

Disruption of early face recognition processes by object substitution masking

Jason E. Reiss and James E. Hoffman

University of Delaware, Newark, DE, USA

Visual stimuli can fail to reach conscious awareness when surrounded by dots sharing a common onset but remaining visible after target removal, a phenomenon known as object substitution masking (OSM). We assessed the locus of substitution by recording the N170 component of the human event-related potential during a face/house classification task. Delayed-offset dots impaired target categorization, confirming the generality of object substitution masking by extending the effect to realistic, complex stimuli. Furthermore, object substitution masking eliminated the N170 amplitude difference between faces and houses, providing the first neural evidence that object substitution masking disrupts structural encoding stages during object recognition.

Single-unit recording in the visual system of the macaque monkey has revealed that neural information processing in the ventral pathway involves a hierarchy of cortical areas beginning in striate cortex and proceeding to higher-level visual areas in temporal cortex (Logothetis & Sheinberg, 1996; Rolls, 2000). In addition to this “feedforward” pathway, there is evidence for “feedback” processes in which information flows from higher areas back to primary visual cortex (Lamme & Roelfsema, 2000). In humans, transcranial magnetic stimulation experiments suggest that these feedback or reentrant processes play an important role in conscious awareness (Tong, 2003). The precise role of feedback in vision remains uncertain but an intriguing phenomenon known as object substitution masking (OSM; Di Lollo, Enns, & Rensink, 2000) may offer important clues.

OSM refers to impaired target identification when the target is accompanied by surrounding dots (the mask) that remain visible after the

Please address all correspondence to Jason E. Reiss, Department of Psychology, University of Delaware, Newark, Delaware, 19716, USA. E-mail: jreiss@udel.edu

This research was supported by National Science Foundation grants (MRI9977628 and BCS 0117475) to JEH. We would also like to thank John Serences for providing the house stimuli as well as Matt Doran, Vince Di Lollo, Geoff Woodman, and an anonymous reviewer for their helpful comments on earlier drafts.

target is removed. Di Lollo and colleagues (2000) point out that OSM appears to be fundamentally different than several other well-known varieties of masking such as metacontrast, integration, and interruption masking (Alpern, 1953; Breitmeyer, 1984; Turvey, 1973). For example, unlike metacontrast, effective masking does not depend on close proximity of mask and target (Di Lollo et al., 2000; Jiang & Chun, 2001; but see Bridgeman, 2006). Similarly, integration and interruption masking, which depend critically on the relative onset of mask and target, cannot explain why delayed *offset* of the mask would produce interference (Di Lollo, Bischof, & Dixon, 1993). Di Lollo et al. (2000) proposed instead that OSM is a new form of masking requiring a different explanation. They suggested that the initial onset of target and mask activates a forward sweep of neural activity resulting in a provisional hypothesis regarding the identity of the target. This initial hypothesis needs to be further refined through feedback or reentrant processes which may, for example, help parse objects from their surrounding context (Lamme, Zipser, & Spekreijse, 2002). When the masking dots persist beyond the offset of the target, these feedback processes are primarily concerned with the dots and the identity and even presence of the target is discarded.

Explanations of OSM in terms of reentrant mechanisms hold that masking occurs at relatively “early” stages in visual processing when the identity of the target is still being determined.¹ However, it is also possible that the masking dots exert their influence late in processing after the target has been identified but before it enters conscious awareness. This appears to be the case for the superficially similar attentional blink (AB) phenomenon in which presentation of two targets in rapid succession results in a failure to report the second target (Raymond, Shapiro, & Arnell, 1992). In this case, there is evidence for successful identification of the second target even when it cannot be reported. For example, the N400 component of the human event-related brain potential (ERP), which appears to be a marker for semantic processing (Kutas & Hillyard, 1980), is unaffected by the AB (Luck, Vogel, & Shapiro, 1996). In contrast, we reported that the N400 is abolished by OSM (Reiss & Hoffman, 2006), indicating that interference by substitution occurs at some stage prior to semantic analysis. This would encompass a variety of visual processing stages from early extraction of primitive features to categorization and identification. However, it does appear that interference in OSM does not occur at the earliest stage of

¹ The term “early processing” refers to a temporal, not anatomical, classification. Feedforward visual processing models define V1 as an early anatomical area whereas a reentrant processing model defines V1 temporally depending on when this area receives and processes information. As such, ERPs in the current study primarily address questions regarding the temporal nature of target processing and OSM.

feature analysis. Woodman and Luck (2003) showed that a target shape that was rendered unreportable by OSM nonetheless triggered a shift of visual attention to its location. In their experiment, targets and distracters were distinguished by basic shape features such as curvature, etc., so it appears that this initial level of feature extraction is unaffected by OSM.

The present study was designed to further bracket which visual processes between basic feature analysis and semantic analysis are affected by OSM. In particular, we examined the N170 component of the ERP which reflects structural encoding mechanisms that discriminate between various *object categories*, such as houses and faces, a process that occurs before information regarding meaning or identity has been determined (Bentin & Deouell, 2000). Evidence that OSM reduces the amplitude of the N170 would be consistent with accounts, such as the reentrance theory, that place the point of interference from the mask at early stages of visual processing concerned with basic categorization of visual objects.

METHOD

Participants

Fifteen native English-speaking, right-handed, neurologically normal volunteers (ages 18–28 years) were paid for their participation. All reported normal or corrected-to-normal colour vision and acuity and provided written informed consent. This experiment was approved by the University of Delaware Human Subjects Review Board. Three participants were removed from all analyses because they failed to demonstrate normal N170 responses (faces larger amplitude than houses) in the simultaneous condition, making it impossible to assess any potential effects of OSM.

Stimuli and procedure

Stimuli were generated on a Dell 3.6 GHz computer running custom software written with Blitz3D (Sibly, 2005) and presented on an UltraScan P990 19-inch CRT screen (1024 × 768 pixel resolution; 75 Hz frame rate). The fixation cross was presented in black (0.11 cd/m²) while the background was white (94.14 cd/m²). The stimulus set comprised of 15 faces selected, with permission, from the face database developed by the Max Planck Institute for Biological Cybernetics in Tübingen, Germany (Troje & Bulthoff, 1996) and 15 house stimuli, which were randomly chosen photographs of single-family homes.

Individuals participated in one 2-hour session, consisting of 30 practice and 480 experimental trials. In addition to discrete trial feedback (described

below), cumulative performance feedback was provided after the end of practice and after every 120 trials during the experiment. Participants were instructed to continually fixate a cross at the centre of the display while viewing the display sequence after which they classified the target as either a face or house. Subjects were cautioned to maintain eye fixation and avoid blinks until they responded. Eye fixation was monitored using a Tobii x50 50-Hz eye tracker (Tobii Technology, Stockholm Sweden; 1.03° eye movement threshold) controlled by a Sony 2.86 GHz computer.

Testing was conducted in a dimly lit, electrically shielded room with a chinrest maintaining a 72-cm viewing distance. A fixation cross was presented until the participant initiated a given trial with a mouse click. Figure 1 illustrates a typical trial sequence. Trials began with a fixation-only display followed by the target display. Target displays consisted of a fixation cross, and four images one of which was cued by eight equally spaced dots forming a rectangle surrounding that image (individual dots $0.20^\circ \times 0.20^\circ$; rectangle $4.52^\circ \times 3.15^\circ$, based on dot centres). The three noncued locations contained three different house images. The cued location contained either an additional, unique, house image, or a face. Pictures were arranged along the vertical (one above/below) and horizontal (one to left/right) meridians. They were cropped to an oval shape ($3.91^\circ \times 2.54^\circ$) and separated from fixation by 7.13° centre-to-centre. The decision to always present house distractors was motivated by two factors. First, we wanted distractors as similar as possible to our targets to avoid rapid attentional deployment to the cued location, a prerequisite for OSM. Second, the inclusion of face distractors would make it difficult to determine how much of the N170 response was generated by the target versus the distractor images. Once the target display was removed, either a 1200-ms fixation-only display was visible (simultaneous mask offset condition) or both masking dots and fixation remained visible for 400 ms, followed by 800 ms of fixation alone (delayed mask offset condition; OSM). Participants were then prompted to enter two unsped responses via mouse clicks. The first response was the target classification ("FACE" or "HOUSE") followed by a confidence rating ("VERY SURE", "SOMEWHAT SURE", or "GUESS"). Feedback regarding target image classification (coloured square; green = correct; red = incorrect) and the presence of eye movements/blinks concluded each trial. Each participant was presented with a unique random trial order in which the 15 face and 15 house images served as a target twice in each combination of target location (4) by mask offset condition (2).

Cases in which a participant made an eye movement/blink were excluded from all behavioural and electrophysiological analyses ($M = 7.88\%$, $SE = 1.29$).

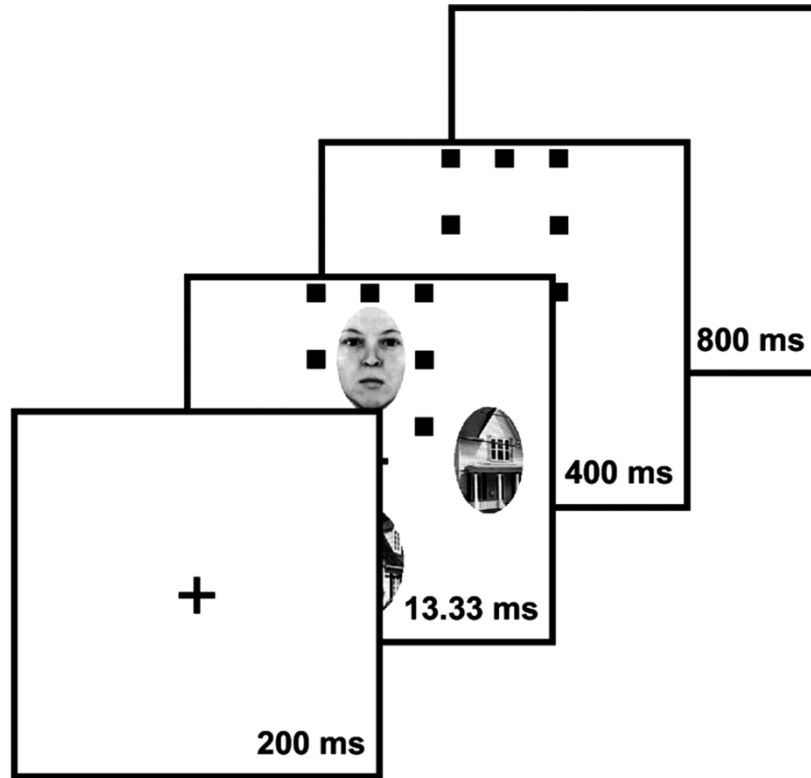


Figure 1. Schematic diagram of one trial sequence. Trials consisted of a fixation-only display followed by a brief target display (a target image, face, or house, surrounded by masking dots and three house distractor images). Next, either only a fixation-only display was visible (simultaneous mask offset condition) or, as seen in the figure, both masking dots and fixation were visible followed by fixation alone (delayed mask offset condition; OSM). Target classification and confidence rating were collected after each trial and feedback provided.

Electrophysiological recording and analysis

Continuous electroencephalogram (EEG) was collected (200-Hz sampling rate; 0.01 to 80 Hz bandpass; vertex reference) using an Electrical Geodesics Inc. (EGI; Eugene, OR) 128-channel Geodesic Sensor Net with individual electrode impedances kept below 50–75 k Ω . Data was stored on a Power Mac G4 computer. EEG was processed offline using EGI Net Station 4.1.2, beginning with a 40-Hz lowpass filter and segmentation time-locked to target display onset (epoch: -100 to 1000 ms). Artifacts detected in individual channels (fast average amplitude > 200 μ V, differential average amplitude > 100 μ V, or zero variance) or segments (greater than 10 bad channels; eye movement or eye blink detected, eye threshold = 70 μ V) were

eliminated from subsequent analyses. Furthermore, channels containing artifacts in greater than 20% of segments were eliminated completely. Bad channels were replaced by data interpolated from surrounding electrode sites. Supplementing the trial rejection criteria mentioned above, segments containing artifacts were eliminated (overall, $M = 11.87\%$, $SE = 2.69$) and bad channels replaced. Following this data processing, segments were averaged, rereferenced to the average reference, and baseline corrected for the 100-ms interval prior to the target onset.

N170 amplitude was defined as the average amplitude for a cluster of six electrodes in the lateral posterior region of the scalp over left and right hemispheres. These two clusters (i.e., 12 electrodes) were selected individually for each subject and consisted of those electrodes that provided the largest difference in N170 amplitude for faces versus houses in the simultaneous mask offset condition during a time window of 150–220 ms post target onset. An additional constraint was that all sensors had to be adjacent to at least one other sensor in the same cluster.² The simultaneous mask offset trials were utilized because they served as the control condition in this experiment.

RESULTS

Delaying the offset of the mask reduced categorization accuracy from 81.45% to 70.40% ($SEs = 2.81$ and 2.81 , respectively), which was significant by a paired-samples t -test, $t(11) = 5.53$, $p < .0001$ one-tailed, confirming the presence of OSM.

Visual inspection of the N170 component for faces and houses in the two masking conditions revealed a pattern consistent with behaviour (see Figure 2). In the simultaneous mask offset condition, where masking should not occur, faces elicited a larger N170 than houses. This difference was dramatically reduced in the delayed offset condition, the same condition that resulted in a reduction in categorization accuracy. N170 amplitude was analysed in a 2 (hemisphere; left vs. right) \times 2 (mask offset; simultaneous vs. delayed) \times 2 (target category; face vs. house) repeated-measures analysis of variance (ANOVA), which revealed a significant effect of mask offset, $F(1, 11) = 6.65$, $MSE = 0.34$, $p < .027$, and target category, $F(1, 11) = 38.51$, $MSE = 0.53$, $p < .00008$, as well as a significant mask offset \times target category interaction, $F(1, 11) = 8.46$, $MSE = 0.72$, $p < .015$. This interaction was caused by a significantly larger N170 amplitude to faces than houses in

² Several alternative approaches to choosing N170 electrodes were also explored, including selection based on grand-averaged data. These approaches produced the same pattern of results reported in the text.

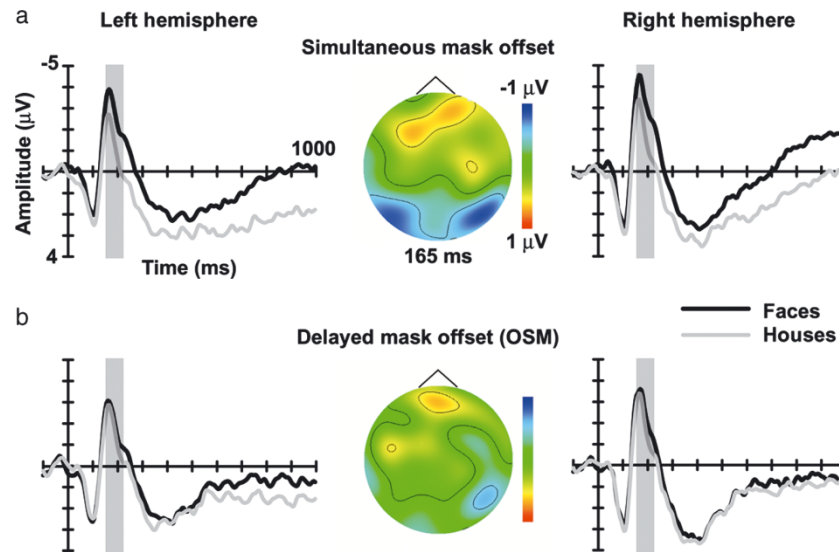


Figure 2. Brain activity in response to target presentation across experimental conditions. (a) Simultaneous mask offset trial group-averaged ERP waveforms for left and right hemisphere N170 sites. Both hemispheres reliably showed more negative mean waveform amplitudes for faces than houses during the N170 component window (shaded region), $t(11) = 7.38$ and 6.50 , $p < .000008$ and $.00003$ one-tailed, $M_s = -1.43$ and -1.42 μV , $SE_s = 0.19$ and 0.22 , respectively. The isopotential voltage scalp map illustrates the peak face-minus-house amplitude at 165 ms post target onset, as a negative bilateral posterior occipital/temporal distribution. (b) Same as Panel A but under delayed mask offset conditions wherein both hemispheres failed to show reliable mean waveform face minus house amplitude differences during the N170 component window, $t(11) = 1.30$ and 1.50 , $p_s = .110$ and $.08$ one-tailed, $M_s = -0.33$ and -0.507 μV , $SE_s = 0.26$ and 0.33 , respectively. The corresponding isopotential voltage scalp map also lacked evidence of any robust N170 component difference. To view this figure in colour, please see the online issue of the Journal.

the simultaneous offset condition, $t(11) = 7.47$, $p < .000007$ one-tailed, $M = -1.43$ μV , $SE = 0.19$, that was not significant with delayed mask offset, $t(11) = 1.59$, $p = .07$ one-tailed, $M = -0.42$ μV , $SE = 0.26$. No significant effects involving hemisphere were found, $F_s < 1.15$, $p_s > .30$.

Finally, an analysis restricted to the left and right target locations revealed that the amplitude of the face minus house N170 response was significantly greater in the hemisphere contralateral versus ipsilateral to the visual field of the target, $t(11) = 3.98$, $p < .002$ one-tailed, hemisphere difference $M = -0.84$ μV , $SE = 0.21$; collapsed across masking conditions.

DISCUSSION

Taken together, the current data offer two important insights into object substitution masking. First, we found that delaying the offset of dots

surrounding a face or house stimulus impaired observers' ability to determine the category of that stimulus, confirming that OSM can be obtained for complex stimuli. Second, we observed the typical amplitude advantage of the N170 component for faces compared to houses, but only when the surrounding masking dots disappeared at the same time as the target stimulus. Delaying the offset of the dots abolished the face–house difference in N170 amplitude, providing an upper bound for the earliest time at which OSM affects visual processing. This relatively short latency has important implications for the stage of processing and hence the mechanisms underlying OSM. Source localization of the N170 component suggests that its generator is located in lateral occipital cortex (Rossion, Joyce, Cottrell, & Tarr, 2003; also see four-dot masking study by Carlson, Rauschenberger, & Verstraten, in press), while other studies point to a generator in the superior temporal sulcus region or fusiform area (Horowitz, Rossion, Skudlarski, & Gore, 2004; Itier & Taylor, 2004; also see metacontrast masking study by Haynes, Driver, & Rees, 2005). These are all areas that have been implicated in the recognition of faces and other objects. Although we cannot determine the precise anatomical source of the suppressed N170, the finding that effects were greater in the contralateral versus ipsilateral hemisphere indicates that this location resides in a spatiotopically organized extrastriate visual area. Thus, our finding indicates that OSM can affect visual processes at an early stage, during or prior to the point at which categorization and identification occur. On the other hand, OSM does not appear to affect earlier stages responsible for processing basic features that may form the building blocks of object recognition (Woodman & Luck, 2003). It is also clear that the face–house difference we observed in the amplitude of the N170 persists well beyond the peak of this component. These sustained differences are not unusual (e.g., Holmes, Winston, & Eimer, 2005; Figure 2, top panel) and possibly reflect later stages of face processing or maintenance of information in working memory.

The present data are remarkable considering that face stimuli are believed to rely on specialized cortical circuitry (Kanwisher & Yovel, 2006) and are often processed effortlessly. For example, faces, unlike many other common objects, capture attention leading to pop out during visual search tasks (Hershler & Hochstein, 2005). Also, recent work has shown that faces are relatively impervious to the AB (Awh et al., 2004). The contrast between face processing immunity to the AB and the deficit we report here underscores the fact that, despite the similarity in subjective experience, OSM and the AB reflect very different mechanisms. Interference in AB appears to operate “late” in processing, after semantic analysis but before conscious awareness. In contrast, OSM affects early processing stages concerned with basic categorization and identification of visual stimuli. Our result supports those theories of OSM, such as reentrance (Di Lollo et al., 2000), that place the

locus of interference in OSM at relatively early visual processing stages (see Fahrenfort, Scholte, & Lamme, in press). Further research will be required to shed additional light on the precise mechanisms that are responsible for this remarkable visual masking phenomenon.

REFERENCES

- Alpern, M. (1953). Metacontrast. *Journal of the Optical Society of America*, *43*, 648–657.
- Awh, E., Serences, J., Laury, P., Dhaliwal, H., van der Jagt, T., & Dassonville, P. (2004). Evidence against a central bottleneck during the attentional blink: Multiple channels for configural and featural processing. *Cognitive Psychology*, *48*, 95–126.
- Bentin, S., & Deouell, L. Y. (2000). Structural encoding and identification in face processing: ERP evidence for separate mechanisms. *Cognitive Neuropsychology*, *17*, 35–54.
- Breitmeyer, B. G. (1984). *Visual masking: An integrative approach*. New York: Oxford University Press.
- Bridgeman, B. (2006). Contributions of lateral inhibition to object substitution masking and attention. *Vision Research*, *46*, 4075–4082.
- Carlson, T. A., Rauschenberger, R., & Verstraten, F. A. J. (in press). No representation without awareness in the lateral occipital cortex. *Psychological Science*.
- Di Lollo, V., Bischof, W. F., & Dixon, P. (1993). Stimulus-onset asynchrony is not necessary for motion perception or metacontrast masking. *Psychological Science*, *4*, 260–263.
- Di Lollo, V., Enns, J. T., & Rensink, R. A. (2000). Competition for consciousness among visual events: The psychophysics of reentrant visual processes. *Journal of Experimental Psychology: General*, *129*, 481–507.
- Fahrenfort, J. J., Scholte, H. S., & Lamme, V. A. F. (in press). Masking disrupts reentrant processing in human visual cortex. *Journal of Cognitive Neuroscience*.
- Haynes, J.-D., Driver, J., & Rees, G. (2005). Visibility reflects dynamic changes of effective connectivity between V1 and fusiform cortex. *Neuron*, *46*, 811–821.
- Hershler, O., & Hochstein, S. (2005). At first sight: A high-level pop out effect for faces. *Vision Research*, *45*, 1707–1724.
- Holmes, A., Winston, J. S., & Eimer, M. (2005). The role of spatial frequency information for ERP components sensitive to faces and emotional facial expression. *Cognitive Brain Research*, *25*, 508–520.
- Horowitz, S. G., Rossion, B., Skudlarski, P., & Gore, J. C. (2004). Parametric design and correlational analyses help integrating fMRI and electrophysiological data during face processing. *Neuroimage*, *22*, 1587–1595.
- Itier, R. J., & Taylor, M. J. (2004). Source analysis of the N170 to faces and objects. *Neuroreport*, *15*, 1261–1265.
- Jiang, Y. H., & Chun, M. M. (2001). The spatial gradient of visual masking by object substitution. *Vision Research*, *41*, 3121–3131.
- Kanwisher, N., & Yovel, G. (2006). The fusiform face area: A cortical region specialized for the perception of faces. *Philosophical Transactions of the Royal Society of London, Series B*, *361*, 2109–2128.
- Kutas, M., & Hillyard, S. A. (1980). Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science*, *207*, 203–205.
- Lamme, V. A. F., & Roelfsema, P. R. (2000). The distinct modes of vision offered by feedforward and recurrent processing. *Trends in Neurosciences*, *23*, 571–579.
- Lamme, V. A. F., Zipser, K., & Spekreijse, H. (2002). Masking interrupts figure-ground signals in V1. *Journal of Cognitive Neuroscience*, *14*, 1044–1053.

- Logothetis, N. K., & Sheinberg, D. L. (1996). Visual object recognition. *Annual Review of Neuroscience, 19*, 577–621.
- Luck, S. J., Vogel, E. K., & Shapiro, K. L. (1996). Word meanings can be accessed but not reported during the attentional blink. *Nature, 383*, 616–618.
- Raymond, J. E., Shapiro, K. L., & Arnell, K. M. (1992). Temporary suppression of visual processing in an RSVP task: An attentional blink? *Journal of Experimental Psychology: Human Perception and Performance, 18*, 849–860.
- Reiss, J. E., & Hoffman, J. E. (2006). Object substitution masking interferes with semantic processing: Evidence from ERPs. *Psychological Science, 17*, 1015–1020.
- Rolls, E. T. (2000). Functions of the primate temporal lobe cortical visual areas in invariant visual object and face recognition. *Neuron, 27*, 205–218.
- Rossion, B., Joyce, C. A., Cottrell, G. W., & Tarr, M. J. (2003). Early lateralization and orientation tuning for face, word, and object processing in the visual cortex. *Neuroimage, 20*, 1609–1624.
- Sibly, M. (2005). *Blitz3D* (Version 1.90) [Computer software]. Auckland, New Zealand: Blitz Research Ltd.
- Tong, F. (2003). Primary visual cortex and visual awareness. *Nature Reviews Neuroscience, 4*, 219–229.
- Troje, N., & Bulthoff, H. H. (1996). Face recognition under varying poses: The role of texture and shape. *Vision Research, 36*, 1761–1771.
- Turvey, M. T. (1973). Peripheral and central processes in vision: Inferences from an information-processing analysis of masking with patterned stimuli. *Psychological Review, 80*, 1–52.
- Woodman, G. F., & Luck, S. J. (2003). Dissociations among attention, perception and awareness during object-substitution masking. *Psychological Science, 14*, 605–611.

Manuscript received October 2006
Manuscript accepted February 2007
First published online July 2007