Learning to See Biological Motion: Brain Activity Parallels Behavior

Emily D. Grossman[1], Randolph Blake[2], and Chai-Youn Kim[2]

Abstract

■ Individuals improve with practice on a variety of perceptual tasks, presumably reflecting plasticity in underlying neural mechanisms. We trained observers to discriminate biological motion from scrambled (nonbiological) motion and examined whether the resulting improvement in perceptual performance was accompanied by changes in activation within the posterior superior temporal sulcus and the fusiform ''face area,'' brain areas involved in perception of biological events. With daily practice, initially naive observers became more proficient at discriminating biological from scrambled animations embedded in an array of dynamic ''noise'' dots, with the extent of improvement varying among observers. Learning generalized to animations never seen before, indicating that observers had not simply memorized specific exemplars. In the same observers, neural activity prior to and following training was measured using functional magnetic resonance imaging. Neural activity within the posterior superior temporal sulcus and the fusiform ''face area'' reflected the participants' learning: BOLD signals were significantly larger after training in response both to animations experienced during training and to novel animations. The degree of learning was positively correlated with the amplitude changes in BOLD signals. ■

INTRODUCTION

In everyday life, perception seems automatic and effortless, yet much of our perceptual proficiency is earned through repeated exposure to environmentally significant objects and events (Goldstone, 1998). Perceptual learning continues throughout our lifetime, modifying our sensitivity and sharpening our discrimination abilities. In the laboratory, perceptual learning can be explored systematically by repeated exposure to stimulus features. A large body of evidence indicates that practice improves performance on a variety of perceptual tasks, ranging from low-level tasks involving detection of simple figures to high-level tasks involving identification of complex, novel objects (Fine & Jacobs, 2002). To give just a few examples drawn from vision, practiced observers can better discriminate subtle differences in the directions of motion of coherently moving fields of dots (Ball & Sekuler, 1982), can more easily group oriented line elements into a global structure (Vidyasagar & Stuart, 1993), and can more easily identify images of faces presented within masking noise (Gold, Bennett, & Sekuler, 1999).

Experience-dependent improvement in perceptual performance undoubtedly involves neural changes within the brain, and in at least some instances, perceptual learning appears to be mediated by neural plasticity within the same brain regions involved in perception of the relevant stimuli. For example, neural responses increase within ventral temporal brain areas involved in face and object perception as observers learn to recognize specific objects and faces in visually degraded images (Dolan et al., 1997). It is worth noting, incidentally, that learning-dependent changes in brain activation do not always involve ''increased'' activation, as exemplified by a study showing decreased cerebral blood flow within the striate and extrastriate cortex of human observers following extensive training on a visual orientation discrimination task (Schiltz et al., 1999).

In this study, we examined perceptual learning and its neural concomitants using a unique class of visual stimuli, namely, point-light animations of biological motion. First popularized by Johansson (1973), these compelling animations depict human activities by the motions of a dozen or so dots strategically placed on the major joints of the body. Most people have little trouble discerning the activities being portrayed by these point-light animations, unless the sequences themselves are displayed upside down (Sumi, 1984) or critical motion tokens are misplaced on the body (Pinto & Shiffrar, 1999). We were motivated to study perceptual learning of biological motion for two reasons. First, it is possible to display point-light sequences under conditions that seriously degrade a naive observer's ability to see biological motion, therefore providing a challenge for perceptual learning. Second, a number of brain imaging studies have measured neural responses

[1]University of California, Irvine, [2]Vanderbilt University

within specific brain areas of both the dorsal and ventral stream pathways selective for the kinematics and form depicted by point-light animations. These brain areas include a dorsal region on the posterior extent of human superior temporal sulcus (STSp) (Beauchamp, Lee, Haxby, & Martin, 2003; Pelphrey et al., 2003; Vaina, Solomon, Chowdhury, Sinha, & Belliveau, 2001; Grossman, Donnelly, et al., 2000); and a ventral region within the fusiform gyrus known for its responsiveness to faces and human figures (fusiform ''face area'' [FFA]) (Santi, Servos, Vatikiotis-Bateson, Kuratate, & Munhall, 2003; Grossman & Blake, 2002). We can thus examine whether experience-dependent changes in discrimination of biological motion are mediated by neural plasticity within brain regions generally believed to be involved in perception of biological motion.

The strategy employed in the present experiment was straightforward. We started with a ''library'' of test animations consisting of two dozen point-light sequences depicting different human activities and two dozen nonbiological sequences created by ''scrambling'' the dot positions of the biological animations (see Figure 1). Viewed on their own, these two types of test animations—biological and scrambled—were easily discriminable. To degrade discriminability, we embedded both types of test animations in dynamic, motion-matched noise consisting of randomly arrayed dots that moved in directions and at speeds that mimicked the dots in the test animations. Using a subset of these noise-masked test animations, we quantified discrimination performance and measured brain activation in a group of naive observers who had never before seen point-light animations. Following these ''pretraining'' measurements, we then gave each observer daily training sessions sufficient to produce significant improvement in the ability to discriminate biological from scrambled animations embedded in noise. Following training, we remeasured discrimination performance and brain activation while participants viewed the biological and nonbiological sequences embedded in the same levels of masking noise employed in the pretraining scanning sessions. During this ''posttraining'' phase, we also tested using a new set of animations not employed during training, to determine whether the effect of training generalized to animations never before seen. The following section describes the changes in perceptual performance and brain activation produced by this training procedure.

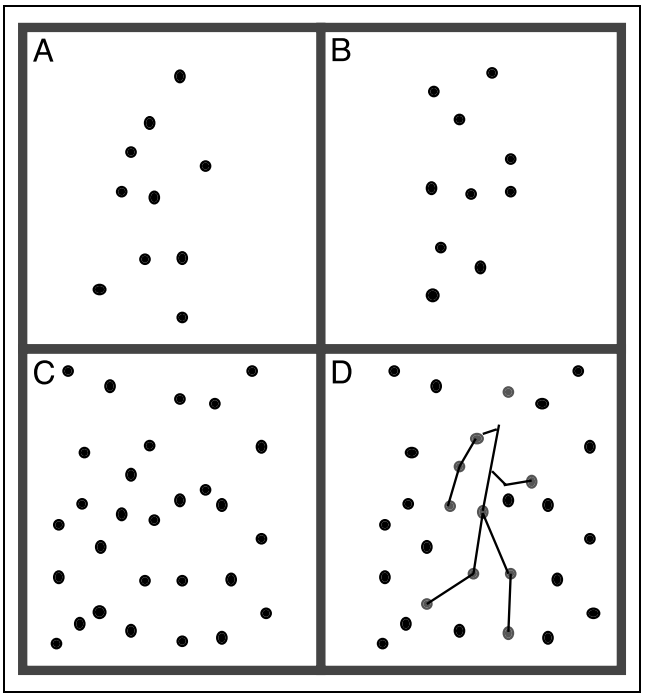

Figure 1. Examples of the point-light biological motion animations used in the study. (A) Single frame of a point-light walker and (B) the same frame scrambled and (C) masked in noise dots. (D) The identical frame as (C), but for visualization the dots are connected to reveal the human form. Observers were not shown animations in which the limbs were connected. Scrambled sequences contain all the same local dot motions as the biological ones, but the initial starting positions of the dots are randomized, thereby destroying the hierarchical structure of the limbs and the biological interpretation of the animations. The noise mask was created by distributing dots across the entire display with the same local motions as the target animation (biological or scrambled). In all instances, the dynamics of the noise array matched those of the target animation. Thus, for example, dots depicting the local motions of a kicker masked the kicker target animation (intact biological or scrambled).

RESULTS

Behavioral Results

The participants in this study were initially naive, having never before viewed point-light biological motion animations. Nonetheless, all participants could discriminate biological from nonbiological animations with perfect accuracy when these animations were presented with small amounts of masking noise, which merely confirms the compelling impression of human activity portrayed by point-light animations. To lower discrimination performance to the 71%-correct level, it was necessary to embed the test animations in a dense array of noise dots, with the criterion noise level varying among observers. (As described in the Methods section, this criterion noise level was determined using a staircase procedure.) It is important to note that the dynamic noise used to mask the target animations contained the same local motions as the target. Thus, for example, the ''jogging'' figure and the scrambled ''jogging'' figure were masked by many spatially scrambled dots drawn from the ''jogging'' animation. Each noise dot in a masked animation also moved in the same phase as the dots comprising the target animation. Noise masks with the same dynamics as biological targets are much more effective in masking point-light biological motion than are noise masks comprising random motion (Bertenthal

& Pinto, 1994). In our experiments, new noise arrays were generated on each trial and contained no information indicating whether the embedded animation depicted biological or scrambled motion.

Using these initial baseline noise levels, we then established for each observer a more difficult "target" level of noise that served as the goal to be met or exceeded with training. A given observer's target noise level was set at a value dependent on his/her level of pretraining performance, with higher target values established for individuals with higher initial performance. We assessed each observer's pretraining performance level using this target noise level, with discrimination performance being indexed by d'—individual pretraining d' values are shown in Table 1. Not surprisingly, performance is very poor as observers were being tested with noise levels substantially higher than their measured thresholds. These pretraining d' values merely confirm that the target noise level provided a lofty performance goal for training.

Each day over the course of about a week, observers returned to the laboratory to participate in the staircase task that estimated the number of noise dots needed to produce 71% correct performance on the scrambled versus biological motion discrimination task (one representative set of staircase results is shown in Figure 2). With practice, all observers improved at the task, with the number of noise dots added to the display matching or exceeding the target level within eight training sessions (the exact number of training sessions varied between five and eight). At the conclusion of the training sessions, d' measures for discriminating biological from scrambled motion had improved by approximately one d' unit (mean = 1.37, SD = 0.40; Table 1). All observers, even those with the highest levels of target noise dots (such as Observers SJ and CL), benefited significantly from the training sessions.

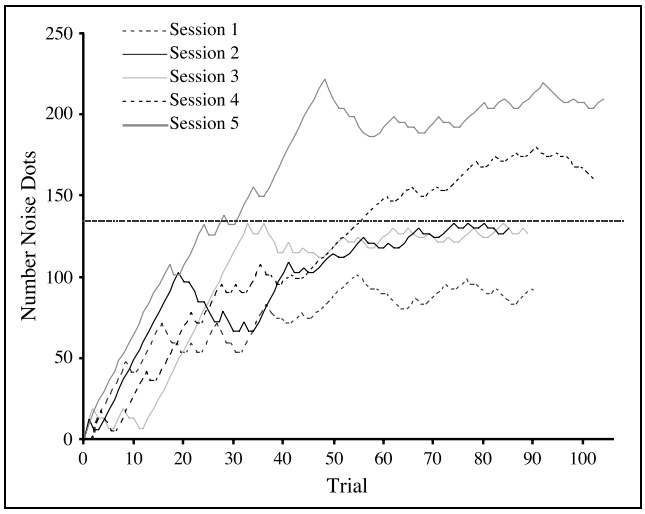

Figure 2. Results from training sessions of a single observer (KJ). Training was implemented using a staircase procedure using a 2–1 rule that converged on 71% correct performance. Dotted horizontal line indicates the target number of noise dots set based on Session 1 threshold and towards which the observer trained. Because the staircase terminated after a predetermined number of reversals in performance, the number of trials for each session varied.

Table 1. Behavioral Results for Individual Observers

	Number Noise Dots		*d′*		
Observer	*Baseline*	*Target*	*Baseline*	*Trained*	*Novel*
SS	40	120	0.18	1.57	2.38
TJ	58	130	0.29	1.07	1.27
JB	99	140	0.89	2.25	1.89
KJ	81	140	0.80	1.27	1.03
JI	111	160	0.15	1.17	1.31
CL	128	175	0.39	1.41	1.52
MK	142	180	0.77	1.17	0.82
SJ	183	210	0.90	1.03	1.22
Average			*0.55*	*1.37*	*1.43*

A staircase procedure measured the number of dots that could mask the animation for the observer to discriminate biological from scrambled motion at 71% accuracy (baseline number noise dots). Based on that number, a target number of noise dots was set at which the observer must learn to discriminate the two kinds of animations (target number noise dots). Baseline d' was measured at the target level of noise but prior to any training (baseline d'). Following the training sessions, d' was again measured using the same animations as in training (trained d') and a novel set of animations (novel d').

We wanted to know whether the posttraining improvements in discrimination performance were confined to the specific animation sequences experienced during training. Thus, following training, we also tested each observer on a new set of animations not used during initial testing or during training. The d' scores for these new animations were also on average one d' unit higher than baseline (mean = 1.43, SD = 0.50), and there was no difference between trained and novel animations in posttraining discrimination performance (p = .66). The equivalence of performance for these two sets of animations strongly implies that training produced a general improvement in visual processing of biological motion displays and not just specific familiarity with a set of items experienced during training.

Brain Imaging Results

Observers participated in two scan sessions, one immediately following the initial, baseline assessment session and one immediately following the termination of training when the observers' performance had reached or exceeded the target level of noise. The STSp region of interest (ROI) was localized on the posterior extent of the superior temporal sulcus of all observers using point-light animations displayed without noise dots (Figure 3;

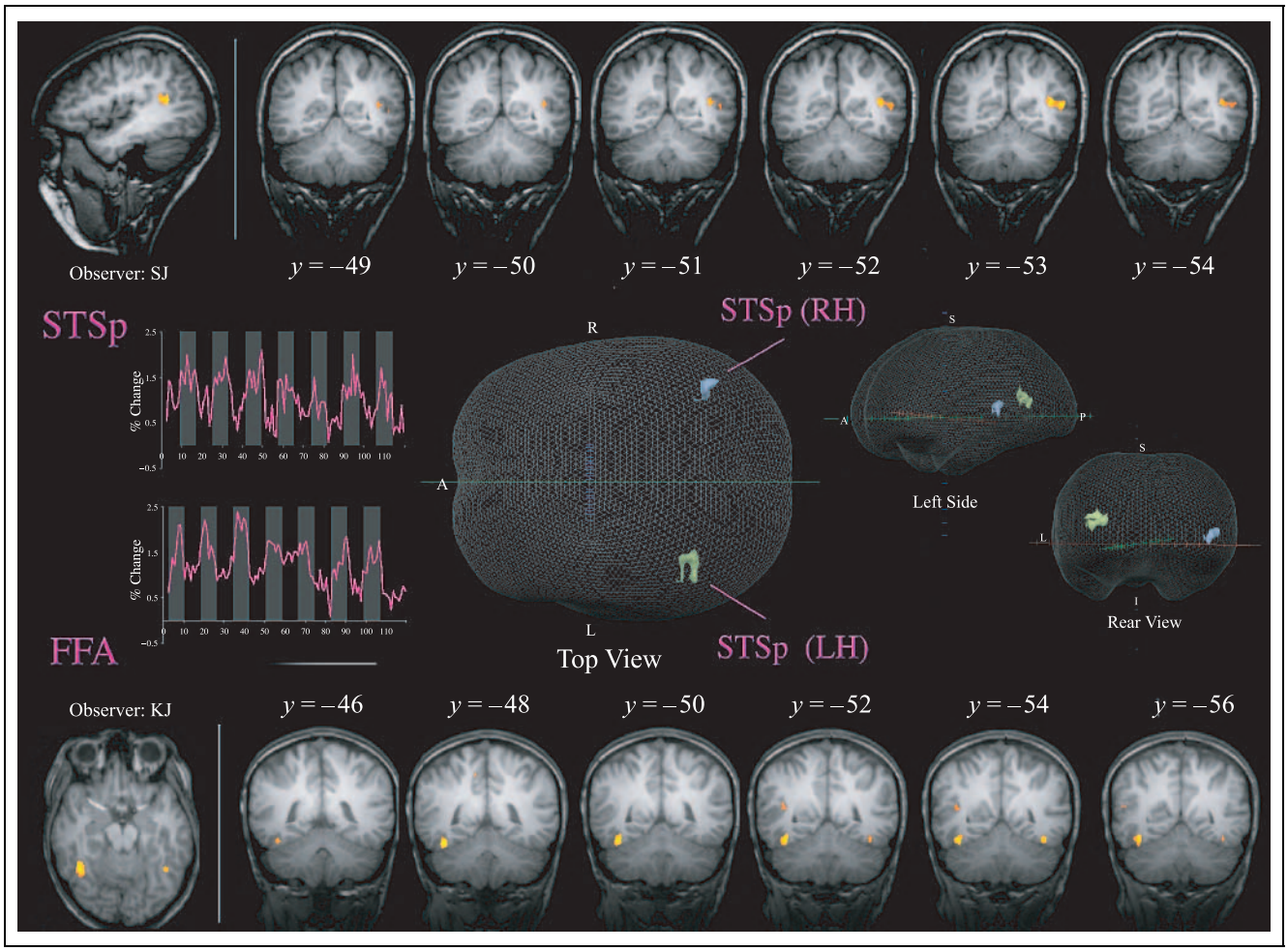

Figure 3. ROIs in two example observers. Top: Sagittal and coronal views of the STSp in the left hemisphere of Observer SJ (displayed in radiological convention, left hemisphere is on right, and right hemisphere on left). Images are consecutive slices from anterior to posterior. Top BOLD activity plot is the average time course from the left STSp ROI of this observer during the biological and scrambled motion localizer. Light bars indicate intervals of biological motion. Middle: Mesh diagrams of the STSp ROI in Observer SJ from the posterior, left, and rear views. The left STSp is colored light green, and the right STSp is colored light blue. Bottom: Axial and coronal views of the FFA ROI in Observer KJ. Lower BOLD activity plot shows the average time course from the right FFA during the biological motion localizer. Light gray bars indicate intervals of biological motion (order of blocks was counterbalanced across observers, and so the biological motion intervals are 180° phase-shifted for these two sample observers).

Table 2). STSp was found unilaterally in the right hemisphere of four of the eight observers, unilaterally in the left hemisphere of two observers, and bilaterally in two observers. The right hemisphere dominance we find in localizing STSp is consistent with previous reports using similar stimulus conditions (Pelphrey et al., 2003; Grossman, Donnelly, et al., 2000). STSp activation was, on average, 0.50% higher during the biological epochs compared to scrambled epochs. The point-light biological versus scrambled motion contrast also revealed a focus of activation in the middle fusiform gyrus on the ventral surface of the temporal cortex in seven of the eight observers (bilateral in six observers, right hemisphere only in one observer). Talairach and Tournoux (1988) coordinates (Table 2) and subsequent scan sessions in which observers viewed images of faces and objects confirmed that this region corresponds to the FFA. Although neural signals within the FFA in response to biological motion were generally weaker than those in response to faces, FFA activations are nonetheless reliably stronger to biological sequences than to scrambled sequences (average difference = 0.33%), replicating earlier results (Grossman & Blake, 2002).

Prior to training, BOLD responses within the STSp and the FFA during viewing of biological motion sequences embedded in noise were quite weak and much smaller compared to responses to the same sequences viewed in the absence of noise (Figure 4). This attenuating effect of masking noise on neural activity to biological sequences was highly significant in both the STSp and the FFA ($p < .00001$, $p < .005$, respectively). The overall weakness of these pretraining BOLD signals to masked sequences is not surprising, as the psychophysical results obtained the day before imply that

Table 2. Talairach Coordinates

	Talairach Coordinates					
	Left Hemisphere			*Right Hemisphere*		
ROI	*x*	*y*	*z*	*x*	*y*	*z*
STSp	−45.0 (7.9)	−51.4 (4.8)	13.2 (5.9)	48.9 (4.6)	−51.6 (8.3)	10.8 (4.6)
FFA	−39.6 (3.1)	−50.1 (4.1)	−11.9 (3.0)	37.4 (6.9)	−50.1 (8.0)	−10.9 (4.4)

Mean coordinates and standard deviations (in parentheses) for the two ROIs: posterior superior temporal sulcus (STSp) and fusiform face area (FFA).

observers were unable to perceive biological motion on the vast majority of presentations during these pretraining scan sequences; from the observer's viewpoint, most sequences—scrambled and biological—were indistinguishable.

Following training, however, differences in neural responses to masked biological and masked scrambled sequences were significantly larger, within both the STSp and the FFA. The difference in BOLD signal between masked biological and scrambled motion doubled within the STSp (0.31% difference compared to 0.16% difference pretraining; $p < .05$), and increased within the FFA (0.22% difference vs. 0.13% difference pretraining; $p < .05$). Moreover, larger magnitude BOLD signals were also measured when observers viewed novel animations (i.e., animations not seen during training) embedded in the same number of noise dots—BOLD response differences increased to 0.25% within the STSp and to 0.21% in the FFA (both significant; $p < .05$).

One could argue that the increased BOLD activity following training arises because after training observers were attending more diligently to the masked animations than they were during the pretraining scan session, not because of training-dependent changes in biological motion processing. We find this argument unpersuasive, however, because observers performed equally well on the same 1-back task during both pretraining and posttraining scanning sessions. This task required judging whether two successive animations portrayed equivalent patterns of motion drawn from the same category of activity (e.g., two successive scrambled joggers in noise). As detailed in the Methods section, to accurately perform this task required attending to the overall patterns of motion presented on each trial, not simply to the movement of selected dots at a given location. Comparison of the pretraining and posttraining accuracy on the 1-back task for masked animations reveals no performance differences—observers exhibited 69.1% accuracy prior to training and 68.8% accuracy after training. (The equivalence before and after training is perhaps not surprising because the 1-back task did not require discriminating biological from scrambled sequences and, thus, was not the task performed during training.) We therefore conclude that the stronger BOLD signals during the masked biological epochs after training arise from genuine learning, and not simply a change in the deployment of attention.

On average, observers' sensitivity to masked biological sequences increased with training by 0.82 d' units (combined across the trained and new animations, for which there was no significant difference), but the extent of improvement varied greatly among observers (e.g., compare Observers SJ and SS). Similarly, the ex-

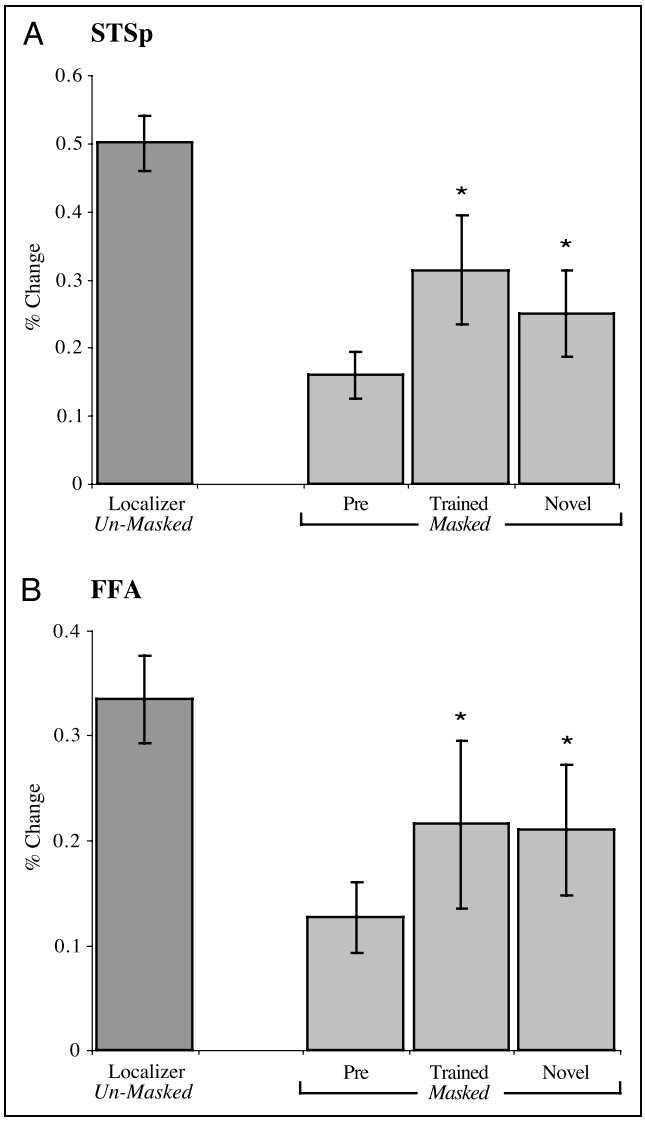

Figure 4. Histograms of BOLD activity averaged across observers. Dark gray bars indicate the unmasked biological motion localizer, and light gray bars indicate masked biological motion scans. (A) STSp ROI. (B) FFA ROI. BOLD signal is shown as percent change during biological motion blocks as compared to during the scrambled motion blocks. Error bars indicate one standard error. (*) Indicates significant difference ($p < .05$) from the pretraining masked scan.

tent to which neural activity was selective for the masked biological motion animations (determined by BOLD signal increase over that resulting from the scrambled animations) increased by varying degree among observers. To learn whether variability within these two datasets is related, we computed the Spearman rank-order correlation between the behavioral and BOLD results. (We elected to use the Spearman correlation because it is a nonparametric measure that is insensitive to metric values and, therefore, is less susceptible to outliers; in addition, the Spearman correlation makes no assumptions about the nature of distribution underlying the sampled data.) This correlation analysis revealed the magnitude of improvement in performance measured behaviorally (expressed as the difference between posttraining and pretraining d' values) was positively correlated (r_s = .90) with the enhanced selectivity for biological motion in the STSp (expressed as the difference between biological and scrambled neural response pre- and posttraining). In other words, those individuals exhibiting the greatest improvement psychophysically also had the largest increases in selectivity for biological motion in the BOLD signal response posttraining (Figure 5). The same was true within FFA: d' difference scores and BOLD signal were positively correlated (r_s = .74).

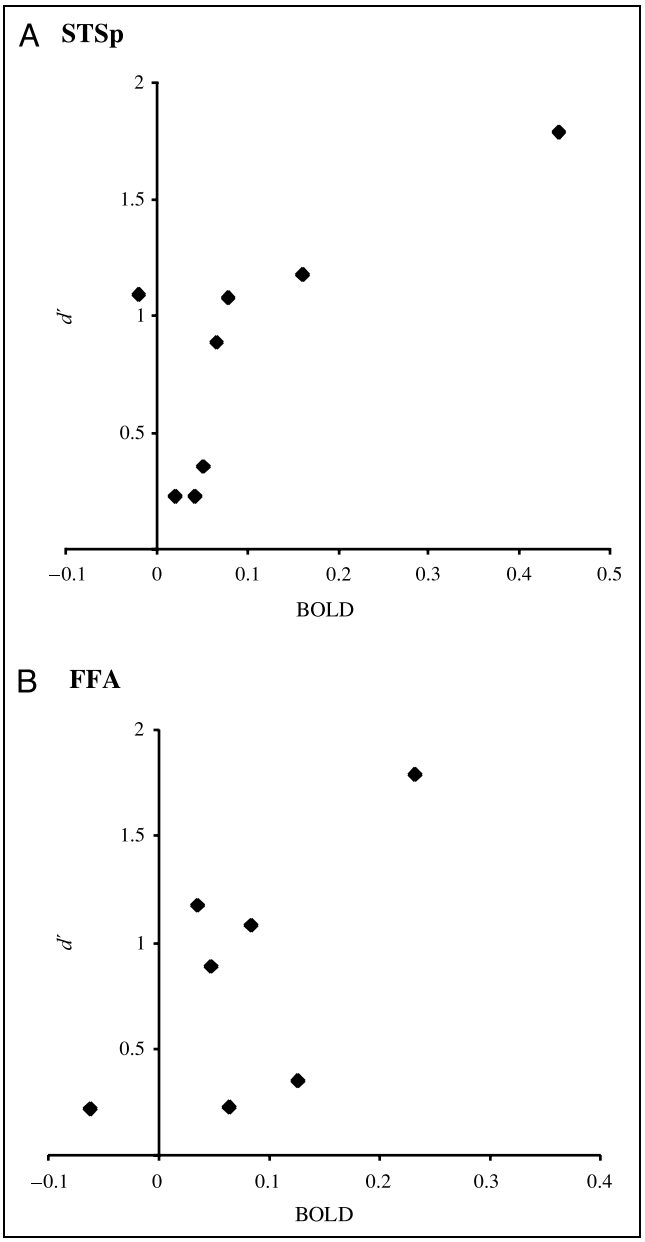

Figure 5. Neural activity versus performance. Scatterplots of BOLD signal in (A) the STSp and (B) the FFA versus behavioral performance. We found no differences between discriminability or BOLD response in the STSp and the FFA following training for the trained and novel animations. In this figure, measurements using the two sets of stimuli are averaged for each individual. The *x*-axis indicates the changes in BOLD response in the posttraining scan as compared to pretraining. Note that in this figure, the BOLD scores reflect a "difference of a difference," that is, the difference between biological and scrambled blocks, before and after training. The *y*-axis indicates the improvements in sensitivity (d') to the biological motion after training as compared to pretraining.

DISCUSSION

Recognizing biological activities portrayed by point-light animations can be a trivially easy task when the sequences are displayed without noise (Ahlstrom, Blake, & Ahlstrom, 1997; Mather, Radford, & West, 1992; Johansson, 1973)—even young infants are sensitive to the kinematics distinguishing biological from nonbiological animations (Fox & McDaniel, 1982), and this sensitivity improves during the first five years of life (Pavlova, Krageloh-Mann, Sokolov, & Birbaumer, 2001). Embedding point-light animations within an array of masking dots, however, can disrupt recognition of biological motion, provided that the masking dots are sufficiently dense and similar in their motion characteristics to the target (Pavlova & Sokolov, 2000; Bertenthal & Pinto, 1994). The results reported here demonstrate that naive observers can learn to see biological motion presented with noise dots that, prior to learning, are sufficiently potent to mask the visibility of the biological activities. Moreover, this perceptual learning transfers completely to new, masked animations never before seen.

What did observers actually learn during these training sessions that allowed them to improve in the ability to perceive biological motion sequences embedded in noise? In many perceptual learning studies, improvements in performance are selective for the particular stimulus features used during training. For example, improvements in motion discrimination are restricted to the directions of motion used during training (Vaina, Belliveau, des Roziers, & Zeffiro, 1998; Ball & Sekuler, 1982), and improved spatial resolution is limited to the contour orientations experienced during training (Fahle, 1997). In addition, visual perceptual learning is often restricted to the specific region of the visual field (and,

hence, region of the retina) where training was experienced (e.g., Ahissar & Hochstein, 1996). These improvements in perceptual performance are often attributed to modifications in the selectivity of neurons tuned for specific stimulus features critical for task performance. Can the same be said for improvements in perception of biological motion?

It is true that there is evidence for the existence of neurons that respond selectively to visual kinematics portraying biological activity (Oram & Perrett, 1994), neurons that could plausibly serve as biological motion ''feature detectors'' (Giese & Poggio, 2003). We find it difficult to imagine, however, how learning in the task used here could involve just the sharpening of the selectivities of neurons responsive to particular patterns of motion. The animations used during training and testing depicted a wide variety of human activities, including walking, kicking, throwing, and jumping. These activities are portrayed by moving dots whose directions and speeds are diverse and whose varied spatial locations encompass a relatively large retinal area. Moreover, the individual motion vectors defining biological activity were also those defining scrambled biological sequences, so local motion signals per se were insufficient cues for correct performance. It is additionally unlikely that performance improved because observers had learned something specific about the noise arrays, because all noise arrays and target dots were drawn from the same individual motion vectors and were spatially randomized on every trial. Lastly and significantly, learning transferred completely to animations never seen during training, which would not be predicted from sharpened responses within neurons tuned to specific biological motion patterns.

Rather than promoting enhanced stimulus selectivity, we believe that training allowed observers to develop keener sensitivity to the coherent kinematics defining biological motion and greater efficiency at segregating those kinematic signals from background signals lacking that coherence. This could be implemented in a strategy that searches for component biological motion within the noise array consisting of only singular local biological motions. Characteristic features of biological motion may consist of local opponent motion that is commonly seen in the left and right sides of body movements, or of local coherent motion signals that are present within individual limbs. Further psychophysical testing will be required to determine the specific nature of point-light animations that produce generalized improvement in discrimination with training.

Whatever the specifics of learning, our results are reminiscent of those reported by Dosher and Lu (1999), who documented enhanced performance with training on a form identification task in which stimuli were perturbed by external noise. They interpreted their findings in terms of experience-dependent noise suppression, not plasticity of basic visual channels or adjustments in cognitive strategies. Both our task and Dosher and Lu's require the extraction of form from noise (although in our case the form was defined by motion signals), and it is entirely feasible that improved performance on the two different tasks arises from computationally comparable operations (albeit operations carried out by different neural areas). In the context of a recent neural model of biological motion perception, these neural operations could involve alterations in the connection strengths among neurons amplifying weak signals from ''motion clutter'' by extracting the strongest features of noise displays while suppressing nonoptimal features (Giese & Poggio, 2003).

Turning next to our brain imaging results, we find that learning is accompanied by significant changes in BOLD signal activation within brain areas thought to be involved in perception of biological motion. Prior to training, when observers were unable to see biological motion figures embedded in noise, neural responses in the STSp and the FFA were weak compared to responses measured when observers viewed unmasked biological motion. Following training, the neural responses to biological motion increased significantly for animations experienced during training and for novel biological motion animations. Moreover, the magnitude of this increase was positively correlated with the degree of improvement in behavioral performance: Individuals with the greatest improvement in discrimination performance were also those showing the largest increases in STSp and FFA activations, and vice versa. It is possible, of course, that changes in neural responsiveness with training also occurred in brain areas other than the STSp and the FFA, for there is evidence for widespread activation when viewing biological motion sequences (Ptito, Faubert, Gjedde, & Kupers, 2003; Grossman & Blake, 2002; Servos, Osu, Santi, & Kawato, 2002; Vaina, Solomon, et al., 2001). In the present study, we concentrated on these two regions because of their previously established, clear involvement in perception of biological motion sequences under nonmasked conditions.

We are not the first, of course, to observe changes in brain activation with perceptual learning. As indicated earlier, neural responses within ventral temporal brain areas associated with face and object perception are larger after observers have learned to recognize objects and faces portrayed by degraded images that were unrecognizable prior to learning (Dolan et al., 1997). Similarly, the spatial extent of activation within visual area MT+, a brain region involved in motion perception, increases following training on a direction discrimination task (Vaina, Belliveau, et al., 1998). Increased BOLD signals in the lateral occipital complex (LOC) result when observers are able to perceptually group line elements into coherent shapes compared to when those same elements are seen as individual, unconnected lines (Murray et al., 2002). Moreover, other studies (e.g.,

Raichle et al., 1994) find that improved performance with practice can be accompanied by ''decreased'' signal strength in given brain regions. In this regard, it should be noted that we are not claiming that activation levels in the STSp and/or the FFA increase as the result of practice—indeed, our results showed no evidence that unmasked biological sequences yield stronger BOLD signals after training in either the STSp or the FFA. What has changed consequent to training in our study is the strength of signals within these brain areas in response to biological sequences viewed under conditions that rendered them difficult, if not impossible, to see prior to training.

It is natural to wonder about the nature of the cellular events underlying these strengthened BOLD signals and the concomitant improvements in perceptual ability. It is true that single-unit studies reveal several different forms of experience-dependent neural plasticity, including recruitment of new populations of neurons (Recanzone, Schreiner, & Merzenich, 1993), sharpening in stimulus selectivity as reflected in the tuning of sensory neurons (Recanzone et al., 1993), and changes in the connection strengths among neurons (Saarinen & Levi, 1995; Zohary, Celebrini, Britten, & Newsome, 1994). But exactly how these forms of neural plasticity relate to changes in BOLD signals we do not know. Indeed, at present, we are still trying to understand the relative contributions of the multiple cellular events underlying the BOLD signal itself (Lauritzen & Gold, 2003; Logothetis, Pauls, Augath, Trinath, & Oeltermann, 2001), so it is premature to relate experience-dependent BOLD signal changes to specific cellular mechanisms. Whatever the particulars of the neural bases of the BOLD signal and its changes with learning, we are encouraged that fMRI is sufficiently sensitive to register modifications in neural activity associated with varying degrees of perceptual learning. And it is noteworthy that those modifications in neural activity include not only the STSp but also the FFA, a ventral stream brain area conventionally associated with form perception.

On a final note, recall that our observers were able to discern biological motion embedded within noise dots following training but not before, and this change in perceptual ability was mirrored by activation changes in the STSp. Earlier work in our laboratory has found that STSp activation is reduced when observers view inverted biological motion sequences that are difficult to recognize (Grossman & Blake, 2001) and abolished when perception of biological motion is disrupted by binocular rivalry (Kim, Blake, & Grossman, submitted). Additionally, a recent study by Pelphrey et al. (2003) finds that the STSp is not simply activating to complex motion patterns, but instead prefers articulated biological motion over complex, nonbiological motion. Considered together, these results provide converging evidence linking brain activation in the STSp to perception of biological motion.

METHODS

Participants

Eight individuals (3 men, 5 women), with normal or corrected-to-normal vision, participated in this study. The observers had no experience viewing point-light animations, and none had participated in a biological motion experiment before. Prior to participation the observers gave informed, written consent as approved by the Vanderbilt University Institutional Review Board.

Animations

Point-light biological motion sequences and scrambled animation sequences were created from videotapes of an individual performing various activities while wearing dark clothing with reflective tape on the joints. The videotapes were digitized and the joint positions encoded as initial positions and vector motions from those starting positions. Small black dots subtending approximately 9 arc min of visual angle replaced the joints and were displayed against a gray background using Matlab (Mathworks, Natick, MA) in conjunction with the Psychophysical Toolbox (Brainard, 1997; Pelli, 1997). Biological figures subtended 6° by 3° of visual angle and were animated for 1 sec. Scrambled animations were created by randomizing the starting x,y position within a region approximating the biological figure. On each presentation, the spatial location of the ''target figure''—whether biological or scrambled—was displaced in a random direction from the center of the screen to prevent observers from basing their judgments on the motions of dots within a particular region of the display. A fixation cross in the center of the field of dots remained visible at all times.

In the masked condition, noise dots identical in size and contrast to the ''figure'' dots were spatially distributed across the entire 19.2° × 14.4° visual display (Figure 1). To effectively camouflage the biological figure, these noise dots moved with the same local motions and in the same phase as the target figure dots. In this arrangement, a target biological kicker is masked by ''kicker'' dots spatially randomized across the screen, a thrower is masked by a ''thrower'' dots, and so forth. While the noise array and scrambled animations contained the same local component motions as the target biological animations, only the dots in the biological sequences conformed to the proper spatio-temporal arrangement required to depict a human figure.

Behavioral Testing Procedure

In an initial psychophysical testing session, a baseline level of discrimination performance was estimated for each observer using a two-alternative forced choice (2AFC) staircase procedure. Observers pressed a key to initiate the 1-sec presentation of an animation com-

prising a figure (biological or scrambled) embedded in noise dots. Following each presentation, observers pressed one of two buttons to indicate whether the sequence depicted a biological figure or a scrambled figure. Over trials, the figure was randomly chosen to be biological or scrambled with equal probability. Using a 2-up/1-down rule, the staircase converged to a noise level corresponding to 71% correct performance. Initially, only 12 dots masked the target figure, and so the first few trials were trivially easy for observers. However, after successive correct responses, the number of noise dots masking the target increased, and eventually the task became quite difficult. In the early trials, the number of dots added or removed was set to a relatively large number (12 dots) so that the staircase would converge relatively quickly to the threshold level of masking noise. Following 12 reversals in the staircase, the number of dots added or removed was lowered to six, to obtain a more refined estimate of the observer's threshold. After 36 reversals, the staircase was terminated and the threshold was defined as the mean number of noise dots in the trials comprising the last 10 reversals.

This estimate of the observer's initial discrimination threshold was used to set a ''target noise level'' that served as the goal to meet or exceed with repeated practice. Pilot experiments with observers who did not participate in this full experiment determined that observers could learn to perform the task at 71% correct with the animation masked in approximately 150 noise dots. With this as a guide, the target number of noise dots was set for each observer based on the threshold level measured the first day. Slightly lower target thresholds were set for those individuals with baseline thresholds far below 150, while higher targets were set for individuals with initial baselines near or greater than 150.

Once the target threshold was set, each observer's sensitivity, as measured by hits and false alarms (*d*-prime: d'), was assessed at this elevated level of difficulty. Observers were administered 200 trials in which a biological or scrambled sequence was masked by the target level of noise dots and indicated with a keypress whether the animation depicted biological or scrambled motion. Observers were instructed to distribute their responses evenly between the two categories when guessing was necessary. This was particularly important for the pretraining assessment, as the noise level was set to exceed the discrimination threshold of the observer, thus rendering the task extremely difficult. Error feedback was not given during the pretraining test session.

Also in this initial experimental session, observers were given practice on a version of the ''1-back'' task they would be performing during the scanning session. In the scanner, observers viewed a series of animations and were required to indicate when two successive sequences portrayed identical patterns of motion. In this practice session observers viewed two intervals of the masking dots alone (no target present) and indicated with a keypress whether the dynamics portrayed in the sequence just seen was identical to the one seen on the previous trial.

On successive training days, observers repeatedly performed the 2AFC staircase procedure described above. The staircase was chosen as the means for promoting perceptual learning because it gave observers initial exposure to easy trials in which the target animation was masked with very few noise dots while still focusing the majority of the trials on difficult levels of noise that, by the very nature of the staircase procedure, were at or near the limits of discriminability. Observers received error feedback on each trial during these training sessions. Because performance on discrimination tasks is sleep-dependent and may actually decline throughout a single day (Mednick et al., 2002), observers completed a single staircase on each day. Observers continued these daily training sessions until the estimated threshold exceeded the target level of noise dots; for no observer was this goal achieved with less than four training sessions. Once an observer reached or exceeded the target threshold, that observer's d' value was measured for that elevated level of noise, in exactly the same way that d' was measured prior to training.

The goal of this portion of the experiment was to determine the extent to which training produced a general improvement in the perception of biological motion rendered incoherent under conditions of masking. It is possible, however, that improvement in performance with training could result simply from increased familiarity with the specific exemplars experienced during training (Nosofsky, 1986). To differentiate between learning and familiarity, we divided the 24 available animations into two sets of 12 animations, a ''trained'' set that was used during initial testing and training, and a ''novel'' set that was used for testing only after training. These trained and novel sets were randomly established for each observer.

Imaging Procedures and Data Analysis

Observers also participated in two scanning sessions. The first scan session occurred following the estimate of the initial threshold of noise tolerance and prior to all training sessions. This scan session served as the baseline to which changes in neural activity associated with training would be compared. Observers viewed blocks of biological motion interleaved with blocks of scrambled motion. In some scans, the animations were masked in noise dots (the target level of noise), while in other scans the animations were presented without noise. The scans without noise dots served as a localizer to identify ROIs (STSp and FFA) in each observer. Each scan was repeated twice, and the two runs were averaged. Following training, observers returned to the scanner and repeated the scans with the same level of masking noise. Four of the eight observers participated in an additional scan

using a more traditional localizer to identify the FFA. In this scan, observers viewed blocks of gray-scale images of faces alternately with blocks of common objects (approximately 6° × 6°) (Kanwisher, McDermott, & Chun, 1997; Haxby et al., 1996; Puce, Allison, Gore, & McCarthy, 1995).

Brain images were collected on a 3 T GE Signa scanner located within the Vanderbilt University Medical School. High-resolution T1-anatomical images were collected for each observer (184 slices, 1.0 × 1.0 × 0.9375 mm). Functional images (single-shot EPI, TR = 2000 msec, TE = 25 msec, flip = 90°) were acquired over the occipital lobe, the posterior parietal cortex and the ventral temporal cortex (14 approximately axial slices, 1.875 × 1.875 mm in-plane, 5 mm thick, no gap).

Visual animations were viewed on MR-compatible LCD monitors mounted inside goggles (Resonance Technology; Northridge, CA). Functional scans lasted 240 sec, the initial 8 sec (4 volumes) of which were discarded prior to analysis to allow for MR stabilization. The 4-min scan was divided into seven blocks of biological and seven blocks of scrambled motion (all either masked or unmasked, depending on the scan). Within each 14-sec block, seven 1-sec animations were presented with an interstimulus interval of 1 sec. A fixation cross remained visible throughout the scan, and observers were instructed to maintain fixation while attending to the entire stimulus. To encourage sustained attention during the scan sessions, observers performed a 1-back task in which they indicated with a button press whenever an animation was the same as the one seen on the previous presentation. To prevent observers from basing their responses on the behavior of one or two individual dots, we spatially varied the exact location of the target (biological or scrambled animation) relative to the fixation point on each trial. Moreover, for blocks that included noise dots, we created new noise dots for each trial, which made it impossible to judge whether two successive sequences were the ''same'' simply based on the particular motions of specific dots at a given location within the display. At all times observers had to attend to the overall patterns of motion within successive presentations in order to perform this task. It is important to note that this 1-back task did not involve discriminating biological from scrambled sequences, the task deployed during behavioral training.

Image analysis was conducted using Brain Voyager 4.4 (Brain Innovations, Maastricht, The Netherlands). All images were detrended to remove any linear drift in time, then multifiltered with a 4-mm FWHM spatial filter (Skudlarski, Constable, & Gore, 1999). Each scan was repeated, and the two runs were averaged.

The STSp ROIs were localized as in previous studies (i.e., Grossman & Blake, 2001). In short, ROIs were created from voxels highly correlated ($p < .01$) with viewing biological versus scrambled motion in the localizer scan (unmasked scan) on the posterior extent of the STS. In most observers, this localizer also revealed a region of correlated activity on the middle fusiform gyrus on the ventral surface of the temporal lobe. The location of this region was verified with other accounts of the face-responsive area on the temporal lobe (Gauthier, Skudlarski, Gore, & Anderson, 2000; Halgren et al., 1999; Kanwisher et al., 1997), as well as within four of our observers that participated in traditional FFA localizer scan (viewed images of faces and objects).

The voxels within each ROI were averaged to create a single time series for each individual. MR signal levels during alike blocks were averaged to create a single estimate of BOLD activity for each stimulus condition. Percent changes in BOLD signal associated with viewing biological motion (either in masked or unmasked) were calculated as the difference between the MR signal during the biological motion blocks and the MR signal during the scrambled motion blocks, divided by the average BOLD signal during the scrambled blocks.

Acknowledgments

These experiments were supported by NSF BCS0079579, NSF BCS0121962, NIH EY07760, NIH EY14437, and NIH EY614437.

Reprint requests should be sent to Emily D. Grossman, Department of Cognitive Sciences, 3151 Social Sciences Plaza, University of California, Irvine, Irvine, CA 92697, or via e-mail: grossman@uci.edu.

The data reported in this experiment have been deposited in the fMRI Data Center (http://www.fmridc.org). The accession number is 2-2004-116YQ.

REFERENCES

Ahissar, M., & Hochstein, S. (1996). Learning pop-out detection: Specificities to stimulus characteristics. *Vision Research, 36,* 3487–3500.

Ahlstrom, V., Blake, R., & Ahlstrom, U. (1997). Perception of biological motion. *Perception,* 1539–1548.

Ball, K., & Sekuler, R. (1982). A specific and enduring improvement in visual motion discrimination. *Science, 218,* 697–698.

Beauchamp, M. S., Lee, K. E., Haxby, J. V., & Martin, A. (2003). FMRI responses to video and point-light displays of moving humans and manipulable objects. *Journal of Cognitive Neuroscience, 15,* 991–1001.

Bertenthal, B., & Pinto, J. (1994). Global processing of biological motion. *Psychological Science, 5,* 221–225.

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision, 10,* 433–436.

Dolan, R. J., Fink, G. R., Rolls, E., Booth, M., Holmes, A., Frackowiak, R. S. J., & Friston, K. J. (1997). How the brain learns to see objects and faces in an impoverished context. *Nature, 389,* 596–599.

Dosher, B. A., & Lu, Z.-L. (1999). Mechanisms of perceptual learning. *Vision Research, 39,* 3197–3221.

Fahle, M. (1997). Specificity of learning curvature, orientation, and vernier discriminations. *Vision Research, 37,* 1885–1895.

Fine, I., & Jacobs, R. A. (2002). Comparing perceptual learning across tasks: A review. *Journal of Vision, 2,* 190–203.

Fox, R., & McDaniel, C. (1982). The perception of biological motion by human infants. *Science, 218,* 486–487.

Gauthier, I., Skudlarski, P., Gore, J. C., & Anderson, A. W. (2000). Expertise for cars and birds recruits brain areas

involved in face recognition. *Nature Neuroscience, 3,* 191–197.
Giese, M. A., & Poggio, T. (2003). Neural mechanisms for the recognition of biological movements. *Nature Reviews Neuroscience, 4,* 179–192.
Gold, J., Bennett, P. J., & Sekuler, A. B. (1999). Signal but not noise changes with perceptual learning. *Nature, 402,* 176–178.
Goldstone, R. L. (1998). Perceptual learning. *Annual Review of Psychology, 49,* 585–612.
Grossman, E., & Blake, R. (2001). Brain activity evoked by inverted and imagined biological motion. *Vision Research, 41,* 1475–1482.
Grossman, E., & Blake, R. (2002). Brain areas active during visual perception of biological motion. *Neuron, 35,* 1157–1165.
Grossman, E., Donnelly, M., Price, R., Pickens, D., Morgan, V., Neighbor, G., & Blake, R. (2000). Brain areas involved in perception of biological motion. *Journal of Cognitive Neuroscience, 12,* 711–720.
Halgren, E., Dale, A. M., Sereno, M. I., Tootell, R. B., Marinkovic, K., & Rosen, B. R. (1999). Location of human face-selective cortex with respect to retinotopic areas. *Human Brain Mapping, 7,* 29–37.
Haxby, J. V., Ungerleider, L. G., Horwitz, B., Maisog, J. M., Rapoport, S. I., & Grady, C. L. (1996). Face encoding and recognition in the human brain. *Proceedings of the National Academy of Science, U.S.A., 93,* 922–927.
Johansson, G. (1973). Visual perception of biological motion and a model for its analysis. *Perception and Psychophysics, 14,* 195–204.
Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The fusiform face area: A module in human extrastriate cortex specialized for face perception. *Journal of Neuroscience, 17,* 4302–4311.
Kim, C. Y., Grossman, E., & Blake, R. (2002). Biologically relevant events are undetectable during suppression phases of binocular rivalry Program No. 161.12.2002. *Abstract Viewer/Itinerary Planner.* Washington, DC: Society for Neuroscience, online.
Lauritzen, M., & Gold, L. (2003). Brain function and neurophysiological correlates of signals used in functional neuroimaging. *Journal of Neuroscience, 23,* 3972–3980.
Logothetis, N. K., Pauls, J., Augath, M., Trinath, T., & Oeltermann, A. (2001). Neurophysiological investigation of the basis of the fMRI signal. *Nature, 412,* 150–157.
Mather, G., Radford, K., & West, S. (1992). Low-level visual processing of biological motion. *Proceedings of the Royal Society of London. Series B: Biological Sciences, 249,* 149–155.
Mednick, S. C., Nakayama, K., Cantero, J. L., Atienza, M., Levin, A. A., Pathak, N., & Stickgold, R. (2002). The restorative effect of naps on perceptual deterioration. *Nature Neuroscience, 5,* 677–681.
Murray, M. M., Wylie, G. R., Higgins, B. A., Javitt, D. C., Schroeder, C. E., & Foxe, J. J. (2002). The spatiotemporal dynamics of illusory contour processing: Combined high-density electrical mapping, source analysis, and functional magnetic resonance imaging. *Journal of Neuroscience, 22,* 5055–5073.
Nosofsky, R. M. (1986). Attention, similarity, and the identification–categorization relationship. *Journal of Experimental Psychology: General, 115,* 39–57.
Oram, M. W., & Perrett, D. I. (1994). Responses of anterior superior temporal polysensory (STPa) neurons to "biological motion" stimuli. *Journal of Cognitive Neuroscience, 6,* 99–116.
Pavlova, M., Krageloh-Mann, I., Sokolov, A., & Birbaumer, N. (2001). Recognition of point-light biological motion displays by young children. *Perception, 30,* 925–933.
Pavlova, M., & Sokolov, A. (2000). Orientation specificity in biological motion perception. *Perception and Psychophysics, 62,* 889–898.
Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision, 10,* 437–442.
Pelphrey, K. A., Mitchell, T. V., McKeown, M. J., Goldstein, J., Allison, T., & McCarthy, G. (2003). Brain activity evoked by the perception of human walking: Controlling for meaningful coherent motion. *Journal of Neuroscience, 23,* 6819–6825.
Pinto, J., & Shiffrar, M. (1999). Subconfigurations of the human form in the perception of biological motion displays. *Acta Psychologica, 102,* 293–318.
Ptito, M., Faubert, J., Gjedde, A., & Kupers, R. (2003). Separate neural pathways for contour and biological-motion cues in motion-defined animal shapes. *Neuroimage, 19,* 246–252.
Puce, A., Allison, T., Gore, J. C., & McCarthy, G. (1995). Face-sensitive regions in human extrastriate cortex studied by functional MRI. *Journal of Neurophysiology, 74,* 1192–1199.
Raichle, M. E., Fiez, J. A., Videen, T. O., MacLeod, A. M., Pardo, J. V., Fox, P. T., & Petersen, S. E. (1994). Practice-related changes in human brain functional anatomy during nonmotor learning. *Cerebral Cortex, 4,* 8–26.
Recanzone, G. H., Schreiner, C. E., & Merzenich, M. M. (1993). Plasticity in the frequency representation of primary auditory cortex following discrimination training in adult owl monkeys. *Journal of Neuroscience, 13,* 87–103.
Saarinen, J., & Levi, D. M. (1995). Perceptual learning in Vernier acuity: What is learning? *Vision Research, 35,* 519–527.
Santi, A., Servos, P., Vatikiotis-Bateson, E., Kuratate, T., & Munhall, K. (2003). Perceiving biological motion: Dissociating visible speech from walking. *Journal of Cognitive Neuroscience, 15,* 800–809.
Schiltz, C., Bodart, J. M., Dubois, S., Dejardin, S., Michel, C., Roucoux, A., Crommelinck, M., & Orban, G. A. (1999). Neuronal mechanisms of perceptual learning: Changes in human brain activity with training in orientation discrimination. *Neuroimage, 9,* 46–62.
Servos, P., Osu, R., Santi, A., & Kawato, M. (2002). The neural substrates of biological motion perception: An fMRI study. *Cerebral Cortex, 12,* 772–782.
Skudlarski, P., Constable, R. T., & Gore, J. C. (1999). ROC analysis of statistical methods used in functional MRI: Individual subjects. *Neuroimage, 9,* 311–329.
Sumi, S. (1984). Upside-down presentation of the Johansson moving light-spot pattern. *Perception, 13,* 283–286.
Talairach, J., & Tournoux, P. (1988). Co-planar stereotaxic atlas of the human brain. New York: Thieme Medical.
Vaina, L. M., Belliveau, J. W., des Roziers, E. B., & Zeffiro, T. A. (1998). Neural systems underlying learning and representation of global motion. *Proceedings of the National Academy of Sciences, U.S.A., 95,* 12657–12662.
Vaina, L. M., Solomon, J., Chowdhury, S., Sinha, P., & Belliveau, J. W. (2001). Functional neuroanatomy of biological motion perception in humans. *Proceedings of the National Academy of Sciences, U.S.A., 98,* 11656–11661.
Vidyasagar, T. R., & Stuart, G. W. (1993). Perceptual learning in seeing form from motion. *Proceedings of the Royal Society of London. Series B: Biological Sciences, 254,* 241–244.
Zohary, E., Celebrini, S., Britten, K. H., & Newsome, W. T. (1994). Neuronal plasticity that underlies improvement in perceptual performance. *Science, 263,* 1289–1292.