The resilience of object predictions: Early recognition across viewpoints and exemplars

Olivia S. Cheung · Moshe Bar

Abstract Recognition of everyday objects can be facilitated by top-down predictions. We have proposed that these predictions are derived from rudimentary image information, or *gist*, extracted rapidly from the low spatial frequencies (LSFs) (Bar Journal of Cognitive Neuroscience 15: 600–609, 2003). Because of the coarse nature of LSF representations, we hypothesized here that such predictions can accommodate changes in viewpoint as well as facilitate the recognition of visually similar objects. In a repetition-priming task, we indeed observed significant facilitation of target recognition that was primed by LSF objects across moderate viewpoint changes, as well as across visually similar exemplars. These results suggest that the LSF representations are specific enough to activate accurate predictions, yet flexible enough to overcome small changes in visual appearance. Such gist representations facilitate object recognition by accommodating changes in visual appearance due to viewing conditions, and help generalize from familiar to novel exemplars.

Humans can identify objects at a glance. It has been proposed that in order to achieve such fast and accurate recognition, the visual system uses coarse information to rapidly activate a

O. S. Cheung · M. Bar
Martinos Center, Massachusetts General Hospital,
Harvard Medical School, Boston, MA, USA

M. Bar
Gonda Center for Multidisciplinary Brain Research,
Bar-Ilan University, Ramat Gan, Israel

O. S. Cheung (✉)
Department of Psychology, Harvard University, 33 Kirkland Street,
Cambridge, MA 02138, USA
e-mail: sccheung.olivia@gmail.com

rough estimate about the content of a visual input (e.g., "a dog?" "a kitchen?"; Bar 2003; Neisser 1967; Oliva 2005; Potter 1975). This approximation of visual input at a glance, or *gist* (Oliva 2005), can be verified and refined when fine image details are processed. By first relying on gist, observers can extract meaningful information from complex visual inputs (Bar 2004; Potter 1975). For instance, observers can categorize scenes or faces on the basis of the overall image structure (Schyns and Oliva 1994; Schyns and Oliva 1999) and can detect whether a particular type of object is present in a scene (Davenport and Potter 2004; Fabre-Thorpe, Delorme, Marlot, and Thorpe 2001; Joubert, Rousselet, Fize, and Fabre-Thorpe 2007; Mack and Palmeri 2010), even when the images are presented briefly (e.g., for 20–80 ms). The concept of gist is central to visual cognition, but it remains elusive how exactly gist may facilitate object recognition.

Objects are often quickly recognized at the basic level (e.g., dogs, cats), which is the level that is most informative for discrimination among categories (as compared with the superordinate level—e.g., living vs. nonliving things—or the subordinate level—e.g., beagle vs. collie; Jolicœur, Gluck, and Kosslyn 1984; Rosch, Mervis, Gray, Johnson, and Boyes-Braem 1976). Because objects within a basic-level category typically share similar features, whereas objects from different basic-level categories have more distinct features (Rosch et al. 1976), it is often possible to recognize objects at the basic level on the basis of the knowledge about the global object shape (Bar 2003; see also Oliva and Torralba 2006). It is also possible that memory representation of an object category evolves from previous encounters with various instances from that category, and that it subserves recognition of new instances of objects from that category. During early processing, a gist representation derived from the *low spatial frequencies* (LSFs) of images may be particularly useful for generating initial guesses about object identity (Bar 2003), since LSFs are often perceptually available faster than fine image details, or *high spatial frequencies*

(HSFs; Breitmeyer 1975; Gish, Shulman, Sheehy, and Leibowitz 1986). We propose here that a gist representation will be flexible and resilient to slight changes in visual details. Specifically, LSF information may accommodate the differences between exemplars of a category and between adjacent views of a single object. Although HSF information is useful for differentiating similar exemplars within a category (e.g., beagle, collie; Collin and McMullen 2005), LSF information may be sufficient for a generic estimate of the category to facilitate recognition (Bar 2003).

The nature of object representation can be examined with a priming paradigm, in which briefly presented images may facilitate the recognition of subsequent stimuli. Specifically, a prime image may activate relevant representations that are useful for rapid recognition of a target image. Facilitation in object priming has been observed across viewpoints and exemplar variations (Biederman and Cooper 1991; Harris and Dux 2005; Harris, Dux, Benito, and Leek 2008; Koutstaal et al. 2001; Simon, Koutstaal, Prince, Wagner, and Schacter 2003). For instance, object representations appears to be invariant to viewpoint changes at early stages of processing (Harris and Dux 2005), although object recognition may be susceptible to viewpoint changes during later processing (Harris et al. 2008). Whereas priming is strongest for identical images (i.e., repetition priming), it can also be observed for different exemplars of a category (Biederman and Cooper 1991; Koutstaal et al. 2001; Simons et al. 2003; but see Harvey and Burgund 2012; Vuilleumier, Henson, Driver, and Dolan 2002).

To test our hypothesis that the facilitation of object processing based on LSF gist can accommodate appearance changes, we asked to what extent briefly presented images with LSF information might produce priming across viewpoints and exemplar variations. In the viewpoint experiment, we tested whether faster performance would be obtained with LSF primes, when the prime and target were of an identical object across slight viewpoint changes, as compared with when the prime and target were different objects. In the exemplar experiment, we measured performance when the prime and target were of different exemplars within a basic-level category (e.g., two types of dogs), as compared with when the prime and target were objects from different basic-level categories, to examine whether LSF facilitation is based on visual or semantic features. If the facilitation is based primarily on the global shape of an object, then facilitation should only be expected for visually similar exemplars (e.g., collie and golden retriever). In contrast, if semantic information about an object category is utilized during early processing, priming should also be found for exemplars that are visually dissimilar (e.g., collie and Chihuahua).

It is important to note that HSF information may also be sufficient to facilitate early object processing. Specifically, a recent study (de Gardelle and Kouider 2010) reported stronger priming for HSF than for LSF in a face judgment task.

However, the HSF stimuli were more visible than the LSF stimuli in that study, and might thus have resulted in stronger priming. Moreover, it is unclear whether any HSF facilitation might be specific to the exact image details, or whether it can also accommodate slight image changes. Because our question of interest was whether object priming across image changes may depend on LSF or HSF, we first equated the visibility of LSF and HSF in a pilot study to prevent any possible confound due to visibility. In the main priming study, we then examined how LSF and HSF information might accommodate variations in image details to facilitate subsequent object processing.

Method

Participants

A group of 24 young adults from the Harvard University community participated in both viewpoint and exemplar experiments for payment. The order of the experiments was counterbalanced across participants. The data from one participant were discarded because the individual's overall RT was two standard deviations slower than the group average.

Stimuli

The viewpoint and exemplar experiments each used 96 everyday objects and 192 abstract sculptures. The viewpoint experiment involved three consecutive views (0º, 30º, or 60º rotation in depth) of 96 objects, obtained from www.tarrlab.org. Stimulus images courtesy of Michael J. Tarr, Center for the Neural Basis of Cognition and Department of Psychology, Carnegie Mellon University. The exemplar experiment was based on three exemplars of 96 objects (one identical, one visually similar, and one visually dissimilar), obtained from http://cvcl.mit.edu/MM/objectCategories.html, which were originally used in Konkle, Brady, Alvarez, and Oliva (2010). Perceptual similarity ratings of the object shapes (on a scale of 1–5), collected from a separate participant group ($n = 22$), showed higher similarity for the similar than for the dissimilar pairs (similar, $M = 3.43$, $SD = 0.57$; dissimilar, 2.10, $SD = 0.71$), $t(21) = 11.86$, $p < .0001$. To evaluate the influence of low-level visual properties, additional analyses of the priming study revealed that regressing out pixelwise similarity between image pairs did not affect the results. Images of abstract 3-D sculptures were obtained from several art websites. All of the images showed minimal occlusion of the object parts and were adjusted to be 256 × 256 pixels in size.

For each participant, different subsets of 32 objects were randomly selected for each spatial-frequency condition, with 24 objects as the *primes* and *targets* for the *same* trials, and the remaining eight objects used as the *targets* for the *different* trials. Sample primes and targets for the two experiments are illustrated in Fig. 1. The prime objects were shown in full spectrum (FS;
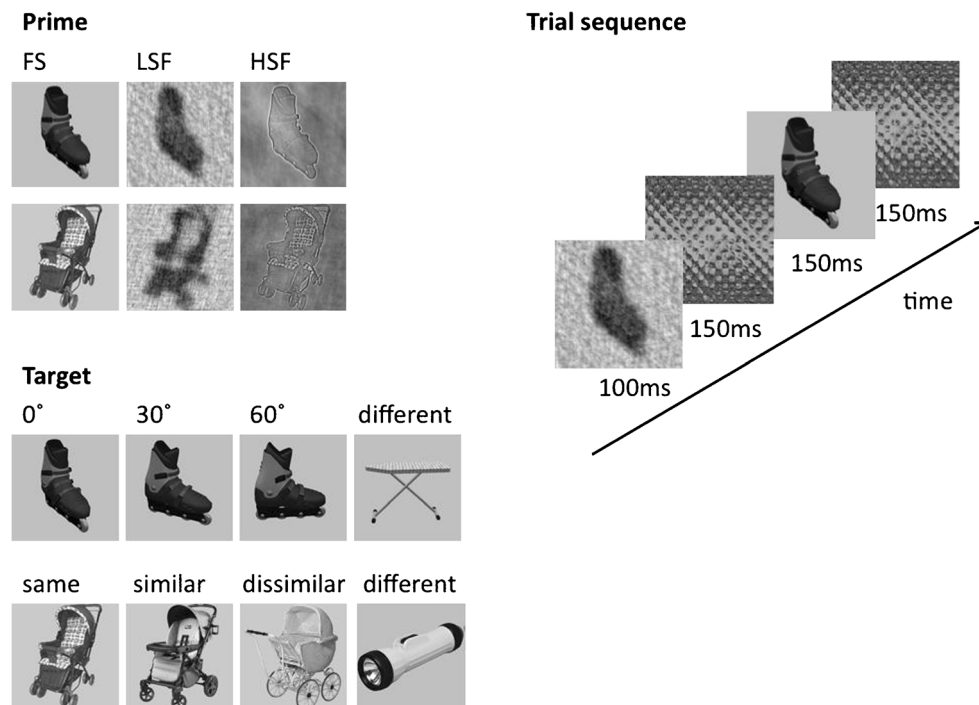
Fig. 1 Sample prime and target objects (left) and the trial sequence of the priming task (right). The prime objects could consist of full-spectrum (FS), low-spatial-frequency (LSF), or high-spatial-frequency (HSF) information. LSF- and HSF-filtered images were first normalized for comparable visibility in a separate experiment. The image contrast here has been adjusted for illustration purposes. In the "object" trials, the target objects could be identical to the prime objects but varied in orientation (in the viewpoint experiment), or they could be exemplars from the same category as the prime object (in the exemplar experiment), or they could be completely different objects from the prime objects. In each "nonobject" trial, the target image showed an abstract 3-D sculpture

unfiltered, containing LSF and HSF), in LSF (<8 cycle per image [cpi] or <1 cycle per degree [cpd]), or in HSF (40–48 cpi, 6–7 cpd). All target objects in both experiments were in FS. For the LSF and HSF objects, the relevant spatial frequencies remained intact, whereas the rest were phase-scrambled. Each participant sat 50 cm away from the monitor and rested his or her head on a chinrest. The primes and targets extended 4º and 5º of visual angle, respectively. In each experiment, each prime was shown six times: once for each of the four "object" conditions, and twice for the "sculpture" condition. Each target was shown once.

As we mentioned above, one possible confound that could potentially affect the priming results was differential visibility of the objects across the LSF and the HSF conditions. In a pilot study ($N = 9$), participants judged whether each LSF or HSF image showed an everyday object or an abstract sculpture, with all of the objects that would be used as primes included. This task allowed for an estimation of general recognition performance with the filtered images, which was appropriate for our goal to test the extent that the gist of visual input would facilitate recognition. Each image was presented for 100 ms and then followed by a 150-ms mask. We found that the HSF objects were more recognizable than the LSF objects [viewpoint: LSF, $d' = 1.45$, RT = 810 ms; HSF, $d' = 2.23$, RT = 707 ms; t tests, d', $t(8) = 3.51$, $p < .01$; RT, $t(8) = -2.14$, $p = .07$; exemplar: LSF, $d' = 0.81$, RT = 738 ms; HSF,

$d' = 1.44$, RT = 738.5 ms; t tests, d', $t(8) = 5.25$, $p < .001$; RT, $t(8) = 0.03$, $p = .98$]. Therefore, we attempted to equate the visibility of objects across the LSF and HSF conditions by adjusting the contrast of the phase-scrambled noise using MATLAB (increasing the noise contrast for the HSF viewpoint and exemplar stimuli by sharpening the extreme 30% and 10% values, respectively, and reducing the noise contrast of the 30% LSF exemplar stimuli). Another group of participants ($N = 16$) performed the object-versus-sculpture judgment task on the adjusted images and showed comparable recognition performance on the LSF and HSF stimuli [viewpoint: LSF, $d' = 1.46$, RT = 691 ms; HSF, $d' = 1.43$, RT = 713 ms; d', $t(15) = 0.39$, $p = .75$; RT, $t(15) = -0.95$, $p = .36$; exemplar: LSF, $d' = 1.34$, RT = 777 ms; HSF, $d' = 1.51$, RT = 774 ms; d', $t(15) = -1.83$, $p = .09$; RT, $t(15) = 0.09$, $p = .93$]. We then used this visibility-adjusted stimulus set in the main priming task.

In the priming task, participants were instructed to look at the first image and judge whether the second image showed a common object or an abstract sculpture (Fig. 1). The spatial-frequency conditions (FS, LSF, or HSF) were blocked. The other conditions (viewpoint: 0º, 30º, 60º, different, or sculpture; exemplar: same, similar, dissimilar, different, or sculpture) were randomized. In each experiment, each of the four "object" conditions (viewpoint: 0º, 30º, 60º, and different;

exemplar: same, similar, dissimilar, and different) had 24 trials, and the "sculpture" condition had 48 trials.

Results

First, discrimination performance in the object/sculpture judgment task was high across all spatial-frequency conditions in both the viewpoint and exemplar experiments (Table 1). One-way analyses of variance (ANOVAs) conducted on sensitivity (d') showed no significant effects of spatial frequency in either the viewpoint experiment, $F(2, 44) = 0.48$, $MSE = .17$, $\eta_p^2 = .02$, $p = .62$, or the exemplar experiment, $F(2, 44) = 1.61$, $MSE = .17$, $\eta_p^2 = .07$, $p = .21$.

Our primary interest was the priming effect within the "object" trials—that is, faster response times (RT) for 0º, 30º, and 60º or for the same, similar, and dissimilar trials, relative to RTs for the different-object trials (see Fig. 2). For RT, trials that required a response of "object" were analyzed in 3 (FS, LSF, HSF) × 4 (viewpoint: 0º, 30º, 60º, different; or exemplar: same, similar, dissimilar, different) ANOVAs. RT outliers (>2.5 SDs within each individual) were excluded (3% of the total trials). Priming effects were revealed by Bonferroni-corrected t tests comparing each of either the 0º, 30º, and 60º or the same, similar, and dissimilar conditions to the *different* conditions.

Viewpoint experiment

The effect of spatial frequency was significant, $F(2, 44) = 4.36$, $MSE = 3,668.8$, $\eta_p^2 = .15$, $p = .02$, revealing faster RTs for FS than for HSF ($p = .015$), whereas we found no difference between FS and LSF ($p > .38$) or between LSF and HSF ($p > .37$). A significant effect of viewpoint was also observed, $F(3, 66) = 26.56$, $MSE = 1,037.8$, $\eta_p^2 = .55$, $p < .0001$. The viewpoint effect was modulated by a Spatial Frequency × Viewpoint interaction, $F(6, 132) = 3.25$, $MSE = 1,270.7$, $\eta_p^2 = .13$, $p = .005$. Bonferroni-corrected comparisons revealed significant priming for FS and LSF across 0º (FS, $p < .0001$; LSF, $p = .001$), 30º (FS, $p < .0001$; LSF, $p = .046$), and 60º (FS, $p < .0001$; LSF, $p = .015$), but not for HSF (ps > .9), primes.

Table 1 Mean sensitivity (d') on the object/sculpture judgment task across the spatial frequency conditions in the viewpoint and exemplar experiments

	Full Spectrum (FS)	Low Spatial Frequency (LSF)	High Spatial Frequency (HSF)
Viewpoint experiment	3.38 (0.12)	3.27 (0.09)	3.35 (0.11)
Exemplar experiment	3.04 (0.12)	2.84 (0.06)	2.99 (0.11)

Standard errors of the means are shown in parentheses.

Exemplar experiment

A significant effect of exemplar was observed, $F(3, 66) = 23.30$, $MSE = 1,743.3$, $\eta_p^2 = .51$, $p < .0001$, but not for spatial frequency, $F(2, 44) = 0.35$, $MSE = 8,193.6$, $\eta_p^2 = .016$, $p = .70$. Although the Spatial Frequency × Exemplar interaction was not significant, $F(6, 132) = 1.35$, $MSE = 1,560.6$, $\eta_p^2 = .06$, $p = .24$, the significant 3 (FS, LSF, HSF) × 2 (same, different) interaction [$F(2, 44) = 3.31$, $MSE = 1,337.5$, $\eta_p^2 = .13$, $p < .05$] replicated a similar interaction in the viewpoint experiment. More importantly, Bonferroni-corrected comparisons revealed that for FS and LSF primes, priming was significant with identical (FS, $p < .0001$; LSF, $p = .003$) and similar (FS, $p = .0002$; LSF, $p = .026$) exemplars, but not with dissimilar exemplars (FS, $p > .87$; LSF, $p > .89$), whereas no significant priming effect emerged in any of the HSF conditions (ps > .79).

Discussion

We observed significant object priming, in accordance with our hypothesis that facilitation can arise from LSF information and that this facilitation accommodates some changes in visual appearance. Specifically, facilitation was found for repeated objects with up to 60º rotation as well as for visually similar exemplars. The lack of facilitation for visually dissimilar exemplars indicates that the priming was based on visual similarity and not on semantic category. All priming effects were observed with both FS primes (which contained LSFs and HSFs), and with the LSF primes. No significant facilitation was observed with HSF primes. These findings are consistent with the idea that the visual similarity of global shapes plays a key role in facilitating recognition processes, and they demonstrate that LSFs are critical elements toward building these representations.

To recognize objects quickly and accurately, it is advantageous to generate a set of limited but flexible predictions regarding the identity of an object during early visual processing (Bar 2003). By using briefly presented object primes, our results reveal that LSF information is involved in activating such processes. LSFs are extracted early (Schyns and Oliva 1994) and are projected rapidly and predominantly to the dorsal stream and also to the ventral stream (Ferrera, Nealey, and Maunsell 1992; Merigan and Maunsell 1993; Shapley 1990). It is possible that LSF information facilitates subsequent recognition either by bottom-up or local feedback process within the ventral stream (Ewbank, Henson, Rowe, Stoyanova, and Calder 2013; Ewbank et al. 2011), or by projecting the input to the orbitofrontal cortex through the dorsal magnocellular pathway to guide subsequent processing in the ventral visual cortex (Bar et al. 2006; Chaumon, Kveraga, Barrett, and Bar 2013; see also Li et al. 2010; Miller, Vytlacil, Fegen, Pradhan, and D'Esposito 2011). Our results
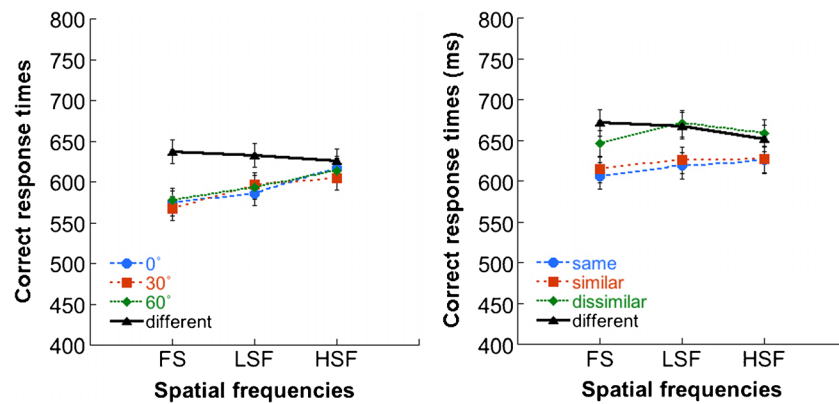
Fig. 2 Correct response times (RT) across spatial frequencies in the viewpoint conditions (left) and the exemplar conditions (right). Error bars represent the 95% confidence intervals of the 3 × 4 interaction

suggest that the mechanisms that support LSF facilitation in object processing should overcome slight changes in visual appearance. In other words, object recognition is facilitated when new or previously encountered objects fall under existing gist representations.

Although we observed priming for LSF primes, we did not find significant facilitation with HSF primes. However, we cannot rule out that facilitation could arise from HSF information, since this would require accepting the null hypothesis. As one piece of evidence in support of a special role for LSFs, we found that the repetition-priming effects were consistently stronger for LSF than for HSF across our experiments: A 2 (experiment: viewpoint/exemplar) × 2 (SF: LSF/HSF) × 2 (condition: same/different) ANOVA revealed a significant Spatial Frequency × Condition interaction, $F(1, 22) = 4.4$, $MSE = 9,671$, $\eta_p^2 = .17$, $p = .048$, showing larger priming for LSF than HSF, and no significant three-way interaction, $F(1, 22) < 1$. However, the relative roles of LSF and HSF may depend on task demands (Oliva and Schyns 1997). It is possible that HSF priming might occur in a task that required discrimination among similar exemplars of a category (Archambault, Gosselin, and Schyns 2000; Collin and McMullen 2005; Schyns, Bonnar, and Gosselin 2002), rather than the basic-level categorization task that we employed. Indeed, such task differences might explain the difference in the findings from a previous study (de Gardelle and Kouider 2010). The findings regarding the relative roles of LSF and HSF priming should require additional work to sort out. Importantly, however, our findings show that when the visibilities of LSF and HSF images are equated, LSFs can activate relevant, flexible representations to facilitate recognition.

Our finding of viewpoint generalization with FS and LSF primes is consistent with previous research revealing viewpoint-invariant performance during early stages of processing (Hamm and McMullen 1998; Harris et al. 2008; Murray 1998). The viewpoint invariance that we observed may stem from the facts that the range of viewpoints that we

tested was significant but limited (up to 60º) and that all main visual features of the objects were clearly shown in most of our stimuli (e.g., Biederman and Gerhardstein 1993; but see Hayward and Tarr 1997). Nonetheless, even though the exact image details varied, a blurry image containing global shape information was sufficient to facilitate subsequent recognition of an object across slight viewpoint changes.

We also found that similar global shapes led to comparable priming effects for identical and visually similar exemplars for both FS and LSF primes, although the local features of the visually similar exemplars were quite distinct. Although semantic information can be extracted from an image relatively quickly (e.g., Dell'Acqua and Grainger 1999; Potter 1975), it appears insufficient to facilitate the recognition of briefly presented exemplars that are visually distinct in shape. It is possible that during early stages of processing, recognizing objects on the basis of matching general visual features of LSF representations is more efficient than relying on semantic processing, and that the visual and semantic processes may rely on different neural substrates; for instance, the oribitofrontal cortex (Bar et al. 2006) and the right fusiform gyrus (Koutstaal et al. 2001; Simons et al. 2003) are more involved in visual processing, whereas the inferior frontal cortex (James and Gauthier 2004) and the left fusiform gyrus (Koutstaal et al. 2001; Simons et al. 2003) are more involved in semantic processing. Future research should further clarify the temporal dynamics of the frontal–temporal areas that may be involved in the processing of visual and semantic information for prediction and recognition (Ghuman, Bar, Dobbins, and Schnyer 2008; Grill-Spector, Henson, and Martin 2006; Segaert, Weber, de Lange, Petersson, and Hagoort 2013).

Taken together, our results suggest that LSF information is important in activating a "visual gist" to facilitate subsequent processing. The results demonstrate that the flexible nature of this representation allows for generalization across adjacent views of an object and across differences between exemplars of a category. By utilizing a generic representation, the visual

system appears to take advantage of the nature of LSF information and uses this information to generate predictions in a fast and flexible manner.

References

Archambault, A., Gosselin, F., & Schyns, P. (2000). A natural bias for the basic level. In L. R. Gleitman & A. K. Joshi (Eds.), *Proceedings of the 22nd Annual Conference of the Cognitive Science Society* (pp. 585–590). Mahwah, NJ: Erlbaum.

Bar, M. (2003). A cortical mechanism for triggering top-down facilitation in visual object recognition. *Journal of Cognitive Neuroscience, 15*, 600–609.

Bar, M. (2004). Visual objects in context. *Nature Reviews Neuroscience, 5*, 617–629. doi:10.1038/nrn1476

Bar, M., Kassam, K. S., Ghuman, A. S., Boshyan, J., Schmid, A. M., Schmidt, A. M., et al. (2006). Top-down facilitation of visual recognition. *Proceedings of the National Academy of Sciences, 103*, 449–454. doi:10.1073/pnas.0507062103

Biederman, I., & Cooper, E. E. (1991). Evidence for complete translational and reflectional invariance in visual object priming. *Perception, 20*, 585–593.

Biederman, I., & Gerhardstein, P. C. (1993). Recognizing depth-rotated objects: evidence and conditions for three-dimensional viewpoint invariance. *Journal of Experimental Psychology: Human Perception and Performance, 19*, 1162–1182. doi:10.1037/0096-1523.19.6.1162

Breitmeyer, B. G. (1975). Simple reaction time as a measure of the temporal response properties of transient and sustained channels. *Vision Research, 15*, 1411–1412.

Chaumon, M., Kveraga, K., Barrett, L.F., & Bar, M. (2013). Visual predictions in the orbitofrontal cortex rely on associative content. *Cerebral Cortex*. Advance online publication. doi:10.1093/cercor/bht146

Collin, C. A., & McMullen, P. A. (2005). Subordinate-level categorization relies on high spatial frequencies to a greater degree than basic-level categorization. *Perception & Psychophysics, 67*, 354–364. doi:10.3758/BF03206498

Davenport, J. L., & Potter, M. C. (2004). Scene consistency in object and background perception. *Psychological Science, 15*, 559–564. doi:10.1111/j.0956-7976.2004.00719.x

de Gardelle, V., & Kouider, S. (2010). How spatial frequencies and visual awareness interact during face processing. *Psychological Science, 21*, 58–66. doi:10.1177/0956797609354064

Dell'Acqua, R., & Grainger, J. (1999). Unconscious semantic priming from pictures. *Cognition, 73*, B1–B15.

Ewbank, M. P., Henson, R. N., Rowe, J. B., Stoyanova, R. S., & Calder, A. J. (2013). Different neural mechanisms within occipitotemporal cortex underlie repetition suppression across same and different-size faces. *Cerebral Cortex, 23*, 1073–1084.

Ewbank, M. P., Lawson, R. P., Henson, R. N., Rowe, J. B., Passamonti, L., & Calder, A. J. (2011). Changes in "top-down" connectivity underlie repetition suppression in the ventral visual pathway. *Journal of Neuroscience, 311*, 5635–5642.

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

Ferrera, V. P., Nealey, T. A., & Maunsell, J. H. R. (1992). Mixed parvocellular and magnocellular geniculate signals in visual area V4. *Nature, 358*, 756–758.

Ghuman, A. S., Bar, M., Dobbins, I. G., & Schnyer, D. M. (2008). The effects of priming on frontal-temporal communication. *Proceedings of the National Academy of Sciences, 105*, 8405–8409.

Gish, K., Shulman, G. L., Sheehy, J. B., & Leibowitz, H. W. (1986). Reaction times to different spatial frequencies as a function of detectability. *Vision Research, 26*, 745–747.

Grill-Spector, K., Henson, R., & Martin, A. (2006). Repetition and the brain: neural models of stimulus-specific effects. *Trends in Cognitive Sciences, 10*, 14–23. doi:10.1016/j.tics.2005.11.006

Hamm, J. P., & McMullen, P. A. (1998). Effects of orientation on the identification of rotated objects depend on the level of identity. *Journal of Experimental Psychology: Human Perception and Performance, 24*, 413–426.

Harris, I. M., & Dux, P. E. (2005). Orientation-invariant object recognition: evidence from repetition blindness. *Cognition, 95*, 73–93.

Harris, I. M., Dux, P. E., Benito, C. T., & Leek, E. C. (2008). Orientation sensitivity at different stages of object processing: evidence from repetition priming and naming. *PLoS ONE, 3*, e2256. doi:10.1371/journal.pone.0002256

Harvey, D. Y., & Burgund, E. D. (2012). Neural adaptation across viewpoint and exemplar in fusiform cortex. *Brain and Cognition, 80*, 33–44.

Hayward, W. G., & Tarr, M. J. (1997). Testing conditions for viewpoint invariance in object recognition. *Journal of Experimental Psychology: Human Perception and Performance, 23*, 1511–1521. doi:10.1037/0096-1523.23.5.1511

James, T. W., & Gauthier, I. (2004). Brain areas engaged during visual judgments by involuntary access to novel semantic information. *Vision Research, 44*, 429–439.

Joliœur, P., Gluck, M. A., & Kosslyn, S. M. (1984). Pictures and names: Making the connection. *Cognitive Psychology, 16*, 243–275. doi:10.1016/0010-0285(84)90009-4

Joubert, O. R., Rousselet, G. A., Fize, D., & Fabre-Thorpe, M. (2007). Processing scene context: Fast categorization and object interference. *Vision Research, 47*, 3286–3297.

Konkle, T., Brady, T. F., Alvarez, G. A., & Oliva, A. (2010). Conceptual distinctiveness supports detailed visual long-term memory. *Journal of Experimental Psychology: General, 139*, 558–578.

Koutstaal, W., Wagner, A. D., Rotte, M., Maril, A., Buckner, R. L., & Schacter, D. L. (2001). Perceptual specificity in visual object priming: Functional magnetic resonance imaging evidence for a laterality difference in fusiform cortex. *Neuropsychologia, 39*, 184–199.

Li, J., Liu, J., Liang, J., Zhang, H., Zhao, J., Rieth, C. A., et al. (2010). Effective connectivities of cortical regions for top-down face processing: A dynamic causal modeling study. *Brain Research, 1340*, 40–51. doi:10.1016/j.brainres.2010.04.044

Mack, M. L., & Palmeri, T. J. (2010). Modeling categorization of scenes containing consistent versus inconsistent objects. *Journal of Vision, 10*(3), 11,1–11. doi:10.1167/10.3.11

Merigan, W. H., & Maunsell, J. H. R. (1993). How parallel are the primate visual pathways? *Annual Review Neuroscience, 16*, 369–402.

Miller, B. T., Vytlacil, J., Fegen, D., Pradhan, S., & D'Esposito, M. (2011). The prefrontal cortex modulates category selectivity in human extrastriate cortex. *Journal of Cognitive Neuroscience, 23*, 1–10.

Murray, J. E. (1998). Is entry-level recognition viewpoint invariant or viewpoint dependent? *Psychonomic Bulletin & Review, 5*, 300–304.

Neisser, U. (1967). *Cognitive psychology*. New York, NY: Appleton-Century-Crofts.

Oliva, A. (2005). Gist of the scene. In L. Itti, G. Rees, & J. K. Tsotsos (Eds.), *Encyclopedia of the neurobiology of attention* (pp. 251–256). San Diego, CA: Elsevier.

Oliva, A., & Schyns, P. G. (1997). Coarse blobs or fine edges? Evidence that information diagnosticity changes the perception of complex visual stimuli. *Cognitive Psychology, 34*, 72–107. doi:10.1006/cogp.1997.0667

Oliva, A., & Torralba, A. (2006). Building the gist of a scene: The role of global image features in recognition. *Progress in Brain Research, 155*, 23–36.

Potter, M. C. (1975). Meaning in visual search. *Science, 187*, 965–966. doi:10.1126/science.1145183

Rosch, E., Mervis, C. B., Gray, W. D., Johnson, D. M., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology, 8*, 382–439. doi:10.1016/0010-0285(76)90013-X

Schyns, P. G., Bonnar, L., & Gosselin, F. (2002). Show me the features! Understanding recognition from the use of visual information. *Psychological Science, 13*, 402–409.

Schyns, P. G., & Oliva, A. (1994). From blobs to boundary edges: Evidence for time-and spatial-scale-dependent scene recognition. *Psychological Science, 5*, 195–200. doi:10.1111/j.1467-9280.1994.tb00500.x

Schyns, P. G., & Oliva, A. (1999). Dr. Angry and Mr. Smile: When categorization flexibly modifies the perception of faces in rapid visual presentations. *Cognition, 69*, 243—265.

Segaert, K., Weber, K., de Lange, F. P., Petersson, K. M., & Hagoort, P. (2013). The suppression of repetition enhancement: A review of fMRI studies. *Neuropsychologia, 51*, 59–66. doi:10.1016/j.neuropsychologia.2012.11.006

Shapley, R. (1990). Visual sensitivity and parallel retinocortical channels. *Annual Review of Psychology, 41*, 635–658.

Simons, J. S., Koutstaal, W., Prince, S., Wagner, A. D., & Schacter, D. L. (2003). Neural mechanisms of visual object priming: Evidence for perceptual and semantic distinctions in fusiform cortex. *NeuroImage, 19*, 613–626.

Vuilleumier, P., Henson, R. N., Driver, J., & Dolan, R. J. (2002). Multiple levels of visual object constancy revealed by event-related fMRI of repetition priming. *Nature Neuroscience, 5*, 491–499.