Spatiotemporal object continuity in human ventral visual cortex

Do-Joon Yi*†, Nicholas B. Turk-Browne‡, Jonathan I. Flombaum‡, Min-Shik Kim*, Brian J. Scholl‡, and Marvin M. Chun‡

*Department of Psychology, Yonsei University, Seoul 120-749, Korea; and ‡Department of Psychology, Yale University, New Haven, CT 06520

Edited by Anne Treisman, Princeton University, Princeton, NJ, and approved April 17, 2008 (received for review March 13, 2008)

Coherent visual experience requires that objects be represented as the same persisting individuals over time and motion. Cognitive science research has identified a powerful principle that guides such processing: Objects must trace continuous paths through space and time. Little is known, however, about how neural representations of objects, typically defined by visual features, are influenced by spatiotemporal continuity. Here, we report the consequences of spatiotemporally continuous vs. discontinuous motion on perceptual representations in human ventral visual cortex. In experiments using both dynamic occlusion and apparent motion, face-selective cortical regions exhibited significantly less activation when faces were repeated in continuous vs. discontinuous trajectories, suggesting that discontinuity caused featurally identical objects to be represented as different individuals. These results indicate that spatiotemporal continuity modulates neural representations of object identity, influencing judgments of object persistence even in the most staunchly ''featural'' areas of ventral visual cortex.

functional MRI | object persistence | repetition attenuation | spatiotemporal continuity | visual tracking

To interact efficiently with our environments, we must parse the world into discrete objects and identify objects as the same over time, for example, as the same predator that we encountered last week or as the same animal we were just tracking. There are two importantly different ways to identify an object as the same over time, however. First, one can note what the object looks like: If you see an animal that is short and brown, and then you see an animal that is tall and yellow, they are unlikely to be the same animal. Second, one can identify objects as the same based on how and where they move through the local visual environment: If you see an animal appear from behind one tree, and then see an animal appear from behind a different tree, they cannot be the same animal, because there was no spatiotemporally continuous path that could link the two encounters. Critically, this second sort of ''sameness'' is independent of the first: If spatiotemporal continuity is violated, then the two animals must be different, even if they look the same.

The use of spatiotemporal continuity to determine object persistence has not only been well documented in many areas of cognitive science, but has been shown to dominate visual similarity in several contexts. In visual processing, for example, spatiotemporal factors can produce the impression of two different objects before and after a brief disappearance, despite visual similarity (1–4). Similarly, research in infant cognition has shown that even 4-month-olds will interpret two featurally identical objects as different individuals if there is no spatiotemporally continuous path to link their appearances (5), and behavioral research with non-human primates has demonstrated that spatiotemporal continuity determines whether one or two objects are inferred from a scene (6). Such results have been used to support several models of midlevel visual processing, wherein objects are tracked over time, regardless of their features, by ''object files'' (7), ''visual indexes'' (8), or ''object tokens'' (9).

Research in cognitive neuroscience, however, has not devoted equal attention to these two ways of determining whether objects are the same over time (10). On one hand, we now know a great deal about the processing of visual features, which drive comparisons of visual similarity. Moreover, visual similarity produces ''repetition attenuation'' in neural responses: a weaker neural response is typically observed to a repeated stimulus compared with a novel stimulus (11–14). For example, presenting two identical objects in rapid succession results in an attenuated fMRI signal in several regions of visual cortex, relative to the rapid presentation of two different objects. Importantly, object-selective cortical regions such as the lateral occipital cortex show such attenuation even though the repeated objects may differ in their sizes, orientations, or perceived depths, revealing cue-independent object representations (15, 16). Thus, like habituation methods classically used in infant cognition (17), repetition attenuation can reveal whether a particular region of the brain treats two stimuli as two different objects or as instances of the same object.

However, there have been very few cognitive neuroscience studies of object persistence and sameness as driven by spatiotemporal factors. Some studies have shown that neural processing does persist even when objects are invisible, as measured with single-unit recording (18, 19), functional magnetic resonance imaging (fMRI) (20, 21), and near-infrared spectroscopy in infants (22). Other studies have demonstrated that the ways in which objects appear and disappear at occluding boundaries will impact whether they are seen to persist over time, as measured with electroencephalography (23) and fMRI (24). However, no previous studies, to our knowledge, have examined the role of spatiotemporal continuity *per se*, despite its status as perhaps the most powerful determinant of object persistence (25, 26).

In the present study, we explored the neural consequences of cues to object persistence with two primary goals. First, we attempt to determine whether spatiotemporally continuous vs. discontinuous object trajectories result in a neural signature of whether a current object is the same individual from an earlier encounter. Second, by exploring the impact of spatiotemporal continuity on processing in ventral visual cortex, we also aimed to determine how such an effect might interact with the better understood processing of objects' surface features.

In our experiments, spatiotemporal continuity was manipulated between two faces that were presented sequentially. Two featurally identical or different faces traced continuous or discontinuous paths of motion either as they emerged from occluders (in experiment 1) or as they were embedded in streams of apparent motion without any occlusion (in experiment 2).

Author contributions: D.-J.Y., N.B.T.-B., J.I.F., B.J.S., and M.M.C. designed research; D.-J.Y., N.B.T.-B., and J.I.F. performed research; D.-J.Y., N.B.T.-B., J.I.F., M.-S.K., B.J.S., and M.M.C. analyzed data; D.-J.Y., N.B.T.-B., J.I.F., M.-S.K., B.J.S., and M.M.C. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Freely available online through the PNAS open access option.

†To whom correspondence should be addressed. E-mail: dojoon.yi@yonsei.ac.kr.

This article contains supporting information online at www.pnas.org/cgi/content/full/0802525105/DCSupplemental.

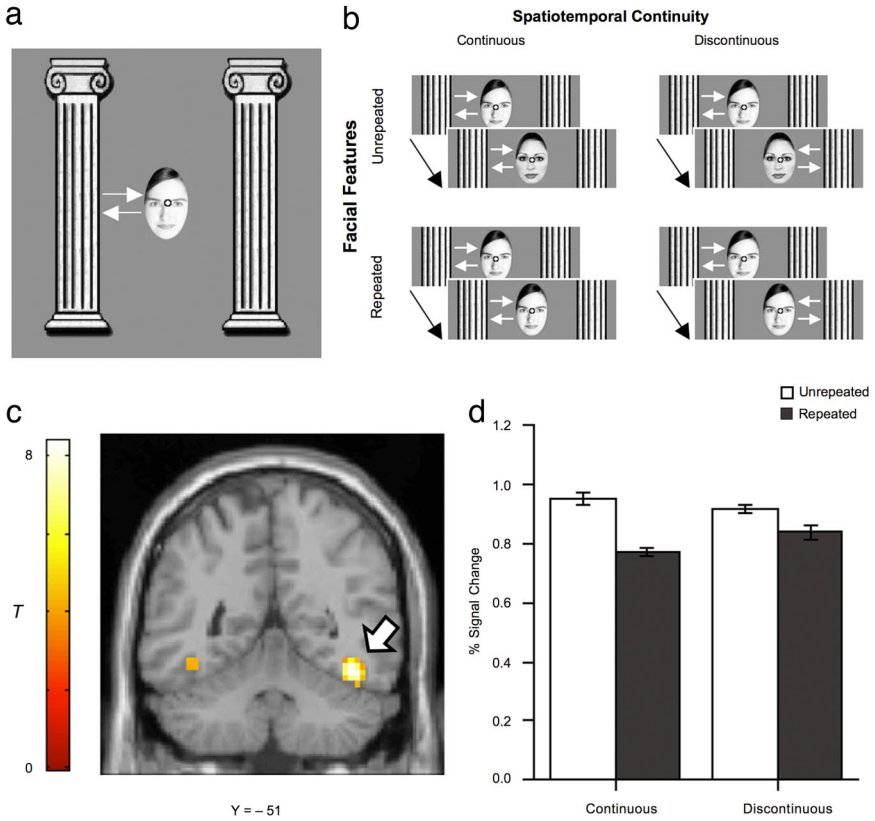

Fig. 1. Methods and results of experiment 1. (*a*) Stimuli and a single event in a trial. A face appeared from behind one of the two columns, moved to fixation, turned back, and disappeared behind that same column. The white arrows indicate motion and were not present in the actual displays. (*b*) The four trial types, constructed from two orthogonally crossed factors: facial features (repeated vs. unrepeated) and spatiotemporal continuity (continuous vs. discontinuous). Black arrows indicate subsequent face animation events from a single trial of each type. For sample animations, see Movies S1-S4. (*c*) A coronal section of the single-subject template superimposed with a representative subject's activation map in a functional localizer run. A white arrow indicates the location of the right FFA. (*d*) fMRI signal changes in the right FFA. Error bars indicate within-subject standard error.

Thus, four types of trials constituted a 2 × 2 design in each experiment: (*i*) a featurally identical face repeatedly appeared in a motion path ("repeated-continuous"), (*ii*) a featurally identical face appeared subsequently in each of the two motion paths ("repeated-discontinuous"), (*iii*) two featurally distinct faces appeared in a single path ("unrepeated-continuous"), and (*iv*) two featurally distinct faces appeared subsequently in different paths ("unrepeated-discontinuous"). Throughout both experiments, subjects performed an inverted face detection task: In some trials, one of two faces was inverted, requiring subjects to make an unspeeded response. The trials were pseudorandomly intermixed in a rapid event-related design.

To test the effects of spatiotemporal continuity, we analyzed fMRI repetition attenuation in face processing, revealing whether two stimuli are treated as the same or different. As noted above, there are two very different ways in which objects may be treated as "the same." Unlike previous studies of repetition attenuation in visual cortex, which focused on similarity in terms of visual features, we manipulated the dynamic context via which the objects appeared. In particular, we asked whether spatiotemporal object discontinuity would affect the representation of objects as the same, even when those objects' visual surface features were (and appeared to be) identical.

Because spatiotemporal factors have been shown to trump visual similarity in behavioral studies with adults (1), infants (5), and nonhuman primates (6), we predicted that spatiotemporal continuity would also influence repetition attenuation in ventral visual cortex, which has typically been associated with object identity. Specifically, fMRI attenuation in the repeated-continuous condition relative to the unrepeated-continuous condition should be larger than the attenuation observed for the repeated-discontinuous relative to the unrepeated-discontinuous condition. However, if spatiotemporal continuity does not affect what it means to be the "same" object in these ventral visual areas, then fMRI signals should show only the main effect of facial features, regardless of spatiotemporal continuity.

Results

Experiment 1: Spatiotemporal Continuity in Dynamic Occlusion. The initial display contained a vertical column on each side of fixation to be used as occluders. Each trial consisted of two subsequent events: In each event, a face appeared from behind one of the two columns, moved to fixation, turned back, and disappeared behind that same column (Fig. 1*a*). Critically, the second face that appeared (*i*) had either the same or different visual features as the first face, and (*ii*) appeared from either the same or different column relative to the first face. Thus, four types of trials constituted a 2 × 2 design (Fig. 1*b*): (*i*) a featurally identical face repeatedly appeared from the same column (repeated-continuous), (*ii*) a featurally identical face appeared subsequently from each of the two columns (repeated-discontinuous), (*iii*) two featurally distinct faces appeared from the same column (unrepeated-continuous), and (*iv*) two featurally distinct faces appeared subsequently from different columns (unrepeated-discontinuous). For an inverted face detection task, 14% of trials presented one inverted face or two. Subjects were instructed to fixate the central point throughout the experiment.

The face movement was very fast in the periphery and slower at fixation [see supporting information (SI) Movies S1–S4]. This pattern of movement, along with our instructions, aimed to discourage eye movements.

Twenty-three right-handed subjects volunteered for monetary compensation. Three subjects were excluded from analyses because of uncorrectable imaging artifacts. The remaining 20 subjects (10 females, mean 22 ± 4 years old) accurately performed the inverted face detection task, committing no misses and very few false alarm errors (on 0.3% of nontarget trials). Subsequent fMRI analyses included only nontarget trials without motor responses. We first analyzed face-selective regions of interest (ROIs) in ventral visual cortex (reported below and in *SI Text, Other Face-Selective ROIs*) and then conducted exploratory whole-brain analyses (reported in *SI Text*, *Exploratory Whole-Brain Analysis* and Fig. S3).

The face-selective regions of interest were localized for each subject with an independent scan (see *Materials and Methods*). The ROIs were bilaterally defined in the fusiform face area (FFA) (27, 28) and in the lateral occipital cortex (LO) (15, 16). The right FFA ($x = 39, y = -50, z = -15$) and the right LO ($x = 46, y = -76, z = -3$) were localized from all 20 subjects. The left FFA ($x = -40, y = -53, z = -14$) and the left LO ($x = -42, y = -79, z = -6$) were localized from 19 and 17 subjects, respectively (Fig. S1). For each ROI, the mean percentage fMRI signal changes were submitted to planned comparisons and a repeated measures ANOVA with two factors: facial features (repeated vs. unrepeated) and spatiotemporal continuity (continuous vs. discontinuous).

The right FFA exhibited robust attenuation effects to repeated faces (Fig. 1 *c* and *d*). The main effect of facial features was significant ($F_{1,19} = 23.440$, $P = 0.0001$). The fMRI signals were lower in the repeated-continuous condition relative to the unrepeated-continuous condition ($t_{19} = 6.023$, $P = 0.000009$) and were also lower in the repeated-discontinuous condition relative to the unrepeated-discontinuous condition ($t_{19} = 2.101$, $P = 0.049$). More importantly, such attenuation was significantly stronger for repetitions occurring along a spatiotemporally continuous path compared with when there was a spatiotemporal discontinuity, as revealed by the two-way interaction ($F_{1,19} = 6.467$, $P = 0.020$). This effect was further highlighted in a planned comparison: The repeated-continuous condition produced significantly weaker fMRI signals than the repeated-discontinuous condition ($t_{19} = 2.183$, $P = 0.042$). These results suggest that the representation of what counts as the "same" face in the right FFA takes into account not only the visual similarity of the faces, but also the spatiotemporal continuity (or lack thereof) by which the two presentations occur.

The main effect of spatiotemporal continuity was not significant in the right FFA ($F_{1,19} = 0.788$, $P > 0.3$), and the fMRI signal strength in the unrepeated-continuous and unrepeated-discontinuous conditions did not differ ($t_{19} = 1.460$, $P > 0.1$). These two results indicate that motion adaptation *per se* cannot be an alternative account for the current results. Rather, our results reflect an interaction between the processing of spatiotemporal and surface features.

The left FFA showed similar results (Fig. S2*a*). The main effect of facial features was significant ($F_{1,18} = 12.781, P = 0.002$), whereas the main effect of spatiotemporal continuity was not ($F_{1,18} = 0.493$, $P = 0.492$). The interaction was marginally significant ($F_{1,18} = 4.372$, $P = 0.051$). The repeated-continuous condition produced significantly weaker fMRI signals than the unrepeated-continuous condition ($t_{18} = 3.410, P = 0.003$) and only marginally lower signals than the repeated-discontinuous condition ($t_{18} = 1.741, P = 0.099$). The left FFA may have been less sensitive to spatiotemporal cues for face stimuli than the right FFA probably because of the well documented lateralization of face processing in the right hemisphere (27–30). Nevertheless, these findings support our hypothesis that spatiotemporal continuity affects ventral visual object processing.

Spatiotemporal continuity modulated repetition attenuation in the lateral occipital cortex as well (Fig. S2 *b* and *c*). The right LO revealed a main effect of facial features ($F_{1,19} = 9.920$, $P = 0.005$) and an interaction between facial features and spatiotemporal continuity ($F_{1,19} = 5.694$, $P = 0.028$) but no main effect of spatiotemporal continuity ($F_{1,19} = 0.025, P > 0.8$). The left LO revealed a main effect of facial features ($F_{1,16} = 12.326$, $P = 0.003$), but neither the main effect of spatiotemporal continuity nor the interaction (P values > 0.1). Importantly, in the continuous conditions, the featurally identical faces produced weaker fMRI signals than the featurally distinct faces in both the right LO ($t_{19} = 5.049$, $P = 0.00007$) and left LO ($t_{16} = 3.590$, $P = 0.002$). The equivalent comparison in the discontinuous conditions failed to reach significance in both LO ROIs (P values > 0.6). Thus, spatiotemporal continuity modulated the neural response in the right LO. Given that the LO is known to be involved in shape processing and to respond selectively to various types of visual objects (15, 16), these results suggest that spatiotemporal cues may affect computations of object identity in visual cortex for various visual categories beyond faces. This possibility, however, needs further evidence.

Experiment 2: Spatiotemporal Continuity in Apparent Motion. Each trial presented four 120-ms frames to induce the perception of either horizontal or vertical apparent motion in two object streams (Fig. 2 *a* and *b*). In each frame, two objects faced each other across fixation. In the first frame, the two objects consisted of scrambled gray-scale faces and were positioned in the periphery to bias the perceived motion directions in subsequent frames (9). In the second frame, the objects approached fixation, and one of them turned into a face. In the third frame, the objects passed the central region, and, again, one of them turned into a face. Critically, the second face that appeared (*i*) had either the same or different visual features as the first face and (*ii*) appeared as part of either the same or different apparent motion stream relative to the first face. Thus, four types of trials constituted a 2×2 design (Fig. 2*c*): (*i*) a featurally identical face repeatedly appeared within the same stream (repeated-continuous), (*ii*) a featurally identical face appeared successively in each of the two streams (repeated-discontinuous), (*iii*) two featurally distinct faces appeared within the same stream (unrepeated-continuous), and (*iv*) two featurally distinct faces appeared successively in different streams (unrepeated-discontinuous). For sample animations, see Movies S5–S8. To enhance the perception of two distinct apparent motion streams, 20% of trials contained only scrambled objects in all four frames. For an inverted face detection task, 16% of trials presented one or two inverted faces. Subjects were instructed to fixate the central point throughout the experiment, and, to rule out any possible effects of anticipatory eye movements, the inverted face target was revealed only very briefly when the objects passed the fixation, and subjects did not know until that point which of the two objects would turn into a face.

Fifteen new subjects volunteered for monetary compensation. One subject was excluded from the analyses because of uncorrectable imaging artifacts and another because of pulse trigger malfunction. The remaining 13 subjects (six females, mean 23 ± 4 years old) missed targets on 11.1% of target trials and committed false alarm errors on 1.2% of nontarget trials, indicating that the inverted face detection was more difficult in experiment 2 than with the longer presentation times in experiment 1.

The fMRI signals were analyzed as in experiment 1. The right FFA ($x = 44, y = -55, z = -20$) and the right LO ($x = 44, y = -75, z = -10$) were localized from 11 and 12 subjects, respectively. The left FFA ($x = -42, y = -52, z = -20$) and the left

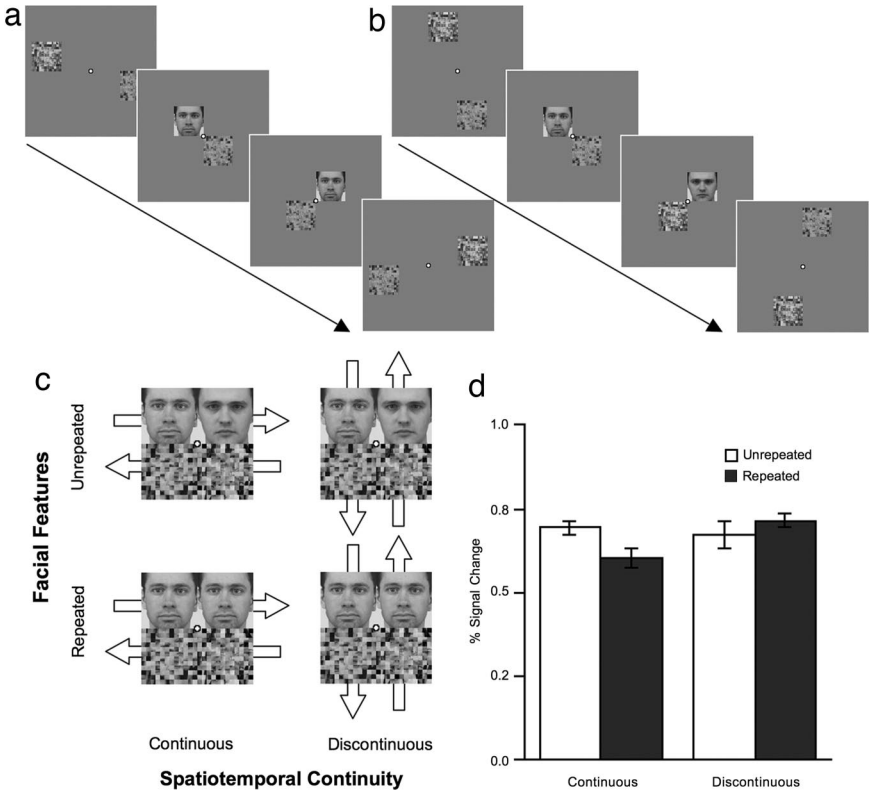

Fig. 2. Methods and results of experiment 2. (*a* and *b*) Examples of four-frame apparent motion. Black arrows indicate four subsequent 120-ms frames in a trial. For sample animations, see Movies S5–S9. (*c*) The four trial types, crossing two critical factors: facial features (repeated vs. unrepeated) and spatiotemporal continuity (continuous vs. discontinuous). Black outlined arrows indicate the perceived direction of apparent motion in the second and third frames. Note that, in the experiment, the direction of apparent motion (horizontal vs. vertical) was orthogonally manipulated with the two critical factors, resulting in the same number of the continuous (or discontinuous) conditions in the vertical (or horizontal) motion. (*d*) fMRI signal changes in the right FFA. Error bars indicate within-subject standard error.

LO ($x = -39, y = -79, z = -12$) were localized from 10 and 12 subjects, respectively.

The right FFA revealed significant effects of spatiotemporal continuity on neural representations of repeated faces (Fig. 2*d*). A significant interaction between facial features and spatiotemporal continuity was found ($F_{1,10} = 9.708, P = 0.011$). In planned comparisons, the repeated-continuous condition produced lower fMRI signals than both the unrepeated-continuous condition ($t_{10} = 4.004, P = 0.003$) and the repeated-discontinuous condition ($t_{10} = 2.645, P = 0.025$). These results were mirrored in the right LO (Fig. S2*e*). The two-way interaction was significant ($F_{1,11} = 15.656, P = 0.002$) and the repeated-continuous condition produced significantly lower fMRI signals than both the unrepeated-continuous condition ($t_{11} = 3.111, P = 0.010$) and the repeated-discontinuous condition ($t_{11} = 2.365, P = 0.038$). In both ROIs, neither the main effects of facial features and spatiotemporal continuity nor the other paired comparisons were significant (P values > 0.2). In contrast to the right hemisphere, the ROIs in the left hemisphere did not show any significant statistical effects (Fig. S2 *d* and *f*). This experiment thus replicated the sensitivity of repetition attenuation to spatiotemporal factors of repeated objects in the ventral visual cortex in the right hemisphere, using a very different display than experiment 1.

Discussion

This study makes three primary contributions to the understanding of how object identity and persistence are neurally represented. First, these results provide a possible neural foundation for the computation of persisting object identity on the basis of spatiotemporal continuity. Given the salience of continuity as a principle in previous theories of visual tracking (31), infant cognition (5, 32), and comparative cognition (6, 33), it is striking that no previous cognitive neuroscience studies (to our knowledge) have directly explored continuity. Our results indicate that signatures of such spatiotemporal processing can be found in relatively well defined regions of visual cortex. Future research may resolve whether object persistence on the basis of spatiotemporal continuity is computed in ventral cortex proper or whether our results reveal the consequences of modulation from dorsal brain regions. Either case would be interesting, given that previous studies have not considered how or whether dynamic contexts can modulate the processing of features and objects in ventral cortex.

Second, the particular brain regions that were implicated in this study—fusiform gyrus and lateral occipital cortex—are also notable. The ventral visual pathway has typically been associated with feature-based processing of "what" an object is (34–36) and so might have been among the least likely places to find effects of spatiotemporal continuity. Thus, the spatiotemporally mediated representation of persisting object identity is a new discovery, and one that must be integrated into classical theories of the nature of ventral cortex, one of the best-understood regions of visual processing. This coupling between ventral cortex and spatiotemporal processing informs psychological theory as well. Theories of object persistence have carefully documented the distinction between spatiotemporal factors and surface features but have remained unable to say much about whether these two sorts of processing are integrated into the same system or reflect completely separate processes (32). The present results, how-

ever, suggest that representations of object identity that are mediated by spatiotemporal continuity are not isolated from other aspects of visual object processing; rather, such spatiotemporal factors impact even the most staunchly "featural" regions of object representation.

Third, these points may also be applied to our understanding of fMRI repetition attenuation itself: Like ventral cortex, the phenomenon of repetition attenuation has been closely linked to the processing of visual object identity, in terms of both visual similarity and categorical identity (12, 37–39). In other words, repetition attenuation has been assumed to reflect processing of an object's identity, although identity in these experiments has typically been manipulated only with respect to objects' surface appearances, not their spatiotemporal histories. The present results demonstrate that objects' spatiotemporal histories also modulate repetition attenuation, independently from their actual and perceived surface appearances. Thus, repetition attenuation is sensitive to spatiotemporal markers of sameness.

Other recent studies have similarly suggested that repetition attenuation is more flexible than previously thought, being affected by attention (14, 40–43), emotion (44), and response mapping (45). Our new finding is qualitatively distinct from these other forms of modulation, however. An alternative account of our findings in experiment 1 might be that attention was captured by the "surprising" conflict between featural and spatiotemporal cues in two of our conditions. In the unrepeated-continuous trials, for example, the novel features on the second face may have broken the expectations set up by the spatiotemporal cues, potentially attracting attention and thereby increasing neural activation. How plausible is this interpretation? Perhaps such surprise might happen on one or two trials in the initial practice session, but it seems unlikely that such events would remain surprising when encountered repeatedly throughout the scanning session. More critically, this explanation does not apply to experiment 2, given the relatively fast stimulus presentations and because the scrambled masks were always involved in two of the four locations of each motion stream (introducing conflicting novel visual features in both continuous and discontinuous motion conditions). Thus, across the two experiments, spatiotemporal continuity is the common factor that modulates repetition attenuation in ventral cortex, distinct from factors such as attention. Repetition attenuation has become a valued methodological tool in cognitive neuroscience, and these results accordingly suggest that this tool may be fruitfully applied to investigate other spatiotemporal principles of object identity independent in principle from what objects look like (26).

These findings may all be summarized by revisiting the notion of what it means to be the "same" persisting individual over time. Nearly all past neuroscience research has focused on a single sense of sameness: what it means to look the same or to be categorized as the same on the basis of visual surface features. In contrast, the present results highlight the importance of complementing this perspective with a second notion of sameness, wherein spatiotemporal continuity exerts a strong influence on the determination of an object's persisting identity over time, independent of the object's visual features, so that two identically looking objects may still be categorized as distinct individuals. Our results suggest that spatiotemporal object continuity of this type is not only a salient aspect of our visual experience of the world, but an important constraint on neural processing of identity in some of the best characterized regions of visual cortex.

Materials and Methods

fMRI Acquisition. The study protocol was approved by the Human Investigation Committees at Yale University and at Yonsei University. Informed consent was obtained from all subjects. Experiments 1 and 2 both used 3T scanners with a standard birdcage head coil and a gradient echo-planar imaging sequence to acquire functional data. In experiment 1, three functional scan runs, each of 165 volumes, were conducted on a Siemens Trio 3T scanner. Each functional volume (2,000-ms repetition time; 25-ms echo time; 90° flip angle; 7-mm thickness with no gap) comprised 19 axial slices parallel to the anterior commissure-posterior commissure line, covering the entire brain. Visual stimuli were projected on a rear LCD-projection screen and seen through an angled mirror attached to the head coil. In experiment 2, a 3T ISOL Forte scanner acquired three functional runs, 190 volumes for an ROI localizer run, and 285 volumes for two main experiment runs. Each functional volume (2,000-ms repetition time; 25-ms echo time; 90° flip angle; 5-mm thickness with no gap) comprised 25 axial slices perpendicular to the orientation of the brainstem, covering the entire brain. Visual stimuli were presented on an LCD-panel attached to the head coil. Both experiments used a magnet-compatible button box to collect responses.

Task and Procedure. In experiment 1, subjects performed an inverted face detection task during the first two functional runs. The runs began with two black and white clip-art column figures presented against a gray background, spanning a central fixation circle (a white-outlined black disk with 0.3° diameter). Each column subtended 4.9° × 19.4°, and its nearest edge was 4.6° from the left or right side of the fixation circle. Trial onset was cued by a 400-ms blink of the fixation mark. Every trial consisted of two subsequent events: In each event, a face appeared from behind one of the two columns, moved to fixation, turned back, and disappeared behind that same column. The second event began as soon as the face from the first event had disappeared. The second face that appeared (*i*) had either the same or different visual features as the first face and (*ii*) appeared from either the same or different column relative to the first face. The animation of two faces was generated by sequential presentation of 62 frames, each for 33 ms (total 2,067 ms). Each face decelerated toward the fixation and accelerated back to a column with a maximum velocity of 117°/sec. For sample animations, see Movies S1–S4. A total of 168 oval-shaped grayscale faces were presented, each subtending 3.1° × 5.1°. All faces were the frontal views of real photos, available from a commercial web site. As critical conditions, four types of 28 trials were randomly intermixed: (*i*) a featurally identical face repeatedly appeared from the same column (repeated-continuous), (*ii*) a featurally identical face appeared subsequently from each of the two columns (repeated-discontinuous), (*iii*) two featurally distinct faces appeared from the same column (unrepeated-continuous), and (*iv*) two featurally distinct faces appeared subsequently from different columns (unrepeated-discontinuous). Four trials of each type contained an inverted face as a target, to which subjects pressed a button with the right index finger in an unspeeded manner. The order of trials was determined by an optimal sequencing program (46). The mean intertrial interval was 5.4 s (with a minimum of 4 s).

Subjects performed a one-back repetition detection task in the third run, which served as a functional localizer for face-selective visual cortex in individual subjects. There were 10 20-s stimulation blocks alternating between face-only and scene-only blocks, interleaved with 10-sec fixation periods. During these blocks, face or scene images, each subtending 10.3° × 10.3°, were sequentially presented every second at the center of the screen (200-ms interstimulus interval). Faces were similar to those used in the preceding runs and scenes depicted various types of the indoor and outdoor spaces (kitchens, offices, the frontal views of buildings, natural landscapes, etc.). There were two or three repetitions per block to which observers made unspeeded responses with the right index finger. The order of face and scene blocks was counterbalanced across subjects.

In experiment 2, subjects performed the one-back repetition detection task as a functional localizer in the first functional run. There were 16 14-s stimulation blocks alternating between face-only and scene-only blocks, interleaved with 8-sec fixation periods. During these blocks, face or scene images, each subtending 7.5° × 7.5°, were sequentially presented every 700 ms at the center of the screen (200-ms interstimulus interval).

Subjects then performed an inverted face detection task in the last two functional runs. These runs began with a fixation circle, which was positioned in the center of an 18° × 18° gray background. Each trial presented two streams of apparent motion, each of which consisted of four 120-ms frames (without interframe intervals). Each frame presented two square objects in opposite locations across fixation. The objects, each subtending 4.5° × 4.5°, were either grayscale faces or scrambled objects, each of which was made of 144 squares constituting a face. In the first frame, the objects were positioned in periphery. One edge of each object touched either the horizontal midline or the vertical midline, and so there were four possible initial configurations, each which biased observers to perceive either horizontal or vertical apparent motion. The objects were translated 4.5° along the vertical or horizontal midline and passed through the central region of the display during the second and third frames, reaching the opposite ends in the fourth frame. For sample animations, see Movies S5–S8. In the face trials (80%), one of two objects turned into a face in the second frame. In the third frame, the object from either that same motion stream or the other motion

stream turned into another face. The second face could be either featurally identical or different from the first face. Therefore, the two faces (*i*) had either the same or different visual features and (*ii*) appeared from either the same or different apparent motion streams. There were 40 randomly intermixed trials of each of the four critical conditions: (*i*) a featurally identical face repeatedly appeared within a motion stream (repeated-continuous), (*ii*) a featurally identical face appeared successively in each of the two separate motion streams (repeated-discontinuous), (*iii*) two featurally distinct faces appeared within the same motion stream (unrepeated-continuous), and (*iv*) two featurally distinct faces appeared successively in different motion streams (unrepeated-discontinuous). Eight trials of each type contained an inverted face as a target, to which subjects pressed a button with the right index finger in an unspeeded manner. There were also additional trials in which the two objects never turned into faces during apparent motion. These ''no-face'' trials (20%) were expected to enhance the perception of two apparent motion streams, so that subjects were less likely to perceive unwanted motion between two faces in the repeated-discontinuous condition. A total of 240 gray-scale faces were presented, all frontal views of real photos.

In both experiments, subjects practiced a block of the inverted face detection tasks before scanning (both outside the scanner and in the scanner) to be familiarized with the displays. The faces used in the practice trials were not presented during the scanning.

fMRI Analyses. Preprocessing and statistical analyses were conducted by using a statistical parametric mapping (SPM2; Wellcome Department of Imaging Neuroscience, University College, London). Because of magnetization equilibration, the first five (experiment 1) or four (experiment 2) volumes of each run were discarded before preprocessing. The remaining volumes were then corrected for slice timing, realigned, normalized (resampling voxel size, $3 \times 3 \times 3$ mm), and smoothed (Gaussian kernel, $8 \times 8 \times 8$ mm). A high-pass frequency filter (cutoff: 128-s period) and an autocorrelation correction were applied to the time series.

The face-selective ROIs were localized for individual subjects from the functional localizer run. Blocks of faces and scenes were separately modeled by canonical hemodynamic response functions (HRFs) with six movement parameters as covariates of no interest. A statistical parametric map of the *t* statistic (SPM{*t*}) was generated from linear contrasts between face and scene blocks. For each individual subject, the maximally face-selective voxel was identified from each of two bilateral ventral cortical regions ($P < 0.0001$, uncorrected; cluster threshold $k = 0$)—the fusiform gyrus and the lateral occipital cortex—and used as the center of a spherical ROI (4-mm radius).

In the region-of-interest analyses, the mean time courses were first extracted from each ROI, using the MarsBar toolbox (http://marsbar.sourceforge.net). Parameter estimates of event-related activity were obtained by using the general linear model with the four critical conditions and a dummy condition. The critical conditions (see *Task and Procedure*) included only nontarget trials without motor responses. The remaining trials with inverted face targets or false alarm responses were treated together as dummy events. Each condition was modeled by a HRF and a temporal derivative to account for variable hemodynamic delays. Six movement parameters were also included as covariates of no interest. The peaks of fitted HRFs for the four conditions were entered into statistical analyses as percentage fMRI signal changes. For further analyses, including voxelwise comparisons, see *SI Text, Other Face-Selective ROIs and Exploratory Whole-Brain Analysis*.

ACKNOWLEDGMENTS. We thank Jenika Beck, Soo-jung Min, and Bo Young Won for recruiting subjects. This work was supported by National Institutes of Health Grants EY014193 and P30 EY000785 (to M.M.C.), National Science Foundation Grant BCS-0132444 (to B.J.S.), a National Science Foundation graduate research fellowship, a National Research Service Award predoctoral fellowship (to J.I.F.), a foreign Natural Sciences and Engineering Research Council of Canada Post-Graduate Scholarship (to N.B.T.-B.), and a Ministry of Science and Technology of the Republic of Korea 21st Century Frontier Research Program Grant (to M.S.K.).

1. Burt P, Sperling G (1981) Time, distance, and feature trade-offs in visual apparent motion. *Psychol Rev* 88:171–195.
2. Flombaum JI, Scholl BJ (2006) A temporal same-object advantage in the tunnel effect: Facilitated change detection for persisting objects. *J Exp Psychol Hum Percept Perform* 32:840–853.
3. Navon D (1976) Irrelevance of figural identity for resolving ambiguities in apparent motion. *J Exp Psychol Hum Percept Perform* 2:130–138.
4. Ramachandran VS, Anstis SM (1983) Perceptual organization in moving patterns. *Nature* 304:529–531.
5. Spelke ES, Kestenbaum R, Simons DJ, Wein D (1995) Spatiotemporal continuity, smoothness of motion and object identity in infancy. *Br J Dev Psychol* 13:113–142.
6. Flombaum JI, Kundey SM, Santos LR, Scholl BJ (2004) Dynamic object individuation in rhesus macaques: A study of the tunnel effect. *Psychol Sci* 15:795–800.
7. Kahneman D, Treisman A, Gibbs BJ (1992) The reviewing of object files: Object-specific integration of information. *Cognit Psychol* 24:175–219.
8. Pylyshyn Z (1989) The role of location indexes in spatial perception: A sketch of the FINST spatial-index model. *Cognition* 32:65–97.
9. Chun MM, Cavanagh P (1997) Seeing two as one: Linking apparent motion and repetition blindness. *Psychol Sci* 8:74–79.
10. Kanwisher N, Driver J (1992) Objects, attributes, and visual attention: Which, what, and where. *Curr Dir Psychol Sci* 1:26–31.
11. Buckner RL, *et al.* (1995) Functional anatomical studies of explicit and implicit memory retrieval tasks. *J Neurosci* 15:12–29.
12. Grill-Spector K, Malach R (2001) fMR-adaptation: A tool for studying the functional properties of human cortical neurons. *Acta Psychol (Amsterdam)* 107:293–321.
13. Schacter DL, Buckner RL (1998) Priming and the brain. *Neuron* 20:185–195.
14. Yi D-J, Chun MM (2005) Attentional modulation of learning-related repetition attenuation effects in human parahippocampal cortex. *J Neurosci* 25:3593–3600.
15. Grill-Spector K, *et al.* (1999) Differential processing of objects under various viewing conditions in the human lateral occipital complex. *Neuron* 24:187–203.
16. Kourtzi Z, Kanwisher N (2001) Representation of perceived object shape by the human lateral occipital complex. *Science* 293:1506–1509.
17. Hunter MA, Ames EW (1988) A multifactor model of infant preferences for novel and familiar stimuli. *Adv Infancy Res* 5:69–95.
18. Assad JA, Maunsell JH (1995) Neuronal correlates of inferred motion in primate posterior parietal cortex. *Nature* 373:518–521.
19. Baker CI, Keysers C, Jellema T, Wicker B, Perrett DI (2001) Neuronal representation of disappearing and hidden objects in temporal cortex of the macaque. *Exp Brain Res* 140:375–381.
20. Hulme OJ, Zeki S (2007) The sightless view: Neural correlates of occluded objects. *Cereb Cortex* 17:1197–1205.
21. Olson IR, Gatenby JC, Leung HC, Skudlarski P, Gore JC (2003) Neuronal representation of occluded objects in the human brain. *Neuropsychologia* 42:95–104.
22. Baird AA, *et al.* (2002) Frontal lobe activation during object permanence: Data from near-infrared spectroscopy. *Neuroimage* 16:1120–1125.
23. Kaufman J, Csibra G, Johnson MH (2005) Oscillatory activity in the infant brain reflects object maintenance. *Proc Natl Acad Sci USA* 102:15271–15274.
24. Shuwairi SM, Curtis CE, Johnson SP (2007) Neural substrates of dynamic object occlusion. *J Cognit Neurosci* 19:1275–1285.
25. Carey S, Xu F (2001) Infants' knowledge of objects: Beyond object files and object tracking. *Cognition* 80:179–213.
26. Spelke ES (2000) Core knowledge. *Am Psychol* 55:1233–1243.
27. Kanwisher N, McDermott J, Chun MM (1997) The fusiform face area: A module in human extrastriate cortex specialized for face perception. *J Neurosci* 17:4302–4311.
28. McCarthy G, Puce A, Gore JC, Allison T (1997) Face-specific processing in the human fusiform gyrus. *J Cognit Neurosci* 9:605–610.
29. Puce A, Allison T, Asgari M, Gore JC, McCarthy G (1996) Differential sensitivity of human visual cortex to faces, letterstrings, and textures: A functional magnetic resonance imaging study. *J Neurosci* 16:5205–5215.
30. De Renzi E (1997) in *Behavioral Neurology and Neuropsychology*, eds Feinberg TE, Farah MJ (McGraw-Hill, New York), pp 245–255.
31. Scholl BJ, Pylyshyn ZW (1999) Tracking multiple items through occlusion: Clues to visual objecthood. *Cognit Psychol* 38:259–290.
32. Xu F, Carey S (1996) Infants' metaphysics: The case of numerical identity. *Cognit Psychol* 30:111–153.
33. Santos LR (2004) ''Core knowledges'': A dissociation between spatiotemporal knowledge and contact-mechanics in a non-human primate? *Dev Psychol* 7:167–174.
34. Ungerleider LG, Mishkin M (1982) in *Analysis of Visual Behaviour*, eds Ingle DJ, Goodale MA, Mansfield RJW (MIT Press, Cambridge, MA), pp 549–586.
35. Desimone R, Albright TD, Gross CG, Bruce C (1984) Stimulus-selective properties of inferior temporal neurons in the macaque. *J Neurosci* 4:2051–2062.
36. Baylis GC, Rolls ET (1987) Responses of neurons in the inferior temporal cortex in short term and serial recognition memory tasks. *Exp Brain Res* 65:614–622.
37. Desimone R (1996) Neural mechanisms for visual memory and their role in attention. *Proc Natl Acad Sci USA* 93:13494–13499.
38. Grill-Spector K, Henson R, Martin A (2006) Repetition and the brain: Neural models of stimulus-specific effects. *Trends Cognit Sci* 10:14–23.
39. Wiggs CL, Martin A (1998) Properties and mechanisms of perceptual priming. *Curr Opin Neurobiol* 8:227–233.
40. Eger E, Henson RN, Driver J, Dolan RJ (2004) BOLD repetition decreases in object-responsive ventral visual areas depend on spatial attention. *J Neurophysiol* 92:1241–1247.
41. Murray SO, Wojciulik E (2004) Attention increases neural selectivity in the human lateral occipital complex. *Nat Neurosci* 7:70–74.
42. Yi D-J, Kelley TA, Marois R, Chun MM (2006) Attentional modulation of repetition attenuation is anatomically dissociable for scenes and faces. *Brain Res* 1080:53–62.
43. Turk-Browne NB, Yi D-J, Chun MM (2006) Linking implicit and explicit memory: Common encoding factors and shared representations. *Neuron* 49:917–927.
44. Ishai A, Pessoa L, Bikle PC, Ungeleider LG (2004) Repetition suppression of faces is modulated by emotion. *Proc Natl Acad Sci USA* 101:9827–9832.
45. Dobbins IG, Schnyer DM, Verfaellie M, Schacter DL (2004) Cortical activity reductions during repetition priming can result from rapid response learning. *Nature* 428:316–319.
46. Dale AM (1999) Optimal experimental design for event-related fMRI. *Hum Brain Mapp* 8:109–114.