Current Biology 23, 1–5, August 5, 2013 ©2013 Elsevier Ltd All rights reserved http://dx.doi.org/10.1016/j.cub.2013.05.065

Report

Shared Representations for Working Memory and Mental Imagery in Early Visual Cortex

Anke Marit Albers,[1] Peter Kok,[1] Ivan Toni,[1]
H. Chris Dijkerman,[2] and Floris P. de Lange[1],*
[1]Donders Institute for Brain, Cognition and Behaviour,
Radboud University Nijmegen, Kapittelweg 29, 6525 EN
Nijmegen, the Netherlands
[2]Experimental Psychology, Helmholtz Institute, Utrecht
University, Heidelberglaan 2, 3584 CS Utrecht, the Netherlands

Summary

Early visual areas contain specific information about visual items maintained in working memory, suggesting a role for early visual cortex in more complex cognitive functions [1–4]. It is an open question, however, whether these areas also underlie the ability to internally generate images de novo (i.e., mental imagery). Research on mental imagery has to this point focused mostly on whether mental images activate early sensory areas, with mixed results [5–7]. Recent studies suggest that multivariate pattern analysis of neural activity patterns in visual regions can reveal content-specific representations during cognitive processes, even though overall activation levels are low [1–4]. Here, we used this approach [8, 9] to study item-specific activity patterns in early visual areas (V1–V3) when these items are internally generated. We could reliably decode stimulus identity from neural activity patterns in early visual cortex during both working memory and mental imagery. Crucially, these activity patterns resembled those evoked by bottom-up visual stimulation, suggesting that mental images are indeed "perception-like" in nature. These findings suggest that the visual cortex serves as a dynamic "blackboard" [10, 11] that is used during both bottom-up stimulus processing and top-down internal generation of mental content.

Results

Here we investigated whether early sensory regions are recruited similarly during the maintenance of previously presented images (i.e., visual working memory [1, 12]), the internal generation of images that have not been presented (i.e., mental imagery [13]), and the perception of visual material [14–17]. We used a multivariate analysis approach [8, 9] to determine the information contained in the spatial patterns of fMRI responses. Participants (N = 24) either kept a grating stimulus in mind (working memory [WM] trials) or internally generated a new stimulus by mentally rotating a grating and subsequently held this new image in their mind's eye for a 10 s period (imagery [IM] trials; see Figure 1). Crucially, during the IM task, the image kept in mind was not a representation of the physically presented grating but was generated de novo by mentally transforming the stimulus material. Behavioral data confirmed that the participants could successfully perform both tasks, with increasing reaction times as a function of the amount of mental transformation (see Figure S1 available online). We defined early visual cortical areas (V1, V2, and V3) using standard retinotopic mapping routines and extracted activity patterns in these regions as the mental imagery process unfolded.

First, we assessed whether the activity pattern in early visual cortex during the working memory period in WM trials reflected the stimulus orientation (three possibilities: 15°, 75°, 135°) that was maintained by the participants, using a WM-trained classifier and a leave-one-run-out cross-validation approach. We found that early visual cortex (V1–V3) indeed contained information about maintained content [WM: decoding accuracy 54%, chance level 33.3%: $t(23) = 5.88$, $p < 1 \times 10^{-5}$] in the period 8–12 s after onset of maintenance). This increase is comparable in size to that observed in earlier studies [1], reflects a medium-to-large effect size (Cohen's $d = 1.21$) [18], and replicates the finding that early visual cortex contains memory representations in the absence of stimulus input [1–4]. To investigate whether the same voxels in early visual cortex also contained information about images that were internally generated and subsequently maintained, we repeated this procedure with an IM-trained classifier applied to IM trials. Indeed, early visual cortex also contained information about internally generated images [IM: decoding accuracy 46%, $t(23) = 3.09$, $p = 0.005$, Cohen's $d = 0.63$], indicating involvement of the visual cortex during mental imagery. Moreover, activity patterns for WM and IM trials were highly similar: when training the multivariate pattern classifier on the delay period during WM and testing on the delay period during IM, we found equally reliable pattern information [WM → IM decoding accuracy 45%, $t(23) = 3.88$, $p < 1 \times 10^{-3}$, Cohen's $d = 0.78$]. Training on IM and testing on WM also resulted in reliable classification [IM → WM decoding accuracy 45%, $t(23) = 4.13$, $p < 1 \times 10^{-3}$, Cohen's $d = 0.83$]. All of these effects were also present when we looked at V1, V2, and V3 separately (Table S1; all accuracies > 39%, all $p < 0.007$).

The similarity between neural representations during WM and IM does not necessarily mean that these representations are "perceptual" in nature (i.e., resemble the bottom-up activity patterns evoked during actual perception), because bottom-up and top-down signals could be encoded differently in early visual cortex [19, 20], or the patterns could reflect some other aspect of the task, such as attention. To test the perceptual nature of these representations, we obtained activity patterns during the actual perception of gratings and trained a classifier to discriminate the orientation of these gratings. Since participants performed a task at fixation during the perception of the gratings, these activity patterns chiefly reflected bottom-up, stimulus-related activity, while the potential effects of top-down attentional processes were reduced by providing subjects with a task at fixation. This "perceptual" classifier could also reliably discriminate between activity patterns in early visual cortex evoked by the different orientations during both WM trials [decoding accuracy 46%, $t(23) = 4.50$, $p < 1 \times 10^{-4}$, Cohen's $d = 0.90$] and IM trials [decoding accuracy 49%, $t(23) = 5.92$ $p < 1 \times 10^{-5}$, Cohen's $d = 1.21$]. This indicates that not only does the early visual cortex contain information about internally generated images during IM, the activity patterns for these images are similar to those evoked

*Correspondence: floris.delange@donders.ru.nl

Please cite this article in press as: Albers et al., Shared Representations for Working Memory and Mental Imagery in Early Visual Cortex, Current Biology (2013), http://dx.doi.org/10.1016/j.cub.2013.05.065

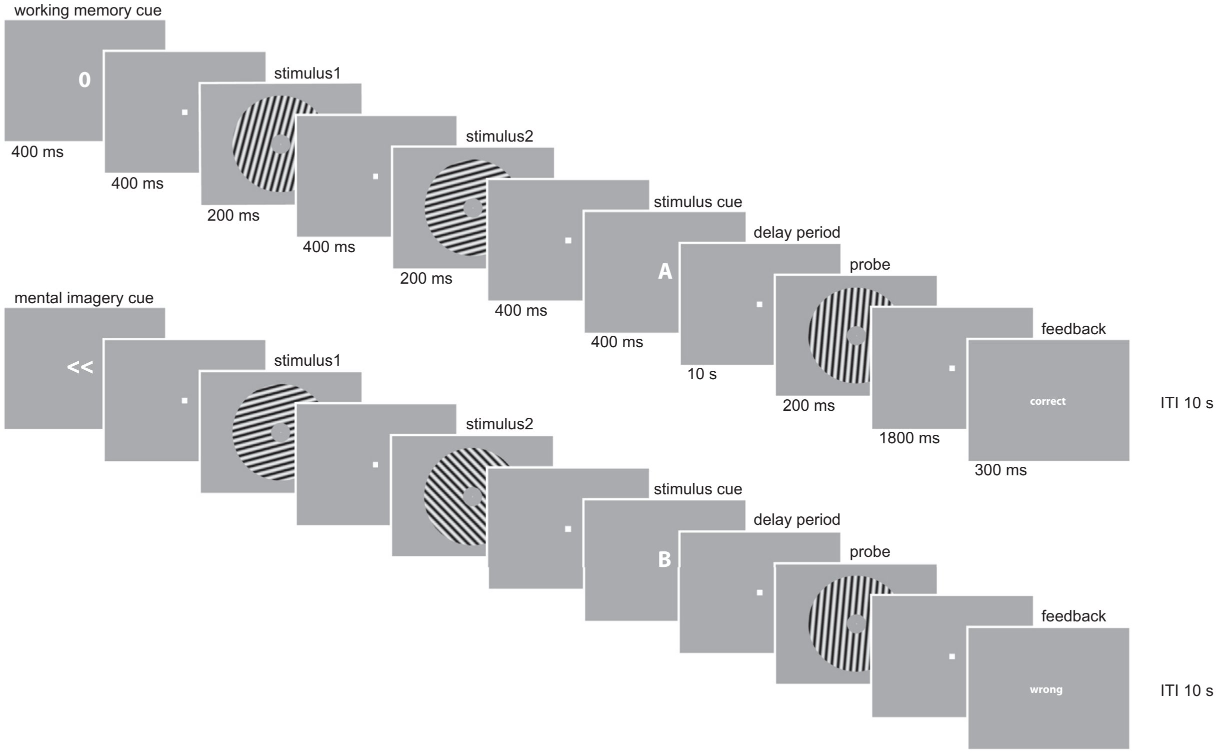

Figure 1. Experimental Design

At the start of each trial, a task cue indicated whether participants had to maintain a stimulus in working memory (WM; top row) or create a new stimulus by imagining rotating the stimulus grating and keeping the ensuing mental image in their mind's eye (mental imagery [IM]; bottom row). During IM trials, mental rotation could be clockwise or counterclockwise (as indicated by arrow direction), and 60° or 120° (as indicated by the number of arrows). After the task cue, two gratings (out of three possible stimuli: 15°, 75°, or 115°) were presented briefly, followed by a second stimulus cue (A or B, denoting the first or second stimulus, respectively) that indicated which stimulus grating to select and maintain (WM) or rotate and then imagine (IM). After a 10 s delay period in which participants were asked to vividly imagine the relevant stimulus, a probe was presented. Participants indicated whether the probe was rotated clockwise or counterclockwise with respect to the stimulus they had kept in mind and received feedback on each trial. See also Figure S1.

during actual perception. Interestingly, decoding accuracy was higher for people who could more accurately form mental images during both tasks (WM: $\rho = -0.51$, $p = 0.0053$; IM: $\rho = -0.37$, $p = 0.039$; Figure S2), providing a strong link between mental imagery ability and the precision of population-level responses [21, 22].

The generalization of the content-specific patterns between bottom-up stimulation and top-down internal generation suggests that similar neural codes are used during these processes in early sensory cortex. To examine the time course of this process and assess whether early visual cortex sequentially represents the initial and final target image (cf. [23]), we analyzed activity patterns in early visual cortex at each time point as the task unfolded, using the independent classifier that was trained on stimulus-driven activity. During WM trials, the classifier was initially at chance, selecting each option approximately one-third of the time. After stimulus presentation, visual cortical (V1–V3) activity patterns reflected a combination of the two presented gratings. Subsequently, the cued (i.e., to be remembered) grating was predominantly selected by the classifier (Figure 2A). Similarly, during IM trials, initial visual cortical activity patterns after stimulus presentation reflected the two presented gratings, but not the unpresented grating (Figures 2B and 2C). Again, shortly after this, the cued (i.e., starting orientation that had to be mentally rotated) grating was predominantly selected by the classifier. Crucially, however, there was a gradual shift from a representation of the cued (starting) grating toward a representation of the internally generated target grating. This target grating was not physically presented on that trial but was mentally created by the participants and after several seconds became the preferred orientation of the classifier. This suggests three sequential stages of representation in early visual cortex during the mental imagery process: first, the physically presented stimuli are represented; second, one of the presented stimuli is selected for transformation; and third, a new representation is formed in early visual cortex.

The time courses of decoding accuracy for the target grating further support this notion (Figure 3A). During WM, the target could be decoded as early as 4 s after delay-period onset, whereas during IM the target could only be decoded from 8 s after delay-period onset. This delay likely reflects a combination of factors. During imagery trials, participants had to not only select the cued grating but also retrieve the task cue, which instructed them about the direction and extent of mental rotation. They subsequently had to perform the mental rotation, with each of these steps contributing to the delay in the formation of the internally generated target image. The patterns in Figures 2 and 3 suggest that participants mentally transformed the image early in the trial, rather than at the time of the probe. Again, similar patterns were present in V1–V3 in isolation.

There was a dissociation between the time course of stimulus representation and the time course of mean neural activity in the early visual regions (Figure 3B). While information about maintained (WM) or internally generated (IM) stimuli increased over the delay interval, overall neural activity decreased, in line with previous work on visual working memory [1]. This stresses the difference between overall activation and information content within activation patterns and puts the results of previous studies that looked only at overall activity levels of sensory cortex during mental imagery in a new perspective. Although early sensory areas did not show robust delay-related activity, there were several other areas outside visual cortex that showed a robust and sustained neural activity increase during the delay period of both IM and WM trials (Figure S3; Table S2), including bilateral parietal and prefrontal cortex, as well as the pre-supplementary motor area. To investigate whether these areas also contained stimulus-related information, we used the same classification approach that we employed for early visual areas. Interestingly, although some of these regions within this network (notably the left parietal cortex and supplementary motor area) showed some evidence of stimulus information when training and testing within the main experiment (Table S2), generalization from the perceptual classifier to the main task resulted in chance-level performance (all p > 0.05).

Discussion

In this study, we used a multivariate pattern analysis approach to directly compare neural representations during visual perception, working memory and mental imagery, in retinotopically defined early visual cortex. We found that activity patterns in early visual areas (V1–V3) could reliably predict which of three oriented gratings was either held in working memory or mentally imagined, even though overall levels of neural activity were low. We observed similar neural activity patterns during periods in which participants either kept visual material in working memory (WM) or internally generated a visual stimulus (IM) by mentally transforming it, as shown by similarly high decoding performance within and between tasks. Crucially, by training on patterns of activity during physical presentation of gratings, we show that activity patterns during mental imagery resemble those elicited by physically presented stimuli, suggesting analogous neural codes for internally generated mental images and stimulus representations. The results are in line with other recent findings of representational content in the visual cortex during high-level cognitive processes [1, 24, 25]

Together, our results suggest that early visual areas may serve as a dynamic "blackboard" that supports information

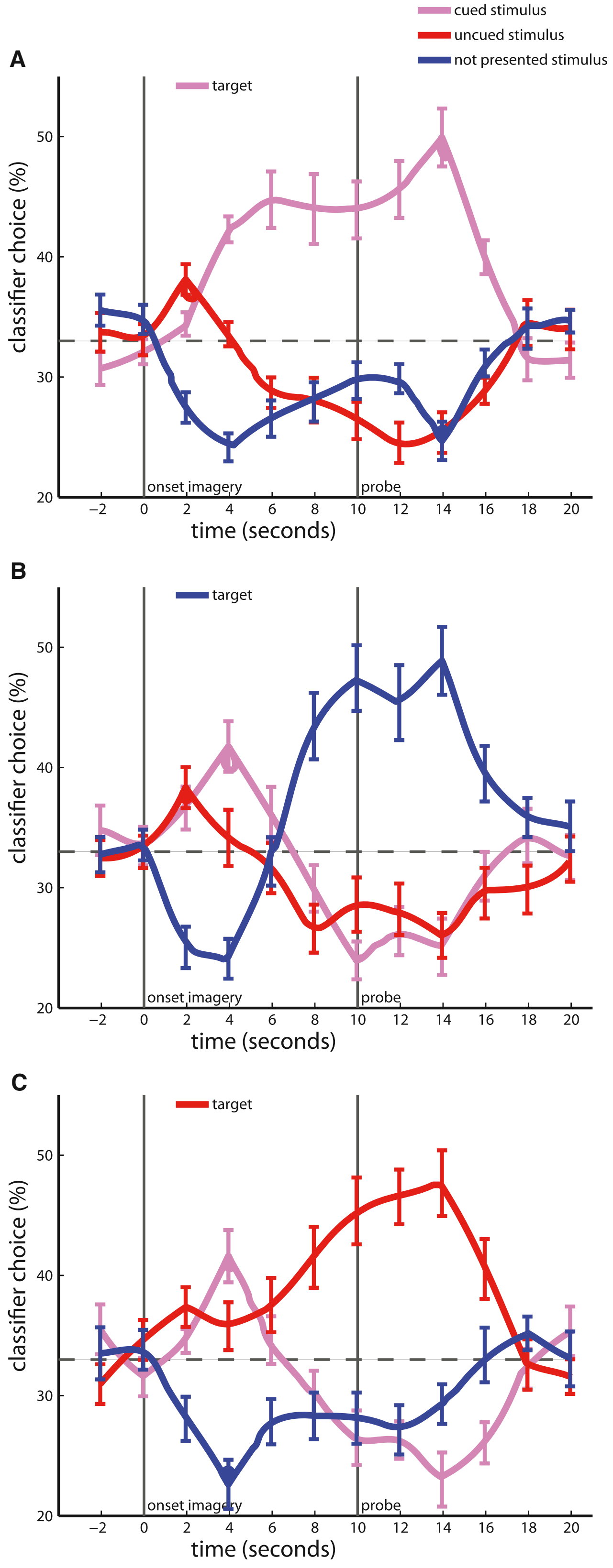

Figure 2. Temporal Unfolding of Mental Representations

Proportion of classifier choice when testing V1–V3 combined (360 voxels), averaged over the 24 participants. Error bars denote SEM; dashed line indicates chance level (33.3%).

(A) Classifier choice over time during WM trials. Activity patterns during the first time point (2 s after WM onset) show a mixture of the two physically presented stimuli (red and pink lines), but not unseen grating, after which the pattern activity was consistently classified as the cued grating (pink line).

(B) Classifier choice over time during IM trials. On these trials, participants mentally rotated the cued stimulus toward the not-physically-presented grating orientation. Again, activity patterns during the first time point (2 s after IM onset) show a mixture of the two physically presented stimuli (red and pink lines), while the not-presented grating is the least selected. Thereafter, there is a gradual switch in classifier choice from the cued grating (pink) to the generated target grating (blue).

(C) Classifier choice over time during IM trials. On these trials, participants mentally rotated the cued stimulus toward the presented but uncued grating orientation. A transition in the representation occurred ~8 s after delay period onset, from the cued grating (pink line) to the created grating that was similar to the presented but uncued grating (red line).

See also Figure S2 and Table S1.

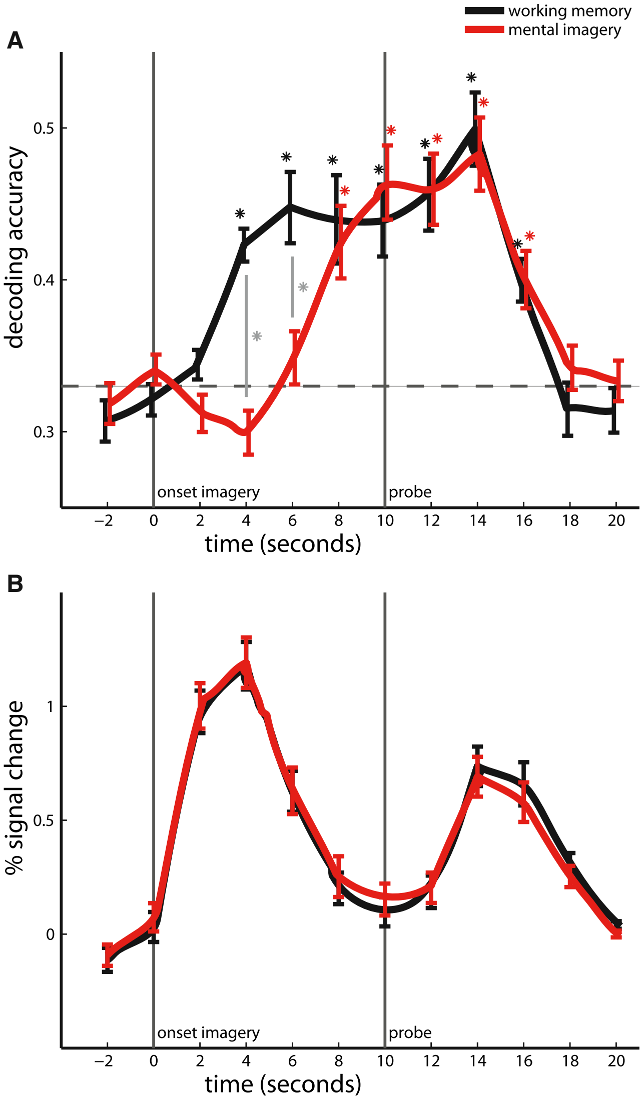

Figure 3. Time Course of Decoding Accuracy and Mean Neural Activity in V1–V3

(A) Time course of decoding accuracy was different for WM (black line) and IM (red line). Accurate decoding of the target image was achieved several seconds later in time for IM than for WM trials, due to the intermediate mental operation. Decoding was significant from 4 to 16 s for WM trials and from 8 to 16 s for IM trials (all $p < 0.001$). Error bars denote SEM, dashed line indicates chance level (33.3%), and asterisks indicate significant decoding accuracy ($p < 0.001$) for WM (black) and IM (red) or a significant difference between the two (gray).

(B) Time course of mean neural activity was indistinguishable between WM (black) and IM (red), as indicated by average blood oxygen level-dependent amplitude time course (averaged over the 360 selected voxels) with respect to average activity immediately preceding trial onset. Neural activity peaked ~4 s after presentation of the stimuli and again ~4 s after presentation of the probe, while activity declined in the delay period between the two presentations. Error bars denote SEM.

See also Figure S3 and Table S2.

processing during both bottom-up and top-down processes [10, 23, 26]. This fits with proposals that view the primary visual cortex not simply as an entry station for subsequent cortical computations in higher-order visual areas but rather as a high-resolution buffer in the visual system that is recruited for several visual computations [11, 24, 26].

These findings also speak to an age-old debate about the nature of mental content [13, 16]. Depictive theories of mental content stress the overlap between representations during perception and mental imagery. Studies that assessed whether mental images activate primary sensory areas, as proposed by "depictive" theorists, have provided mixed results [5–7, 27]. By showing that there is content-specific overlap of activation patterns during mental imagery and bottom-up visual stimulation in primary visual cortex, we show that mental imagery partly depends on the same mechanisms as visual perception, in line with depictive accounts of mental representations [20].

An open issue relates to the role of nonsensory areas in the maintenance and internal generation of sensory material. Although we observed strong increases in activity in a specific set of regions in prefrontal and parietal cortex [28], we and others [3] did not find reliable encoding of stimulus-related information in these areas when training on perceptual input. This suggests that although it is very possible that these nonvisual areas contain stimulus representations, their format appears distinct from the automatic, bottom-up representation evoked by visual stimulation. Studies using neural recordings in monkeys have observed coding of individual stimuli in prefrontal cortex [29] and content-specific synchronization of activity across the frontal parietal network [30] during visual working memory. Interestingly, we also obtained evidence for some stimulus-related information in parietal and frontal regions when comparing stimulus-specific patterns within the main tasks (IM and WM). Together, these results suggest complementary roles for early visual cortex and frontoparietal regions [30–32], whereby frontoparietal regions create flexible stimulus representations that are in line with behavioral goals [3, 33]. However, the exact role and representational content of the frontoparietal regions during mental imagery remain to be determined.

The generalization of stimulus information from stimulus-driven activity patterns to mental imagery-induced activity patterns in early visual cortex suggests a common representation of bottom-up and top-down signals in these cortical areas. It should be noted here that generalization was robust but not perfect, which may be due to the fact that internally generated images can lead to less robust and more variable activation patterns than bottom-up visual stimulation, due to internal fluctuations in attentional state. These fluctuations are likely reduced during the perceptual localizer, although it is also possible here that subjects still paid some attention to the stimuli (even though they performed a task at fixation). The current task design makes it unlikely that eye movements contributed to the decoding of imagined orientations. First, we trained the classifier on the independent localizer, during which participants had to perform a task at fixation. Second, the stimulus gratings were presented very briefly (200 ms), too short for systematic eye movement preparation and execution. Additionally, the relevant grating was only cued after the stimulus presentation [1].

It may seem surprising that overall neural activity levels appeared low during mental imagery and working memory, even though the patterns in early visual cortex carried stimulus information during this period. One reason for this may be that visual areas also exhibit an overall high level of spontaneous activity during rest [34], the functional significance of which may be quite similar to mental imagery [35]. Indeed, a recent developmental study [36] showed that spontaneous fluctuations in visual regions become increasingly similar to stimulus-evoked patterns, suggesting that activity patterns in visual cortex may constitute an internal model that continuously adapts to expected upcoming input. Evidence for such an internal, predictive model of the world in early visual regions has also been obtained recently in humans [37]. Building on this, mental imagery might entail the generation of such an internal model, with top-down biasing signals projecting to visual areas in order to sharpen upcoming perception, leading to a similar overall level of activation in visual regions during imagery and rest. The idea that mental imagery plays a functional role in facilitating future perception is supported by a recent study that found that mental imagery biases subsequent perception in a binocular rivalry task [35, 38], as well as by the correlation between IM performance and representational precision (Figure S2).

In conclusion, we observed analogous sensory representations during visual working memory and mental imagery in early visual cortex. Crucially, these activity patterns resembled those evoked by bottom-up visual stimulation, suggesting that mental images are "perception-like" in nature. These findings provide empirical support for the notion that visual cortex acts as a blackboard that is used during both bottom-up stimulus processing and top-down internal generation of mental content.

Experimental Procedures

The experimental procedures are summarized briefly throughout the Results and are presented in complete detail in the Supplemental Experimental Procedures.

Supplemental Information

Supplemental Information includes three figures, two tables, and Supplemental Experimental Procedures and can be found with this article online at http://dx.doi.org/10.1016/j.cub.2013.05.065.

Acknowledgments

This study was supported by the Netherlands Organisation for Scientific Research (NWO Brain & Cognition 433-09-248).

Received: February 22, 2013
Revised: May 10, 2013
Accepted: May 31, 2013
Published: July 18, 2013

References

1. Harrison, S.A., and Tong, F. (2009). Decoding reveals the contents of visual working memory in early visual areas. Nature *458*, 632–635.
2. Christophel, T.B., Hebart, M.N., and Haynes, J.D. (2012). Decoding the contents of visual short-term memory from human visual and parietal cortex. J. Neurosci. *32*, 12983–12989.
3. Riggall, A.C., and Postle, B.R. (2012). The relationship between working memory storage and elevated activity as measured with functional magnetic resonance imaging. J. Neurosci. *32*, 12990–12998.
4. Serences, J.T., Ester, E.F., Vogel, E.K., and Awh, E. (2009). Stimulus-specific delay activity in human primary visual cortex. Psychol. Sci. *20*, 207–214.
5. Klein, I., Paradis, A.-L., Poline, J.-B., Kosslyn, S.M., and Le Bihan, D. (2000). Transient activity in the human calcarine cortex during visual-mental imagery: an event-related fMRI study. J. Cogn. Neurosci. *12*(*Suppl 2*), 15–23.
6. Knauff, M., Kassubek, J., Mulack, T., and Greenlee, M.W. (2000). Cortical activation evoked by visual mental imagery as measured by fMRI. Neuroreport *11*, 3957–3962.
7. Kaas, A., Weigelt, S., Roebroeck, A., Kohler, A., and Muckli, L. (2010). Imagery of a moving object: the role of occipital cortex and human MT/V5+. Neuroimage *49*, 794–804.
8. Haynes, J.-D., and Rees, G. (2006). Decoding mental states from brain activity in humans. Nat. Rev. Neurosci. *7*, 523–534.
9. Kamitani, Y., and Tong, F. (2005). Decoding the visual and subjective contents of the human brain. Nat. Neurosci. *8*, 679–685.
10. Bullier, J. (2001). Feedback connections and conscious vision. Trends Cogn. Sci. *5*, 369–370.
11. Mumford, D. (1991). On the computational architecture of the neocortex. I. The role of the thalamo-cortical loop. Biol. Cybern. *65*, 135–145.
12. Baddeley, A. (2003). Working memory: looking back and looking forward. Nat. Rev. Neurosci. *4*, 829–839.
13. Kosslyn, S.M., Thompson, W.L., and Ganis, G. (2006). The Case for Mental Imagery, *Volume 39* (Oxford: Oxford University Press).
14. Kosslyn, S.M., Ganis, G., and Thompson, W.L. (2001). Neural foundations of imagery. Nat. Rev. Neurosci. *2*, 635–642.
15. Slotnick, S.D., Thompson, W.L., and Kosslyn, S.M. (2005). Visual mental imagery induces retinotopically organized activation of early visual areas. Cereb. Cortex *15*, 1570–1583.
16. Pylyshyn, Z.W. (2003). Return of the mental image: are there really pictures in the brain? Trends Cogn. Sci. *7*, 113–118.
17. Pylyshyn, Z.W. (2002). Mental imagery: in search of a theory. Behav. Brain Sci. *25*, 157–182, discussion 182–237.
18. Friston, K. (2012). Ten ironic rules for non-statistical reviewers. Neuroimage *61*, 1300–1310.
19. Stokes, M., Thompson, R., Cusack, R., and Duncan, J. (2009). Top-down activation of shape-specific population codes in visual cortex during mental imagery. J. Neurosci. *29*, 1565–1572.
20. Lee, S.-H., Kravitz, D.J., and Baker, C.I. (2012). Disentangling visual imagery and perception of real-world objects. Neuroimage *59*, 4064–4073.
21. Emrich, S.M., Riggall, A.C., Larocque, J.J., and Postle, B.R. (2013). Distributed patterns of activity in sensory cortex reflect the precision of multiple items maintained in visual short-term memory. J. Neurosci. *33*, 6516–6523.
22. Ester, E.F., Anderson, D.E., Serences, J.T., and Awh, E. (2013). A neural measure of precision in visual working memory. J. Cogn. Neurosci. *25*, 754–761.
23. Moro, S.I., Tolboom, M., Khayat, P.S., and Roelfsema, P.R. (2010). Neuronal activity in the visual cortex reveals the temporal order of cognitive operations. J. Neurosci. *30*, 16293–16303.
24. Pasternak, T., and Greenlee, M.W. (2005). Working memory in primate sensory systems. Nat. Rev. Neurosci. *6*, 97–107.
25. Horikawa, T., Tamaki, M., Miyawaki, Y., and Kamitani, Y. (2013). Neural decoding of visual imagery during sleep. Science *340*, 639–642.
26. 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.
27. Kosslyn, S.M., and Thompson, W.L. (2003). When is early visual cortex activated during visual mental imagery? Psychol. Bull. *129*, 723–746.
28. Duncan, J. (2010). The multiple-demand (MD) system of the primate brain: mental programs for intelligent behaviour. Trends Cogn. Sci. *14*, 172–179.
29. Rainer, G., Rao, S.C., and Miller, E.K. (1999). Prospective coding for objects in primate prefrontal cortex. J. Neurosci. *19*, 5493–5505.
30. Salazar, R.F., Dotson, N.M., Bressler, S.L., and Gray, C.M. (2012). Content-specific fronto-parietal synchronization during visual working memory. Science *338*, 1097–1100.
31. Miller, E.K., and Cohen, J.D. (2001). An integrative theory of prefrontal cortex function. Annu. Rev. Neurosci. *24*, 167–202.
32. Ridderinkhof, K.R., Ullsperger, M., Crone, E.A., and Nieuwenhuis, S. (2004). The role of the medial frontal cortex in cognitive control. Science *306*, 443–447.
33. Çukur, T., Nishimoto, S., Huth, A.G., and Gallant, J.L. (2013). Attention during natural vision warps semantic representation across the human brain. Nat. Neurosci. *16*, 763–770.
34. Kenet, T., Bibitchkov, D., Tsodyks, M., Grinvald, A., and Arieli, A. (2003). Spontaneously emerging cortical representations of visual attributes. Nature *425*, 954–956.
35. Kosslyn, S.M., Thompson, W.L., Kim, I.J., and Alpert, N.M. (1995). Topographical representations of mental images in primary visual cortex. Nature *378*, 496–498.
36. 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.
37. 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.
38. Pearson, J., Clifford, C.W., and Tong, F. (2008). The functional impact of mental imagery on conscious perception. Curr. Biol. *18*, 982–986.

Current Biology, Volume *23*

Supplemental Information

Shared Representations for Working Memory and Mental Imagery in Early Visual Cortex

Anke Marit Albers, Peter Kok, Ivan Toni, H. Chris Dijkerman, and Floris P. de Lange

Supplemental Inventory

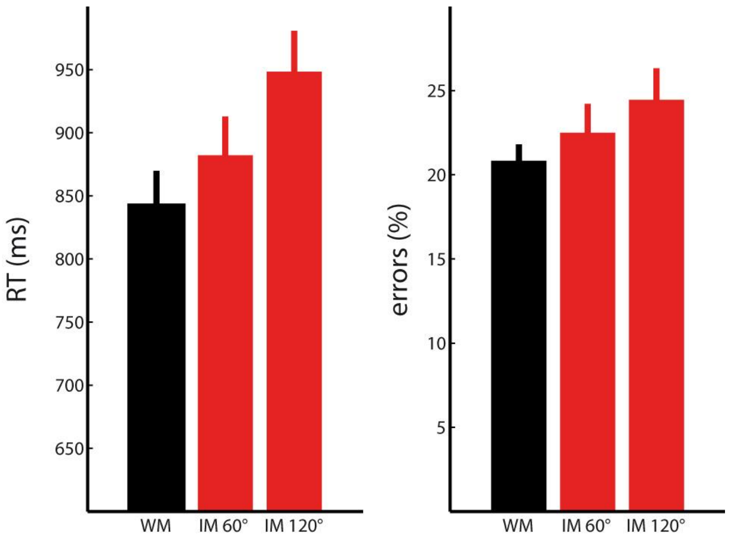

Figure S1. Reaction Times and Accuracy Scores
Reaction times (left) and error rates (right), split for working memory (WM, in black), and mental imagery (IM, either 60 degrees or 120 degrees rotation). Error bars denote standard error of the mean. Reaction times increased significantly with rotation angle (*F* (2, 23) = 19.45, *p* < 0.001), while there were no significant differences in error rate between conditions (*F* (2, 23) = 1.57, *p* = 0.22).

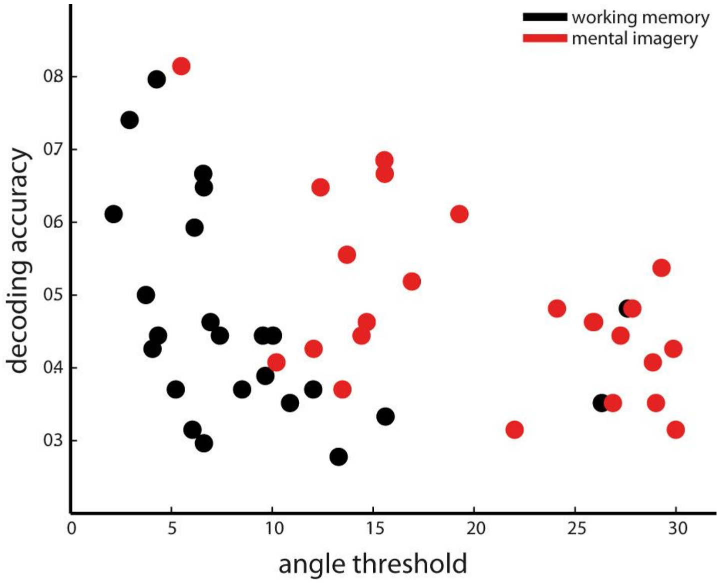

Figure S2. Correlation between Angle Threshold and Decoding Accuracy
There was a monotonic relationship between inter-individual differences in decoding accuracy and task performance (measured as threshold angle difference with the probe (staircase procedure) for both working memory and mental imagery. Decoding accuracy was higher for people that were more accurate when maintaining (WM: Spearman's rho = -0.51, *p* = 0.0053) or imagining (IM: rho = -0.37, *p* = 0.039) mental content. Dots represent single subjects, with one black (WM) and one red (IM) dot per participant.

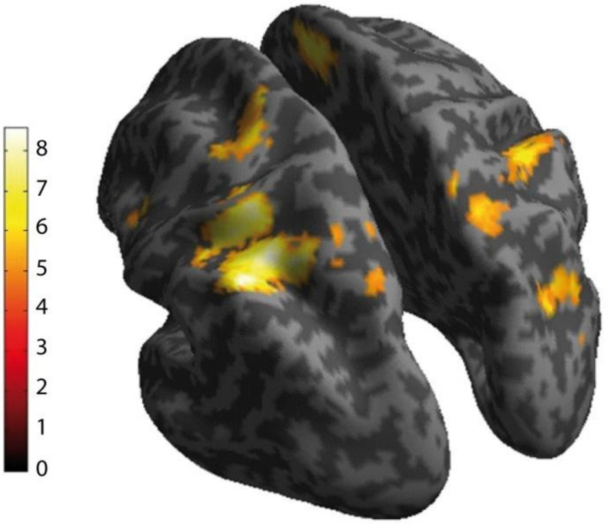

Figure S3. Areas with Increased BOLD Response during the Delay Period
Areas with a sustained BOLD response over the delay period during both imagery and working memory (WM delay & IM delay > baseline). To investigate whether information was restricted to visual areas or present throughout the brain, we created ROIs from these regions that showed elevated neural activity during the delay and tried to classify stimulus identity in these areas. Size and coordinates of the different areas, as well as decoding accuracies in these areas can be found in Table S2.

Table S1. Decoding Accuracies for the Different Classifiers in V1, V2, V3, and V1-V3

Area	Decoding within condition		Decoding across conditions		Decoding based on perception	
	WM-WM	IM-IM	IM-WM	WM-IM	VS-WM	VS-IM
V1	44.6%**	41.6%**	41.3%***	38.7%**	40.1**	43.2%***
V2	50.8%***	44.0%**	43.6%***	41.7%**	45.2%***	46.1%***
V3	52.4%***	46.0%**	44.6%***	42.5%**	44.4%***	46.1%***
V1-V3	54.2%***	46.1%**	45.5%***	45.2%***	46.4%***	48.5%***

Decoding accuracies calculated over 8-10 seconds after onset of the delay period. Significance levels: *p<0.05; **p<0.01; ***p<0.001; non-significant classification in grey. To check for differences between areas in perception-based decoding, we performed a 2-way repeated-measures ANOVA with area and condition (WM, IM) as factors. Although there were overall differences in decoding accuracy between areas (F (2,46) = 6.48, p = 0.0033; due to overall slightly better decoding accuracy in V2), there was no overall difference in decoding accuracy between tasks (F (1,23) = 0.95, p = 0.34), nor an interaction between area and task (F (2,46) = 0.40, p = 0.67). This lack of interaction indicates that there were no differences between IM and WM tasks that were specific to particular visual areas.

Table S2. Decoding Accuracies in the Regions Showing Delay-Related Activation

Area [MNI coordinates]	Z (k)	Decoding within condition		Decoding across conditions		Decoding based on perception	
		WM-WM	IM-IM	IM-WM	WM-IM	VS-WM	VS-IM
L intraparietal sulcus [-42, -46, 52]	6.66 (620)	35.5%	36.6%*	36.5%*	37.8%***	33.1%	33.8%
Superior frontal gyrus [9, 14, 49]	5.81 (470)	36.0%*	38.6%**	35.5%	37.4%**	33.0%	33.1%
R supramarginal gyrus [39, -37, 46]	5.61 (170)	36.9%*	36.3%	34.8%	34.8%	32.9%	33.9%
R inferior frontal gyrus [51, 8, 19]	5.52(81)	34.2%	35.7%	34.0%	33.9%	34.4%	33.5%
R superior parietal gyrus [9, -64, 64]	5.47 (76)	34.4%	37.0%*	35.4%	36.0%**	32.9%	33.7%
L precentral sulcus [-51, 5, 31]	5.32 (72)	35.1%	32.9%	36.3%*	35.3%	33.1%	33.0%
R intraparietal sulcus [27, -70, 22]	5.23 (60)	36.1%	34.8%	35.8%*	35.8%*	33.3%	33.3%

Decoding accuracies for the different classifiers in areas that showed significant activation (*p* < 0.05 FWE corrected, cluster size > 50 voxels) during the 8-10 seconds after onset of the delay period of both working memory [WM > baseline] and mental imagery [IM > baseline] trials. MNI coordinates are indicated between square brackets, Z-score (Z) and extent (k) are indicated in the second column. Significance levels: *p<0.05; **p<0.01; ***p<0.001; non-significant classification in grey.

Supplemental Experimental Procedures

Participants

Thirty right-handed participants with normal or corrected-to-normal vision were recruited from the student population at the Radboud University in Nijmegen. Participants gave written informed consent in accordance with the institutional guidelines of the local ethical committee (CMO region Arnhem-Nijmegen, The Netherlands) and were paid for their participation. All participants were trained on the task in a behavioral setting for approximately one hour. After training, six participants were excluded from further participation due to a failure to understand the task or to reach sufficient performance (inclusion threshold: 75% correct with a difference between probe and target of < 30°). The remaining 24 participants (10 male, ages 18-30) participated in the scanning session and were included in all analyses.

Stimuli

Stimuli were grayscale luminance-defined sinusoidal gratings generated using MATLAB (MathWorks, Natick, MA) in conjunction with the Psychophysics Toolbox [1]. The gratings were presented in an annulus (outer diameter: 15° of visual angle, inner diameter: 3° of visual angle) surrounding a central fixation point. The gratings had a spatial frequency of 1 cpd, a Michelson contrast of 80% and orientation of either 15°, 75° and 135° degrees from the vertical axis. Stimuli were displayed on a rear-projection screen using an EIKI (EIKI, Rancho Santa Margarita, CA) projector (1,024 x 768 resolution, 60 Hz refresh rate).

As there were 3 possible starting orientations, and, for mental imagery (IM) trials, two directions (clockwise, counter-clockwise) and rotation magnitudes (60° or 120°), there were also 3 final grating orientations: 15°, 75° and 135°. Due to the large number of stimulus, task and cue combinations, there were 60 unique trials (5 different task cues (0, >, >>, <, <<; representing the amount and direction of rotation), 6 combinations of presented stimuli ([15 75], [15 135], [75 15], [75 135], [135 0], [135 75]) and 2 retro-cues (select either the first or the second stimulus). There were 20 trials for each grating orientation per task (WM, IM), resulting in a total of 120 trials, which were randomly intermixed and divided over 6 runs (20 trials per run). All trials were included in the analyses. To investigate whether participants became more familiar with the stimuli over time, we split the experiment in half and investigated the separate decoding accuracies. There was no difference in decoding accuracy of the neural representations for stimuli presented in the first 3 blocks compared to the last 3 blocks in any of the visual areas (all $p > 0.05$).

Staircase Procedure

We used a staircase procedure to ensure equal task difficulty levels for WM and IM trials. After each WM or IM delay period, subjects had to indicate whether a probe stimulus was rotated slightly clockwise or counterclockwise with respect to the internal image. The staircase procedure estimated the difference between probe and mental image that ensured 75% performance, using QUEST [1]. The staircase was seeded with an orientation difference of 15° and dynamically adapted based on subjects' accuracy. We imposed an upper limit on the orientation difference, ensuring a maximal difference between internal image and probe of 30°.

At roughly matched performance levels (WM: 77% ± 7.1%; IM: 79% ± 4.8%; t (23) =-1.57,p=0.13), the average orientation difference between mental image and probe was smaller for WM trials (9.0° ± 1.5°) than for IM trials (20.4° ± 2.8°; t(23) = -7.4, p < 0.001), indicating that participants were a bit more accurate when they had to maintain the presented image than when they had to mentally rotate and generate a new image.

Additional Localizer Scans

After the main experiment, participants underwent two additional scanning runs. To obtain data for the perceptual classifier, the same gratings that were used during the main experiment were presented for longer durations (12 seconds), during which each grating was flashed at 4 Hz. We collected fifteen blocks, with pseudo-random order of the orientations. The fixation dot changed 8-10 times per block at random time points, leading to an average number of changes of 150 (range 142-159), to which participants responded on average 94+/-4% (mean +/- SD) of the time. After each block of three orientations, a baseline period of 15 seconds was presented. Throughout the localizer, participants had to monitor the fixation dot for occasional brief changes in color, to which they had to respond with a button press. The same task was applied during the retinotopic mapping session, in which subjects viewed a wedge, consisting of a flashing black-and-white checkerboard pattern (3 Hz), first rotating clockwise for 9 cycles and then anticlockwise for another 9 cycles (at a rotation speed of 24 s/cycle).

fMRI Acquisition Parameters

Functional images were acquired using a 3T Trio MRI system (Siemens, Erlangen, Germany), using a T2*-weighted gradient-echo EPI sequence (FA = 80 degrees, FOV = 64x64x31 voxels, voxels size 3x3x3 mm, TR/TE = 2000/30 ms). Structural images were acquired using a T1-weighted MP-Rage sequence (FA = 8 degrees, FOV = 192x256x256, voxel size 1x1x1, TR/TE = 2300/3/03ms).

Data Extraction

All analyses were performed on an individual, per subject basis. We used Freesurfer (http://surfer/nmr/mgh/harvard/edu/) to draw the borders of V1, V2 and V3 [2-5]. Regions of interest (ROIs) were created for each early visual area. Due to confluent foveal representations for V1-V3 [5], we excluded the foveal representation from our regions of interest. Within each ROI, the 120 most active voxels were selected based on the independent localizer. For the combined early visual cortex ROI, all the voxels from the individual V1-V3 ROIs were selected. We extracted the BOLD time course for each voxel in the ROIs and high pass-filtered the data (removing signal with f<1/128 Hz). All reported results are based on V1-V3 combined, unless specified otherwise. Over all participants, slightly more voxels were selected from dorsal than ventral V2 and V3 (V2d: 64 voxels, V2v: 56 voxels, $t(23) = 1.69$, $p = 0.105$; V3d: 65 voxels, V3v: 55 voxels, $t(23) = 2.60$, $p = 0.016$).

Classification Analysis

For all multivariate pattern analyses (MVPA) we trained linear support vector machines (SVMs) to discriminate between the three grating orientations based on the pattern of BOLD activity over voxels. Classification accuracy can be seen as an indication of the amount of orientation information available in the BOLD signal, such that relative changes therein can be informative about the stimulus being maintained (WM) or imagined (IM) [6, 7]. When training and testing on WM and IM trials we averaged the activity over time points 8-12 seconds after onset of the delay period of the working memory trials (2 separate scans). We selected 8-12 seconds to maximize the temporal distance from stimulus related activity evoked by the two stimuli presented at the start of each trial, yet avoid contamination with activity elicited by the probe presentation. After averaging, data were normalized using a Z transformation. We trained the classifier using a leave-one-run-out procedure where we trained on stimuli in 5 out of 6 runs and tested on stimuli in the remaining run when training on WM, IM, or performed generalization between WM and IM. For generalization from perception to IM and WM we trained the 'perceptual classifier' using the average BOLD signal 5-13 s after the onset of each block in the localizer. When investigating the temporal unfolding of the representations, we tested each time point independently, without any averaging.

Given the three-class problem, we combined the classification results from independent support vector machine classifiers for each pair of gratings, following the procedure described by Kamitani and Tong [8, 9]. To test for significant decoding accuracy we performed one-sample t-tests over participants against chance level (33.3%). We also calculated the effect size of all our effects using Cohen's d. To compare the different areas and conditions (WM, IM), we performed a 2-way repeated-measures ANOVA over participants, with area and condition as factors. To investigate whether decoding accuracy was higher for people that could more accurately form mental images, we calculated the one-tailed Spearman's correlation coefficient between threshold angle and decoding accuracy for WM and IM separately.

Mean Activity Analysis

To investigate which areas were active during both working memory and mental imagery periods, we additionally performed a whole-brain univariate analysis, in the framework of the General Linear Model (GLM). Individual data were realigned, co-registered, normalized, smoothed and high-pass filtered at 128 Hz, using SPM8 (http://www.fil.ion.ucl.ac.uk/spm, Wellcome Trust Centre for Neuroimaging, London, UK). We separately modeled the onset of each trial, the maintenance/imagery (delay) period, and the response period, for IM and WM separately. To quantify activity during the maintenance/imagery period, we created contrasts between these regressors and the implicit baseline (intertrial interval). Head-motion parameters were included as nuisance regressors. Second level analysis consisted of a conjunction analysis testing for common activity during the delay period of WM and IM trials, compared to baseline. We thresholded this map using stringent methods for multiple comparisons (voxel-wise family-wise error correction p<0.05) and considered only clusters with a spatial extent of >50 voxels. To investigate whether there was stimulus information in the pattern of activity in the regions that showed activity during this delay period, we created ROIs from these active regions, and converted them back to native space for each participant, after which we performed the classification analysis as described above.

Supplemental References

1. Brainard, D.H. (1997). The Psychophysics Toolbox*. Spatial Vision *10*, 433-436.
2. Engel, S.A., Glover, G.H., and Wandell, B.A. (1997). Retinotopic organization in human visual cortex and the spatial precision of functional MRI. Cereb. Cortex *7*, 181-192.
3. Sereno, M.I., Dale, A.M., Reppas, J.B., Kwong, K.K., Belliveau, J.W., Brady, T.J., Rosen, B.R., and Tootell, R.B.H. (1995). Borders of multiple visual areas in humans revealed by functional magnetic resonance imaging. Science *268*, 889-893.
4. DeYoe, E.A., Carman, G.J., Bandettini, P., Glickman, S., Wieser, J., Cox, R., Miller, D., and Neitz, J. (1996). Mapping striate and extrastriate visual areas in human cerebral cortex. Proc. Natl. Acad.Sci. USA *93*, 2382-2386.
5. Wandell, B.A., Dumoulin, S.O., and Brewer, A.A. (2007). Visual field maps in human cortex. Neuron *56*, 366-383.
6. Kok, P., Jehee, Janneke F.M., and de Lange, Floris P. (2012). Less is more: Expectation sharpens representations in the primary visual cortex. Neuron *75*, 265-270.
7. Jehee, J.F.M., Brady, D.K., and Tong, F. (2011). Attention improves encoding of task-relevant features in the human visual cortex. J. Neurosci. *31*, 8210-8219.
8. Kamitani, Y., and Tong, F. (2005). Decoding the visual and subjective contents of the human brain. Nat. Neurosci. *8*, 679-685.
9. Kamitani, Y., and Tong, F. (2006). Decoding seen and attended motion directions from activity in the human visual cortex. Curr. Biol. *16*, 1096-1102.