configuration and simulation assumptions, this shows that both lateral inhibition and recurrent feedback are factors in perception during metacontrast backward masking.

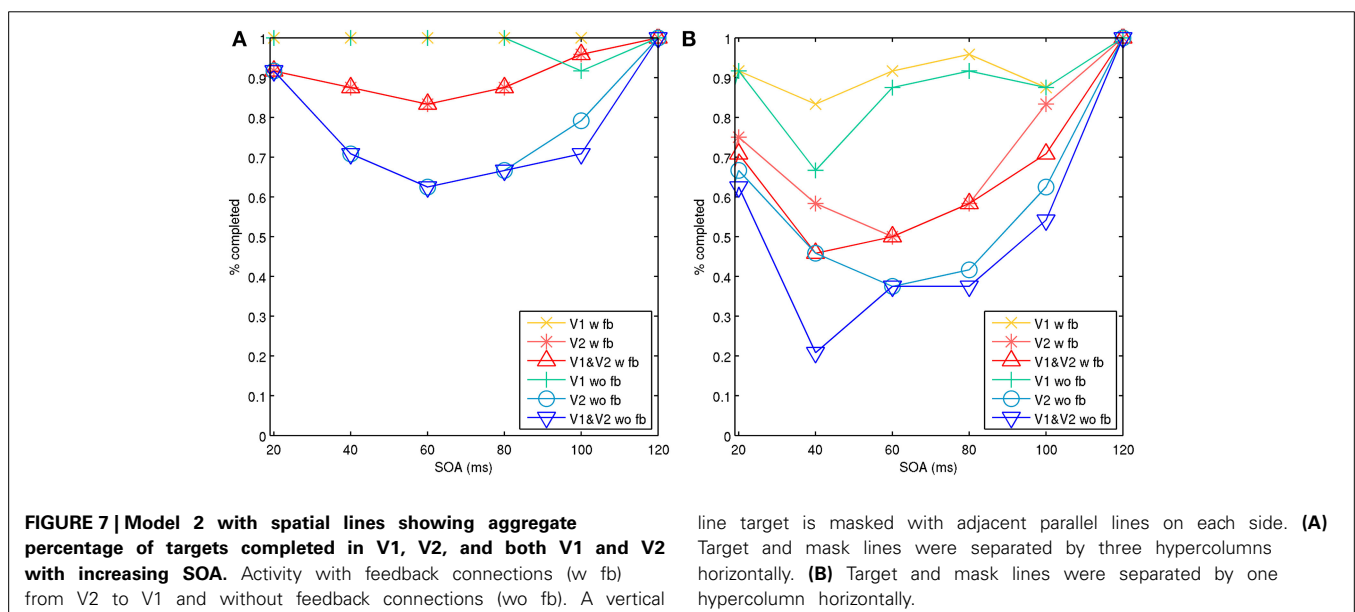
The model 2 configuration with the spatial line representations exhibited a type-B masking behavior or U-shaped function. Results can be seen in **Figure 7**, which shows simulations with the target and mask separated by a spatial distance of 1 and 3 hypercolumns. Results for each were aggregated across two sets of 25 trials, one with and one without feedback projections. The masking effects decreased with spatial distance, similar to psychophysical findings in Growney et al. (1977). Regional activity in layer 2/3 across both V1 and V2 produced type-B masking, as did analyzing activity in V2 alone. Feedback also diminished the masking effect on V2 alone, likely from boosted and recurrent feedforward activity from V1. However, when analyzing activity in V1 alone, activity appeared more monotonic when feedback is present.

DISCUSSION

The simulations showed that lateral, feedforward and feedback activity within V1 and V2 are all factors in activating and recognizing target patterns, in the presence of masks. Feedforward with feedback activity can also provide target reinforcement before lateral processing completes. This suggests that feedback processing reduces masking effects and correspondingly that masking effects may increase without the presence of feedback projections. This process of iterative reinforcement may occur among pairs of areas along the ventral stream. For example, V1 and V4 are also recurrently connected, and because of longer projection lengths,

likely provides feedback with longer latencies. However, should higher level feature detectors be trained through experience or expectation to activate or reinforce an alternative lower representation, feedback interference could cause masking effects to increase on partial or ambiguous target stimuli. There is ongoing debate on the role of feedback processing on observed properties of backward masking (Di Lollo et al., 2002; Francis and Hermens, 2002; Pöder et al., 2014), with object substitution in particular. The results here suggest there is a role, which might be more highlighted by contrasting expected sparse target recognition with ambiguous or conflicting (either primed or trained) higher level representations. On object substitution as defined by Di Lollo et al. (2000), feedback interference from larger set-sizes and distractors could be computationally explored with extensions to the existing model, by biasing or weakly stimulating higher-level attractor memories.

This study utilized a biophysical model, with equations for representing neural and synaptic properties, as well as microcircuits and network connectivity, from which characterized backward masking behaviors might emerge. Previous work has defined quantitative mathematical descriptions of backward masking behaviors from the top down. Quantitative mathematical methods known as efficient masking, mask blocking and target blocking have been described by Francis (2000) to account for type-B masking effects in metacontrast masking. Efficient masking refers to greater efficiency when masking at later SOAs when the target stimulus is weaker. The presented model did capture aspects of this behavior, because as the target attractor adapted through calcium dynamics and synaptic depression, lateral inhibition from the mask was more efficient at suppressing it. Mask blocking occurs if the target signal can block a weaker masking signal. This was observed as well, particularly at short SOAs. It may also have been a contributing factor to a sometimes observed target strength increase at an SOA of 40 ms, as seen in **Figure 6A**. A stimulated minicolumn in a target attractor memory is itself a small-world network and mini-attractor, which is



more resilient to inhibition during stimulation and early activation. This resilience could be one explanation for sometimes observed higher target visibility during common-onset masking (Enns and Di Lollo, 2000), because the effective inhibition from the mask during target stimulation is lost, reducing the effective mask exposure length. Target blocking occurs when the mask is so strong, that the target signal cannot produce a percept. In the models, this can occur when lateral inhibition is high enough that not even the minicolumns can become recurrently active. Without active minicolumns, patch-level attractors cannot activate and complete.

Among computational models for backward masking, The BCS (Francis, 1997) and Bugmann and Taylor (2005) also used detailed neural representations. The BCS represents a complex hierarchy of feature detectors along the visual ventral stream with abstract non-spiking neurons, representing functional classifications of cells, including simple cells with two orientation preferences, as well as complex and hypercomplex cells. The BCS has been able to reproduce a broad range of psychophysical phenomena, including backward masking. It also has recurrent feedback and resonance with erosion, which may be a more abstract representation of distributed neural attractors and associated adaptation and dwell times. The spiking neural attractor model presented here is at a lower level of abstraction, representing various neural types, with functional activity and microcircuits determined by cell behavior and distributed synaptic connectivity. Functionally, it represents V1 and V2 and cannot yet reproduce the same level of behaviors as the BCS can. However, it likely has closer correspondence to spiking activity observed in electrophysiological studies of early visual cortex. It also has the potential of representing a large number of feature detectors for complex spatial representations, by scaling up the number of neurons and training the feature detectors as sparse, distributed neural codes. Bugmann and Taylor (2005) also developed a neural model for backward masking composed of a 5-level hierarchy of integrate-and-fire pyramidal cells. After initial stimulation of the LGN, each level extending across V1 and V2 received feedforward input. It did not have inhibitory neurons or feedback except for self-connections at the highest level, but was able to reproduce a U-shaped behavior response under some conditions, using this simplified model.

More biophysically detailed models can provide some unique advantages. They can allow for the exploration of some neural effects and relationships which cannot be easily investigated in electrophysiology experiments. The role of microcircuits in behavior can be investigated, as well as the effects of psychotropic drugs. For instance, the existence of synaptic channels in the models could enable the simulation of drug effects such as benzodiazepine on backward masking. Benzodiazepines have been found to slow down cortical processing and extend the attentional blink and other visual processing, both experimentally (Giersch and Herzog, 2004) and in computation models (Silverstein and Lansner, 2011). Thus, it could be predicted that benzodiazepines and other GABA agonists, which slow down cortical processing and feedback, would also increase the temporal window and SOA lengths when backward masking occurs. They may also amplify the depth of the masking function in type-B masking.

However, biophysically detailed neural models such as presented here have limitations and require a considerable amount of assumptions. These models can be very computationally intensive and may require parameter tuning. While some neural network parameters can be obtained from the literature, not all are well characterized, but the expectation is that biological plausibility constrains the hypotheses and parameter values enough that some evidence is gained on how the neural circuits might work. Some model assumptions were necessary, due to the limited electrophysiological data on primates and humans. In particular, the conductance strengths and ratios of excitatory and inhibitory feedforward and especially feedback projections are not well understood yet. This could be investigated further by computationally by varying the conductance strengths and excitatory/inhibitory ratios of these projections and observing changes in the masking function. Cell, synaptic and microcircuit parameters defined in the Appendix are based on experimental electrophysiology, but are simplified. In the models presented, not all neocortical layers and projections were represented in V1 and V2. Layer 6 was not implemented. Nor were there feedback connections between layer 2/3 and layer 5. In addition, because there were no areas represented downstream of V2, layer 5 of V2 did not have higher level feedback. To compensate for this, V2 layer 4 to layer 5 and V2 layer 4 to layer 2/3 conductance was boosted to provide a higher activity level. But regardless, the recurrent feedback did reduce the effects of backward masking, by making full target attractor activation more likely. A competing mask was also used with model 1, with slightly different results. At low SOAs, the target was usually quiesced as well, but at higher SOAs it was likely that both the target and competing mask would become active. However, target activity would be truncated after the competing metacontrast mask became active. This could be investigated further, as well as the effects of masks with partially overlapping features with the target. Such masks might have the effect of diminishing the masking effect, because the target attractor would receive more stimulation.

One weakness of the existing models is the limited spatial representation of feature detectors. Including biophysical feature detectors for various orientation preferences and contours is a challenging problem and an area for future work. Model 2 included spatial representations for lines as a step towards that goal. Extensions of the line representations may be applicable for computational investigations of the shine-through paradigm as discussed earlier, which is primarily based on the use of vertical lines. Model 1 often produced type-A masking, perhaps because the metacontrast noise mask was strong and in close spatial proximity. However, when the mask was weaker, it did sometimes produce aspects of type-B masking as well. Model 2 produced type-B masking under more parameter regimes, which may have occurred because the stimulated mask minicolumns were spatially farther away than in the model 1 configuration and therefore the lateral inhibition was weaker. When observing activity in V1 independently, type-A masking was more often produced. Yet, observing V2 alone more often demonstrated type-B masking behavior, as did co-activation of both V1 and V2. This may indicate type-B masking is a property of propagating attractor activity between V1 and V2. If so, stronger masks as used in model 1

may cause type-A masking overall because V1 is more strongly affected, causing highly diminished feedforward activity for propagation to V2. Weaker masks may allow more complex dynamics between V1 and V2, resulting in the emergence of type-B masking. Part of the U-shaped function may have occurred because of activity in layer 4, where the memory pattern long-range connections are weaker due to reduced lateral connectivity. This meant that activated minicolumns in layer 4 had shorter dwell times, and were more vulnerable sooner when the mask was presented.

It was also observed that lags in the inhibitory responses from the target and mask presentation during short SOAs can affect target salience. Adding a 3 ms delay on basket to pyramidal synapses made target pattern completion at short SOAs more likely. Lags in inhibitory populations can occur, because interneurons such as martinotti cells have facilitating synapses (Krishnamurthy et al., 2012) and gap junctions in basket cells can leak excitatory potentials to other basket cells. This might be a factor in common-onset masking (Enns and Di Lollo, 2000), since inhibitory populations can be largely silent before the common-onset stimulus.

ACKNOWLEDGMENTS

This work was supported by grants from the Swedish Foundation for Strategic Research (through the Stockholm Brain Institute; www.stockholmbrain.se), the Swedish Science Council (Vetenskapsrådet, VR-621-2004-3807; www.vr.se) and the Swedish e-Science Research Centre (SeRC). Neural simulations were performed at the PDC Center for High Performance Computing at KTH. The author wishes to thank the reviewers and editor for the useful comments.

REFERENCES

- Albrecht, T., and Mattler, U. (2012). Individual differences in subjective experience and objective performance in metacontrast masking. *J. Vis.* 12, 5. doi: 10.1167/12.5.5
- Amit, D. (1989). *Modeling Brain Function: The World of Attractor Neural Networks*. Cambridge: University Press.
- Breitmeyer, B. G., and Ganz, L. (1976). Implications of sustained and transient channels for theories of visual pattern masking, saccadic suppression, and information processing. *Psychol. Rev.* 83, 1–36.
- Breitmeyer, B. G., and Ogmen, H. (2000). Recent models and findings in visual backward masking: a comparison, review, and update. *Percept. Psychophys.* 62, 1572–1595. doi: 10.3758/BF03212157
- Breitmeyer, B. G., and Ogmen, H. (2006). *Visual Masking: Time Slices Through Conscious and Unconscious Vision*. Oxford: Oxford University Press. doi: 10.1093/acprof:oso/9780198530671.001.0001
- Bugmann, G., and Taylor, J. (2005). A model of visual backward masking. *Biosystems* 79, 151–158. doi: 10.1016/j.biosystems.2004.09.015
- Carnevale, N. T., and Hines, M. L. (2006). *The Neuron Book*. Cambridge, UK: Cambridge University Press. doi: 10.1017/CBO9780511541612
- Dehaene, S., and Changeux, J. P. (2011). Experimental and theoretical approaches to conscious processing. *Neuron* 70, 200–227. doi: 10.1016/j.neuron.2011.03.018
- Di Lollo, V. (1980). Temporal integration in visual memory. *J. Exp. Psychol. Gen.* 109, 75–97. doi: 10.1037/0096-3445.109.1.75
- Di Lollo, V., Enns, J. T., and Rensink, R. A. (2000). Competition for consciousness among visual events: the psychophysics of reentrant visual processes. *J. Exp. Psychol. Gen.* 129, 481–507. doi: 10.1037/0096-3445.129.4.481
- Di Lollo, V., Enns, J. T., and Rensink, R. A. (2002). Object substitution without reentry? *J. Exp. Psychol. Gen.* 131, 594–596. doi: 10.1037/0096-3445.131.4.594
- Ekeberg, Ö., Wallén, P., Lansner, A., Träven, H., Brodin, L., and Grillner, S. (1991). A computer based model for realistic simulations of neural networks. *Biol. Cybern.* 65, 81–90. doi: 10.1007/BF00202382
- Enns, J., and Di Lollo, V. (2000). What's new in visual masking? *Trends Cogn. Sci.* 4, 345–352. doi: 10.1016/S1364-6613(00)01520-5
- Federer, F., Williams, D., Ichida, J., Merlin, S., and Angelucci, A. (2013). Two projection streams from macaque V1 to the pale cytochrome oxidase stripes of V2. *J. Neurosci.* 33, 11530–11539. doi: 10.1523/JNEUROSCI.5053-12.2013
- Francis, G. (1997). Cortical dynamics of lateral inhibition: metacontrast masking. *Psychol. Rev.* 104, 572–594. doi: 10.1037/0033-295X.104.3.572
- Francis, G. (2000). Quantitative theories of metacontrast masking. *Psychol. Rev.* 107, 768–785. doi: 10.1037/0033-295X.107.4.768
- Francis, G. (2009). Cortical dynamics of figure-ground segmentation: shine-through. *Vision Res.* 49, 140–163. doi: 10.1016/j.visres.2008.10.002
- Francis, G., and Hermens, F. (2002). Comment on “Competition for consciousness among visual events: the psychophysics of reentrant visual processes” (Di Lollo, Enns, & Rensink 2000). *J. Exp. Psychol. Gen.* 131, 590–593. doi: 10.1037/0096-3445.131.4.590
- Fransén, E., and Lansner, A. (1998). A model of cortical associative memory based on a horizontal network of connected columns. *Network* 9, 235–264. doi: 10.1088/0954-898X/9/2/006
- Giersch, A., and Herzog, M. (2004). Lorazepam strongly prolongs visual information processing. *Neuropsychopharmacology* 29, 1386–1394. doi: 10.1038/sj.npp.1300429
- Girard, P., Hupe, J., and Bullier, J. (2001). Feedforward and feedback connections between areas V1 and V2 of the monkey have similar rapid conduction velocities. *J. Neurophysiol.* 85, 1328–1331.
- Grossberg, S. (1997). Cortical dynamics of three-dimensional figure-ground perception of two-dimensional pictures. *Psychol. Rev.* 104, 618–658. doi: 10.1037/0033-295X.104.3.618
- Grossberg, S., and Mingolla, E. (1985). Neural dynamics of perceptual grouping: textures, boundaries, and emergent segmentations. *Percept. Psychophys.* 38, 141–171. doi: 10.3758/BF03198851
- Growney, R., Weisstein, N., and Cox, S. (1977). Metacontrast as a function of spatial separation with narrow line targets and masks. *Vision Res.* 17, 1205–1210. doi: 10.1016/0042-6989(77)90155-9
- Hermens, F., Luksys, G., Gerstner, W., Herzog, M., and Ernst, U. (2008). Modeling spatial and temporal aspects of visual backward masking. *Psychol. Rev.* 115, 83–100. doi: 10.1037/0033-295X.115.1.83
- Hertz, J., Krogh, A., and Palmer, R. G. (1991). *Introduction to the Theory of Neural Computation*. Redwood City, CA: Addison-Wesley.
- Herzog, M. H., Ernst, U. A., Eitzold, A., and Eurich, C. W. (2003). Local interactions in neural network explain global effects in gestalt processing and masking. *Neural Comput.* 15, 2091–2113. doi: 10.1162/08997660322297304
- Herzog, M. H., Fahle, M., and Koch, C. (2001). Spatial aspects of object formation revealed by a new illusion, shine-through. *Vision Res.* 41, 2325–2335. doi: 10.1016/S0042-6989(01)00122-5
- Hodgkin, A. L., and Huxley, A. F. (1952). A quantitative description of membrane current and its application to conduction and excitation in nerve. *J. Physiol.* 117, 500–544. doi: 10.1113/jphysiol.1952.sp004764
- Hopfield, J. J. (1982). Neural networks and physical systems with emergent collective computational abilities. *Proc. Natl. Acad. Sci. U.S.A.* 79, 2554–2558. doi: 10.1073/pnas.79.8.2554
- Kravits, D., Kadharbatcha, S., Baker, C., Ungerleider, L., and Mishkin, M. (2013). The ventral visual pathway: an expanded neural framework for the processing of object quality. *Trends Cogn. Sci.* 17, 26–49. doi: 10.1016/j.tics.2012.10.011
- Krishnamurthy, P., Silberberg, G., and Lansner, A. (2012). A cortical attractor network with martinotti cells driven by facilitating synapses. *PLoS ONE* 7:e30752. doi: 10.1371/journal.pone.0030752
- Lamme, V. A., and Roelfsema, P. (2000). The distinct models of vision offered by feedforward and recurrent processing. *Trends Neurosci.* 23, 571–579. doi: 10.1016/S0166-2236(00)01657-X
- Lamme, V. A., Zipser, K., and Spekreijse, H. (2002). Masking interrupts figure-ground signals in V1. *J. Cogn. Neurosci.* 14, 1044–1053. doi: 10.1162/089892902320474490
- Lundqvist, M., Rehn, M., Djurfeldt, M., and Lansner, A. (2006). Attractor dynamics in a modular network model of neocortex. *Network* 17, 253–276. doi: 10.1080/09548980600774619
- Lytton, W. (1996). Optimizing synaptic conductance calculation for network simulations. *Neural Comput.* 8, 501–509. doi: 10.1162/neco.1996.8.3.501

- Macknick, S. (2006). Visual masking approaches to visual awareness. *Prog. Brain Res.* 155, 177–215. doi: 10.1016/S0079-6123(06)55011-3
- Macknick, S., and Martinez-Conde, S. (2007). The role of feedback in visual masking and visual processing. *Adv. Cogn. Psychol.* 3, 125–152. doi: 10.2478/v10053-008-0020-5
- Nowak, L., and Bullier, J. (1997). The timing of information transfer in the visual system. *Cereb. Cortex* 12, 205–241. doi: 10.1007/978-1-4757-9625-4_5
- Ogmen, H. (1993). A neural theory of retino-cortical dynamics. *Neural Netw.* 6, 245–273. doi: 10.1016/0893-6080(93)90020-W
- Pöder, E. (2014). The changing picture of object substitution masking: reply to Di Lollo. *Front. Psychol.* 5:1004. doi: 10.3389/fpsyg.2014.01004
- Rehn, M., Silverstein, D., Olmäs, L., and Lansner, A. (2011). A hybrid model of the primary visual cortex. *BMC Neurosci.* 12(Suppl. 1):P184. doi: 10.1186/1471-2202-12-S1-P184
- Rolls, E. T. (2004). Consciousness absent and present: a neurophysiological exploration. *Prog. Brain Res.* 144, 95–106. doi: 10.1016/S0079-6123(03)14406-8
- Rüter, J., Francis, G., Frehe, P., and Herzog, M. (2011). Testing dynamical models of vision. *Vision Res.* 51, 343–351. doi: 10.1016/j.visres.2010.11.013
- Sato, T., Nauhaus, I., and Carandini, M. (2012). Traveling waves in visual cortex. *Neuron* 75, 218–219. doi: 10.1016/j.neuron.2012.06.029
- Shao, Z., and Burkhalter, A. (1996). Different balance of excitation and inhibition in forward and feedback circuits of rat visual cortex. *J. Neurosci.* 16, 7353–7365.
- Silverstein, D., and Lansner, A. (2011). Is attentional blink a byproduct of neocortical attractors? *Front. Comput. Neurosci.* 5:13. doi: 10.3389/fncom.2011.00013
- Sincich, L., and Horton, J. (2005). The circuitry of V1 and V2: integration of color, form and motion. *Annu. Rev. Neurosci.* 28, 303–326. doi: 10.1146/annurev.neuro.28.061604.135731
- Sperling, G. (1963). A model for visual memory tasks. *Hum. Factors* 5, 19–31.
- Sugihara, T., Qui, F., and von der Heydt, R. (2011). The speed of context integration in the visual cortex. *J. Neurophysiol.* 106, 374–385. doi: 10.1152/jn.00928.2010
- Super, H., Spekreijse, H., and Lamme, V. (2001). Two distinct models of sensory processing observed in monkey primary visual cortex (V1). *Nat. Neurosci.* 4, 304–310. doi: 10.1038/85170
- Szalisznyo, K., Silverstein, D., Duffau, H., and Smits, A. (2013). Pathological neural attractor dynamics in slowly growing gliomas supports an optimal time frame for white matter plasticity. *PLoS ONE* 8:e69798. doi: 10.1371/journal.pone.0069798
- Ts'O, D., Zarella, M., and Burkitt, G. (2009). Wither the hypercolumn? *J. Physiol.* 587, 2791–2805. doi: 10.1113/jphysiol.2009.171082
- Varela, J., Sen, K., Gibson, J., Fost, J., Abbott, L., and Nelson, S. (1997). A qualitative description of short-term plasticity at excitatory synapses in layer 2/3 of rat primary visual cortex. *J. Neurosci.* 17, 7926–7940.
- Wyatte, D., Jilk, D., and O'Reilly, R. (2014). Early recurrent feedback facilitates visual object recognition under challenging conditions. *Front. Psychol.* 5:674. doi: 10.3389/fpsyg.2014.00674

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.

Received: 08 July 2014; accepted: 05 January 2015; published online: 24 February 2015.

Citation: Silverstein DN (2015) A computational investigation of feedforward and feedback processing in metacontrast backward masking. *Front. Psychol.* 6:6. doi: 10.3389/fpsyg.2015.00006

This article was submitted to Perception Science, a section of the journal *Frontiers in Psychology*.

Copyright © 2015 Silverstein. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

APPENDIX

CELL MODELS

The single cell models were described previously in Silverstein and Lansner (2011), where the implementation of the Hodgkin Huxley formalism (1952) was based on Ekeberg et al. (1991). With the membrane potential V and the Nernst potential E_i for $i \in \{Na, K, Ca, K_{Ca}\}$ and given Ohm's law: $I_i = g_i(V - E_i)$ combined with Kirchoff's laws, yields:

$$I_m = C_m \frac{dV}{dt} + g_{Na}(V, t)(V - E_{Na}) + g_K(V, t)(V - E_K) + g_{Ca}(V, t)(V - E_{Ca}) + g_{K_{Ca}}(V, t)(V - E_{K_{Ca}}) + g_L(V - E_L)$$

where g_L is a constant leak conductance. The dynamic conductance $g_i(V, t)$ can be expressed with a gating model for individual ion channels. For modeling the for Na^+ and K^+ ion channel dynamics, Hodgkin and Huxley framework was employed.

$$I_m = C_m \frac{dV}{dt} + \bar{g}_{Na} m^3 h (V - E_{Na}) + \bar{g}_K n^4 (V - E_K) + g_L(V - E_L)$$

where \bar{g}_i with $i \in \{Na, K\}$ is the maximal conductance when a channel is open, and gating variable m is Na^+ channel activation, n is K^+ channel activation h and is Na^+ channel inactivation. The gating variables can be expressed as the following differential equations:

$$\frac{dm}{dt} = \alpha_m(1 - m) - \beta_m m \quad \text{with} \quad \alpha_m = \frac{A(V - B)}{1 - e^{(B-V)/C}}$$

$$\text{and} \quad \beta_m = \frac{A(B - V)}{1 - e^{(V-B)/C}}$$

$$\frac{dh}{dt} = \alpha_h(1 - h) - \beta_h h \quad \text{with} \quad \alpha_h = \frac{A(B - V)}{1 - e^{(V-B)/C}}$$

$$\text{and} \quad \beta_h = \frac{A}{1 + e^{(B-V)/C}}$$

$$\frac{dn}{dt} = \alpha_n(1 - n) - \beta_n n \quad \text{with} \quad \alpha_n = \frac{A(V - B)}{1 - e^{(B-V)/C}}$$

$$\text{and} \quad \beta_n = \frac{A(B - V)}{1 - e^{(V-B)/C}}$$

A , B and C are parameters and independently specified for α and β of each channel. Ca^{2+} is treated differently, because Ca^{2+} pools are assumed to be inside the cell near the cell membrane and can activate Ca^{2+} gated K^+ channels to achieve hyperpolarization. Using q to represent Ca^{2+} activation, a relation similar to the Na^+ channel activation (m) holds:

$$\frac{dq}{dt} = \alpha_q(1 - q) - \beta_q q \quad \text{with} \quad \alpha_q = \frac{A(V - B)}{1 - e^{(B-V)/C}}$$

$$\text{and} \quad \beta_q = \frac{A(B - V)}{1 - e^{(V-B)/C}}$$

with the Ca^{2+} current into the cell being $I_{Ca} = \bar{g}_{Ca} q^5 (V - E_{Ca})$. Channel equation parameters used in the simulations are specified in Table A1.

If we denote Ca^{2+} entering the cell as entering the Ca_{AP} pool, then the change in concentration of $[Ca_{AP}]$ is equivalent to the rate of ions entering the pool and less the ions leaving the pool:

$$\frac{d[Ca_{AP}]}{dt} = \varphi_{AP} q^5 (V - E_{Ca}) - \delta_{AP} [Ca_{AP}],$$

where φ_{AP} is the rate of Ca^{2+} influx and δ_{AP} is the rate of decay. The concentration $[Ca_{AP}]$ will activate Ca^{2+} gated K^+ channels inside the cell membrane with the following current:

$$I_{K_{Ca}} = \bar{g}_{K_{Ca}} (V - E_K) [Ca_{AP}]$$

Table A1 | Hodgkin-Huxley and NMDA ion channel parameters based on equations from Ekeberg et al. (1991) and values from Fransén and Lansner (1998).

	Na^+ M	Na^+ H	K^+ N	Ca^{2+} q	NMDA p
α A ($mV^{-1} ms^{-1}$)	0.58	0.232	0.058	0.232	2.03 ms^{-1}
B (mV)	-50	-50	-50	10	-
C (mV)	1	1	0.8	11	17
β A ($mV^{-1} ms^{-1}$)	0.174	1.16 (ms^{-1})	0.0145	0.0029	0.029 ms^{-1}
B (mV)	-59	-46	-40	10	-
C (mV)	20	2	0.4	0.5	17

Table A2 | Neuron parameters.

Parameter	Pyramidal	Basket	RSNP	Unit
E_{leak}	-64	-65	-65	mV
E_{Na}	50	50	50	mV
E_{Ca}	150	150	150	mV
E_K	-80	-80	-80	mV
$E_{Ca(NMDA)}$	20	20	20	mV
C_m	0.01	0.01	0.01	$\mu F/mm^2$
g_m	0.74	0.44	0.74	$\mu S/mm^2$
g_{Na} soma	150	150	150	$\mu S/mm^2$
g_K soma	250 \pm 2%	1000 \pm 2%	1000 \pm 2%	$\mu S/mm^2$
g_{Na} initial segment	2500	2500	2500	$\mu S/mm^2$
g_K initial segment	83	5010	5010	$\mu S/mm^2$
Ca_{hh} influx rate	1.00	1.00	1.00	$mV^{-1} ms^{-1} mm^{-2}$
Ca_{hh} decay rate	6.3	9	30	s^{-1}
Ca_{NMDA} influx rate	3.0	-	0.01	$s^{-1} mV^{-1} \mu S^{-1}$
Ca_{NMDA} decay rate (soma)	1	-	3	s^{-1}
Ca_{NMDA} decay rate (dend)	2	-	-	s^{-1}
g_K (Ca)	2.9	0.15	0.29	pS / mM
soma diameter \pm stdev	21 \pm 2.1	7 \pm 0.7	7 \pm 0.7	μm
Total compartments	4	3	3	
Dendritic area (rel. soma)	4	4	4	
Initial seg area (rel. soma)	0.1	0.1	0.1	

Table A3 | Parameters for synaptic dynamics.

Pre-Post	Type	$C_{dur} sec$	$\tau_{nraise} sec$	$\tau_{decay} sec$	$E_{rev} mV$	D_{fast}	D_{slow}	$\tau_{rec} ms$	$\tau_{rec} ms$	$E_{slow} mV$
Pyr-Pyr	Kainate/AMPA	0.0	0.0	0.006	0	0.78	0.98	634	9200	–
Pyr-Pyr	NMDA	0.02	0.005	0.150	0	0.78	–	634	–	0.020
Pyr-Basket	Kainate/AMPA	0.0	0.0	0.006	0	0.78	0.98	634	9200	–
Basket-Pyr	GABA _A	0.0	0.0	0.006	–85	0.94	–	1900	–	–
Pyr-Rsnp	Kainate/AMPA	0.0	0.0	0.006	0	0.75	0.98	575	9200	–
Pyr-Rsnp	NMDA	0.02	0.005	0.150	0	0.75	–	575	–	0.020
Noise	Kainate/AMPA	0.0	0.0	0.01	0	–	–	–	–	–

Table A4 | V1 and V2 neuron counts.

layer	Neuron	No. in minicolumn	No. in model 1 hypercolumn	No. in model 2 hypercolumn	Model 1 total	Model 2 total
L 2/3	pyramidal	20	320	160	5120	12960
L 2/3	basket	–	32	16	512	1296
L 2/3	rsnp	2	32	16	512	1296
L 4	pyramidal	20	320	160	5120	12960
L 4	lg. basket	–	32	16	512	1296
L 4	sm. basket	2	32	16	512	1296
L 5	pyramidal	20	320	160	5120	12960
L 5	basket	–	32	16	512	1296
L5	rsnp	2	32	16	512	1296

After an increased neural firing rate, calcium buildup in the cell will cause hyperpolarization and a reduction in the firing rate. **Table A2** specifies neuron parameters and calcium dynamics. The $[Ca_{AP}]$ pool flux rates originate from either Ca^{2+} membrane channels (hh) or NMDA channels (NMDA).

SYNAPTIC EQUATIONS

For implementing the synaptic coupling, neurotransmitter gated ionotropic synapses were modeled, where the channels conduct ionic current produced by a voltage driving force and channel conductance. AMPA and GABA_A currents are governed by:

$$I_{syn} = (E_{syn} - V) G_{syn} s \quad 0 \leq s \leq 1$$

Where s is the level of synaptic activation, with 1 being the most active. All synapses are consolidating and saturating as defined by Lytton (1996) and depressing as defined by Varela et al. (1997). Every synaptic spike results in neurotransmitter release for duration C_{dur} when it binds to receptors with binding rate α and unbinding rate β . Saturation occurs because any spike following another spike by less than C_{dur} extends neurotransmitter release for another C_{dur} interval. W_{sum} is the sum of all synaptic weights currently active within C_{dur} . After each spike and during C_{dur} , W_{sum} is incremented by the synaptic weight W_{syn} and after C_{dur} , W_{sum} is decremented by W_{syn} . Consolidation occurs by summing across synaptic activations into state variables R_{on} and R_{off} , which have the following dynamics:

$$\frac{dR_{on}}{dt} = \frac{W_{sum} R_{inf} - R_{on}}{R_{tau}} \quad \frac{dR_{off}}{dt} = -\beta R_{off} \quad R_{inf} = \frac{\alpha}{\alpha + \beta}$$

The consolidated level of synaptic activation is represented by $s = R_{on} + R_{off}$. For synaptic depression, W_{syn} is decreased during C_{dur} according to recent short-term pre-synaptic activity with: $W_{syn} = W_{syn} d_{fast} d_{slow}$, where depression variable $d_i = d_i D_i$ after a spike occurs, which then decays to 1 with $d_i = 1 - (1 - d_i) e^{-t/\tau_i}$. NMDA synapses are similar to AMPA and GABA_A but with additional dynamics for the Mg^{2+} block. Parameters for synaptic dynamics are specified in **Table A3**.

$$I_{NMDA} = (E_{NMDA} - V) G_{NMDA} p s \quad 0 \leq s \leq 1 \quad 0 \leq p \leq 1$$

Where p is the voltage gated variable for the Mg^{2+} block with the following dynamics:

$$\frac{dp}{dt} = \alpha_p (1 - p) - \beta_p p \quad \text{with} \quad \alpha_p = A \alpha e^{\frac{V}{C}} \quad \beta_p = A e^{-V/C}$$

The parameters A and C are independently specified in **Table A1** for α and β of channel p . All neurons but the LGN relay cells receive noise input from an excitatory synapse driven by a 300 Hz Poisson process. The pyramidal cell has the noise synapse on the apical dendrite and the basket and RSNP cells have it on the basal dendrite. The noise synapse is identical to the AMPA synapse but without synaptic depression, and a decay time constant of 10 ms.

NETWORK MODEL

The network architecture is organized into interconnected patches for LGN, V1 and V2, with some differences between models 1 and 2. The fixed neuron counts for patches V1 and V2 can be found in **Table A4**. Within patches V1 and V2, individual minicolumns span across layers 2/3, 4, and 5. Local populations of pyramidal cells in each of these layers are interconnected with local populations in the other layers, with the pre-synaptic to post-synaptic connection probabilities listed in **Table A5**. These probabilities were partially determined by tuning for plausible attractor activity levels in individual layers when stimulated with targets.

Between patches V1 and V2 are feedforward and feedback memory pattern projections, which can be excitatory or inhibitory. Excitatory projections connect a subset of minicolumns within two individual attractor memories across two

regions. Inhibitory projections connect one attractor memory to other attractor memories in common hypercolumns that also receive an excitatory projection from the originating attractor memory, potentially inhibiting these other attractor memories through di-synaptic inhibition. The expected synaptic counts in projections between pairs of attractor memories and opposing attractor memories (within common hypercolumns) are listed in **Table A6**. The excitatory synaptic counts were tuned to provide plausible activity transfer from stimulated attractor memories. The inhibitory synaptic counts were generally assumed to be about 20% of excitatory synaptic counts for feedforward projections and about 40–60% of excitatory synaptic counts for feedback projections.

Table A5 | Synaptic connection probabilities between pyramidal cells in the different layers within the minicolumns.

src. layer	dst. layer	conn. prob. model 1 (%)	conn. prob. model 2 (%)
V1L4	V1L2/3	13	8
V1L4	V1L5	26	16
V1L5	V1L2/3	6	3
V2L4	V2L2/3	13	12
V2L4	V2L5	34	21
V2L5	V2L2/3	6	3

Table A6 | Expected approximate feedforward (ff) and feedback (fb) projection synapse counts between attractor memories.

proj. src.	proj. dst.	dir.	exc. syn. model 1	inh. syn. model 1	exc. syn. model 2	inh. syn. model 2
V1L4	V2L4	ff	560	112	800	160
V1L2/3	V2L2/3	ff	100	20	100	20
V2L2/3	V1L2/3	fb	32	16	32	14
V2L4	V1L2/3	fb	32	16	32	14
V2L4	V1L5	fb	256	128	256	102
V2L4	V2L5	fb	512	256	512	205

Includes both excitatory (exc) synapses on pyramidal cells and inhibitory (inh) synapses on RSNP interneurons. The V2L4 to V2L5 projection partially compensates for the lack of feedback to V2 from higher areas.



Contributions of cortical feedback to sensory processing in primary visual cortex

Lucy S. Petro*, Luca Vizioli and Lars Muckli*

Centre for Cognitive Neuroimaging, Institute of Neuroscience and Psychology, University of Glasgow, Glasgow, UK

Edited by:

Hulusi Kafaligonul, Bilkent University, Turkey

Reviewed by:

Peter Kok, Radboud University Nijmegen, Netherlands
Jonathan Nassi, Salk Institute for Biological Studies, USA

*Correspondence:

Lucy S. Petro and Lars Muckli, Centre for Cognitive Neuroimaging, Institute of Neuroscience and Psychology, University of Glasgow, 58 Hillhead Street, Glasgow G12 8QB, Scotland, UK
e-mail: lucy.petro@glasgow.ac.uk;
lars.muckli@glasgow.ac.uk

Closing the structure-function divide is more challenging in the brain than in any other organ (Lichtman and Denk, 2011). For example, in early visual cortex, feedback projections to V1 can be quantified (e.g., Budd, 1998) but the understanding of feedback function is comparatively rudimentary (Muckli and Petro, 2013). Focusing on the function of feedback, we discuss how textbook descriptions mask the complexity of V1 responses, and how feedback and local activity reflects not only sensory processing but internal brain states.

Keywords: V1, feedback, fMRI, vision, electrophysiology

IS V1 (SOMETIMES) AT THE TOP OF THE HIERARCHY?

The era of Mountcastle, Hubel and Wiesel had “profound physiological implications” for the study of cortical processing (see Kandel, 2014). Hubel and Wiesel (1959) characterized the response properties of visual cortical neurons in columns: V1 neurons respond to their selective stimulus (e.g., a line of a certain orientation), and are embedded in a cortical architecture that exposes a functional map of columnar orientation preference and ocular dominance. These milestone findings furnished the (still current) textbook accounts of V1, which are dedicated to the feedforward cascade of processing and biased to neuronal spiking as recorded in electrophysiology. However, owing to increasingly sophisticated methodologies to assess functional responses, such as high-resolution magnetic resonance imaging or optogenetics combined with electrophysiology, this feedforward model of V1 can be updated to incorporate the rich response properties conferred by cortical feedback.

Neurons in early visual areas do not act as linear feature detectors when faced with complex inputs such as natural scenes, emphasizing the contribution of response modulation beyond the classical receptive field (Kayser et al., 2004). For example, non-linear receptive field models using natural stimuli predict V1 activity more optimally than a model fit using grating stimuli (David et al., 2004); V1 responses to bars embedded in a natural scene are reduced compared to bars on a uniform background (MacEvoy et al., 2008); and during natural scene viewing, the surround, local field potential (LFP) and spike history contribute to V1 spiking almost as much as the classical receptive field (Haslinger et al., 2012). Furthermore, V1 neurons are active even during occlusion (Sugita, 1999; Lee and Nguyen, 2001), revealing that non-stimulus-driven inputs allow early neurons to respond even to stimuli which are inferred but not directly presented to the retina. Early visual neurons therefore do not only transform retinal

signals, but integrate top-down and lateral inputs, which convey prediction, memory, attention, reward, task, expectation, locomotion, learning, and behavioral context. Such higher processing is fed back (monosynaptically or otherwise) to V1 from cortical and subcortical sources (Muckli and Petro, 2013). Understanding the function of feedback has implications not only for vision, but for structural and dynamic networks for cognition and behavior (Harris and Mrsic-Flogel, 2013). Indeed, Gilbert and Li (2013) suggest that each cortical neuron is a “microcosm of the brain as a whole, with synapses carrying information originating from far flung brain regions.” Top-down influences modulate feedforward (classical) receptive fields and also many of the contextual interactions performed by intrinsic V1 neurons. Here, we discuss some effects of top-down inputs to V1, culminating in the tempting speculation that V1 is misplaced as merely the earliest, sensory stage of the visual cortical hierarchy.

NON-GENICULATE INPUT TO V1 – INTERNAL PROCESSING

Aptly, the visual brain is classically studied by presenting it with visual stimuli, revealing extrinsically driven receptive fields in V1. However: (1) sensory areas are neither monomodal (e.g. Vetter et al., 2014) nor immune to higher processes; (2) feedback and lateral inputs outnumber feedforward inputs and (3) the brain is now more commonly referred to as a parallel rather than serial processor (Singer, 2013). Much can therefore be learned about intrinsically driven “response fields” in V1 (Muckli, 2010), and there is abundant evidence that V1 is involved in processing distinct from the classical feedforward activation that defines its position as the first cortical stage of vision. The reciprocal nature of the visual system suggests that in fact, in an inversion of sensory processing, visual scenes can be back-projected to V1 (Harth et al., 1987). If so, intuitively this “internal vision” would be accessible in V1 during sleep or mental processing, i.e., when

there is no feedforward input. It is possible to study internal processing by examining V1 in the absence of feedforward activation, such as in visual occlusion (Smith and Muckli, 2010) or illusion (Lee and Nguyen, 2001; Muckli et al., 2005; Murray et al., 2006; Weigelt et al., 2007; Maus et al., 2010; Kok and de Lange, 2014), in the blind (e.g., Amedi et al., 2004), blindfolded (Vetter et al., 2014) or sleeping (Horikawa et al., 2013), and during working memory (Harrison and Tong, 2009), imagery (Albers et al., 2013) and expectation (Kok et al., 2014). During eyes-closed, resting state functional magnetic resonance imaging (fMRI), hyperactive V1 has been observed in individuals with posttraumatic stress disorder who score highly on scales for re-experiencing (Zhu et al., 2014). In addition to feedback from higher visual areas, such as during occlusion or illusion, top-down influences signal behavioral context so that V1 neurons respond adaptively to the functional state of the brain (Gilbert and Li, 2013). We discuss higher processing that can be read out in V1, and suggest that not only is V1 activity linked to higher vision, but to brain states such as attention or expectation (that are determined by network interactions, Park and Friston, 2013) and tasks (Petro et al., 2013).

PREDICTION

A great deal is established about which external inputs make visual neurons spike. In contrast, less is known about the inputs which do not directly signal feedforward information transmission. One eminent theory is that feedback is actively involved in the analysis of feedforward signaling. Feedback may perform hypothesis-testing by transmitting Bayesian priors generated from memory or internal models down the visual hierarchy (e.g. Lee and Mumford, 2003). For example, one candidate mechanism for perceptual inference is that of predictive coding, in which descending predictions arising from deep pyramidal cells are compared to incoming sensory signals, and the computed mismatch (prediction error) is transferred in the feedforward stream of the superficial pyramidal cells up to the next higher cortical level to update internal models (reviewed in detail Friston, 2005; Clark, 2013). Several models in which neurons engage in probabilistic processing in order to infer the causes of their inputs have been proposed (e.g., Rao and Ballard, 1999; George and Hawkins, 2009; Lochmann and Deneve, 2011; Dura-Bernal et al., 2012), posing a challenge to feedforward theories of vision. The role of internal models in mediating predictive processing has been suggested by data from ferret V1, where, over development, spontaneous activity becomes increasingly similar to the activation induced by natural scenes (Berkles et al., 2011). This indicates that intrinsic (spontaneous) activity is akin to the responses that were previously experienced. Furthermore, when the visual flow of grating stimuli is selectively de-coupled to the rate at which a mouse runs on a ball, neurons in layer II/III of V1 signal the mismatch between actual visual flow feedback and that predicted by locomotion (Keller et al., 2012), which could be the putative error signal in V1. Visual evoked potentials in mouse V1 have been shown to be specific to previously learned spatiotemporal sequences of grating stimuli, and are even predictive of individual sequence elements during omissions (Gavornik and Bear, 2014). Furthermore, there are experimental observations indicating cortical

prediction in human V1. Using fMRI, Alink et al. (2010), measured a reduction in blood oxygen level dependent (BOLD) signal to spatiotemporally predictable stimulation. This reduction is consistent with the suppression of predictable inputs in lower levels by feedback from higher areas (in this instance, V5; Vetter et al., 2013). Such observations are tailored to the assumptions made by predictive coding, and it is known that the hemodynamic signal is sensitive to top-down afferents to V1 (Logothetis, 2008; Muckli, 2010). However, relating theoretical models with empirical data will require more invasive strategies. Techniques such as optogenetic fMRI (ofMRI, which permits the study of neuronal function whilst measuring brain activity, Lee et al., 2010) promise to shed light on how to extrapolate from the macroscopic level of the BOLD signal to the microscopic level of neurons prescribed in the predictive coding framework, during testable visual stimulation.

Theories of cortical prediction are elegant, biologically conceivable and mathematically valid, however, they remain data-modest in early visual cortex. We identify at least two key areas that require substantiation: (1) How are predictions and errors implemented by V1 neurons? Models of prediction are constrained by anatomy (cortical laminae, feedback/feedforward projections, cell subtype, e.g. local GABAergic inhibitory interneurons and long-range glutamatergic excitatory neurons, and synaptic physiology), but it remains theoretical to what extent or how V1 neurons implement prediction in their ion channels, membrane voltage, and synapses (see Fiorillo, 2008). Furthermore, (2) how does the abstract language spoken by higher areas translate to the detailed language of V1 neurons? V1 projects upwards a fine-grained representation, which becomes increasingly invariant as it advances the hierarchy, but it is unclear how abstract representations are transmitted back down the hierarchy. If feedback contains probabilities or predictions of sensory inputs, and V1 assimilates these with the actual sensory inputs, then V1 is best conceptualized as an interactive hierarchical loop and not as a “first pass analysis” (Lee et al., 1998). How sensory inputs, which signal detail, are combined with internal templates, which may signal predicted means or variances of sensory details, needs to be tested further. A candidate for the integration of feedforward and feedback signals is back-propagation-activated calcium signaling (BAC; Larkum, 2013). The anatomical substrate of this “BAC” mechanism is the layer I tuft dendrites of the pyramidal cells which reside in layer V. Vast feedback inputs arrive to these tuft dendrites, triggering Ca^{2+} spikes proximal to the apical dendrites. The consequence of these dendritic Ca^{2+} spikes is that feedback inputs may dictate the firing of the pyramidal neuron far more than was previously thought. Via this Ca^{2+} spiking mechanism, the response to feedforward somatic input (or sensory signals) is strengthened if it matches the contextual inputs or internal predictions to the tuft dendrites, e.g., it can convert a single somatic output spike into a 10 ms burst containing 2–4 spikes. The discovery of this associative mechanism illuminates one “crowning mystery” of cortex, that is, layer I (Hubel, 1982).

Observations of BAC firing impose constraints on models of how pyramidal neurons accomplish predictive coding. During BAC signaling, the predictable information is *amplified*. However,

under rules of predictive coding, feedback acts to *suppress* activity in the preceding area of cortex. The common detail between predictive coding and BAC signalling is apparent in the laminar organization of predictive coding: deep layer 5 pyramidal neurons are the “prediction units,” the same as is described in mechanisms of BAC signaling. However, BAC signaling suggests that predictable inputs are amplified within a single neuron, whereas predictive coding may engage computations within columnar circuitry for an overall effect to silence predictable inputs to an area. Therefore, although predictive coding and BAC overlap insofar as deep pyramidal neurons signal predictions, it remains to be seen how these amplified predictions within a layer 5 neuron contribute within a column (or area) to suppress prediction error in layer 2/3 (the “prediction error units” in predictive coding) before residual errors are sent up the hierarchy. Preparations which can measure dendritic signaling will contribute to resolving this question, and more generally are an exciting prospect for future explorations of V1 neurons which receive only feedback inputs, i.e., during occlusion or expectation (prior to stimulation).

MEMORY

Given the fine-grained and retinotopic nature of V1, it is a candidate region for the maintenance of high-resolution information during working memory or reactivation during episodic memory. Spatially specific working memory representations in V1 have been demonstrated by the successful decoding of grating stimuli during a retention period in the cortical location of their original representation (Pratte and Tong, 2014). The information maintained in working memory that is represented in V1 reflects the relevance of items, and this can be causally interrupted using transcranial magnetic stimulation (TMS) (Zokaei et al., 2014). In a memory-color paradigm, successful cross-classification of V1 activity patterns between colored hues and gray scale objects associated with those hues, was interpreted as the result of the feedback of prior knowledge to V1 (Bannert and Bartels, 2013). The capacity of visual memory for object details is great (Brady et al., 2008); these details may be stored as early as V1 and reactivated by feedback, contingent on behavioral demands. The reactivation of V1 may be related to top-down influences from the hippocampus for successful memory consolidation during sleep. Firing sequences evoked during awake experience are replayed in both V1 and the hippocampus during sleep phases in the mouse (Ji and Wilson, 2007). Furthermore, human hippocampus activity covaries with early visual activity, which predicts the information that subjects retrieve from memory (Bosch et al., 2014). The hippocampus exerts top-down effects on early vision during scene extrapolation (Chadwick et al., 2013), prompting new theories of hippocampal memory whereby it constructs the “world beyond the immediate sensorium” (Maguire and Mullally, 2013). The recruitment of V1 by the hippocampus to construct the world would appear functional, given that V1 depicts the visual environment with the highest resolution.

REWARD

V1 was not classically thought to play an essential role in reward processing. However, a number of studies indicate that reward modulates the representation of features in V1. V1 neurons in

the rat have been shown to signal value (Shuler and Bear, 2006), and more recent calcium-imaging data from mouse V1 reveals that the association between stimulus and reward alters response amplitude in stimulus-specific assemblies (Goltstein et al., 2013). Neurons in macaque V1 that signal value also exhibit strong attentional effects (Stănişor et al., 2013) and future studies will clarify the role of feedback in this overlap. Cholinergic input to V1 from the basal forebrain of the rat modulates specifically the learning of reward timing, but not the expression of previously learned cue-reward intervals (Chubykin et al., 2013). In human early visual cortex, value is encoded across populations of neurons, in which response profiles are sharpened (Serences and Saproo, 2010). Anticipatory activity in V1 may in some instances be driven by dopaminergic input directly from the ventral tegmental area (Phillipson et al., 1987; Tan, 2009) or indirectly from the prefrontal cortex (Noudoost and Moore, 2011). Anticipatory haemodynamic signals in V1 are found even without feedforward stimulation (Sirotin and Das, 2009). Such baseline shifts point to the “dark matter” of the brain, that is, much can be learned from the substantial energy consumption of neurons even during resting states (Shoham et al., 2006; Raichle, 2011). In an elegant design to exclude the effects of anticipation (as well as attention and expectation), it was shown that the effects of dopaminergic reward on V1 can decrease its activity (Arsenault et al., 2013). Further experiments will elucidate if this decrease equates to a sharpened representation of rewarding stimuli, and more generally, the role of cholinergic, dopaminergic, and feedback mechanisms in reward effects in V1.

VISION FOR ACTION AND VISUAL PERCEPTION

Feedback to V1 has a role in how we perceive and interact with the visual world. For example, reciprocal feedback from parietal portions of the dorsal stream to early visual areas is likely involved in visuospatial processing, although the function of these networks remains to be fully elucidated. The dorsal stream is activated during reaching and grasping, and Ban et al. (2013) offer the thought-provoking idea that early visual cortex interacts with other sensory modalities (e.g. tactile or motor), as an implicit representation of an occluded object in visual cortex could facilitate the touching or grasping of the occluded portion of the object. In addition, top-down input, likely from auditory cortex or association areas, leads to categorical activation in early visual cortex (Vetter et al., 2014) during natural sound processing in blindfolded subjects. Such activity in visual cortex could be biased by higher areas to the feature content or localisation of content in a visual scene, or, with motor guidance, aid in visually orienting to the source of auditory signals. Motor inputs related to locomotion are sufficient to drive V1 activity; Keller et al. (2012) observed motor-related activity in mouse V1, without any visual input. V1 neurons responded when the mouse was running on a ball during complete darkness, and this activity was comparable to that evoked by visual stimulation with gratings. Further studies will clarify the involvement of cortical feedback in visuo-motor processing and the sensory guidance of movement, and the recruitment of V1 in these processes.

The ventral visual stream is concerned with detailed form representation, and the importance of feedback in the ventral stream

or “recurrent occipitotemporal network” (Kravitz et al., 2013) is linked to the retinotopic organization of V1. For example, objects presented in the periphery trigger feedback to foveal V1 where object detail is processed (Williams et al., 2008), and interrupting this feedback using TMS at a relatively late time interval impairs peripheral object perception (Chambers et al., 2013). In contrast, feedback related to scene processing is back-projected to the periphery of V1 (Smith and Muckli, 2010). A causal role of recurrent processing in the ventral stream suggests that late activation in V1 contributes to scene categorization (Koivisto et al., 2011). During face processing, feedback (putatively from temporal cortex) task-dependently biases retinotopic sub-regions of V1 responding to certain features (Petro et al., 2013). Perceptual expectation can enhance the representation of stimuli in V1 whilst at the same time suppressing V1 (Kok et al., 2012), in line with predictive coding theories of dampening predicted inputs. Within both dorsal and ventral streams, recurrent feedback loops might be critical for conscious processing (Lamme and Roelfsema, 2000; Dehaene and Changeux, 2011).

Effects of feedback to V1 on visual perception have also been studied more invasively, contributing to the mechanistic understanding of feedback. Enhanced visual discrimination is seen in awake behaving mice after optogenetically activating cholinergic neurons projecting to V1 from the basal forebrain (Pinto et al., 2013), with a probable role in attentional function. During scene processing, population codes in mouse V1 become increasingly sparse compared to viewing control scenes lacking statistical regularities (Froudarakis et al., 2014). This encoding by a smaller set of neurons only when the scenes were not phase-scrambled fits with theories of back-projected predictions suppressing feedforward processing, or could be related to microcircuits within V1. The study of Berkes et al. (2011) mentioned previously hints that the cortex utilizes a strategy of decreased processing for experienced or expected signals. In ferret V1, it was found that across development spontaneous activity begins to reflect the activity evoked by natural scenes, and therefore prior expectations. This increasing similarity between evoked and spontaneous activity reveals that the cortex updates its internal model with experience, with predictive coding theories suggesting that these internal models are used to generate predictions of sensory input, which can suppress activity at early cortical levels. Intra-areal, inhibitory interneurons can also modulate visual perception. By optogenetically targeting parvalbumin-positive interneurons in V1, Lee et al. (2012) revealed that these neurons are involved in sharpening feature selectivity and in perceptual orientation discrimination. Parvalbumin neurons are targeted by feedback, and top-down connections putatively control inhibitory activity as dictated by behavioral demands (possibly by inhibiting predictable inputs). In neuropathophysiology, *N*-methyl-D-aspartate (NMDA) receptor hypofunction on parvalbumin neurons interferes with gamma oscillations, which is linked to schizophrenia and depression (Gonzalez-Burgos and Lewis, 2012; Phillips and Silverstein, 2013). Attenuated visual illusion effects observed in schizophrenia might relate to an interruption of top-down predictions (see Notredame et al., 2014). These predictions are maintained in healthy populations who experience the illusions.

Of the aforementioned studies, the data on visual perception are conceivably related to predictions from higher cortical visual areas. This predictive processing may be temporally discernible from that of attention, and have sources in independent regions from those that allocate attention. In contrast, it is less intuitive to associate feedback with prediction during vision for action, without knowing more about the cortico-cortical connections crossing domains from motor to visual. However, prediction is assumed to be a general function of the cortex and motor actions are often highly repetitive and structured (and therefore predictable). Anatomical connections reveal that it is essential to include the contribution of subcortical pathways and the cerebellum in predictive feedback during sensorimotor processing. For example, the cerebellum is involved in generating predictions of the sensory consequences of actions (Kawato and Wolpert, 1998), which may also be represented in V1. The cerebellum is also involved in predictions of perception (Roth et al., 2013), suggesting that, like cortex, the cerebellum's role in prediction is unspecific to any one processing domain. It remains unknown how V1 and the cerebellum interact during perception, and what role feedback has in this processing.

A NEW LANDSCAPE OF V1

For several years, neuroscience has yielded abundant data on the intricate workings of V1. Yet, this unique cortical area remains, in many ways, a mystery. The gain of modern experimentation is that, with advancing imaging and recording techniques, we can understand the (cellular) mechanics of V1. The reward for venturing “under the hood,” will be to learn if theoretical concepts of cortical feedback can be realized in corresponding biological substrates. Hence, decades after the revolutionary work of Hubel and Wiesel (1959), there are continued efforts to understand V1. For example, it was found that in the macaque, the majority of feedback to V1 arises from V2, where axons arborize in supragranular layers I and II, and infragranular layer V (Rockland and Virga, 1989), and more recently it has been shown that these axons fed back from V2 to V1 differ in their bouton morphology (axons forming bouton clusters or studded continuously with boutons in layer I, or forming en passant boutons in layers III and V) and postsynaptic density size (Anderson and Martin, 2009). Rodent models of V1 lend themselves to innovative invasive approaches, allowing counter-current visual processing streams to be studied on the cellular level. For example, using subcellular Channelrhodopsin-2-assisted circuit mapping and patch clamp recordings, Yang et al. (2013) showed that depolarizing feedback input is balanced between parvalbumin interneurons and pyramidal neurons in layer II/III of mouse V1. This balance is in contrast to feedforward pathways (which provide substantially more depolarizing input to layer II/III parvalbumin neurons than to excitatory pyramidal cells) and therefore has implications for pathway-specific excitation/inhibition. In mouse layer V, feedback input to tuft dendrites leads to NMDA spikes which (by supporting calcium spikes, which are “tremendously explosive”) are thought to be critical for the integration of top-down inputs in cortex (Larkum et al., 2009; Larkum, 2013). Furthermore, it becomes increasingly clear that we must “reach beyond the classical receptive field” (Angelucci and Bullier, 2003) because feedback inputs bestow the full range

of center and surround receptive field properties to V1 neurons (in combination with feedforward and lateral inputs; Angelucci and Bressloff, 2006). For example, in the macaque, surround suppression in V1 is reduced when feedback is eliminated (Nassi et al., 2013), feedback augments V1 responses to collinear contours in the owl monkey (Shmuel et al., 2005), and during pattern motion processing, feedback modulates subthreshold influences beyond the classical receptive field, facilitating global constructs from local features represented in V1 (Schmidt et al., 2011). Moreover, the spatiotemporal receptive fields in layer II/III of macaque V1 may be best characterized by their intracortical inputs and not by their visual inputs (Yeh et al., 2009).

Markov et al. (2014) suggest that actually we have only a rudimentary understanding of connectional rules of feedback and feedforward projections. Indeed, there are thought to be two systems of feedback and feedforward projections: supragranular and infragranular (Markov and Kennedy, 2013), and future investigations will enlighten how these pathways constrain models of cortical information processing in more detail (e.g., Bastos et al., 2012). Studies of neuronal synchrony suggest gamma-band phase coherence is restricted to supragranular and beta-band to infragranular layers (see Buffalo et al., 2011; Xing et al., 2012). Top-down input in the beta or alpha band to deep layers modulates gamma activity (associated with bottom-up processing) in more superficial layers (see Spaak et al., 2012; Bastos et al., 2014, bioRxiv). In monkey V1, distinct multiunit profiles in layers corresponding to feedforward and feedback processing can be seen during the perception of figure-ground segregation (Self et al., 2013). Laminar analysis of V1 in humans, using fMRI with multivoxel pattern analysis (MVPA), shows that contextual feedback arrives to the superficial layers of cortex (Muckli, OHBM conference abstract, 2014). MVPA scrutinizes information in the multivariate pattern of activity across an array of voxels, to discriminate between stimuli or states that are potentially neglected by conventional analysis involving spatial averaging (Kriegeskorte and Bandettini, 2008). The spectrally symmetric encoding models (Gourtzelidis et al., 2005; Thirion et al., 2006; Dumoulin and Wandell, 2008; Jerde et al., 2008; Kay et al., 2008; Mitchell et al., 2008; Naselaris et al., 2009; Schönwiesner and Zatorre, 2009) can explicitly quantify the information contained in individual voxels (Naselaris et al., 2011), thus providing insights about the preferred features coded by a given voxel. Aiding in our understanding of visual cortex, these techniques have thus far been optimized within the feedforward framework. With a clearer understanding of their advantages and limitations, these approaches combined with layer-resolution fMRI have the potential to unveil a wealth of information about the functional role of feedback activity in V1 (Muckli, OHBM conference abstract, 2014; Morgan et al., OHBM conference abstract, 2014). Layer-resolution fMRI can also be combined with pharmacological intervention to assess laminar differences during tasks that are dependent on top-down processing. For example, texture discrimination, which relies on recurrent processing, is impaired subsequent to ketamine administration (Meuwese et al., 2013). Ketamine blocks NMDA receptors which are implicated in feedback processing due to their higher concentration in supragranular layers (Rosier et al., 1993), their modulatory function (Collingridge and Bliss, 1987)

and their contribution to figure-ground segregation (Self et al., 2012).

CONCLUSION

V1 is one of the best studied cortical areas in terms of its robust stimulus-response relationship. This fine-grained, feedforward propagation of the visual world is V1's principle function. However, increasing evidence reveals a more complex and comprehensive account of V1: through intrinsic and feedback connections, V1 neurons are also capable of complex visual (scene analysis) and non-visual (cognitive) responses. One function of feedback may be to flexibly "set the system" according to present behavioral requirements, i.e., distribute top-down influences even to the earliest sensory areas. We have highlighted some higher-order processes that can be read-out from V1. During active vision, feedback may transmit Bayesian inferences of forthcoming inputs to V1, to facilitate perception. Feedback may also sharpen the representation of rewarding stimuli in V1. During sleep, V1 might be involved in the higher-order replay of experienced events, for memory consolidation. With growing capabilities to study the brain at molecular, cellular, systems, behavioral and cognitive levels, one hopes that future developments will clarify the role of V1 neurons as adaptive responders, and elucidate how internal brain states regulate sensory processing in V1.

ACKNOWLEDGMENTS

We thank Bill Philips and Philippe Schyns for insightful discussions. We also thank the European Research Council for generous support.

REFERENCES

- Albers, A. M., Kok, P., Toni, I., Dijkerman, H. C., and de Lange, F. P. (2013). Shared representations for working memory and mental imagery in early visual cortex. *Curr. Biol.* 23, 1427–1431. doi: 10.1016/j.cub.2013.05.065
- Alink, A., Schwiedrzik, C. M., Kohler, A., Singer, W., and Muckli, L. (2010). Stimulus predictability reduces responses in primary visual cortex. *J. Neurosci.* 30, 2960–2966. doi: 10.1523/JNEUROSCI.3730-10.2010
- Amedi, A., Floel, A., Knecht, S., Zohary, E., and Cohen, L. G. (2004). Transcranial magnetic stimulation of the occipital pole interferes with verbal processing in blind subjects. *Nat. Neurosci.* 7, 1266–1270. doi: 10.1038/nn1328
- Anderson, J. C., and Martin, K. A. (2009). The synaptic connections between cortical areas V1 and V2 in macaque monkey. *J. Neurosci.* 29, 11283–11293. doi: 10.1523/JNEUROSCI.5757-08.2009
- Angelucci, A., and Bressloff, P. C. (2006). Contribution of feedforward, lateral and feedback connections to the classical receptive field center and extra-classical receptive field surround of primate V1 neurons. *Prog. Brain Res.* 154, 93–120. doi: 10.1016/S0079-6123(06)54005-1
- Angelucci, A., and Bullier, J. (2003). Reaching beyond the classical receptive field of V1 neurons: horizontal or feedback axons? *J. Physiol. Paris* 97, 141–154. doi: 10.1016/j.jphysparis.2003.09.001
- Arsenault, J. T., Nelissen, K., Jarraya, B., and Vanduffel, W. (2013). Dopaminergic reward signals selectively decrease fMRI activity in primate visual cortex. *Neuron* 77, 1174–1186. doi: 10.1016/j.neuron.2013.01.008
- Ban, H., Yamamoto, H., Hanakawa, T., Urayama, S., Aso, T., Fukuyama, H., et al. (2013). Topographic representation of an occluded object and the effects of spatiotemporal context in human early visual areas. *J. Neurosci.* 33, 16992–17007. doi: 10.1523/JNEUROSCI.1455-12.2013
- Bannert, M. M., and Bartels, A. (2013). Decoding the yellow of a gray banana. *Curr. Biol.* 23, 2268–2272. doi: 10.1016/j.cub.2013.09.016
- Bastos, A. M., Usrey, W. M., Adams, R. A., Mangun, G. R., Fries, P., and Friston, K. J. (2012). Canonical microcircuits for predictive coding. *Neuron* 76, 695–711. doi: 10.1016/j.neuron.2012.10.038

- Bastos, A. M., Vezoli, J., Bosman, C. A., Schoffelen, J.-M., Oostenveld, R., Dowdall, J. R., et al. (2014). *Visual Areas Exert Feedforward and Feedback Influences Through Distinct Frequency Channels*. Available at: <http://biorxiv.org/content/early/2014/05/06/004804>
- Berkes, P., Orbán, G., Lengyel, M., and Fiser, J. (2011). Spontaneous cortical activity reveals hallmarks of an optimal internal model of the environment. *Science* 331, 83–87. doi: 10.1126/science.1195870
- Bosch, S. E., Jehee, J. F., Fernández, G., and Doeller, C. F. (2014). Reinstatement of associative memories in early visual cortex is signaled by the hippocampus. *J. Neurosci.* 34, 7493–7500. doi: 10.1523/JNEUROSCI.0805-14.2014
- Brady, T. F., Konkle, T., Alvarez, G. A., and Oliva, A. (2008). Visual long-term memory has a massive storage capacity for object details. *Proc. Natl. Acad. Sci. U.S.A.* 105, 14325–14329. doi: 10.1073/pnas.0803390105
- Budd, J. M. (1998). Extrastriate feedback to primary visual cortex in primates: a quantitative analysis of connectivity. *Proc. Biol. Sci.* 265, 1037–1044. doi: 10.1098/rspb.1998.0396
- Buffalo, E. A., Fries, P., Landman, R., Buschman, T. J., and Desimone, R. (2011). Laminar differences in gamma and alpha coherence in the ventral stream. *Proc. Natl. Acad. Sci. U.S.A.* 108, 11262–11267. doi: 10.1073/pnas.1011284108
- Chadwick, M. J., Mullally, S. L., and Maguire, E. A. (2013). The hippocampus extrapolates beyond the view in scenes: an fMRI study of boundary extension. *Cortex* 49, 2067–2079. doi: 10.1016/j.cortex.2012.11.010
- Chambers, C. D., Allen, C. P., Maizy, L., and Williams, M. A. (2013). Is delayed foveal feedback critical for extra-foveal perception? *Cortex* 49, 327–335. doi: 10.1016/j.cortex.2012.03.007
- Chubykin, A. A., Roach, E. B., Bear, M. F., and Shuler, M. G. (2013). A cholinergic mechanism for reward timing within primary visual cortex. *Neuron* 77, 723–735. doi: 10.1016/j.neuron.2012.12.039
- Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behav. Brain Sci.* 36, 181–204. doi: 10.1017/S0140525X12000477
- Collingridge, G. L., and Bliss, T. V. P. (1987). NMDA receptors – their role in long-term potentiation. *Trends Neurosci.* 10, 288–293. doi: 10.1016/0166-2236(87)90175-5
- David, S. V., Vinje, W. E., and Gallant, J. L. (2004). Natural stimulus statistics alter the receptive field structure of V1 neurons. *J. Neurosci.* 24, 6991–7006. doi: 10.1523/JNEUROSCI.1422-04.2004
- Dehaene, S., and Changeux, J. P. (2011). Experimental and theoretical approaches to conscious processing. *Neuron* 70, 200–227. doi: 10.1016/j.neuron.2011.03.018
- Dumoulin, S. O., and Wandell, B. A. (2008). Population receptive field estimates in human visual cortex. *Neuroimage* 39, 647–660. doi: 10.1016/j.neuroimage.2007.09.034
- Dura-Bernal, S., Wennekers, T., and Denham, S. L. (2012). Top-down feedback in an HMAX-like cortical model of object perception based on hierarchical Bayesian networks and belief propagation. *PLoS ONE* 7:e48216. doi: 10.1371/journal.pone.0048216
- Fiorillo, C. D. (2008). Towards a general theory of neural computation based on prediction by single neurons. *PLoS ONE* 3:e3298. doi: 10.1371/journal.pone.003298
- Friston, K. (2005). A theory of cortical responses. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 360, 815–836. doi: 10.1098/rstb.2005.1622
- Froudarakis, E., Berens, P., Ecker, A. S., Cotton, R. J., Sinz, F. H., Yatsenko, D., et al. (2014). Population code in mouse V1 facilitates readout of natural scenes through increased sparseness. *Nat. Neurosci.* 17, 851–857. doi: 10.1038/nn.3707
- Gavornik, J. P., and Bear, M. F. (2014). Learned spatiotemporal sequence recognition and prediction in primary visual cortex. *Nat. Neurosci.* 17, 732–737. doi: 10.1038/nn.3683
- George, D., and Hawkins, J. (2009). Towards a mathematical theory of cortical micro-circuits. *PLoS Comput. Biol.* 5:e1000532. doi: 10.1371/journal.pcbi.1000532
- Gilbert, C. D., and Li, W. (2013). Top-down influences on visual processing. *Nat. Rev. Neurosci.* 14, 350–363. doi: 10.1038/nrn3476
- Goltstein, P. M., Coffey, E. B., Roelfsema, P. R., and Pennartz, C. M. (2013). In vivo two-photon Ca^{2+} imaging reveals selective reward effects on stimulus-specific assemblies in mouse visual cortex. *J. Neurosci.* 33, 11540–11555. doi: 10.1523/JNEUROSCI.1341-12.2013
- Gonzalez-Burgos, G., and Lewis, D. A. (2012). NMDA receptor hypofunction, parvalbumin-positive neurons, and cortical gamma oscillations in schizophrenia. *Schizophr. Bull.* 38, 950–957. doi: 10.1093/schbul/sbs010
- Gourtzelidis, P., Tzarakakis, C., Lewis, S. M., Crowe, D. A., Auerbach, E., Jerde, T. A., et al. (2005). Mental maze solving: directional fMRI tuning and population coding in the superior parietal lobule. *Exp. Brain Res.* 165, 273–282. doi: 10.1007/s00221-005-2298-6
- Harris, K. D., and Mrsic-Flogel, T. D. (2013). Cortical connectivity and sensory coding. *Nature* 503, 51–58. doi: 10.1038/nature12654
- Harrison, S. A., and Tong, F. (2009). Decoding reveals the contents of visual working memory in early visual areas. *Nature* 458, 632–635. doi: 10.1038/nature07832
- Harth, E., Unnikrishnan, K. P., and Pandya, A. S. (1987). The inversion of sensory processing by feedback pathways: a model of visual cognitive functions. *Science* 237, 184–187. doi: 10.1126/science.3603015
- Haslinger, R., Pipa, G., Lima, B., Singer, W., Brown, E. N., and Neuenschwander, S. (2012). Context matters: the illusive simplicity of macaque V1 receptive fields. *PLoS ONE* 7:e39699. doi: 10.1371/journal.pone.0039699
- Horikawa, T., Tamaki, M., Miyawaki, Y., and Kamitani, Y. (2013). Neural decoding of visual imagery during sleep. *Science* 340, 639–642. doi: 10.1126/science.1234330
- Hubel, D. H. (1982). Cortical neurobiology: a slanted historical perspective. *Annu. Rev. Neurosci.* 5, 363–370. doi: 10.1146/annurev.ne.05.030182.002051
- Hubel, D. H., and Wiesel, T. N. (1959). Receptive fields of single neurones in the cat's striate cortex. *J. Physiol.* 148, 574–591.
- Jerde, T., Lewis, S., Goerke, U., Gourtzelidis, P., Tzarakakis, C., Lynch, J., et al. (2008). Ultra-high field parallel imaging of the superior parietal lobule during mental maze solving. *Exp. Brain Res.* 187, 551–561. doi: 10.1007/s00221-008-1318-8
- Ji, D., and Wilson, M. A. (2007). Coordinated memory replay in the visual cortex and hippocampus during sleep. *Nat. Neurosci.* 10, 100–107. doi: 10.1038/nn1825
- Kandel, E. R. (2014). Republication of The Journal of Physiology (2009) 587, 2733–2741: an introduction to the work of David Hubel and Torsten Wiesel. *J. Physiol.* 559, 2–10. doi: 10.1113/jphysiol.2013.268060_2
- Kawato, M., and Wolpert, D. (1998). Internal models for motor control. *Novartis Found. Symp.* 218, 291–304; discussion 304–307.
- Kay, K. N., Naselaris, T., Prenger, R. J., and Gallant, J. L. (2008). Identifying natural images from human brain activity. *Nature* 452, 352–355. doi: 10.1038/nature06713
- Kayser, C., Körding, K. P., and König, P. (2004). Processing of complex stimuli and natural scenes in the visual cortex. *Curr. Opin. Neurobiol.* 14, 468–473. doi: 10.1016/j.conb.2004.06.002
- Keller, G. B., Bonhoeffer, T., and Hübner, M. (2012). Sensorimotor mismatch signals in primary visual cortex of the behaving mouse. *Neuron* 74, 809–815. doi: 10.1016/j.neuron.2012.03.040
- Koivisto, M., Railo, H., Revonsuo, A., Vanni, S., and Salminen-Vaparanta, N. (2011). Recurrent processing in V1/V2 contributes to categorization of natural scenes. *J. Neurosci.* 31, 2488–2492. doi: 10.1523/JNEUROSCI.3074-10.2011
- Kok, P., and de Lange, F. P. (2014). Shape perception simultaneously up- and down-regulates neural activity in the primary visual cortex. *Curr. Biol.* 24, 1531–1535. doi: 10.1016/j.cub.2014.05.042
- Kok, P., Failing, M. F., and de Lange, F. P. (2014). Prior expectations evoke stimulus templates in the primary visual cortex. *J. Cogn. Neurosci.* 26, 1546–1554. doi: 10.1162/jocn_a_00562
- Kok, P., Jehee, J. F., and de Lange, F. P. (2012). Less is more: expectation sharpens representations in the primary visual cortex. *Neuron* 75, 265–270. doi: 10.1016/j.neuron.2012.04.034
- Kravitz, D. J., Saleem, K. S., Baker, C. I., Ungerleider, L. G., and Mishkin, M. (2013). The ventral visual pathway: an expanded neural framework for the processing of object quality. *Trends Cogn. Sci.* 17, 26–49. doi: 10.1016/j.tics.2012.10.011
- Kriegeskorte, N., and Bandettini, P. (2008). Analyzing for information, not activation, to exploit high-resolution fMRI. *Neuroimage* 38, 649–662. doi: 10.1016/j.neuroimage.2007.02.022
- Lamme, V. A., and Roelfsema, P. R. (2000). The distinct modes of vision offered by feedforward and recurrent processing. *Trends Neurosci.* 23, 571–579. doi: 10.1016/S0166-2236(00)01657-X
- Larkum, M. A. (2013). Cellular mechanism for cortical associations: an organizing principle for the cerebral cortex. *Trends Neurosci.* 36, 141–151. doi: 10.1016/j.tins.2012.11.006
- Larkum, M. E., Nevian, T., Sandler, M., Polsky, A., and Schiller, J. (2009). Synaptic integration in tuft dendrites of layer 5 pyramidal neurons: a new unifying principle. *Science* 325, 756–760. doi: 10.1126/science.1171958
- Lee, J. H., Durand, R., Gradinaru, V., Zhang, F., Goshen, I., Kim, D. S., et al. (2010). Global and local fMRI signals driven by neurons defined optogenetically by type and wiring. *Nature* 465, 788–792. doi: 10.1038/nature09108

- Lee, S. H., Kwan, A. C., Zhang, S., Phoumthipphavong, V., Flannery, J. G., Masmanidis, S. C., et al. (2012). Activation of specific interneurons improves V1 feature selectivity and visual perception. *Nature* 488, 379–383. doi: 10.1038/nature11312
- Lee, T. S., and Mumford, D. (2003). Hierarchical Bayesian inference in the visual cortex. *J. Opt. Soc. Am. A Opt. Image Sci. Vis.* 20, 1434–1448. doi: 10.1364/JOSAA.20.001434
- Lee, T. S., Mumford, D., Romero, R., and Lamme, V. A. (1998). The role of the primary visual cortex in higher level vision. *Vision Res.* 38, 2429–2454. doi: 10.1016/S0042-6989(97)00464-1
- Lee, T. S., and Nguyen, M. (2001). Dynamics of subjective contour formation in the early visual cortex. *Proc. Natl. Acad. Sci. U.S.A.* 98, 1907–1911. doi: 10.1073/pnas.98.4.1907
- Lichtman, J. W., and Denk, W. (2011). The big and the small: challenges of imaging the brain's circuits. *Science* 334, 618–623. doi: 10.1126/science.1209168
- Lochmann, T., and Deneve, S. (2011). Neural processing as causal inference. *Curr. Opin. Neurobiol.* 21, 774–781. doi: 10.1016/j.conb.2011.05.018
- Logothetis, N. K. (2008). What we can do and what we cannot do with fMRI. *Nature* 453, 869–878. doi: 10.1038/nature06976
- MacEvoy, S. P., Hanks, T. D., and Paradiso, M. A. (2008). Macaque V1 activity during natural vision: effects of natural scenes and saccades. *J. Neurophysiol.* 99, 460–472. doi: 10.1152/jn.00612.2007
- Maguire, E. A., and Mullally, S. L. (2013). The hippocampus: a manifesto for change. *J. Exp. Psychol. Gen.* 142, 1180–1189. doi: 10.1037/a0033650
- Markov, N. T., and Kennedy, H. (2013). The importance of being hierarchical. *Curr. Opin. Neurobiol.* 23, 187–194. doi: 10.1016/j.conb.2012.12.008
- Markov, N. T., Vezoli, J., Chameau, P., Falchier, A., Quilodran, R., Huissoud, C., et al. (2014). Anatomy of hierarchy: feedforward and feedback pathways in macaque visual cortex. *J. Comp. Neurol.* 522, 225–259. doi: 10.1002/cne.23458
- Maus, G. W., Weigelt, S., Nijhawan, R., and Muckli, L. (2010). Does area V3A predict positions of moving objects? *Front. Psychol.* 1:186. doi: 10.3389/fpsyg.2010.00186
- Meuwese, J. D., van Loon, A. M., Scholte, H. S., Lirk, P. B., Vulink, N. C., Hollmann, M. W., et al. (2013). NMDA receptor antagonist ketamine impairs feature integration in visual perception. *PLoS ONE* 8:e79326. doi: 10.1371/journal.pone.0079326
- Mitchell, T. M., Shinkareva, S. V., Carlson, A., Chang, K., Malave, V. L., Mason, R. A., et al. (2008). Predicting human brain activity associated with the meanings of nouns. *Science* 320, 1191–1195. doi: 10.1126/science.1152876
- Muckli, L. (2010). What are we missing here? Brain imaging evidence for higher cognitive functions in primary visual cortex V1. *IJIST* 20, 131–139. doi: 10.1002/ima.20236
- Muckli, L., Kohler, A., Kriegeskorte, N., and Singer, W. (2005). Primary visual cortex activity along the apparent-motion trace reflects illusory perception. *PLoS Biol.* 3:e265. doi: 10.1371/journal.pbio.0030265
- Muckli, L., and Petro, L. S. (2013). Network interactions: non-geniculate input to V1. *Curr. Opin. Neurobiol.* 23, 195–201. doi: 10.1016/j.conb.2013.01.020
- Murray, S. O., Boyaci, H., and Kersten, D. (2006). The representation of perceived angular size in human primary visual cortex. *Nat. Neurosci.* 9, 429–434. doi: 10.1038/nn1641
- Naselaris, T., Kay, K. N., Nishimoto, S., and Gallant, J. L. (2011). Encoding and decoding in fMRI. *Neuroimage* 56, 400–410. doi: 10.1016/j.neuroimage.2010.07.073
- Naselaris, T., Prenger, R. J., Kay, K. N., Oliver, M., and Gallant, J. L. (2009). Bayesian reconstruction of natural images from human brain activity. *Neuron* 63, 902–915. doi: 10.1016/j.neuron.2009.09.006
- Nassi, J. J., Lomber, S. G., and Born, R. T. (2013). Corticocortical feedback contributes to surround suppression in V1 of the alert primate. *J. Neurosci.* 33, 8504–8517. doi: 10.1523/JNEUROSCI.5124-12.2013
- Notredame, C.-E., Pins, D., Deneve, S., and Jardri, R. (2014). What visual illusions teach us about schizophrenia. *Front. Integr. Neurosci.* 8:63. doi: 10.3389/fnint.2014.00063
- Noudoost, B., and Moore, T. (2011). Control of visual cortical signals by prefrontal dopamine. *Nature* 474, 372–375. doi: 10.1038/nature09995
- Park, H. J., and Friston, K. (2013). Structural and functional brain networks: from connections to cognition. *Science* 342, 1238411. doi: 10.1126/science.1238411
- Petro, L. S., Smith, F. W., Schyns, P. G., and Muckli, L. (2013). Decoding face categories in diagnostic subregions of primary visual cortex. *Eur. J. Neurosci.* 37, 1130–1139. doi: 10.1111/ejn.12129
- Phillips, W. A., and Silverstein, S. M. (2013). The coherent organization of mental life depends on mechanisms for context-sensitive gain-control that are impaired in schizophrenia. *Front. Psychol.* 4:307. doi: 10.3389/fpsyg.2013.00307
- Phillipson, O. T., Kilpatrick, I. C., and Jones, M. W. (1987). Dopaminergic innervation of the primary visual cortex in the rat, and some correlations with human cortex. *Brain Res. Bull.* 18, 621–633. doi: 10.1016/0361-9230(87)90132-8
- Pinto, L., Goard, M. J., Estandian, D., Xu, M., Kwan, A. C., Lee, S. H., et al. (2013). Fast modulation of visual perception by basal forebrain cholinergic neurons. *Nat. Neurosci.* 16, 1857–1863. doi: 10.1038/nn.3552
- Pratte, M. S., and Tong, F. (2014). Spatial specificity of working memory representations in the early visual cortex. *J. Vis.* 14, 22. doi: 10.1167/14.3.22
- Raichle, M. E. (2011). The restless brain. *Brain Connect.* 1, 3–12. doi: 10.1089/brain.2011.0019
- 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. doi: 10.1038/4580
- Rockland, K. S., and Virga, A. (1989). Terminal arbors of individual “feedback” axons projecting from area V2 to V1 in the macaque monkey: a study using immunohistochemistry of anterogradely transported *Phaseolus vulgaris*-leucoagglutinin. *J. Comp. Neurol.* 285, 54–72. doi: 10.1002/cne.902850106
- Rosier, A. M., Arckens, L., Orban, G. A., and Vandesande, F. (1993). Laminar distribution of NMDA receptors in cat and monkey visual cortex visualized by [3H]-MK-801 binding. *J. Comp. Neurol.* 335, 369–380. doi: 10.1002/cne.903350307
- Roth, M. J., Synofzik, M., and Lindner, A. (2013). The cerebellum optimises perceptual predictions about external sensory events. *Curr. Biol.* 23, 930–935. doi: 10.1016/j.cub.2013.04.027
- Schmidt, K. E., Lomber, S. G., Payne, B. R., and Galuske, R. A. (2011). Pattern motion representation in primary visual cortex is mediated by transcortical feedback. *Neuroimage* 54, 474–484. doi: 10.1016/j.neuroimage.2010.08.017
- Schönwiesner, M., and Zatorre, R. J. (2009). Spectro-temporal modulation transfer function of single voxels in the human auditory cortex measured with high-resolution fMRI. *Proc. Natl. Acad. Sci. U.S.A.* 106, 14611–14616. doi: 10.1073/pnas.0907682106
- Shmuel, A., Korman, M., Sterkin, A., Harel, M., Ullman, S., Malach, R., et al. (2005). Retinotopic axis specificity and selective clustering of feedback projections from V2 to V1 in the owl monkey. *J. Neurosci.* 25, 2117–2131. doi: 10.1523/JNEUROSCI.4137-04.2005
- Self, M. W., Kooijmans, R. N., Supér, H., Lamme, V. A., and Roelfsema, P. R. (2012). Different glutamate receptors convey feedforward and recurrent processing in macaque V1. *Proc. Natl. Acad. Sci. U.S.A.* 109, 11031–11036. doi: 10.1073/pnas.1119527109
- Self, M. W., van Kerkoerle, T., Supér, H., and Roelfsema, P. R. (2013). Distinct roles of the cortical layers of area V1 in figure-ground segregation. *Curr. Biol.* 23, 2121–2129. doi: 10.1016/j.cub.2013.09.013
- Serences, J. T., and Saproo, S. (2010). Population response profiles in early visual cortex are biased in favor of more valuable stimuli. *J. Neurophysiol.* 104, 76–87. doi: 10.1152/jn.01090.2009
- Shoham, S., O'Connor, D. H., and Segev, R. (2006). How silent is the brain: is there a “dark matter” problem in neuroscience? *J. Comp. Physiol. A Neuroethol. Sens. Neural Behav. Physiol.* 192, 777–784. doi: 10.1007/s00359-006-0117-6
- Shuler, M. G., and Bear, M. F. (2006). Reward timing in the primary visual cortex. *Science* 311, 1606–1609. doi: 10.1126/science.1123513
- Singer, W. (2013). Cortical dynamics revisited. *Trends Cogn. Sci.* 17, 616–626. doi: 10.1016/j.tics.2013.09.006
- Sirotni, Y. B., and Das, A. (2009). Anticipatory haemodynamic signals in sensory cortex not predicted by local neuronal activity. *Nature* 457, 475–479. doi: 10.1038/nature07664
- Smith, F. W., and Muckli, L. (2010). Nonstimulated early visual areas carry information about surrounding context. *Proc. Natl. Acad. Sci. U.S.A.* 107, 20099–20103. doi: 10.1073/pnas.1000233107
- Spaak, E., Bonnefond, M., Maier, A., Leopold, D. A., and Jensen, O. (2012). Layer-specific entrainment of gamma-band neural activity by the alpha rhythm in monkey visual cortex. *Curr. Biol.* 22, 2313–2318. doi: 10.1016/j.cub.2012.10.020
- Stănişor, L., van der Togt, C., Pennartz, C. M., and Roelfsema, P. R. (2013). A unified selection signal for attention and reward in primary visual cortex. *Proc. Natl. Acad. Sci. U.S.A.* 110, 9136–9141. doi: 10.1073/pnas.1300117110
- Sugita, Y. (1999). Grouping of image fragments in primary visual cortex. *Nature* 401, 269–272. doi: 10.1038/45785

- Tan, C. O. (2009). Anticipatory changes in regional cerebral hemodynamics: a new role for dopamine? *J. Neurophysiol.* 101, 2738–2740. doi: 10.1152/jn.00141.2009
- Thirion, B., Duchesnay, E., Hubbard, E., Dubois, J., Poline, J. B., Lebihan, D., et al. (2006). Inverse retinotopy: inferring the visual content of images from brain activation patterns. *Neuroimage* 33, 1104–1116. doi: 10.1016/j.neuroimage.2006.06.062
- Vetter, P., Grosbras, M. H., and Muckli, L. (2013). TMS Over V5 disrupts motion prediction. *Cereb. Cortex* doi: 10.1093/cercor/bht297 [Epub ahead of print].
- Vetter, P., Smith, F. W., and Muckli, L. (2014). Decoding sound and imagery content in early visual cortex. *Curr. Biol.* 24, 1256–1262. doi: 10.1016/j.cub.2014.04.020
- Weigelt, S., Singer, W., and Muckli, L. (2007). Separate cortical stages in amodal completion revealed by functional magnetic resonance adaptation. *BMC Neurosci.* 8:70. doi: 10.1186/1471-2202-8-70
- Williams, M. A., Baker, C. I., Op de Beeck, H. P., Shim, W. M., Dang, S., Triantafyllou, C., et al. (2008). Feedback of visual object information to foveal retinotopic cortex. *Nat. Neurosci.* 11, 1439–1445. doi: 10.1038/nn.2218
- Xing, D., Yeh, C. I., Burns, S., and Shapley, R. M. (2012). Laminar analysis of visually evoked activity in the primary visual cortex. *Proc. Natl. Acad. Sci. U.S.A.* 109, 13871–13876. doi: 10.1073/pnas.1201478109
- Yang, W., Carrasquillo, Y., Hooks, B. M., Nerbonne, J. M., and Burkhalter, A. (2013). Distinct balance of excitation and inhibition in an interareal feedforward and feedback circuit of mouse visual cortex. *J. Neurosci.* 33, 17373–17384. doi: 10.1523/JNEUROSCI.2515-13.2013
- Yeh, C. I., Xing, D., Williams, P. E., and Shapley, R. M. (2009). Stimulus ensemble and cortical layer determine V1 spatial receptive fields. *Proc. Natl. Acad. Sci. U.S.A.* 106, 14652–14657. doi: 10.1073/pnas.0907406106
- Zhu, H., Zhang, J., Zhan, W., Qiu, C., Wu, R., Meng, Y., et al. (2014). Altered spontaneous neuronal activity of visual cortex and medial anterior cingulate cortex in treatment-naïve posttraumatic stress disorder. *Compr. Psychiatry* 55, 1688–1695. doi: 10.1016/j.comppsy.2014.06.009
- Zokaei, N., Manohar, S., Husain, M., and Feredoes, E. (2014). Causal evidence for a privileged working memory state in early visual cortex. *J. Neurosci.* 34, 158–162. doi: 10.1523/JNEUROSCI.2899-13.2014

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.

Received: 30 June 2014; accepted: 09 October 2014; published online: 06 November 2014.

Citation: Petro LS, Vizioli L and Muckli L (2014) Contributions of cortical feedback to sensory processing in primary visual cortex. *Front. Psychol.* 5:1223. doi: 10.3389/fpsyg.2014.01223

This article was submitted to Perception Science, a section of the journal *Frontiers in Psychology*.

Copyright © 2014 Petro, Vizioli and Muckli. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Visual crowding illustrates the inadequacy of local vs. global and feedforward vs. feedback distinctions in modeling visual perception

Aaron M. Clarke^{1*}, Michael H. Herzog¹ and Gregory Francis^{1,2}

¹ Laboratory of Psychophysics, Brain, Mind Institute, Science Vie, École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland

² Department of Psychological Sciences, Purdue University, West Lafayette, IN, USA

Edited by:

Hulusi Kafaligonul, Bilkent University, Turkey

Reviewed by:

Thomas S. A. Wallis, The University of Tübingen, Germany

Ronald Van Den Berg, University of Cambridge, UK

*Correspondence:

Aaron M. Clarke, Laboratory of Psychophysics, Brain, Mind Institute, Science Vie, École Polytechnique Fédérale de Lausanne, Station 19, CH-1015 Lausanne, Switzerland
e-mail: aaron.clarke@epfl.ch

Experimentalists tend to classify models of visual perception as being either local or global, and involving either feedforward or feedback processing. We argue that these distinctions are not as helpful as they might appear, and we illustrate these issues by analyzing models of visual crowding as an example. Recent studies have argued that crowding cannot be explained by purely local processing, but that instead, global factors such as perceptual grouping are crucial. Theories of perceptual grouping, in turn, often invoke feedback connections as a way to account for their global properties. We examined three types of crowding models that are representative of global processing models, and two of which employ feedback processing: a model based on Fourier filtering, a feedback neural network, and a specific feedback neural architecture that explicitly models perceptual grouping. Simulations demonstrate that crucial empirical findings are not accounted for by any of the models. We conclude that empirical investigations that reject a local or feedforward architecture offer almost no constraints for model construction, as there are an uncountable number of global and feedback systems. We propose that the identification of a system as being local or global and feedforward or feedback is less important than the identification of a system's computational details. Only the latter information can provide constraints on model development and promote quantitative explanations of complex phenomena.

Keywords: feed-forward, hierarchical models, feedback, object recognition, scene processing

1. INTRODUCTION

A common approach to understanding vision is to identify whether a particular aspect of visual perception involves “local” or “global” processing. Local processing suggests that the information needed for some behavioral task is determined predominantly by information that is spatially close to the target stimulus. Global processing suggests that information processing is influenced by elements that may be distant from the target. Distinguishing between visual processing as being local or global has long been an important aspect of the Gestalt approach to perception (see the review by Wagemans et al., 2012). The local vs. global distinction also plays an important role in characterizing the flow of information in visual cortex (e.g., Altmann et al., 2003) and identifying the order of processing for natural scenes (e.g., Rasche and Koch, 2002; Cesarei and Loftus, 2011).

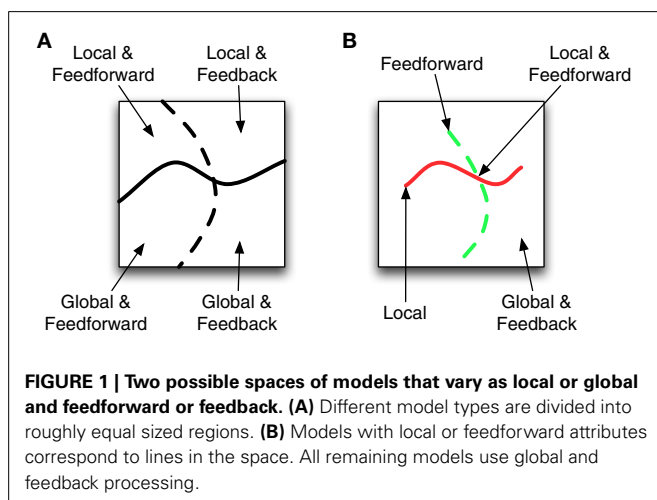
Likewise, many investigations try to identify whether visual processing involves “feedforward” or “feedback” processing. In a feedforward system the information flows in one direction, while in a feedback system the information flowing back and forth within and between areas can alter the processing at a given cortical location. In neuroanatomical studies, feedback processing is sometimes referred to as recurrent processing or re-entrant processing (especially when it involves information from higher

cortical areas projecting to lower visual areas). Since feedforward processing tends to be easier to model, interpret, and compute than feedback processing, it is often the starting point for computational and neurophysiological theories and serves as a standard comparison for subsequent studies that explore feedback effects. For example, Hubel and Wiesel (1962) proposed a local feedforward model that accounted for the properties of simple and complex cell receptive fields, and subsequent studies then proposed the existence of non-classical receptive fields by demonstrating effects of feedback or global processing (e.g., Von der Heydt et al., 1984; DeAngelis et al., 1994; Freeman et al., 2001; Harrison et al., 2007). Likewise, a popular theory of visual processing proposed that both a rapid feedforward sweep and a slower recurrent process is involved in different behavioral tasks to different degrees (Lamme and Roelfsema, 2000; Lamme, 2006), and many studies have explored whether particular phenomena depend on one or the other processing approach. Examples include Altmann et al. (2003) reporting evidence for feedback processing in an fMRI study of perceptual organization; Enns and Di Lollo (2000) arguing that some forms of visual masking require re-entrant signals that represent objects; Juan and Walsh (2003) using TMS to argue that the representation of information in area V1 is influenced by feedback from other areas; and Keil et al. (2009) reaching a

similar conclusion for emotionally arousing stimuli using an ERP study.

Experimental vision science is full of many other examples of investigations into local vs. global and feedforward vs. feedback processing, and we generally agree with their methods and conclusions. However, we are less convinced that these characterizations are especially useful for developing models of visual perception that might account for observed behavioral phenomena, and we suspect that the benefits of the local vs. global and feedforward vs. feedback dichotomies have been somewhat overstated. The seeming appeal of investigations that distinguish between local vs. global and feedforward vs. feedback processing may derive from a misunderstanding about the general properties of complex systems. **Figure 1A** schematizes one way of conceptualizing model space. The solid wavy line separates local models from global models while the dashed line separates feedforward models from feedback models. Under such a model space, identifying whether a system requires local or global processing divides the possible number of models nearly in half. Likewise identifying whether a system requires feedforward or feedback processing again divides the number of possible models in half. If the model space were as dichotomous as in **Figure 1A**, then investigations about the local vs. global or feedforward vs. feedback nature of visual processing would be very beneficial to modelers.

However, the characterization in **Figure 1A** cannot be correct because there must necessarily be fewer feedforward and local systems than feedback or global systems (e.g., every feedforward system can be augmented with multiple types of feedback), so the model space depicted in **Figure 1B** is closer to reality. Here the local models are characterized by a thin red line and the feedforward models are characterized by a thin dashed green line. The class of local and feedforward models is the small intersection of these lines, while global and feedback models correspond to almost everything else. If this perspective of the model space is correct, then scientists gain a lot of information by knowing a system uses local (Weisstein, 1968) or feedforward processing (VanRullen et al., 2001), but they gain very little information by knowing the model uses global and feedback processing.



Our argument is not that distinctions between local and global or feedforward and feedback processing provide *no* information about the properties of the visual system; but if **Figure 1B** is correct, then such distinctions will not generally provide sufficient constraints to promote model development for the identified effects. While this limitation may already be clear to many modelers, it seems that some experimentalists do not fully understand that such distinctions provide very little guidance for model development. Part of the problem is the underlying textbook assumption that there is *one* standard feedforward model and another standard feedback model, which implies that all we have to do is perform an experiment to see which type of model better describes task performance. It is indeed true that there are successful and popular feedforward and feedback models. The feedforward model of Riesenhuber and Poggio (1999), for example, has been used successfully for things like fast-feedforward object recognition or scene classification (e.g., Hung et al., 2005; Serre et al., 2005, 2007a,b; Poggio et al., 2013). Similarly, the feedback model of Grossberg (e.g., Grossberg and Mingolla, 1985) has spawned a multitude of subsequent publications (e.g., Grossberg and Todorovic, 1988; Grossberg and Rudd, 1989; Grossberg, 1990; Francis et al., 1994; Francis and Grossberg, 1995; Dresch and Grossberg, 1997; Grossberg, 2003; Grossberg and Howe, 2003; Grossberg and Yazdanbakhsh, 2003; Grossberg et al., 2011; Foley et al., 2012). Clearly there is an important role for both types of model architectures. However, the success of these models is not simply because of their feedforward or feedback architecture. Even these “popular” models involve parameter variations and additional stages from one paper to the next that make them suitable for modeling one experimental data set, but not another. Moreover, there exists a broad continuum of models that are designed to model various phenomena and include various amounts of feedforward and feedback processing, or local and global processing, and that are all different. In this sense, there is not really a “standard” model for the visual system. Even V1 receptive field models are vast and varied, including such models as Gabors (Gabor, 1946; Jones and Palmer, 1987), balanced Gabors (Cope et al., 2008, 2009), difference of Gaussians (Sceniak et al., 1999), oriented difference of Gaussians (Blakeslee and McCourt, 2004; Blakeslee et al., 2005), the log-Gaussian in the Fourier domain (Field, 1987), and many more, all of which produce similar, but distinctly different effects when applied to natural images and lab illusions. Moreover, V1 receptive fields comprise just the first step in a model of visual cortex. Thus, no “standard” models exists for either feedforward or for feedback architectures, and similarly for a local or a more global connection architecture. Simply specifying one or the other type of architecture is not helpful for many modeling projects. To demonstrate our point, we consider empirical data from studies of visual crowding that show a clear non-local effect, and that likely require feedback mechanisms to enable perceptual grouping. We then describe the properties of three plausible models: one that can be considered to be feedforward and global, one that can be considered to be feedback and global, and one that can be considered to be feedback and global with a clear interpretation of perceptual grouping. We show through computer simulations that none of these models can account for the empirical findings

that motivated them. This result suggests that we need to stop focusing on unhelpful dichotomies such as local vs. global and feedforward vs. feedback and instead should explore other properties of visual perception that help identify robust computational principles.

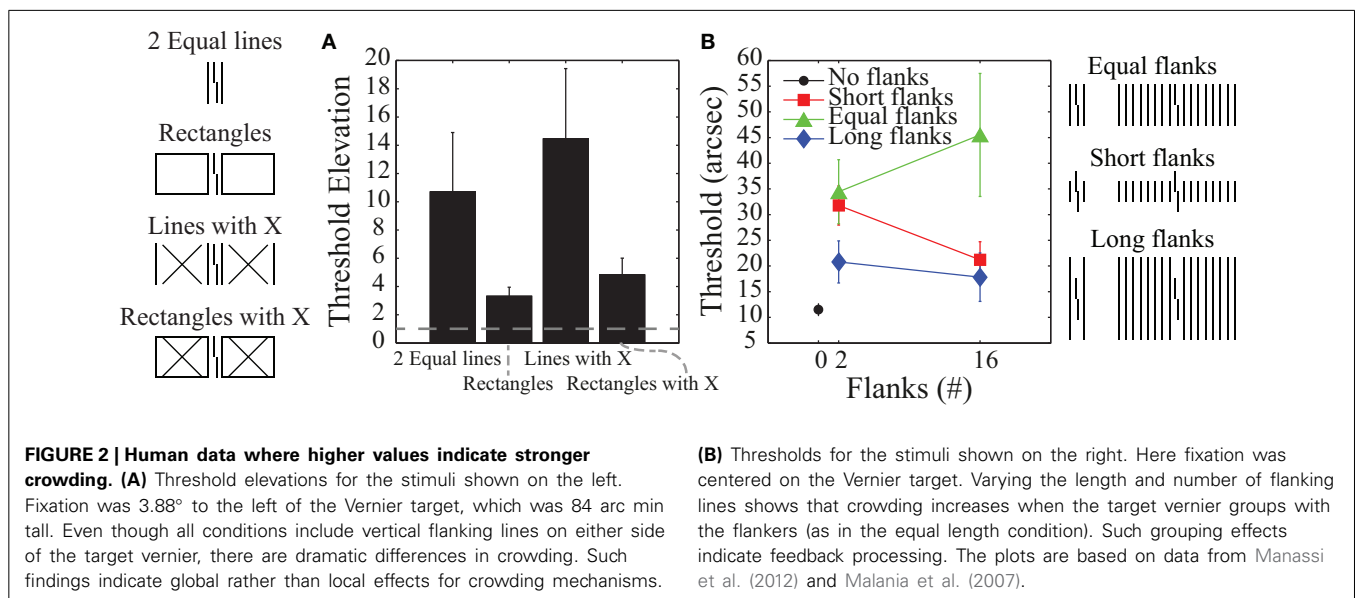
2. VISUAL CROWDING AS AN EXAMPLE

In visual crowding the discrimination of a target stimulus is impaired by the presence of neighboring elements. Crowding is ubiquitous in human environments. Even while you read these words, the letters appearing in the periphery of your visual field are crowded and largely unintelligible. Crowding can even be life-threatening in driving situations where a pedestrian can become unidentifiable by standing amongst other elements in the visual scene (Whitney and Levi, 2011). Moreover, visual crowding has been used to investigate many other aspects of perceptual and cognitive processing including visual acuity (Atkinson et al., 1988), neural competition (Keyser and Perrett, 2002), and awareness (Wallis and Bex, 2011).

The most popular models of crowding are local and feedforward models in which deteriorated target processing is due to information about the target being pooled with information about the flankers (e.g., Parkes et al., 2001). Although such pooling mechanisms are the default interpretation of crowding effects, recent studies have suggested that crowding involves global (rather than local) and feedback (rather than feedforward) processing (Malania et al., 2007; Levi and Carney, 2009; Sayim et al., 2010; Livne and Sagi, 2011; Manassi et al., 2012, 2013). **Figure 2** schematizes eleven different types of stimuli where the task is always to identify the offset direction of a central target vernier. **Figure 2A** shows human vernier offset discrimination threshold elevations (relative to a no-flanker case), where larger threshold elevations indicate more crowding (from Manassi et al., 2012). The stimuli used are depicted on the far left-hand side of the figure. In all cases, the vernier is flanked by two vertical lines whose length matches the vertical extent of the vernier. The data

in **Figure 2A** indicate that the different flanker types do not produce equivalent crowding despite the identical neighboring lines. Although the flanking lines alone or with an “X” produce substantial crowding, there is very little crowding when the very same lines are part of a larger structure. A “local” mechanism, such as pooling, would predict similar (or stronger) crowding with the additional contours in the rectangle configurations. The observed decrease in crowding suggests that the phenomenon cannot be explained by local interactions between stimuli.

Figure 2B shows human vernier thresholds (Malania et al., 2007) that have also been used to argue for feedback processing. Different experimental conditions varied the lengths of the flanking lines (shorter than, equal to, or longer than the vernier) and the number of flanking lines (0, 2, or 16). For the equal-length flankers, an increase in the number of flankers leads to stronger crowding, while for the short- and long-flanker lines, an increase in the number of flankers either reduced crowding or produced essentially no change. The argument for feedback processing has two parts. First, the data for the different conditions in **Figure 2B** suggest that crowding is strongest when the target vernier perceptually groups with the flankers (e.g., 16 equal-length flankers) and it is weakest when the target is perceptually segmented from the flankers (e.g., 16 short or long flankers). A sense of these grouping effects can be gained by looking at the schematized stimuli at the far right of **Figure 2B**. Second, perceptual grouping seems to require systems with feedback processing (e.g., Grossberg and Mingolla, 1985; Herzog et al., 2003; Craft et al., 2007; Hermens et al., 2008; Francis, 2009; Kogo et al., 2010). In particular, as Manassi et al. (2013) noted, the properties of crowding seem to defy low-level feedforward models based on stimulus energy or similar concepts (although they did not attempt to model their results). In their experiments they had subjects perform Vernier offset discrimination tasks and showed that when holding local information constant, global stimulus information still influenced thresholds. Thus, local information must have been propagated globally. Further experiments showed that



this local-to-global information propagation takes time, implying feedback and recurrent processing.

Since the crowding data in **Figure 2** indicate a role for global rather than local and feedback rather than feedforward processing, we wanted to use this knowledge to help develop a model of visual processing that accounted for crowding. Several models for crowding exist in the literature (e.g., Wilkinson et al., 1997; Balas et al., 2009; Greenwood et al., 2009; Van den Berg et al., 2010; Freeman and Simoncelli, 2011). Our intent here is not to classify these models as feedforward/feedback or local/global, and see how well they work, but rather to examine some clear examples of models using various amounts of feedforward/feedback and local/global processing and demonstrate the utility (or lack thereof) of knowing that a phenomena requires feedforward/feedback or local/global processing for modeling behavioral results. As the following sections demonstrate, we found this knowledge to be inadequate, and we believe that the modeling challenges here reflect issues that also apply to other phenomena and modeling efforts. Although it is possible that we happened to simulate models that are poor fits for the phenomena, we deliberately investigated models that have successfully modeled similar stimuli and phenomena, so we believed that they might also be able to account for the empirically observed crowding effects.

3. A FEEDFORWARD GLOBAL MODEL: FOURIER ANALYSIS

Researchers have suggested that it is useful to describe visual processing in terms of Fourier components (Campbell and Robson, 1968; De Valois et al., 1982). Luminance values at different (x, y) coordinates in the pixel plane can be converted to weights for different sine wave frequencies' amplitudes and phases. In principle, such a transformation does not lose any information, so if the luminance image contains information about the offset direction of a vernier, then so does the Fourier representation. However, such information can become degraded or lost when important frequencies or phases are filtered out of the representation. Such filtering can be justified on neurophysiological grounds (Campbell and Robson, 1968; De Valois et al., 1982) or be chosen to explain perceptual phenomena. For example, multi-scale filtering can explain a variety of brightness illusions (Blakeslee and McCourt, 1999, 2001, 2004; Blakeslee et al., 2005).

Fourier decomposition can be considered to be a feedforward process, with a bank of filters that are tuned to different frequencies, orientations, and phases (Fourier, 1822), and such an interpretation is a common first-approximation to cortical visual processing (Campbell and Robson, 1968). On the other hand, Fourier analysis is decidedly global rather than local in the sense that the weights assigned to different frequencies are based on the pattern of luminance values across the entire image plane (Rasche and Koch, 2002; Cesarei and Loftus, 2011). It is also global in the sense that a filter that suppresses some frequencies will influence representations of luminance values across the entire image plane when the frequency weights are converted back to an image representation.

We developed a model that applies a Fourier analysis to the image, filters out a subset of spatial frequencies, applies a Fourier synthesis to construct a filtered version of the image, and then compares the output with a template for discriminating right-

from left-offset verniers. The difference between template matching results for the left- and right-offset verniers are subtracted and the difference is then inverted and linearly scaled to the range of the human data. Model details are provided in the Supplementary Material. To try to match the empirical data, we examined various filtering schemes, including high-pass filtering, band-pass filtering, and low-pass filtering. **Figure 3** shows a representative selection of results for the stimuli used to produce the data in **Figure 2**. Even though they all allow for global processing, many of these frequency filtering functions produce results that differ dramatically from the human data. Within each filtering scheme we identified the filter parameters that yielded the smallest sum of squared residuals between the model and human data from **Figure 2** by exhaustive, brute-force search over the entire parameter space. **Figure 3C**, shows the best fit overall, which was obtained with a band-pass filter.

This best filter mask for the data used in **Figure 3C** does a reasonably good job at reproducing the human data from **Figure 2B**, but it does a poor job reproducing the human data shown in **Figure 2A**. Although the model roughly follows the pattern of the data for the two-line flanker and rectangle conditions, it predicts very little threshold elevation (and even threshold improvement) for the conditions with an "X" superimposed over the flanker regions. These predictions do not match the empirical data. The other filter functions also fail to reproduce the human data for these flanker conditions.

Moreover, the best filter is fragile in that small changes in bandwidth and/or center frequency lead to very different model predictions. This fragility is demonstrated in **Figure 4**, which shows model performance for band pass filters that are only slightly different from the filter that produces the best fit to the empirical data in **Figure 2B**. This behavior is surprising since Fourier models generally tend to fail gracefully with small deviations from the optimal filter parameters. The wildly varying model behavior suggests that the good fit exhibited in **Figure 3C** reflects over-fitting rather than a mechanistic explanation of the behavior. Overall, the model fails to account for the human data in a robust way. Such a failure occurs even though the model is inherently global in terms of processing, and thus satisfies one of the requirements seemingly needed to account for crowding effects. We cannot definitively claim that all Fourier-type models cannot account for crowding effects, but it seems that a good model does not easily appear simply because it has global processing.

4. A FEEDBACK MODEL: WILSON-COWAN NEURAL NETWORK

We next considered a model that derives its key properties from the recurrent nature of information processing in a cooperative-competitive neural network. Variations of this kind of model have successfully accounted for visual masking data (Hermens et al., 2008) using stimuli very similar to those in **Figure 2**. The model first convolves the input image with an on-center, off-surround receptive field mimicking processing by the LGN. Next, the input activations are fed into both an excitatory and an inhibitory layer of neurons. Each layer convolves the input activations with a Gaussian blurring function and propagates activity

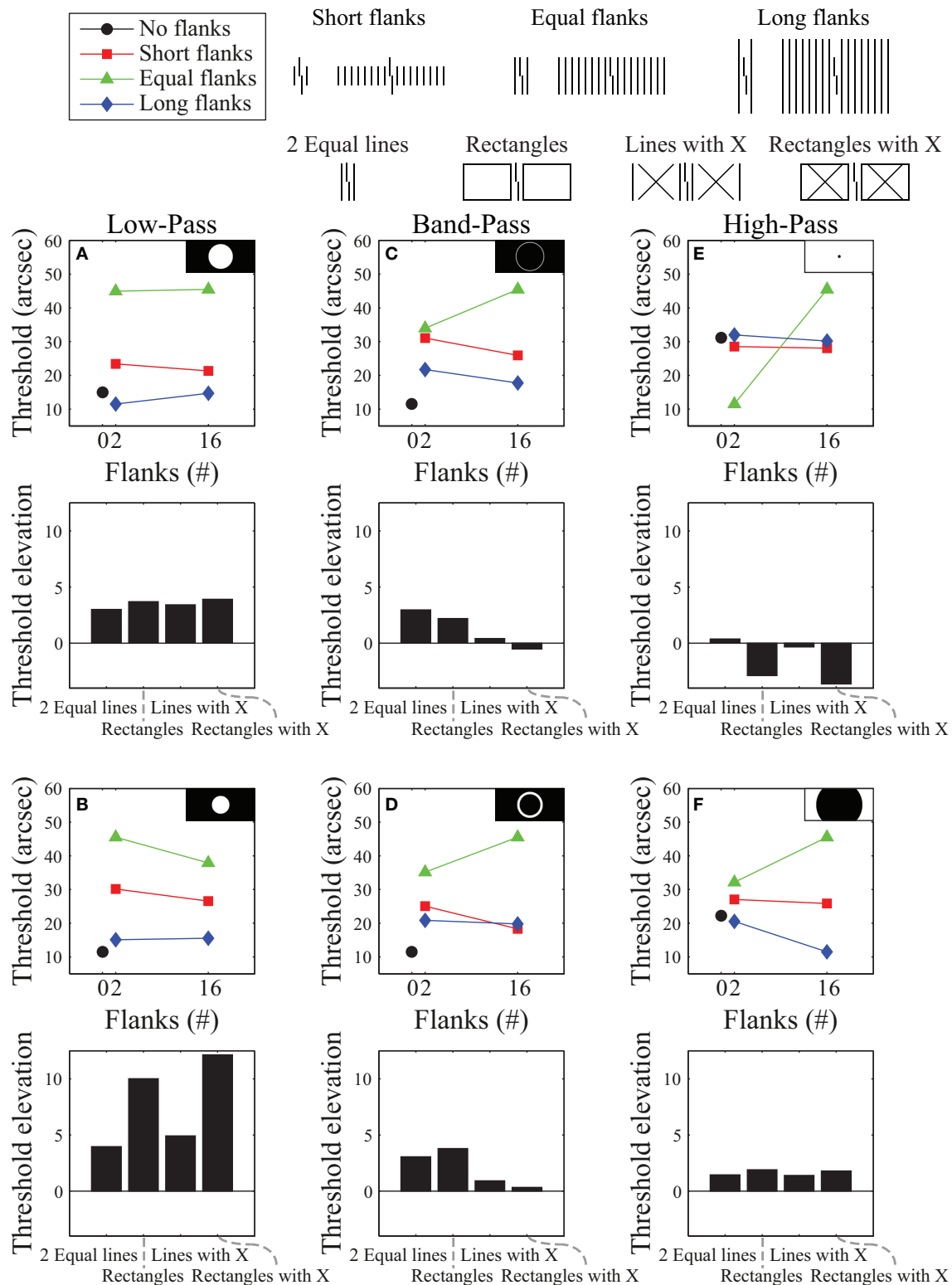


FIGURE 3 | Simulation results using a Fourier model for the stimuli that produced the data presented in Figure 2. Model results are plotted for representative low-pass filters (A,B), band-pass filters (C,D), and high-pass filters (E,F). Black and white insets show which frequencies were passed (white areas) and which frequencies were suppressed (black areas) in Fourier space (with lower frequencies in the center and higher frequencies near the

edges). The top row of subplots shows the best performance obtainable (using brute-force exhaustive search for the smallest sum of squared residuals against the human data) with each filter type. The bottom row shows results for filtering functions selected from different parts of the space - illustrating the variability of results obtainable with each filter type. The best overall performance we could obtain with this model is shown in (C).

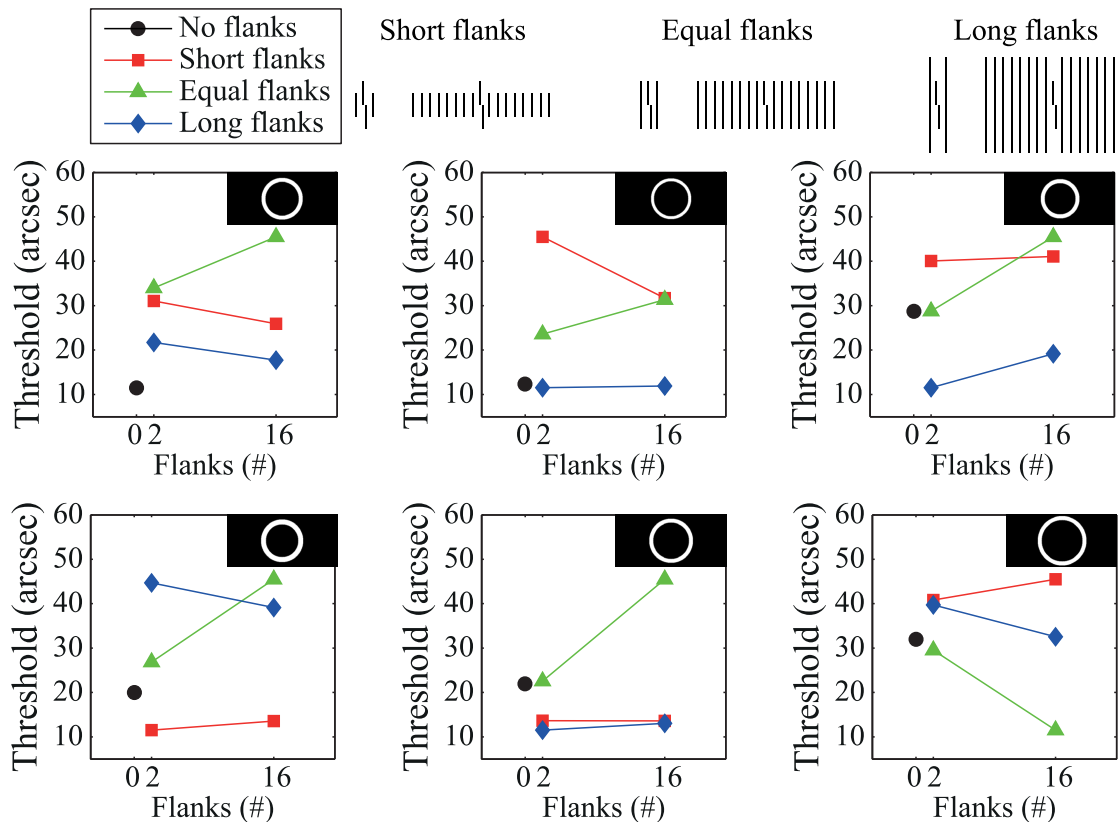


FIGURE 4 | Predicted behavior of the Fourier model for filters that slightly differ from the optimal band pass filter (shown in the upper left graph). Small changes in the band pass filter's center frequency and/or bandwidth lead to dramatic changes in the model's behavior.

over space with increasing time. The layers are reciprocally connected such that the excitatory units excite the inhibitory units and the inhibitory units inhibit the excitatory units. Details of the model, its filters, and its parameters can be found in Hermens et al. (2008) and Panis and Hermens (2014). Although the filters are local, the strength of activity at any given pixel location partly depends on the global pattern of activity across the network because of the feedback connections. When played out over time in a backward masking situation with stimuli similar to those in Figure 2, Hermens et al. (2008) showed that masking strength decreased as the number of flanking elements increased. More generally, the feedback in the network functions somewhat like a discontinuity detector by enhancing discontinuities and suppressing regularities. Panis and Hermens (2014) showed similar behavior for stimuli that produce crowding.

Since the model includes lateral feedback that promotes global processing, it satisfies the requirements identified above as “necessary” to explain crowding's effects. Moreover, the model's parameters were previously optimized for one stimulus, and then the model was validated by applying it to novel stimuli without further parameter optimization (Hermens et al., 2008). Thus, we would expect that any additional stimulus conditions that we apply this model to should require no further parameter optimization. We analyzed the model's behavior in response to the stimuli used to generate the findings

in Figure 2 but found that the model performs poorly overall (Figure 5). In particular, the model produces virtually no difference between any of the conditions shown in Figure 5A. Figure 5B shows that the model also fails to reproduce the human data plotted in Figure 2B. Here, the model produces no substantial differences between the different flanker length conditions, it produces no crowding for the case where there are two flanking lines (thresholds are the same as in the unflanked case), and model thresholds always go up as an increasing function of the number of flankers (contrary to the human data).

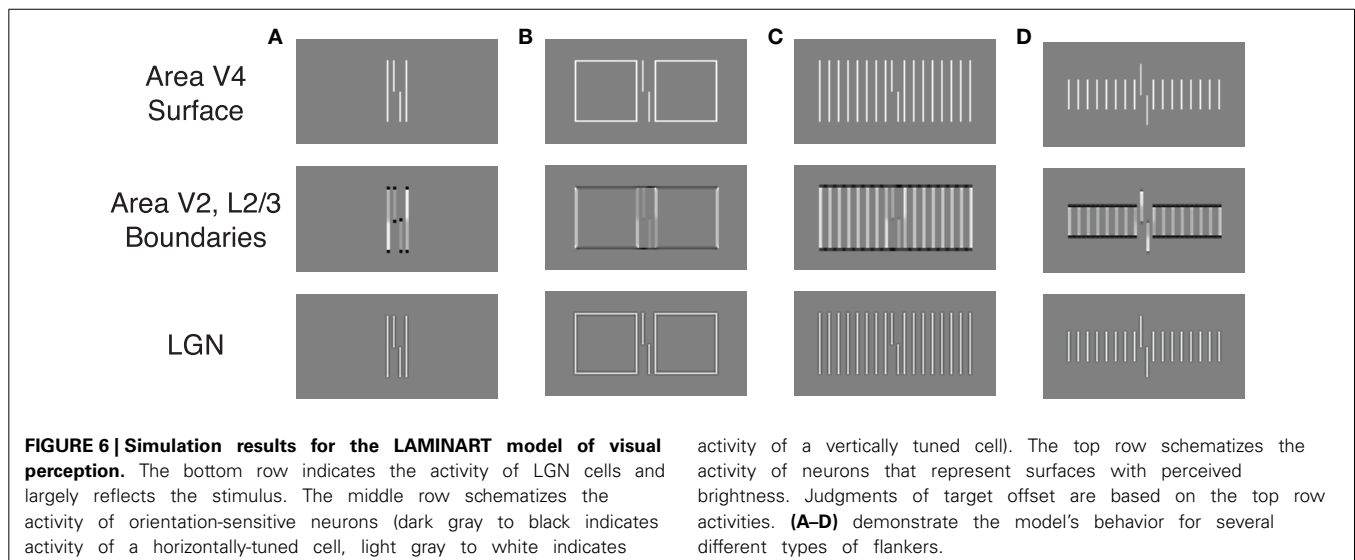
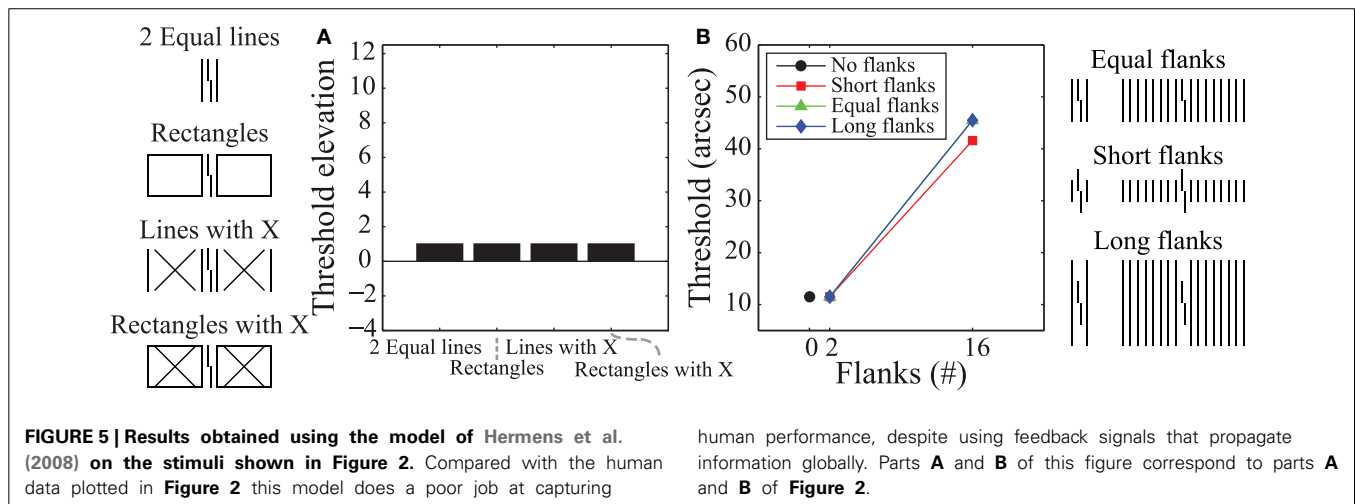
Even though the model has previously accounted for perceptual effects with similar kinds of stimuli and has strong feedback and global effects, the model simulations reported here do not account for the crowding effects in Figure 2. We cannot claim that the model architecture is fully rejected, as different filters and parameters may produce different model behaviors. Nevertheless, it is clear that global and feedback processing by themselves do not sufficiently constrain model properties relative to the observed crowding effects.

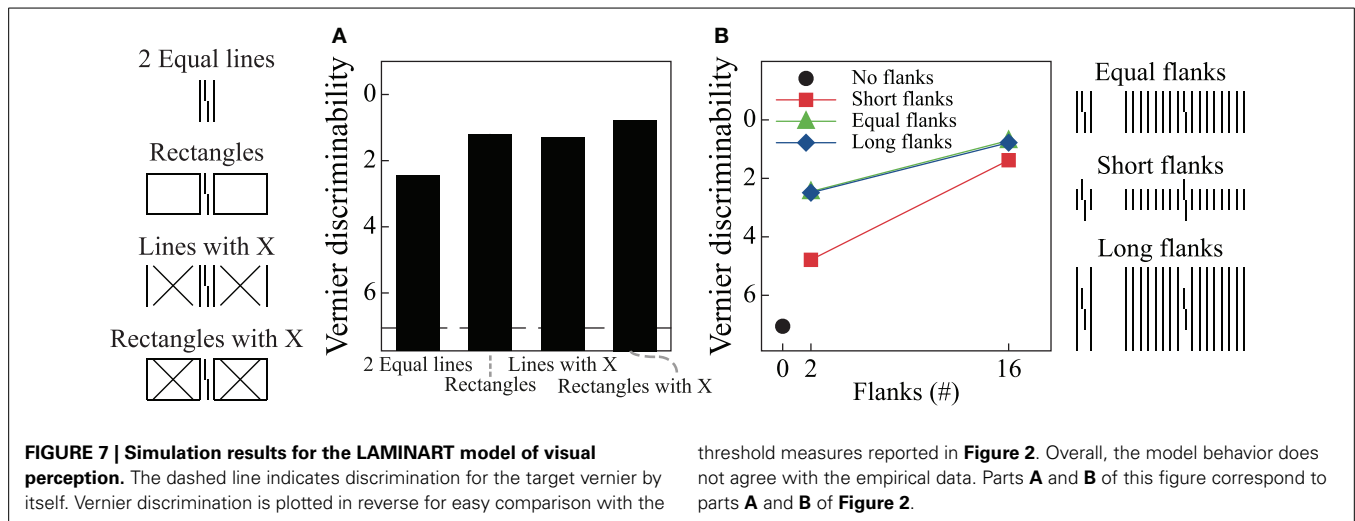
5. A FEEDBACK MODEL WITH PERCEPTUAL GROUPING: LAMINART NEURAL NETWORK

The previous simulations indicate that a model needs additional constraints beyond just feedback and global processing.

We next consider a model that has many additional constraints, the LAMINART model that has been proposed by Grossberg and colleagues (Raizada and Grossberg, 2001). The model is very complex and involves neural signals that interact across retinotopic coordinates, across laminar layers within a cortical area, and across cortical areas V1, V2, and V4. Various forms of the model account for neurophysiological and behavioral data related to depth perception (Grossberg, 1990; Grossberg and Howe, 2003), brightness perception (Grossberg and Todorovic, 1988), illusory contours (Grossberg and Mingolla, 1985), backward masking (Francis, 1997), and many other effects (Grossberg, 2003; Grossberg and Yazdanbakhsh, 2003). In particular, model simulations in Francis (2009) used stimuli very similar to those in **Figure 2** to successfully account for a variety of backward masking effects. An integral part of the model explanations involved a form of perceptual grouping, which was indicated by the presence of illusory contours connecting elements within a group. Consistent with the ideas derived from **Figure 2**, these model grouping processes use feedback to generate global effects.

The model proposes separate processing streams for boundary and surface information. Grouping effects mostly occur in the boundary system through formation of illusory contours that connect nearly collinearly oriented edges, and **Figures 6A,B** show simulation results for two of the stimulus conditions in **Figure 2A**. When the flankers are two lines, the model generates boundary signals that represent each stimulus line (Area V2, Layer 2/3) and these boundary signals constrain brightness signals that pass from the LGN to Area V4. As a result, the Area V4 representation is essentially veridical relative to the original stimulus. As described in the Supplementary Material, the model signals are connected to human performance with a template matching process that tries to distinguish between verniers shifted to the left or right. Crowding effects occur because the vernier template (whose width is five times the spacing between stimulus elements) integrates information from both the flankers and the vernier target, thereby reducing the signal-to-noise ratio for vernier discrimination. In this way, the model matches the empirical finding that two flanking lines can produce crowding. **Figure 7A** shows





vernier discriminability (plotted in reverse for comparison with the threshold data) for this simulation, and it indicates that it is harder to identify a vernier with two flankers than to identify a vernier by itself (the dashed line).

Figure 6B shows the model's behavior when the flanking elements are rectangles. Although the local information is similar to that in the case of two flanking lines, the Area V2, Layer 2/3 cells respond quite differently by producing illusory contours that connect the two rectangles and the target vernier. Nevertheless, at the V4 filling-in stage, the perceptual representation is nearly veridical, and crowding occurs because the flanking elements again interfere with the vernier template matching calculations. Although the model has perceptual grouping, it incorrectly produces strong crowding where the empirical data indicate only weak crowding effects. These effects are indicated in Figure 7A, where the rectangle flankers condition indicates worse vernier discrimination than does the two equal-length flankers condition. Using flankers with an "X" produces the same pattern as for the conditions without an "X," and the rectangles provide the strongest masking. The data in Figure 2A shows the opposite pattern for the rectangles.

Similar properties exist for the stimuli producing the findings in Figure 2B. Figures 6C,D show the model's behavior in response to sixteen equal and short flankers. Consistent with the arguments about grouping described above, in the equal-length case the model generates illusory contours that connect the flankers with the target, thereby collectively grouping the flankers and target together. At the filling-in stage, all of the elements are represented and there is strong crowding. Also consistent with the above arguments, grouping is different for the short flankers (similar behavior would occur for the long flankers), such that the flanking elements are connected by illusory contours but the target remains separate. However, such grouping does not lead to a release from crowding in the model. At the Area V4 filling-in stage the flanking elements still interfere with the vernier discrimination process, even though the boundaries indicate that the flankers and target are part of different perceptual groups. Figure 7B shows that the LAMINART model does

not do a good job of matching the behavioral data in Figure 2B. This failure occurs even though the model includes feedback, has global effects, and contains grouping mechanisms that seem to operate much as recommended. Our claim is not that the model can be fully rejected by this failure, but we want to emphasize that a model with feedback, global processing, and mechanisms for perceptual grouping is not necessarily able to account for the observed human data.

6. WHAT CONSTRAINTS DOES A MODEL NEED?

The model simulations of crowding demonstrate that identification of global vs. local and feedback vs. feedforward processing does not necessarily promote the development of models that can account for human performance. We suspect the same kind of conclusion applies to models for many other visual phenomena. Although quantitative models of visual perception that account for visual processing often do include feedback and global processing (e.g., Bridgeman, 1971; Grossberg and Mingolla, 1985; Francis, 1997; Roelfsema, 2006; Craft et al., 2007; Kogo et al., 2010), this inclusion is often because such mechanisms provide specific computational properties that are needed to produce a functional visual system. The failure of the models discussed here relative to their success for other phenomena (e.g., backward masking) encourages a consideration of what kinds of constraints are useful for model development. It is unlikely that there is one single answer to this question, but we are willing to propose some ideas.

6.1. GLOBAL vs. LOCAL IS ABOUT INFORMATION REPRESENTATION

All models of visual processing involve encoding and representing information about the stimulus, and such a representation changes at various model stages so that some information is explicitly represented, other information is only implicitly represented, and some information is absent. A local model is one where the encoding of information about a certain position in visual space is modified only by information at nearby positions in space. In the case of crowding, the argument against local processing is that explicit or implicit information about the target

vernier appears to be affected by stimulus characteristics that are spatially far away in an unexpected way (e.g., two flanking squares produce less crowding than two flanking lines).

Even when the argument for non-local effects is convincing, it does not specify exactly how information about the target should be represented in a global-effects model. The crowding models described here include different types of information representation and different types of global effects. The Fourier model transforms spatial information into spectra and then applies a filtering step that loses some information about the target (as well as information about the flankers). The Wilson-Cowan model represents visual information in spatial (retinotopic) coordinates and introduces global effects via recurrent lateral inhibition. The LAMINART model also represents information in spatial coordinates, and it generates global effects via long-range illusory contours that connect spatially disparate boundaries, which can alter the boundary representations of the target. In practice, none of these global mechanisms produce crowding effects that emulate the behavioral data, at least in the instantiations considered here.

It seems to us that the global vs. local processing issue is something of a “red herring” that ignores deeper questions about the representation of visual information. A model must encode visual information in a way that allows for local or global processing, and identification of this encoding and its representation is the real model challenge. For example, in the LAMINART model, the information at the V4 surface stage provides a representation of information that (for the stimuli considered here) is essentially the same as the stimulus. Although there are groupings among boundaries, they do not modify the representation of visual information that is involved in the vernier offset judgment. What appears to be needed is for the boundary groupings to segment the visual information so that the target is represented separately from the flankers. In this way, the target’s offset could be discriminated with less interference from background elements. Francis (2009) described how such segmentation can occur for some visual masking situations that encode information about the target at the V4 surface representations in a different depth plane than information about the flankers. Such segmentation promotes good discrimination of the vernier offset. Foley et al. (2012) demonstrated that attentional effects could also produce similar segmentations in crowding conditions.

6.2. FEEDFORWARD vs. FEEDBACK IS ABOUT MODEL FUNCTION

Many of the discussions about feedforward vs. feedback processing seem predicated on the notion that if information is available at a model stage, then it can be used for a relevant task. For example, if binocular disparity information is available at V1, then it can be used for making depth discriminations at this stage. However, this attitude does not consider the many ways that feedback processing can influence information processing. In general, feedback processing tends to produce one of five robust model functions.

1. Completion: Excitatory feedback can “fill-in” missing information and thereby make explicit information that is implicitly represented by other aspects of an input pattern. One example of such completion is the generation of illusory contours in the LAMINART model, where the model explicitly represents “missing” contours that are justified by the co-occurrence of appropriate contours that are physically present. Another example of such completion is in the convergence of a Hopfield (1982) network to states with active neurons that were not directly excited by the input but are justified by their association with other active neurons.
2. Competition: A combination of excitatory and inhibitory feedback can enhance differences in neural activity and, in extreme forms, generate winner-take-all behavior in a network of neurons (Grossberg, 1973). Such networks can suppress noisy or irrelevant information and enhance the representation of other information in the system. For example, Wilson et al. (1992) proposed a competitive neural network to explain vector summation in motion perception, where units tuned to a particular motion direction inhibit units tuned to the orthogonal motion direction.
3. Preservation: Excitatory feedback can allow signals to persist well beyond the physical offset of a stimulus (e.g., Ögmen, 1993; Francis et al., 1994). Inhibitory feedback can also play a role in preservation of information by suppressing incoming signals that might alter the current pattern (Francis, 1997, 2000). A combination of excitatory and inhibitory linear feedback can also preserve pattern representations even with large changes in overall intensity (Grossberg, 1973). Grossberg et al. (2011), for example, used preservation to extend flickering stimuli long enough through time to allow their temporally integrated signals to generate smooth motion percepts.
4. Uniformization: Some types of non-linear feedback can diminish differences in neural activity and lead to uniform activity (Grossberg, 1973). Such information-losing feedback is not commonly used in neural models.
5. Comparison: Appropriately structured excitatory and inhibitory feedback can generate a signal that indicates the degree to which two neural activity patterns differ. Such signals are helpful for larger architectures that need to detect changes or stabilize learning (e.g., Grossberg, 1980; Sutton and Barto, 1998; Di Lollo et al., 2000).

These different functions often require rather different feedback mechanisms that involve the distribution of excitatory and inhibitory relations, the relative strength of feedback and feedforward signals, and the form of signal transformation between neurons. Thus, model development requires a characterization of function in order to be able to properly implement feedback. Characterizing model function is, of course, very challenging and generally requires some kind of over-arching theoretical framework to guide the computational goals of the model. For example, a model of crowding that theorizes a role for perceptual grouping needs to indicate how elements in a scene are identified as being “grouped,” explain the mechanisms by which such distinctions are generated, and characterize how such representations influence target processing and decision making. A focus on such functional details may reveal that a certain form of feedback processing is critical for the model to reproduce the human behavior (Raizada and Grossberg, 2001), or it may reveal that the

feedforward vs. feedback distinction is not as relevant as it first appeared (e.g., Francis and Hermens, 2002; Poder, 2013).

7. CONCLUSIONS

If the starting point of theorizing is that visual processing involves local interactions in a feedforward system, then it makes sense that investigations should explore whether such systems are sufficient to account for a given phenomenon. However, the modeling efforts presented here suggest that clear evidence of a role for global and feedback processing does not sufficiently constrain a model. At best, such investigations are only the starting point for model development, and further considerations are required concerning the details of information representation and model function. It might be easier to initiate theorizing by assuming global and feedback processing and then look for other more informative constraints such as task optimality, or perceptual completion.

AUTHOR CONTRIBUTIONS

Aaron M. Clarke coded the simulations for the Fourier and Wilson-Cowan models. Gregory Francis coded the simulations for the LAMINART model. All authors contributed to the text.

ACKNOWLEDGMENTS

Aaron Clarke was funded by the Swiss National Science Foundation (SNF) project “Basics of visual processing: what crowds in crowding?” (Project number: 320030_135741). For Greg Francis, the research leading to these results has received funding from the European Union Seventh Framework Programme (FP7/2007-2013) under grant agreement n° 604102 (HBP).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://www.frontiersin.org/journal/10.3389/fpsyg.2014.01193/abstract>

REFERENCES

- Altmann, C. F., Bulthoff, H. H., and Kourtzi, Z. (2003). Perceptual organisation of local elements into global shapes in the human visual cortex. *Curr. Biol.* 13, 342–349. doi: 10.1016/S0960-9822(03)00052-6
- Atkinson, J., Anker, S., Evans, C., Hall, R., and Pimm-Smith, E. (1988). Visual acuity testing of young children with the cambridge crowding cards at 3 and 6 m. *Acta Ophthalmol.* 66, 505–508. doi: 10.1111/j.1755-3768.1988.tb04371.x
- Balas, B., Nakano, L., and Rosenholtz, R. (2009). A summary-statistic representation in peripheral vision explains visual crowding. *J. Vis.* 9, 13. doi: 10.1167/9.12.13
- Blakeslee, B., and McCourt, M. E. (1999). A multiscale spatial filtering account of the white effect, simultaneous brightness contrast and grating induction. *Vision Res.* 39, 4361–4377. doi: 10.1016/S0042-6989(99)00119-4
- Blakeslee, B., and McCourt, M. E. (2001). A multiscale spatial filtering account of the wertheimer-benary effect and the corrugated mondrian. *Vision Res.* 41, 2487–2502. doi: 10.1016/S0042-6989(01)00138-9
- Blakeslee, B., and McCourt, M. E. (2004). A unified theory of brightness contrast and assimilation incorporating oriented multiscale spatial filtering and contrast normalization. *Vision Res.* 44, 2483–2503. doi: 10.1016/j.visres.2004.05.015
- Blakeslee, B., Pasioka, W., and McCourt, M. E. (2005). Oriented multiscale spatial filtering and contrast normalization: a parsimonious model of brightness induction in a continuum of stimuli including White, Howe and simultaneous brightness contrast. *Vision Res.* 45, 607–615. doi: 10.1016/j.visres.2004.09.027
- Bridgeman, B. (1971). Metacontrast and lateral inhibition. *Psychol. Rev.* 78, 528–539. doi: 10.1037/h0031782
- Campbell, F. W., and Robson, J. G. (1968). Application of fourier analysis to the visibility of gratings. *J. Physiol.* 197, 551–566.
- Cesare, A. D., and Loftus, G. R. (2011). Global and local vision in natural scene identification. *Psychon. Bull. Rev.* 18, 840–847. doi: 10.3758/s13423-011-0133-6
- Cope, D., Blakeslee, B., and McCourt, M. E. (2008). Structural theorems for simple cell receptive fields. *J. Vis.* 8:802. doi: 10.1167/8.6.802
- Cope, D., Blakeslee, B., and McCourt, M. E. (2009). Simple cell response properties imply receptive field structure: balanced gabor and/or bandlimited field functions. *J. Opt. Soc. Am. A Opt. Image Sci. Vis.* 26, 2067–2092. doi: 10.1364/JOSAA.26.002067
- Craft, E., Schuetze, H., Niebur, E., and von der Heydt, R. (2007). A neural model of figure-ground organization. *J. Neurophysiol.* 97, 4310–4326. doi: 10.1152/jn.00203.2007
- DeAngelis, G. C., Ohzawa, I., and Freeman, R. D. (1994). Length and width tuning of neurons in the cat's primary visual cortex. *J. Neurophysiol.* 71, 347–374.
- De Valois, R. L., Albrecht, D. G., and Thorell, L. G. (1982). Spatial frequency selectivity of cells in macaque visual cortex. *Vision Res.* 22, 545–559. doi: 10.1016/0042-6989(82)90113-4
- Di Lollo, V., Enns, J. T., and Rensink, R. A. (2000). Competition for consciousness among visual events: the psychophysics of reentrant visual processes. *J. Exp. Psychol. Gen.* 129, 481–507. doi: 10.1037/0096-3445.129.4.481
- Dresp, B., and Grossberg, S. (1997). Contour integration across polarities and spatial gaps: from local contrast filtering to global grouping. *Vision Res.* 37, 913–924. doi: 10.1016/S0042-6989(96)00227-1
- Enns, J. T., and Di Lollo, V. (2000). What's new in visual masking? *Trends Cogn. Sci.* 4, 345–352. doi: 10.1016/S1364-6613(00)01520-5
- Field, D. J. (1987). Relations between the statistics of natural images and the response properties of cortical cells. *J. Opt. Soc. Am. A* 4, 2379–2394. doi: 10.1364/JOSAA.4.002379
- Foley, N., Grossberg, S., and Mingolla, E. (2012). Neural dynamics of object-based multifocal visual spatial attention and priming: object cueing, useful-field-of-view, and crowding. *Cogn. Psychol.* 65, 77–117. doi: 10.1016/j.cogpsych.2012.02.001
- Fourier, J. B. J. (1822). *La Théorie Analytique De La Chaleur*. Paris: Firmin Didot.
- Francis, G. (1997). Cortical dynamics of lateral inhibition: metacontrast masking. *Psychol. Rev.* 104, 572. doi: 10.1037/0033-295X.104.3.572
- Francis, G. (2000). Quantitative theories of metacontrast masking. *Psychol. Rev.* 107, 768. doi: 10.1037/0033-295X.107.4.768
- Francis, G. (2009). Cortical dynamics of figure-ground segmentation: shine-through. *Vision Res.* 49, 140–163. doi: 10.1016/j.visres.2008.10.002
- Francis, G., and Grossberg, S. (1995). *Cortical Dynamics of Boundary Segmentation and Reset: Persistence, Afterimages, and Residual Traces*. Technical Report, Boston University Center for Adaptive Systems and Department of Cognitive and Neural Systems.
- Francis, G., Grossberg, S., and Mingolla, E. (1994). Cortical dynamics of feature binding and reset: control of visual persistence. *Vision Res.* 34, 1089–1104. doi: 10.1016/0042-6989(94)90012-4
- Francis, G., and Hermens, F. (2002). Comment on: competition for consciousness among visual events: The psychophysics of reentrant visual processes, by di lollo, enns and rensink (2000). *J. Exp. Psychol. Gen.* 131, 590–593. doi: 10.1037/0096-3445.131.4.590
- Freeman, J., and Simoncelli, E. P. (2011). Metamers of the ventral stream. *Nat. Neurosci.* 14, 1195–1201. doi: 10.1038/nn.2889
- Freeman, R. D., Ohzawa, I., and Walker, G. (2001). Beyond the classical receptive field in the visual cortex. *Prog. Brain Res.* 134, 157–170. doi: 10.1016/S0079-6123(01)34012-8
- Gabor, D. (1946). Theory of communication. *J. Inst. Electr. Eng.* 93, 429–457.
- Greenwood, J. A., Bex, P. J., and Dakin, S. C. (2009). Positional averaging explains crowding with letter-like stimuli. *Proc. Natl. Acad. Sci. U.S.A.* 106, 13130–13135. doi: 10.1073/pnas.0901352106
- Grossberg, S. (1973). Contour enhancement, short term memory, and constancies in reverberating neural networks. *Stud. Appl. Math.* 52, 213–257.
- Grossberg, S. (1980). How does a brain build a cognitive code? *Psychol. Rev.* 87, 1–51. doi: 10.1037/0033-295X.87.1.1

- Grossberg, S. (1990). Neural facades: visual representations of static and moving form-and-color-and-depth. *Mind Lang.* 5, 411–456. doi: 10.1111/j.1468-0017.1990.tb00171.x
- Grossberg, S. (2003). Laminar cortical dynamics of visual form perception. *Neural Netw.* 16, 925–931. doi: 10.1016/S0893-6080(03)00097-2
- Grossberg, S., and Howe, P. D. (2003). A laminar cortical model of stereopsis and three-dimensional surface perception. *Vision Res.* 43, 801–829. doi: 10.1016/S0042-6989(03)00011-7
- Grossberg, S., Léveillé, J., and Versace, M. (2011). How do object reference frames and motion vector decomposition emerge in laminar cortical circuits? *Atten. Percept. Psychophys.* 73, 1147–1170. doi: 10.3758/s13414-011-0095-9
- Grossberg, S., and Mingolla, E. (1985). Neural dynamics of form perception: boundary completion, illusory figures, and neon color spreading. *Psychol. Rev.* 92, 173–211. doi: 10.1037/0033-295X.92.2.173
- Grossberg, S., and Rudd, M. E. (1989). A neural architecture for visual motion perception: group and element apparent motion. *Neural Netw.* 2, 421–450. doi: 10.1016/0893-6080(89)90042-7
- Grossberg, S., and Todorovic, D. (1988). Neural dynamics of 1-d and 2-d brightness perception: a unified model of classical and recent phenomena. *Percept. Psychophys.* 43, 241–277. doi: 10.3758/BF03207869
- Grossberg, S., and Yazdanbakhsh, A. (2003). Laminar cortical dynamics of 3-d surface stratification, transparency, and neon spreading. *J. Vis.* 3, 247. doi: 10.1167/3.9.247
- Harrison, L. M., Stephan, K. E., Rees, G., and Friston, K. J. (2007). Extra-classical receptive field effects measured in striate cortex with fmri. *Neuroimage* 34, 1199–1208. doi: 10.1016/j.neuroimage.2006.10.017
- Hermens, F., Luksys, G., Gerstner, W., Herzog, M. H., and Ernst, U. (2008). Modeling spatial and temporal aspects of visual backward masking. *Psychol. Rev.* 115, 83–100. doi: 10.1037/0033-295X.115.1.83
- Herzog, M. H., Ernst, U. A., Etzold, A., and Eurich, C. W. (2003). Local interactions in neural networks explain global effects in gestalt processing and masking. *Neural Comput.* 15, 2091–2113. doi: 10.1162/089976603322297304
- Hopfield, J. J. (1982). Neural networks and physical systems with emergent collective computational abilities. *Proc. Natl. Acad. Sci. U.S.A.* 79, 2554–2558. doi: 10.1073/pnas.79.8.2554
- Hubel, D. H., and Wiesel, T. N. (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *J. Physiol.* 160, 106–154.
- Hung, C. P., Kreiman, G., Poggio, T., and DiCarlo, J. J. (2005). Fast readout of object identity from macaque inferior temporal cortex. *Science* 310, 863–866. doi: 10.1126/science.1117593
- Jones, J., and Palmer, L. (1987). An evaluation of the two-dimensional gabor filter model of simple receptive fields in cat striate cortex. *J. Neurophysiol.* 58, 1233–1258.
- Juan, C.-H., and Walsh, V. (2003). Feedback to V1: a reverse hierarchy in vision. *Exp. Brain Res.* 150, 259–263. doi: 10.1007/s00221-003-1478-5
- Keil, A., Sabatinelli, D., Ding, M., Lang, P. J., Hsien, N., and Heim, S. (2009). Re-entrant projections modulate visual cortex in affective perception: evidence from granger causality analysis. *Hum. Brain Mapp.* 30, 532–540. doi: 10.1002/hbm.20521
- Keyser, C., and Perrett, D. I. (2002). Visual masking and {RSVP} reveal neural competition. *Trends Cogn. Sci.* 6, 120–125. doi: 10.1016/S1364-6613(00)01852-0
- Kogo, N., Strelha, C., Gool, L. V., and Wagemans, J. (2010). Surface construction by a 2-d differentiation-integration process: a neurocomputational model for perceived border ownership, depth, and lightness in kanizsa figures. *Psychol. Rev.* 117, 406–439. doi: 10.1037/a0019076
- Lamme, V. A. F. (2006). Towards a true neural stance on consciousness. *Trends Cogn. Sci.* 10, 494–501. doi: 10.1016/j.tics.2006.09.001
- Lamme, V. A. F., and Roelfsema, P. R. (2000). The distinct modes of vision offered by feedforward and recurrent processing. *Trends Neurosci.* 23, 571–579. doi: 10.1016/S0166-2236(00)01657-X
- Levi, D. M., and Carney, T. (2009). Crowding in peripheral vision: why bigger is better. *Curr. Biol.* 19, 1988–1993. doi: 10.1016/j.cub.2009.09.056
- Levi, D. M., Klein, S. A., and Aitsebaomo, A. (1985). Vernier acuity, crowding and cortical magnification. *Vision Res.* 25, 963–977. doi: 10.1016/0042-6989(85)90207-X
- Livne, T., and Sagi, D. (2011). Multiple levels of orientation anisotropy in crowding with gabor flankers. *J. Vis.* 11, 1–10. doi: 10.1167/11.13.18
- Malania, M., Herzog, M. H., and Westheimer, G. (2007). Grouping of contextual elements that affect vernier thresholds. *J. Vis.* 7, 1–7. doi: 10.1167/7.2.1
- Manassi, M., Sayim, B., and Herzog, M. H. (2012). Grouping, pooling, and when bigger is better in visual crowding. *J. Vis.* 12, 1–14. doi: 10.1167/12.10.13
- Manassi, M., Sayim, B., and Herzog, M. H. (2013). When crowding of crowding leads to uncrowding. *J. Vis.* 13, 1–10. doi: 10.1167/13.13.10
- Ögmen, H. (1993). A neural theory of retino-cortical dynamics. *Neural Netw.* 6, 245–273. doi: 10.1016/0893-6080(93)90020-W
- Panis, S., and Hermens, F. (2014). Time course of spatial contextual interference: event history analysis of simultaneous masking by nonoverlapping patterns. *Am. Psychol. Assoc.* 40, 129–144. doi: 10.1037/a0032949
- Parkes, L., Lund, J., Angelucci, A., Solomon, J. A., and Morgan, M. (2001). Compulsory averaging of crowded orientation signals in human vision. *Nat. Neurosci.* 4, 739–744. doi: 10.1038/89532
- Poder, E. (2013). Attentional gating models of object substitution masking. *J. Exp. Psychol. Gen.* 142, 1130–1141. doi: 10.1037/a0030575
- Poggio, T., Mutch, J., Anselmi, F., Tacchetti, A., Rosasco, L., and Leibo, J. Z. (2013). “Does invariant recognition predict tuning of neurons in sensory cortex?” in *MIT: Computer Science and Artificial Intelligence Laboratory Technical Report* (Cambridge, MA), 1–12.
- Raizada, R. D., and Grossberg, S. (2001). Context-sensitive binding by the laminar circuits of V1 and V2: A unified model of perceptual grouping, attention, and orientation contrast. *Vis. Cogn.* 8, 431–466. doi: 10.1080/13506280143000070
- Rasche, C., and Koch, C. (2002). Recognizing the gist of a visual scene: possible perceptual and neural mechanisms. *Neurocomputing* 44–46, 979–984. doi: 10.1016/S0925-2312(02)00500-3
- Riesenhuber, M., and Poggio, T. (1999). Hierarchical models of object recognition in cortex. *Nat. Neurosci.* 2, 1019–1025.
- Roelfsema, P. R. (2006). Cortical algorithms for perceptual grouping. *Annu. Rev. Neurosci.* 29, 203–227. doi: 10.1146/annurev.neuro.29.051605.112939
- Sayim, B., Westheimer, G., and Herzog, M. H. (2010). Gestalt factors modulate basic spatial vision. *Psychol. Sci.* 21, 641–644. doi: 10.1177/0956797610368811
- Sceniak, M. P., Ringach, D. L., Hawken, M. J., and Shapley, R. (1999). Contrast's effect on spatial summation by macaque v1 neurons. *Nat. Neurosci.* 2, 733–739. doi: 10.1038/11197
- Schwartz, E. L. (1980). Computational anatomy and functional architecture of striate cortex: a spatial mapping approach to perceptual coding. *Vision Res.* 20, 645–669. doi: 10.1016/0042-6989(80)90090-5
- Serre, T., Oliva, A., and Poggio, T. (2007a). A feedforward architecture accounts for rapid categorization. *Proc. Natl. Acad. Sci. U.S.A.* 104, 6424–6429. doi: 10.1073/pnas.0700622104
- Serre, T., Wolf, L., Bileschi, S., Riesenhuber, M., and Poggio, T. (2007b). Robust object recognition with cortex-like mechanisms. *IEEE Trans. Pattern Anal. Mach. Intell.* 29, 411–426. doi: 10.1109/TPAMI.2007.56
- Serre, T., Wolf, L., and Poggio, T. (2005). “Object recognition with features inspired by visual cortex,” in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2005 Vol. 2*, (San Diego, CA), 994–1000.
- Sutton, R. S., and Barto, A. G. (1998). *Reinforcement Learning: An Introduction*. Cambridge, MA: The MIT Press.
- Van den Berg, R., Roerdink, J. B., and Cornelissen, F. W. (2010). A neurophysiologically plausible population code model for feature integration explains visual crowding. *PLoS Comput. Biol.* 6:e1000646. doi: 10.1371/journal.pcbi.1000646
- VanRullen, R., Delorme, A., and Thorpe, S. J. (2001). Feed-forward contour integration in primary visual cortex based on asynchronous spike propagation. *Neurocomputing* 38–40, 1003–1009. doi: 10.1016/S0925-2312(01)00445-3
- Von der Heydt, R., Peterhans, E., and Baumgartner, G. (1984). Illusory contours and cortical neuron responses. *Science* 224, 1260–1262. doi: 10.1126/science.6539501
- Wagemans, J., Elder, J. H., Kubovy, M., Palmer, S. E., Peterson, M. A., Sing, M., et al. (2012). A century of gestalt psychology in visual perception: I. perceptual grouping and figure-ground organization. *Psychol. Bull.* 138, 1172–1217. doi: 10.1037/a0029333
- Wallis, T. S., and Bex, P. J. (2011). Visual crowding is correlated with awareness. *Curr. Biol.* 21, 254–258. doi: 10.1016/j.cub.2011.01.011
- Weinstein, E. (1968). A Rashevsky-Landahl neural net: Simulation of metacontrast. *Psychol. Rev.* 75, 494–521. doi: 10.1037/h0026769

- Whitney, D., and Levi, D. M. (2011). Visual crowding: a fundamental limit on conscious perception and object recognition. *Trends Cogn. Sci.* 15, 160–168. doi: 10.1016/j.tics.2011.02.005
- Wilkinson, F., Wilson, H. R., and Ellemberg, D. (1997). Lateral interaction in peripherally viewed texture arrays. *J. Optic. Soc. Am. A* 14, 2057–2068. doi: 10.1364/JOSAA.14.002057
- Wilson, H. R., Ferrera, V. P., and Yo, C. (1992). A psychophysically motivated model for two-dimensional motion perception. *Vis. Neurosci.* 9, 79–97. doi: 10.1017/S0952523800006386

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.

Received: 20 June 2014; accepted: 02 October 2014; published online: 21 October 2014.
Citation: Clarke AM, Herzog MH and Francis G (2014) Visual crowding illustrates the inadequacy of local vs. global and feedforward vs. feedback distinctions in modeling visual perception. *Front. Psychol.* 5:1193. doi: 10.3389/fpsyg.2014.01193
This article was submitted to Perception Science, a section of the journal *Frontiers in Psychology*.

Copyright © 2014 Clarke, Herzog and Francis. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



A hidden ambiguity of the term “feedback” in its use as an explanatory mechanism for psychophysical visual phenomena

Talis Bachmann *

Institute of Public Law, University of Tartu, Tartu, Estonia

*Correspondence: talis.bachmann@ut.ee

Edited by:

Haluk Ogmen, University of Houston, USA

Reviewed by:

Frouke Hermens, University of Aberdeen, UK

Stephanie Goodhew, The Australian National University, Australia

Keywords: awareness, reentrant processing, feedforward processing, non-specific thalamocortical modulation, masking

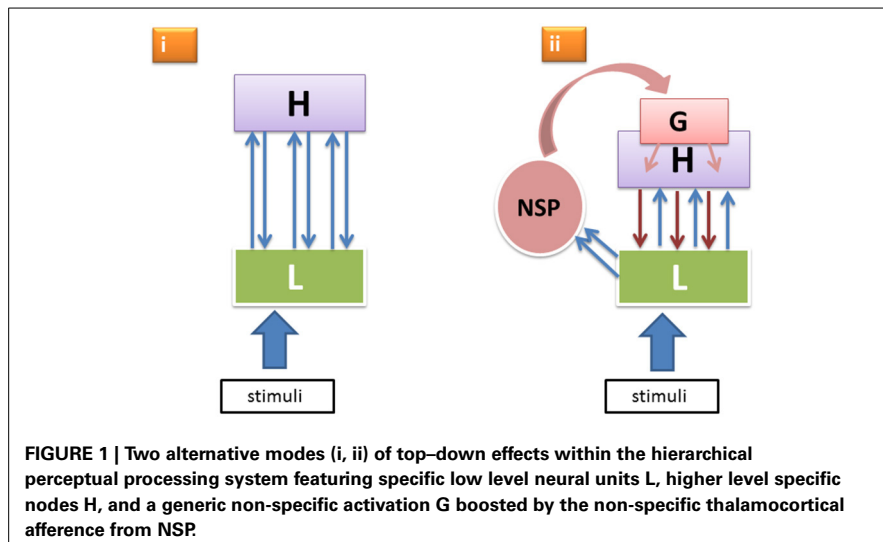
Over the last decades many researchers have used concepts like “feedback,” “reentrance,” “backpropagation,” “top–down (modulation),” or “reverse hierarchy” to specify the mechanisms that underlie various visual phenomena (e.g., Di Lollo et al., 2000; Lamme and Roelfsema, 2000; Pascual-Leone and Walsh, 2001; Supér et al., 2001; Ro et al., 2003; Ahissar and Hochstein, 2004; Bar et al., 2006; Fahrenfort et al., 2007; Koivisto, 2012). An incomplete list of these phenomena includes visual (object substitution) masking, shape discrimination, illusory contours, illusory motion, priming effects, etc. Empirical evidence or theoretical argumentation in favor of the suggested mechanistic explanations mainly consists in finding or postulating an association between a temporally delayed, secondary activation of lower level neural units with correct reports of target stimuli, even though the higher level neural units in the processing hierarchy were already activated earlier. On that basis, feedforward processing has been argued to be insufficient for target perception. However, in most of the studies the relative *temporal order* of activity at different levels alone is taken as proof of reentrant modulation without precisely measuring the *neural sources* of this top–down effect. In principle, it is equally possible that the source of the higher level activity from which the top–down signals are sent back to earlier feature-encoding neural units (i) is specifically linked to those features by virtue of constituting the higher level nodes associated with specific attributes

of the target stimulus (thus mediating feature-binding for object integration) or (ii) is not specifically linked in this manner. In the latter case, the *source* of top–down modulation may be the result of the arousal or alerting boost triggered by the target stimulus via feedforward collateral activation of subcortical reticulo-thalamic units, which in turn is followed by the cortical spread of the thalamocortical activation, including the downpropagation of the non-specific wave of modulation to the early cortical areas. The non-specific system functions include arousal, attentional modulation, intercortical synchronization of neural activity, bringing the preconsciously processed specific content to awareness, “event-holding” the content in working memory, and alerting subjects to newly appearing objects and changes (Magoun, 1958; Purpura, 1970; Purpura and Schiff, 1997; Jones, 2001; Llinás and Ribary, 2001; Van der Werf et al., 2002; Ribary, 2005; Schiff et al., 2013; Saalmann, 2014). This non-specific system (NSP) targets layer-1 apical dendrites of the layer-5 and -6 pyramidal neurons. But since NSP-modulation is directed at the cortical neurons with specific representational functions, its function may go unacknowledged because the cortical units, when activated by NSP-modulation, can produce content-specific subjective effects misleading us to believe that the entire process has been specific throughout.

The focus of the present paper will be on the experimental-behavioral and neurobiological evidence in comparing the two processing modes, (i) and (ii), with

arguments from computational modeling left for some other occasion.

It is known that reticulo-thalamic, intralaminar and other matrix cells of the NSP project more heavily to lateral and frontal cortical areas and less so to the primary visual areas. (Even when rare examples of direct intralaminar-thalamic input to V1 were documented, these afferents were found to be much sparser than the more frontal ones—Miller and Benevento, 1979.) Moreover, this more rostrally directed thalamo-cortical flow can cause cortical responses as fast as or even faster than the afferent volleys through the specific geniculo-cortical pathways ignite primary visual cortical responses strongly enough (Kennedy and Baleyrier, 1977; Kaufman and Rosenquist, 1985; Herkenham, 1986; Cruikshank et al., 2012; Liang et al., 2013; Saalmann, 2014). Thus, the primary cortical areas receive NSP-modulation not directly, but via the higher level cortical neurons that project onto apical parts of the layer-5 pyramidal neurons in the lower cortical areas. Consequently, as illustrated in **Figure 1**, we have two principal modes through which lower level neural units L responsible for encoding sensory features of perceptual objects receive top–down input from higher levels H: (i) from the specific nodes in H that were previously activated by L in a cortical feedforward manner and that now send reentrant signals back to L (here the feedforward-reentrant loop pertains to the specific sensory-perceptual attributes constituting a perceptual object LH); and (ii) from the



generic nodes G that were activated by the boost of the NSP directed at the more frontal and mid-level cortical neurons that now send their downpropagating wave to the lower level visual areas, including L.

When analyzing the experimental data from most of the studies that propose *specific* top-down linkages (i), there is no direct evidence that would invalidate the alternative, *non-specific* theory of down-propagation (ii). The specificity of visual experiences is due to the fact that the NSP-modulation arrives at specific early units L and may not be due to the specificity of the higher level from where this modulation arrives. Although the direct input from NSP to L may be weak, the top-down input from higher levels H/G driven by NSP may be strong enough to emphasize the specificity of the visual experience encoded in L. The pending task should be to try disentangle these two explanations experimentally. The experiments should ascertain whether the two modes of top-down modulation are incompatible or mutually complementary. In the latter case—how the two types of down-propagation are specifically combined and what relative roles each of them has? It is also possible that the standard views of reentrance (e.g., Di Lollo et al., 2000; Lamme and Roelfsema, 2000) may be valid in some empirical instances, difficult to ascertain in some other cases, and incompatible with the neurophysiological realities of processing in different experiments. Let me comment on some examples of

typical experiments aimed at supporting the standard views of reentrance listed below and see whether version (i) should be exclusively preferred or whether versions (i) and (ii) both are compatible with the experimental results.

1. In typical object substitution masking (OSM) a target stimulus (e.g., a Landolt C) is presented together with four dots that surround the target. When after a very brief delay the target is switched off, the four dots either are also switched off or remain displayed for varying duration acting as a post-mask (the simultaneous onset, asynchronous offset condition.) The delayed-offset condition leads to strong masking but in the simultaneous-offset condition masking is weak. The classic theory of OSM (Di Lollo et al., 2000, but see Pöder, 2013) explains this by a reentrant model (a variety of model i) according to which target-activated units at level H activated by the target send reentrant signals back to level L in order to test whether levels H and L are consistent in representing the target. If mismatch is registered (e.g., when target signals do not arrive anymore and mask signals arrive instead), the iterative feedforward-reentrant cycles are interrupted and new iterative “hypothesis testing” begins for the new object—the mask. Because cycles of reentrance are necessary for registration of the stimulus in awareness, the target is not consciously perceived when reentrant

testing is prematurely interrupted by the stronger top-down mask signal. However, when mask’s offset is synchronous with that of the target, the target-plus-mask is a composite object that provides both level L and H contents; hence, the target can be extracted from the composite representation that is maintained through the feedforward-reentrant cycles. Let us see how the model (ii) works for OSM. Presentation of target evokes specific signaling along L-H vertical axis and also a collaterally ignited boost of NSP modulation. (NSP is necessary for awareness of the specific contents represented by L and H.) When asynchronous-offset mask remains in view and target signals do not arrive anymore, the top-down activation G that was initiated fast at higher levels, but takes time to become active at lower levels “finds” mask related activity in L, but the target related activity has decayed already relative to the mask activity, because the target was switched off earlier. Although the level G activity is non-specific, when its downpropagating generic influence reaches L it helps emphasize mask features because level L units themselves *are* specific. The mask-object representation becomes consciously perceived instead of target. Thus, models (i) and (ii) both are usable. At this point one may ask why not follow Ockham’s rule and take the simpler one (i), i.e., the one with fewer hypotheses? However, the G units are important because neurobiological evidence has overwhelmingly shown that NSP is necessary for awareness of the specific contents represented by L and H.

2. In Lamme et al. (2000) monkeys were trained to discriminate visual targets. V1 responses began to differentiate the “seen” from the “unseen” trials after 125 ms. In subsequent studies occipital ERPs in humans differentiated visibility of masked targets after 109–141 ms or peaked at about 160 ms (Fahrenfort et al., 2007, 2008). Again, a variety of model (i) was used for explaining the results because specifically the temporally late target related activity at level L (which followed earlier time epochs sufficient for level H to have become

active in target processing) were associated with correct discrimination. And again, model (ii) can explain these empirical results: the late part of neural activity at L which is enhanced in trials where target is successfully discriminated may be modulated by the top-down process G passed down through levels H (or even bypassing stimulus-specific level H units either via direct fibers or level H units different from the stimulus-related ones).

3. Temporal precedence of high-level MEG activity which discriminated correct and incorrect target processing over low level activity in the study by Bar et al. (2006) was also interpreted as a variety of model (i). However, if activity of G at frontal sites fluctuates (fluctuation of the thalamocortical NSP activity is a norm rather than an exception) and dictates whether the top-down modulation is stronger or weaker, these experimental results can be interpreted also according to model (ii).
4. Ro et al. (2003) utilized transcranial magnetic stimulation (TMS) in a metacontrast masking paradigm and showed that TMS of visual cortex, when timed to produce visual suppression of an annulus (a metacontrast mask), induced recovery of a target disc which was imperceptible when TMS was not used. Moreover, TMS suppression of an annulus was more pronounced when a disk preceded it than when an annulus was presented alone. The authors assume that when the later activity, supposedly reflecting the reentrant effects is suppressed then target perceptibility can be reinstated. They argue that a prior visual stimulus can influence subsequent perception at early stages of visual encoding via feedback projections, supporting model (i). Alternatively, model (ii) can also be applied. It is known that a preceding brief stimulus (e.g., target) speeds up perception of the following stimulus (e.g., mask) (Bachmann, 1989; Scharlau, 2007). When target disc was presented before mask it may have speeded up masking annulus processing by presetting NSP modulation for its signals. This in turn may have optimized the effective processing delay

so as to coincide with the maximal TMS effect.

5. Up to now, both models appear to be equally applicable, but model (ii) provides an explanation of the results of an elegant experiment carried out by Wu et al. (2009) that model (i) cannot as readily provide. Capitalizing on the motion-induced blindness (MIB) phenomenon (Bonneh et al., 2001), where a static visual target-object continuously presented on a rotating background periodically disappears from awareness, they showed that a flashed stimulus that *caused* reappearance in awareness of the target was perceived *after* the reappearance of the target in consciousness. (The temporal value of reversal was about 100 ms, which is the value assumed to characterize the full cycle of reentrance based visual processing for awareness.) The temporal advantage of updating the conscious representation from the preexisting unconscious representation of the invisible static target was explained by a version of model (i), invoking reentry of neural signals after the first feed-forward sweep for a stimulus to be consciously perceived. Thus, MIB, by blocking reentry signals, prevents awareness. In Bachmann and Aru (2009) we pointed out some inconsistencies of this explanation and offered an explanation in terms of model (ii). When an object fades from awareness by MIB, its L and H level activity will be sustained because cortical specific signals are constantly present, but now it is dissociated from NSP-activity. When the flashed object is presented, the L/H process for representation of the flash occurs in parallel with a boost of the NSP-process igniting G. G leads to binding of the already present preconscious L/H-activity of the target with global consciousness-level representation. This process takes little time, because there is no need for build-up of the content-specific L/H representation of the target; consequently, its rapid reappearance in consciousness. The flashed object appears in consciousness not as fast, because its corresponding coherent L/H-representation must be built up, which takes time. The G that services *target* awareness has L/H

content of the target ready on the “waiting list” but the G process has to wait as a “dummy process” until the L/H contents of the *flashed object* are ready to be modulated.

It appears that experiments have difficulty in distinguishing between the two models. This raises the question whether a computational/mathematical argument could be developed that allows to test different predictions about experimental data on the basis of the two models. Sadly, space does not allow me to dwell into this important perspective which must be dealt with in future research.

CONCLUSION

In this opinion paper I argued for the view that in the majority of the standard experimental studies set to support the model of top-down processing featuring exclusively the specific system components also the combined non-specific/specific model seems equally valid.

ACKNOWLEDGMENTS

I thank also Bruno Breitmeyer for his useful suggestions and editorial help in preparing the final version of this article.

REFERENCES

- Ahissar, M., and Hochstein, S. (2004). The reverse hierarchy theory of visual perceptual learning. *Trends Cogn. Sci.* 8, 457–464. doi: 10.1016/j.tics.2004.08.011
- Bachmann, T. (1989). Microgenesis as traced by the transient paired-forms paradigm. *Acta Psychol.* 70, 3–17. doi: 10.1016/0001-6918(89)90056-5
- Bachmann, T., and Aru, J. (2009). Mechanisms of consciousness: why the effect can precede its cause in awareness – Commentary on Wu et al. *Curr. Biol.* Available online at: <http://www.cell.com/current-biology/comments/S0960-9822%2809%2901836-3>
- Bar, M., Kassam, K. S., Ghuman, A. S., Boshyan, J., Schmid, A. M., Dale, M., et al. (2006). Top-down facilitation of visual recognition. *Proc. Natl. Acad. Sci. U.S.A.* 103, 449–454. doi: 10.1073/pnas.0507062103
- Bonneh, Y. S., Cooperman, A., and Sagi, D. (2001). Motion-induced blindness in normal observers. *Nature* 411, 798–801. doi: 10.1038/35081073
- Cruikshank, S. J., Ahmed, O. J., Stevens, T. R., Patrick, S. L., Gonzalez, A. N., Elmaleh, M., et al. (2012). Thalamic control of Layer 1 circuits in prefrontal cortex. *J. Neurosci.* 32, 17813–17823. doi: 10.1523/JNEUROSCI.3231-12.2012
- Di Lollo, V., Enns, J. T., and Rensink, R. A. (2000). Competition for consciousness among visual events: the psychophysics of reentrant visual processes. *J. Exp. Psychol. Gen.* 129, 481–507. doi: 10.1037/0096-3445.129.4.481

- Fahrenfort, J. J., Scholte, H. S., and Lamme, V. A. F. (2007). Masking disrupts reentrant processing in human visual cortex. *J. Cogn. Neurosci.* 19, 1488–1497. doi: 10.1162/jocn.2007.19.9.1488
- Fahrenfort, J. J., Scholte, H. S., and Lamme, V. A. F. (2008). The spatiotemporal profile of cortical processing leading up to visual perception. *J. Vis.* 8, 12.1–12.12. doi: 10.1167/8.1.12
- Herkenham, M. (1986). “New perspectives on the organization and evolution of nonspecific thalamocortical projections,” in *Sensory-Motor Areas and Aspects of Cortical Connectivity*, Chapter 11, eds E. G. Jones and A. Peters (New York, NY: Plenum Press), 403–423. doi: 10.1007/978-1-4613-2149-1_11
- Jones, E. G. (2001). The thalamic matrix and thalamocortical synchrony. *Trends Neurosci.* 24, 595–601. doi: 10.1016/S0166-2236(00)01922-6
- Kaufman, E. F. S., and Rosenquist, A. C. (1985). Efferent projections from thalamic intralaminar nuclei in the cat. *Brain Res.* 335, 257–279. doi: 10.1016/0006-8993(85)90478-0
- Kennedy, H., and Baleydir, C. (1977). Direct projections from thalamic intralaminar nuclei to extra-striate visual cortex in the cat traced with horseradish peroxidase. *Exp. Brain Res.* 28, 133–139. doi: 10.1007/BF00237091
- Koivisto, M. (2012). Is reentry critical for visual awareness of object presence? *Vis. Res.* 63, 43–49. doi: 10.1016/j.visres.2012.05.001
- Lamme, V. A. F., Supér, H., Landman, R., Roelfsema, P. R., and Spekreijse, H. (2000). The role of primary visual cortex (V1) in visual awareness. *Vis. Res.* 40, 1507–1521. doi: 10.1016/S0042-6989(99)00243-6
- Lamme, V. A., and Roelfsema, P. R. (2000). The distinct modes of vision offered by feedforward and recurrent processing. *Trends Neurosci.* 23, 571–579. doi: 10.1016/S0166-2236(00)01657-X
- Liang, M., Mouraux, A., and Iannetti, G. D. (2013). Bypassing primary sensory cortices—A direct thalamocortical pathway for transmitting salient sensory information. *Cereb. Cortex* 23, 1–11. doi: 10.1093/cercor/bhr363
- Llinás, R., and Ribary, U. (2001). Consciousness and the brain. The thalamocortical dialogue in health and disease. *Ann. N.Y. Acad. Sci.* 929, 166–175. doi: 10.1111/j.1749-6632.2001.tb05715.x
- Magoun, H. W. (1958). *The Waking Brain*. Springfield, Ill: C.C.Thomas. doi: 10.1037/11149-000
- Miller, J. W., and Benevento, L. A. (1979). Demonstration of a direct projection from the IL central lateral nucleus to the primary visual cortex. *Neurosci. Lett.* 14, 229–234. doi: 10.1016/0304-3940(79)96153-6
- Pascual-Leone, A., and Walsh, V. (2001). Fast back-projections from the motion to the primary visual area necessary for visual awareness. *Science* 292, 510–512. doi: 10.1126/science.1057099
- Pöder, E. (2013). Attentional gating models of object substitution masking. *J. Exp. Psychol. Gen.* 142, 1130–1141. doi: 10.1037/a0030575
- Purpura, D. (1970). “Operations and processes in thalamic and synaptically related neural subsystems,” in *The Neurosciences. Second Study Program*, ed F. O. Schmitt (New York, NY: Rockefeller University Press), 458–470.
- Purpura, K. P., and Schiff, N. D. (1997). The thalamic intralaminar nuclei: a role in visual awareness. *Neuroscientist* 3, 8–15.
- Ribary, U. (2005). “Dynamics of thalamo-cortical network oscillations and human perception,” in *Progress in Brain Research* 150, ed S. Laureys (Amsterdam: Elsevier), 127–142. doi: 10.1016/S0079-6123(05)50010-4
- Ro, T., Breitmeyer, B., Burton, P., Singhal, N. S., and Lane, D. (2003). Feedback contributions to visual awareness in human occipital cortex. *Curr. Biol.* 11, 1038–1041. doi: 10.1016/S0960-9822(03)00337-3
- Saalmann, Y. B. (2014). Intralaminar and medial thalamic influence on cortical synchrony, information transmission and cognition. *Front. Syst. Neurosci.* 8:83. doi: 10.3389/fnsys.2014.00083
- Scharlau, I. (2007). Perceptual latency priming: a measure of attentional facilitation. *Psychol. Res.* 71, 678–686. doi: 10.1007/s00426-006-0056-4
- Schiff, N. D., Shah, S. A., Hudson, A. E., Nauvel, T., Kalik, S. F., and Purpura, K. P. (2013). Gating of attentional effort through the central thalamus. *J. Neurophysiol.* 109, 1152–1163. doi: 10.1152/jn.00317.2011
- Supér, H., Spekreijse, H., and Lamme, V. A. F. (2001). Two distinct modes of sensory processing observed in figure-ground activity of monkey primary visual cortex (V1). *Nat. Neurosci.* 4, 304–310. doi: 10.1038/85170
- Van der Werf, Y. D., Witter, M. P., and Groenewegen, H. J. (2002). The intralaminar and midline nuclei of the thalamus. Anatomical and functional evidence for participation in processes of arousal and awareness. *Brain Res. Brain Res. Rev.* 39, 107–140. doi: 10.1016/S0165-0173(02)00181-9
- Wu, C.-T., Busch, N. A., Fabre-Thorpe, M., and VanRullen, R. (2009). The temporal interplay between conscious and unconscious perceptual streams. *Curr. Biol.* 19, 2003–2007. doi: 10.1016/j.cub.2009.10.017

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.

Received: 29 May 2014; accepted: 02 July 2014; published online: 22 July 2014.

Citation: Bachmann T (2014) A hidden ambiguity of the term “feedback” in its use as an explanatory mechanism for psychophysical visual phenomena. *Front. Psychol.* 5:780. doi: 10.3389/fpsyg.2014.00780

This article was submitted to Perception Science, a section of the journal Frontiers in Psychology.

Copyright © 2014 Bachmann. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

ADVANTAGES OF PUBLISHING IN FRONTIERS



FAST PUBLICATION

Average 90 days
from submission
to publication



COLLABORATIVE PEER-REVIEW

Designed to be rigorous –
yet also collaborative, fair and
constructive



RESEARCH NETWORK

Our network
increases readership
for your article



OPEN ACCESS

Articles are free to read,
for greatest visibility



TRANSPARENT

Editors and reviewers
acknowledged by name
on published articles



GLOBAL SPREAD

Six million monthly
page views worldwide



COPYRIGHT TO AUTHORS

No limit to
article distribution
and re-use



IMPACT METRICS

Advanced metrics
track your
article's impact



SUPPORT

By our Swiss-based
editorial team