Perceptual History Influences Neural Responses to Face and Body Postures

Tjeerd Jellema[1] and David I. Perrett[2]

Abstract

■ We show that under natural viewing, the responses of cells in the temporal lobe of the macaque to the sight of static head and body postures is controlled by the sight of immediately preceding actions. Cells in the anterior part of the superior temporal sulcus responded vigorously to the sight of a face or body posture that followed a particular body action, but not when it followed other actions. The effective action or posture presented in isolation or in different sequences failed to produce a response. Our results demonstrate that cells in the temporal cortex could support the formation of expectations about impending behavior of others. ■

INTRODUCTION

The interpretation of a visual scene is profoundly influenced by information other than that in current visual input. One important source of information is the perceptual history (events that have been recently witnessed), which builds up expectations and can help resolve the interpretation of ambiguous or degraded visual stimuli (Dolan et al., 1997). It may also play a role in creating ''object permanence'' (Baker, Keysers, Jellema, Wicker, & Perrett, 2001) and might allow one to anticipate sensory events (Mistlin & Perrett, 1990). Our understanding of how the brain mediates these effects is, however, still rudimentary.

The stream of processing from occipital to temporal cortex (the ventral stream) is characterized by a gradual increase in the complexity of stimuli analyzed by cells, which could support the recognition of complex objects (Kobatake & Tanaka, 1994; Perrett & Oram, 1993; Goodale & Milner, 1992). The anterior part of the superior temporal sulcus (STSa) in the macaque monkey is considered part of the ventral stream, as STSa cells often maximally respond to the shape and motions of animate objects, most notably conspecifics and humans (Jellema & Perrett, 2002; Kendrick & Baldwin, 1987; Perrett, Smith, Potter, et al., 1985; Perrett, Smith, Mistlin, et al., 1985; Perrett et al., 1989; Desimone, Albright, Gross, & Bruce, 1984; Gross, Rocha-Miranda, & Bender, 1972). Such findings have led to the idea that STSa is primarily involved in the visual analysis of actions of others (Oram & Perrett, 1996; Perrett et al., 1989). This view is supported by recent brain imaging studies (for a review, see Allison, Puce, & McCarthy, 2000), which show activation of the human STS for the perception of ''biological motion'' of human figures (Bonda, Petrides, Ostry, & Evans, 1996), hand actions (Grafton, Arbib, Fadiga, & Rizzolatti, 1996; Rizzolatti et al., 1996), static faces (Allison et al., 2000), eye gaze and eye motion (Hoffman & Haxby, 2000; Puce, Allison, Bentin, Gore, & McCarthy, 1998), and meaningful actions (Decety et al., 1997).

The traditional view is that in passive viewing situations, the responses of visual temporal cells are straightforwardly related to the current visual stimulation of the retina; the response selectivity is determined by the ''geometrical'' characteristics of the stimuli. However, studies of delayed matching-to-sample (Miller, Li, & Desimone, 1993; Miyashita, 1988) and paired-association tasks (Naya, Sakai, & Miyashita, 1996; Sakai & Miyashita, 1991) have shown that neurons in the macaque anterior temporal cortex can acquire stimulus selectivity through learning and are able to link the representations of temporally associated stimuli, even though the stimuli are geometrically dissimilar. The temporal cortex may therefore be involved in forming and storing long-term visual associative memories (Booth & Rolls, 1998). The sensitivity to multiple stimuli that are conceptually related, even though visually distinct, is echoed in the tuning of cells to multiple views of the same object (Jellema, Baker, Oram, & Perrett, 2002; Logothetis, Pauls, & Poggio, 1995; Perrett, Mistlin, & Chitty, 1987) and in the tuning for multiple body signals of directed attention (Oram & Perrett, 1992; Perrett, Hietanen, Oram, & Benson, 1992; Perrett, Smith, Mistlin, et al., 1985).

Here, we report that the responsiveness of some cells in STSa to faces and bodies can be best explained by postulating a sensitivity to the perceptual history, that is, to specific visual patterns that were on the retina in

[1]Utrecht University, [2]University of St. Andrews

the immediate past. In effect, the perceptual history can enable or prevent a cell's response to the current retinal input. We argue that, as a result, the ''vocabulary'' of actions and body postures coded for by single STSa cells is much larger than previously thought, also encompassing specific action–posture sequences. The findings are discussed in relation to their role in comprehending complex actions and behavior. Preliminary results of this study were reported in abstract form (Jellema & Perrett, 2001).

RESULTS

In this study, 272 cells in STSa were found to be visually responsive and were examined further. For the majority of the cells (208/272, 76%), maximal responses were obtained when the visual object was a conspecific or a human (i.e., an animate object) performing a particular action, or assuming a particular body posture. Only these cells were subjected to detailed analysis. The other 66 responded less selectively to various visual stimuli and motions. Typically, the visual stimulus consisted of a sequence of first a bodily action, directly followed by a static body posture. The latter formed the natural, or most likely, ''end-point'' of the action. Examples are extension of the arm resulting in a static extended arm; backward bending of the upper body resulting in a convex body posture with attention directed upward; downward motion of the bipedal body involving bending of the knees and resulting in a crouched position; and whole-body walking followed by stopping, resulting in the static view of that individual. Some movements do not have a natural end-point, in which case the end-point was defined as the point in the trajectory of an action at which the cell stopped responding if the action continued. For example, a cell might respond during rotation of the body toward the observer but would cease to respond to rotation once the body begun to turn away from the observer.

Each of the 208 cells was allocated to one of three cell categories according to their responses during the motion and static phases of the visual stimuli. (1) Cells active during both the motion and the static phase (66/208, 32%). (2) Cells exclusively active during the motion phase and silent during the static phase (125/208, 60%). (3) Cells silent during the motion phase and exclusively active during the static phase (17/208, 8%).

Further testing suggested that this division of cells did not fully indicate the response properties of subsets of the cells that responded to the static phase. It became clear that for some cells the responses to the static body posture were influenced by the preceding body actions. We investigated these properties for 54 cells responding to static postures of the body and/or face, which formed a subset of Categories 1 and 3 (thus, excluding Category 2).

Three experimental conditions were employed using live and/or video stimuli. (1) Static body posture views were presented preceded by various body actions, each of which led in a natural (consistent) way to that body posture view. (2) The same static body posture views were presented preceded by a blank (closed shutter), that is, without a preceding action. (3) The body actions were directly followed by a visual stimulus (moving or static) different to the static posture. This could either be a continuation of the action, or a static image in which the agent was occluded.

The nature of the movement preceding a static posture proved critical to the modulation of cell responses. All 54 cells were tested with a blank and at least one body action as perceptual histories, whereas 35 of them were tested with up to four different body actions leading up to one and the same static body posture in a natural (consistent) way. For 18 of these 35 cells (51%), significant differences were found between responses to a particular static body posture when preceded by different actions. In total, we found 29 cells sensitive to perceptual history, the remaining 25 cells were not sensitive to the perceptual history.

Figure 1 shows the population response of the cells for which the response during the static phase depended on the preceding motion phase. Individual mean cell responses were entered for an ''effective'' and for an ''ineffective'' perceptual history in 100-msec time bins (the two histories consisted of either two different actions leading up to the same posture, or an action leading to a posture and a blank preceding the same posture). Two-way ANOVA with history and time

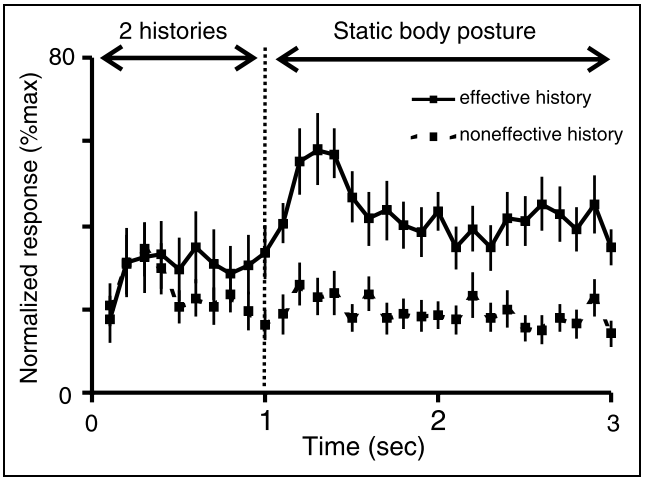

Figure 1. Population response to static body postures dependent on perceptual history. The responses of 25 cells sensitive to perceptual history were pooled. Cells were tested with two different perceptual histories, directly followed by the same static body posture. The two histories were either two different bodily actions (n = 18), or one body action and a blank (n = 7). Only cells recorded for at least 2 sec of the static phase were included. For each cell, the mean spike count in 100-msec time bins was expressed as a percentage of the maximal response of that cell during the 3-sec recording period.

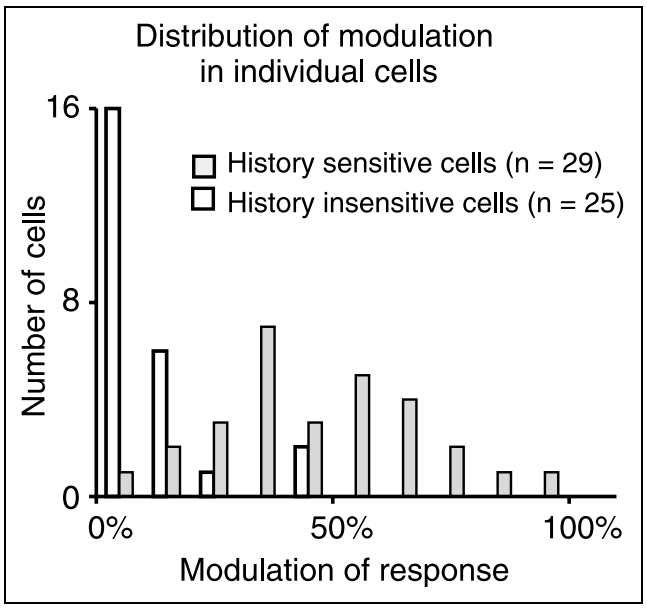

Figure 2. The distribution of the response modulations of individual cells sensitive and insensitive to perceptual history. For each cell, the modulation was calculated as [(Max − Min)/Max] × 100.

(consisting of three levels: seconds 1, 2, and 3, starting at the start of the motion phase) as factors showed a main effect of history, $F(1,48) = 16.4$, $p = .0003$, no main effect of time, $F(2,96) = 1.98$, $p = .15$, and an interaction effect of history and time, $F(2,96) = 4.7$, $p = .012$. Post hoc testing showed that the responses to the static posture when preceded by the ''effective'' history were significantly larger than when preceded by the ''ineffective'' history, both during second 1 ($p < .0002$) and during second 2 ($p < .0005$) of the static phase. The responses during the two histories did not differ ($p = .2$).

There was a considerable gitter in the onset latencies of the response to the static phase in the individual cells contributing to the population response in Figure 1. The gitter resulted from variations in onset latencies between different cell types (compare Figures 3 and 5 to Figure 4), and from variations in the exact duration of the action phase and, thus, in the start of the static phase, when the stimuli were presented live.

It was surprising to find that the responses during the static phase when preceded by an ''effective'' action were often quite prolonged. At the level of individual cells, we found in 14/25 cells (56%) that the response during second 2 of the static phase was still significantly enhanced (in eight cells, this lasted for up to 10 sec). The relatively long duration of response to static stimuli was surprising because the ''classical'' response of STSa cells to, for example, a static face, presented on video or by the opening of a shutter (live), typically shows a strong transient elevation of discharge, with the response rapidly declining thereafter (Oram & Perrett, 1992).

Figure 2 shows that the degree of modulation, defined as [(Max − Min)/Max] × 100 (see Methods), varied between individual cells that were sensitive ($n = 29$) and not sensitive ($n = 25$) to perceptual history. For the cells sensitive to perceptual history, modulation was most often in the 40–60% range; for cells insensitive to history, modulation was lower.

The sensitivity to perceptual history applied to virtually all actions that were found to be effective in evoking STSa cell responses (Table 1), stressing the pervasive influence of the immediate history on perceptual processing. The body actions were divided into four types: rotations of the head/body in the horizontal plane, rotations of the head/body in the vertical plane, walking, and ''other actions.'' The latter group contained various actions such as bending or stretching of the knees, reaching with the arm, and extension of the leg.

Some cells ($n = 13$) did not respond during the sight of the particular action that was required to produce the response during the subsequent static phase. Thus, although the type of preceding motion was crucial for evoking a response during the static phase, the cells did not respond to or discriminate between the different action types. Such cells are easily overlooked, simply because the one ''effective'' action was not presented, and, therefore, their frequency of occurrence is likely to be underestimated. Figure 3 gives an example of such a cell. This cell belonged to a subpopulation of cells ($n = 6$), which responded selectively to the cessation of walking. None of these cells responded following cessation of translation movements of control objects. The cell illustrated in Figure 3 responded to the static front view of the body following cessation of walking forward toward the subject. However, when preceded by other types of body actions, the identical static front view of the body,

Table 1. Subdivision of Cells Sensitive to the Perceptual History According to the Type of ''Effective'' Preceding Body Action

Preceding Body Action	*No. of Cells Sensitive to Perceptual History*	*Percentage of Cells*
Rotation of head/body in horizontal plane	9/25	36
Rotation of head/body in vertical plane	10/14	71
Walking	6/8	75
Other actions	4/7	57
Total	29/54	54

Cells ($n = 54$) responsive to the sight of static body postures tested for their sensitivity to perceptual history. Four body action types are discriminated (far left). The middle column gives the number of cells sensitive to perceptual history out of the total number of cells in the category.

at the identical location in the testing room, hardly evoked a response. Examples of such alternative actions were backward walking away from the subject (keeping the front view of the body directed towards the subject; Figure 3B) and body rotation through 180° from back to front view (data not shown). Presentation of the static front view following a blank also evoked a significantly smaller response (Figure 3C), $F(2,17) = 13.3$, $p < .0005$ (post hoc, $p < .001$ for each comparison with A;5-sec period of assessment). The responses to the static posture following walking backward away from the subject (B) and following a blank (C) were not different from each other ($p > .05$). Note the virtual absence of spikes during each of the body actions and during shutter closure. The response began after a delay of about 500 msec following the onset of the static phase (Figure 3A), indicating that the response was not related to the deceleration of the motion. Thus, the body actions can have highly specific consequences, even though the cell does not respond during the actions.

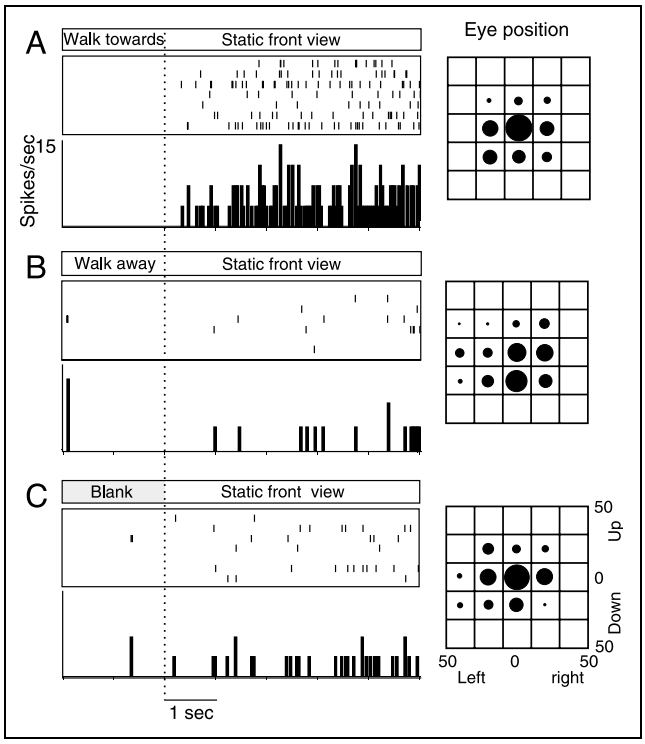

Figure 3. Cell response sensitive to perceptual history. The cell responded to the static front view of the body following walking toward the subject. (A) The actor walks toward the subject, stops at a distance of 2 m, and remains there static for 5 sec. (B) The actor walks backward away from the subject (i.e., maintaining a front view of the body to the subject) and stops at the same position in the room (2-m distance). (C) The static front view of the actor at a distance of 2 m is presented directly following shutter opening. The different phases of the visual stimulus are indicated at the top of each panel. The gray box in the visual stimulus (C) indicates a closed shutter prior to its opening to reveal a static body view. The vertical dotted lines separate the motion and static phases. The middle section of each panel shows the spike rastergrams, and the lower section shows the peristimulus time histograms (bin width 20 msec). At the right-hand side of each panel, the positions of eye fixations during the static phase (5 sec) are shown. The matrix of horizontal and vertical positions represents the subject's visual field (maximal visual angle is 100°, column and row width is 20°). The surface area of each black circle is proportional to the duration of fixation within that area of the visual field.

The percentage of time the subject spent fixating with different eye positions (Figure 3, right) was analyzed with a two-way ANOVA (trial type or perceptual history, 3 levels: walk toward, walk away, static; visual field location, 15 levels, excluding the top and bottom rows where no fixations were recorded). The analysis for this cell showed a significant main effect of visual field location, $F(14,238) = 13.6$, $p < .00001$, reflecting more frequent fixation of the central position and the experimenter than other positions, but no difference in fixation pattern across the different perceptual histories [Perceptual History × Visual Field Location interaction, $F(28,238) = 1.27$, $p = .17$].

Another class of body action frequently found to activate cells was formed by articulations of the upper torso in the vertical plane (Table 1). The cell illustrated in Figure 4 is a typical example of this category. This cell required a rotation of the upper body in the forward direction starting from a backward bent posture, in order to respond to the static upright body view (Figure 4A,B, Motion 1). The action itself did not evoke a response. The response lasted for at least 2 sec after the static upright posture was assumed. Responses to the same posture were significantly smaller when the posture was preceded by a backward directed movement starting from the forward bent position (Figure 4C, Motion 2), $F(1,12) = 29.7$, $p < .0002$ (2-sec period of assessment following cessation of motion). During Motions 1 and 2, the cell produced little spike activity. Two-way ANOVA of the type of perceptual history (2 levels: Motion 1, Motion 2) and visual field location position (15 levels) for this cell again showed a significant main effect of visual field location on eye position during the static phase of trials, $F(14,182) = 30.4$, $p < .00001$, but no difference in fixation pattern across the different perceptual histories [Perceptual History × Visual Field Location interaction, $F(14,182) = .50$, $p = .93$]. Very similar results were obtained when the experimenter was observed not from the front view, but from the left profile, the right profile, or the back views, while making the same set of movements (Figure 4D). For each view, the cell responded to the static upright posture after one type of articulation (Motion 1) but not to the same posture following a second type of articulation (Motion 2). Two-way ANOVA of the type of perceptual history (two levels: Motion 1, Motion 2) and view (four levels: front, left, right, back view) showed for this cell a significant main effect of perceptual history, $F(1,48) = 78.8$, $p < .00001$, but did not show a main effect of view, $F(3,48) = 1.3$, $p = .30$, nor an interaction effect, $F(3,48) = .23$, $p = .23$. This clearly indicates that

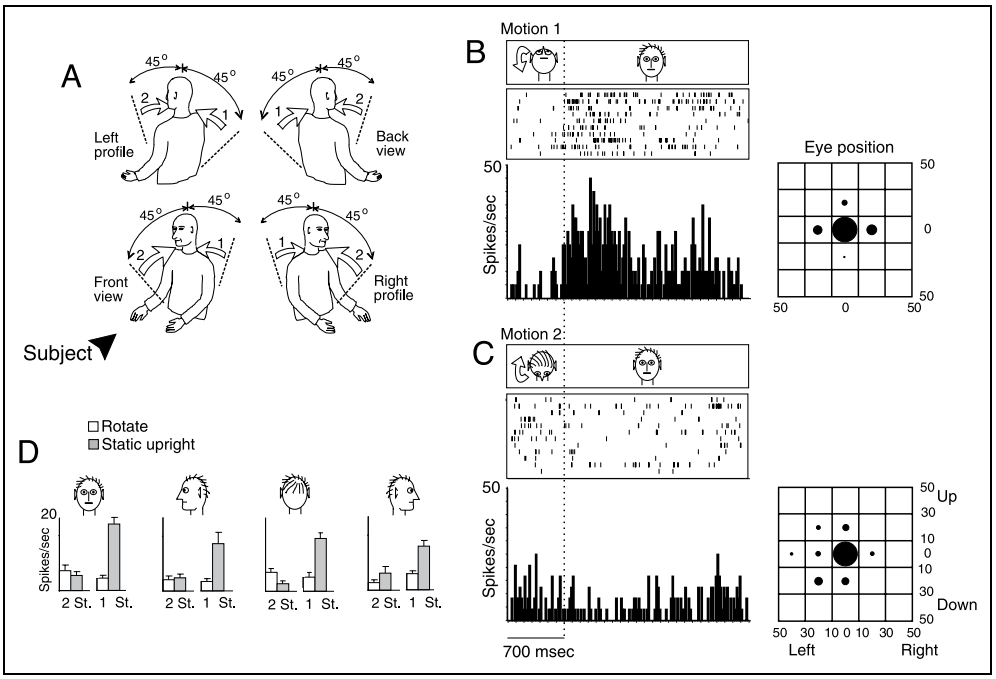

Figure 4. Cell response to the static upright body view following forward rotation of the upper torso. (A) Schematic representation of the motions and postures used to test the cell. The subject's viewing position and orientation are indicated by arrowhead at bottom left. (B) Cell response to the front view of the upright body when preceded by a forward rotation of the upper torso (Motion 1). (C) Cell response to the same static body view but preceded by a backward rotation of the upper torso (Motion 2). At the right-hand side, the mean durations of eye fixations during the static phase (2 sec) are shown (see legend to Figure 3 for description). (D) The actor was observed from different perspectives (front, left profile, back, and right profile view). Mean (± *SEM*) of responses from eight trials; clear bars = response during rotation; shaded bars = response during static upright posture (St).

the responses were not related to the direction of motion per se nor to the body view per se. Other cells in this category were selectively responsive to the upright body view when preceded by Motion 2 and not when preceded by Motion 1 (data not shown).

The cells that generalize across the viewpoints of the observer are said to code for the action–posture sequence in an object-centered manner. Previous recording in the STS has emphasized the prevalence of coding that does not generalize across viewpoint, that is, viewer-centered coding (Oram & Perrett, 1996; Perrett, Smith, Potter, et al., 1985, Perret et al., 1989). In this study, however, performed at a more anterior STS site than previous studies, a relatively large proportion of the cells (24/54 cells, 44%) exhibited object-centered coding, the other 30 cells exhibited viewer-centered coding. For the subpopulation of 29 cells found sensitive to perceptual history, 19 (65%) exhibited object-centered coding.

A third condition was employed to investigate whether the response obtained during presentation of the static body view, immediately following an ''effective'' action, depended on the actual presence of the static body view or would have happened anyway, irrespective of the stimulus that followed the ''effective'' action. In this condition, the static body view was replaced by a static scene, which did not contain the body (Condition 3i), or by another action (Condition 3ii). In Condition 3i, the body was occluded from sight at the moment in time where the static posture would otherwise have been assumed. This test was performed for the cells responding to stopping following walking forward ($n = 6$). Occlusion was achieved when the agent walked behind a screen positioned at the location where otherwise the static posture was assumed (Figure 5). The response of the cell in Figure 5 increased dramatically about 500 msec after the agent had stopped walking (Figure 5A). The cell displayed object-centered coding and was sensitive to the static posture adopted after cessation of walking forward in any direction, but not following cessation of walking backward. When the agent became occluded from sight after walking, this increase in response was absent (Figure 5C). All six cells showed this lack of response during occlusion. Presentation of the identical static body view following opening of the shutter also did not excite the cell (Figure 5B), $F(2,27) = 36.4$, $p < .00001$ (4-sec period of assessment following termination of motion phase or blank). Post hoc testing showed that the response during the static back view of the body following cessation of walking (A) was larger than following a blank (B) and when the body was hidden from sight after walking forward (c) ($p < .0002$ each comparison with (A); (B) vs. (C), $p > .05$).

In Condition 3ii, the ''effective'' action continued beyond the point where otherwise the static posture would have been assumed. This condition was applied

to three types of cell: cells responding to stopping after walking (n = 6) and cells responding to head/body rotations in the horizontal plane (n = 9) and vertical plane (n = 6). Continuation of the ''effective'' action did not result in a change in spiking frequency in these cells, whereas presentation of the static body view did enhance the response. The results from these two controls (Conditions 3i and 3ii) strongly indicate that the response during the static phase requires the presence of the particular static body view and cannot be explained by cessation of the visibility of the effective action or by a delayed, or continued response to the effective action.

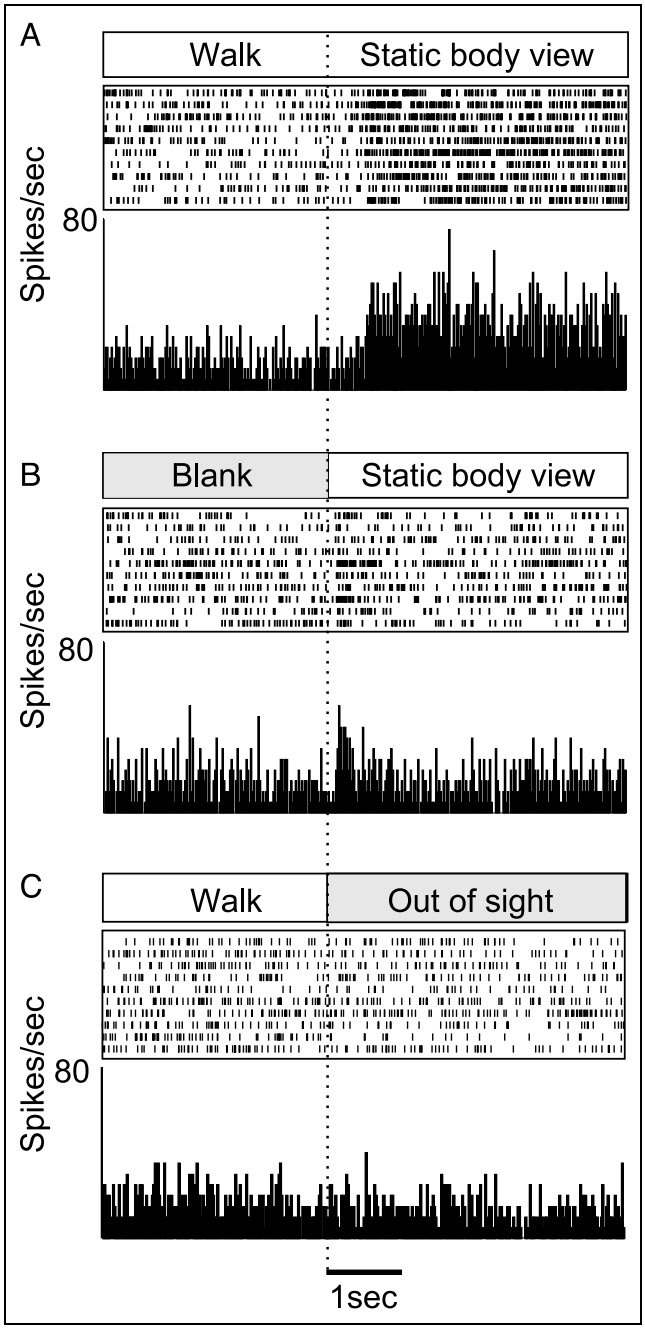

Figure 5. Cell response sensitive to a walking–static posture sequence, but not to isolated walking and static sequence elements. (A) Vigorous response to the static view of the experimenter after the experimenter had stopped walking. (B) The cell gave a small transient response to the identical static view when presented immediately following opening of the shutter. (C) The experimenter did not stop walking while in view as in (A), but kept on walking until he completely disappeared from view behind an occluding screen. The activity during full occlusion of the experimenter did not differ from the activity during walking (p > .1).

The magnitudes of responses during the motion phase and the subsequent static phase were positively correlated both for cells sensitive to the history (R^2 = .66) and for cells insensitive to history (R^2 = .74). Because the relationship between the responses in the motion and static phases is equivalent for both types of cell, ''low-level'' interactions between activity in the motion and static phases cannot predict the modulation of response by perceptual history. The distributions of the correlation values for individual cells of the history sensitive (mean = +0.20) and history insensitive (mean = +0.24) cell types were statistically equivalent (t test, p > .7).

DISCUSSION

We provided the first direct evidence that the response selectivity of a substantial number of cells in STSa in passive viewing situations cannot be fully understood on the basis of merely the current visual stimulation pattern on the retina: images from the immediate past control the response to the current stimulus. We presented temporal sequences of body actions and postures and showed that both the particular body action and the particular static body posture that followed the action were necessary to evoke a response to the body posture.

The necessity of the particular action was demonstrated by testing a range of body actions all leading to the same posture in a natural way. This showed that for each cell, just one of several actions was able to enhance the response. Because all actions tested were consistent with the posture, but only one action produced the effect, consistency may be necessary but not sufficient to evoke the response. Moreover, in the absence of any preceding action, the posture did not evoke a response.

The necessity of the particular static posture was demonstrated by replacing it by other stimuli (continued actions or static views of the laboratory with the agent hidden behind an occluding screen). Although these alternative stimuli also followed the effective action in a natural way, they did not evoke responses. The long latency (500 msec, see Figures 3 and 5) and long duration (up to 10 sec) of the responses during the static phase when preceded by an ''effective'' action are also incompatible with the possibility that the response constituted a continued response to the action itself.

Previous studies indicate that cells in temporal cortex represent particular attributes of objects and are activated when these attributes are present and attended to in

the visual scene, or are recalled from memory through associations. Such a description fails to predict the visual selectivity for prior events that we describe. As a result, the view in which the cells in temporal cortex hold a ''one-to-one'' relation with the current or recollected retinal pattern (Kobatake & Tanaka, 1994) needs to be extended. These previous studies might lead one to believe that a cell in STSa either does or does not code for the sight of a complex stimulus, for example, a face. It now emerges that this activity may depend entirely on what happened to the perceived face in the preceding second or so.

The sensitivity for perceptual history applied to various types of body actions (Table 1) and is, therefore, likely to be a fundamental property of some populations of cells in STSa. From an ethological perspective, sensitivity to perceptual history is understandable. In the natural world, objects show temporal continuity when the observer or the objects move: A rotating head does not move instantaneously from the front view to the back view, skipping over the intermediate profile view. Such consistencies allow us to predict the most likely next move or posture and may underlie the neuronal sensitivity to stimuli associated in time (Naya et al., 1996; Miller et al., 1993; Sakai & Miyashita, 1991; Miyashita, 1988).

Using a delayed matching-to-sample task, Miyashita (1988) showed that cells responding optimally to a particular computer-generated image from a fixed, repeatedly presented, sequence of 97 such images, were more likely to respond also to visually dissimilar images from neighboring positions in the sequence, than to images resembling the optimal one but presented further away in the sequence. In pair-association learning tasks, the strongest and second strongest responses of temporal cortex cells were evoked by particular pairs of images with a learnt association, even though the two images had no apparent geometrical similarity (Sakai & Miyashita, 1991). This is similar to findings of cells responding to visually dissimilar stimuli, which may have the same conceptual meaning (e.g., gaze, head and body posture, indicating the same direction of attention, Perrett, Smith, Potter, et al., 1985; Perrett et al., 1992). Thus, the only similarities among the particular collections of stimuli selected by particular temporal cortex neurons in these studies were that the stimuli were often presented sequentially in time or were associated with the same conceptual outcome. However, our results are fundamentally different from temporal cortex cell responses showing sensitivity to associated pairs of static images (Naya et al., 1996; Miller et al., 1993; Sakai & Miyashita, 1991; Miyashita, 1988). For the cells that we described as sensitive to action–posture sequences, we show that component elements of the sequence (the action alone or the posture alone) do not by themselves produce responses (e.g., Figure 5). The cells showing paired associate sensitivity described previously do show responses to one of the images from the pair when this is presented in isolation.

Studies using delayed matching-to-sample paradigms have also shown that cells in inferior temporal (IT) cortex respond differently to the same stimulus depending on the behavioral context. For example, Miller et al. (1993) showed that the responses of cells in the anterior–ventral part of IT were suppressed when a test stimulus matched the sample stimuli, even when up to four intervening stimuli were presented between sample and matching test stimulus. Thus, the responses of IT neurons to the current stimulus can be influenced by traces of that stimulus held in memory. These characteristics are, however, quite different from the ones that we described in STSa and cannot explain our results. In our data, all preceding actions ''match'' the static posture to the same degree, yet, only one of them is able to enhance the response to the posture (Figure 1), whereas in the Miller study, the degree of similarity between sample and test stimulus determined the response. Moreover, in our study, unexpected stimuli (presented following a blank) did not produce any responses; in the Miller study, unexpected, new, stimuli consistently produced maximal responses. The contingencies that we observed in STSa cell responses presumably reflect temporal associations the subjects have witnessed. The contingencies in the Miller paradigm need not be induced by temporal associations but may reflect habituation to repeated stimuli. Cells involved in the linking of temporally associated motor events are also found in the monkey supplementary motor area (SMA) (Tanji & Shima, 1994). When monkeys were trained to perform three different actions (push, pull, or turn a manipulandum) in different orders, the activity of many SMA cells was found to be dependent on the specific temporal sequence of the upcoming movements. Low-level influence of the activity from the preceding motion phase on response during the static phase can also be discarded as an explanation for our results. We show that the correlation between the responses in the motion and subsequent static phase is unrelated to the modulation by the visual sequence.

Psychophysical studies (Wallis & Bülthoff, 2001; Stone, 1998) have also stressed the importance of the temporal order of image presentation for recognition. For example, Stone (1998) found that novel objects seen rotating exclusively in one direction were more difficult to discriminate when tested with the opposite rotation direction. Wallis and Bülthoff (2001) presented image sequences of rotating heads in which the identity of the face changed as the head rotated, which did not prevent observers from treating the views as if they were from the same person. Such findings suggest that the spatiotemporal sequence of information acquired when an object is encountered or explored forms an integral part of the neural representation of that object (Wallis & Bülthoff, 2001; Stone, 1998; Koenderink & Van Doorn,

1979). Given that visual communication systems throughout the animal kingdom heavily rely on specific sequences of actions (Tinbergen, 1951), it should not be surprising that motion sequences influence primate recognition mechanisms.

The notion of the contiguity between different views of moving objects is encapsulated in the ''temporal association hypothesis'' (Perrett & Oram, 1993; Rolls, 1992), which has been successfully implemented in neural network models of recognition, especially those models that seek to explain transform-invariant object recognition (Földiák, 1991; Poggio & Edelman, 1990; Koenderink & Van Doorn, 1979). Many studies indicate that the stimulus selectivity of temporal cortex neurons is not fixed but is acquired via learning throughout life (Ashbridge, Perrett, Oram, & Jellema, 2000; Booth & Rolls, 1998; Kobatake, Wang, & Tanaka, 1998; Sakai & Miyashita, 1991). The linking or association of representations of the different consecutive momentary views as the body or head moves can result in a distinct representation for that particular behavioral event. The cells described in this report are likely to be constituents of such representations.

The neural representations for sequences of events may play a role in predicting or anticipating the next move or posture of the animate object. For example, the sight of a body that has just stopped walking forward may invoke an expectation that, should walking commence again, it is likely to resume forward direction. The same view of a static body that has just stopped walking backward, by contrast, may be expected to move in a backward direction should walking resume. A mechanism for ''predicting'' the most likely next stage of an observed action would also be useful for keeping ''track'' of actions that are temporarily obscured from view (Assad & Maunsell, 1995). Recent studies show that neural responses in STSa (Baker et al., 2001) and prefrontal cortex (Umilta et al., 2001) to actions persist when the action becomes occluded from view. It should be noted that information about the most likely next motion can also be derived from a still image of a person performing a motor act. Such implied motion can activate the posterior motion sensitive areas in humans (Kourtzi & Kanwisher, 2000). Experience of a full perceptual history prior to viewing a static body, however, allows greater certainty in predicting the likelihood and nature of the body's future movement.

The question whether the action prediction itself is also computed within STS remains open. Our data strongly suggest that the building blocks for the representations of complex action–static sequences are represented within STSa, but the prediction might be computed elsewhere. The relatively long onset latency (~500 msec) of some cells sensitive to perceptual history (see, e.g., the cells in Figures 3 and 5) allows ample time for top–down input to exert its effect, yet, other cells responded to the static phase without an obvious delay (see, e.g., the cell in Figure 4). Due to the variations in the onset latencies of the cell responses to the static phase, a more accurate analysis at the population level of the onset latencies could not be made.

In conclusion, it is widely accepted that temporal lobe cells are selective in their response among visual stimuli: Only some patterns evoke a response, other patterns do not. We find that for a population of STSa cells there is an additional (second order) selectivity for the sequence of two stimuli. Classically, cells respond to a particular static stimulus X and not to other static stimuli (Y, Z). For the cells that we described, a response to X occurs only when it follows movement A, and not movements B or C. The visual tuning to postures and preceding actions that we reported here contributes to our understanding of the cell's sensitivity to the visual characteristics of bodies and faces. We showed that subpopulations of STSa cells are capable of coding sequences of events that can help the understanding of natural body actions. Such natural actions are not isolated postures, but they are continuous and complex sequences of postures with linking movements.

METHODS

Subjects and Physiological Procedures

The experiments were performed on two awake rhesus macaque monkeys (*Macaca mulatta*, age 4–6 years). A detailed description of the surgical and recording procedures can be found elsewhere (Oram & Perrett, 1996). Animal care and experimental procedures were performed in accordance with UK Home Office guidelines.

Recording

Spikes were captured online onto a PC (CED1401plus, Cambridge Electronic Design, UK). Additionally, spikes were stored on an audio track of a HiFi videotape recorder. The stimulus events (seen from the subject's perspective) were recorded with a video camera and stored simultaneously on the video track of the same tape. Eye movements were recorded with a second (infrared sensitive) video camera. The signals from the two cameras were integrated (Panasonic VHS video mixer, WJAVE7) prior to recording. The signal from the eye movement camera was also recorded separately on a second video tape recorder, synchronized with a time-code generator and frame counter (VITC Horita VG50), for off-line analysis of eye position (Iview, Sensomotoric Instruments).

Stimuli

Stimuli consisted of a wide range of bodily actions followed by static body postures, which formed the

natural, or most likely, ''end-point'' of the action. As soon as a response was obtained to a static posture following an action, three experimental conditions were employed: (1) Different actions also naturally leading to that posture were tested. (2) The static posture was presented without any preceding action (following opening of the shutter). (3) The action was presented directly followed by continuation of the action, or by a static image in which the agent was occluded from view.

Stimuli were presented either on film projected onto a screen at life size, or live from behind a fast rise-time liquid crystal shutter (aperture 20 × 20 cm at a distance of 15 cm). Video stimuli consisted of either monkeys or people; most cells do not discriminate between these two species (Perrett, Smith, Potter, et al., 1985; Desimone et al., 1984). The shutter introduced some increase in luminance when going from closed to open state, but this was not observed to affect the response selectivity of the cells. STSa cells tolerate large changes in luminance (Gross et al., 1972). In some cases, a mechanical shutter with a larger aperture was used to avoid narrowing the scope of view of the subject. Retinal images of live presented bodies varied from about 67° × 23° (vertically × horizontally) at 1.5 m distance, to 28° × 9° at 4 m distance. Five to twelve repetitions were tested per stimulus condition in pseudorandom order. Control stimuli included objects of comparable size moved in comparable ways (e.g., at the same velocity).

Analysis

Off-line spike sorting (template matching) was routinely performed (Spike2, Cambridge Electronic Design, UK) and established whether or not spikes recorded during both motion and static phases were, indeed, generated by one and the same cell, instead of two cells (one for each phase). Spike counts were performed during presentation of the ''static'' stimuli, starting 100 msec after the shutter became transparent (Oram & Perrett, 1996), or 300 msec after the cessation of the preceding motion phase (to allow for an ''overshoot'' into the static phase). Responses were regarded as ceased when the response rate declined to spontaneous level. Cell responses were analyzed using ANOVAs and Newman–Keuls post hoc testing (significance level at $p < .05$). The modulation of response by the perceptual history was expressed as [(Max − Min)/Max] × 100, where Max and Min are the mean response rates during the static phase following two different perceptual histories.

To perform the eye position analysis, the subject's visual field was subdivided in a matrix of horizontal and vertical sectors, with a maximal visual angle in both planes of 100°, and sector column and row width of 20° (see Figures 3 and 4). The time (number of frames) during which eye position fell in each visual field sector of the matrix was assessed for the static phases of each trial, expressed as a percentage of the total duration, and entered into the two-way ANOVA (with perceptual history and visual field sector as factors).

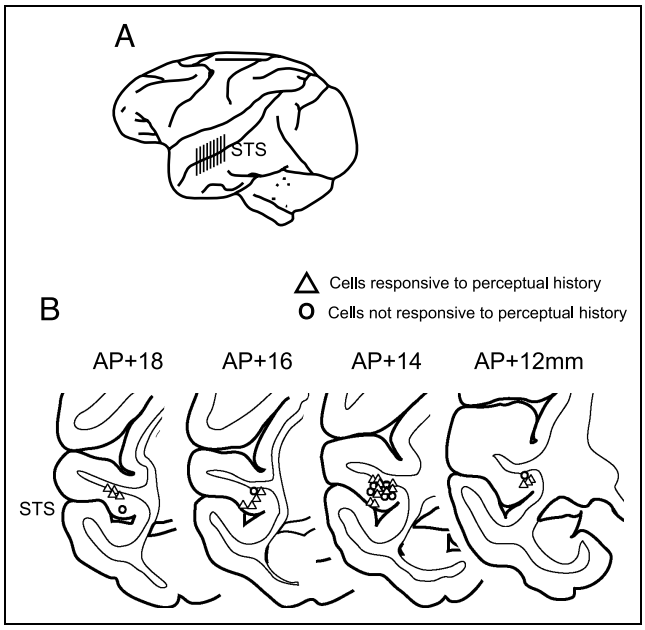

Figure 6. Histological reconstruction of cell locations. (A) Left side view of the macaque brain. Cells were recorded in the banks of the STSa, between 11 and 19 mm anterior to the interaural plane (indicated by vertical bars). (B) Reconstruction of coronal sections of the left hemisphere taken at 18, 16, 14, and 12 mm anterior to the interaural plane. Each section represents a 2-mm-thick slice. The cells are marked according to whether their responses were influenced by the perceptual history (triangles) or not (circles). All cells for which reconstructions were made were located in the upper bank of STSa. Thick line, cortical surface; thin lines, edge of gray matter.

Cell Localization

A detailed description can be found elsewhere (Jellema, Baker, Wicker, & Perrett, 2000). At completion of each experiment, frontal and lateral X-ray photographs were taken with the electrode still in place to locate the electrode and the recorded cells with respect to specific bone landmarks. During the final experiment, electrolytic microlesions were produced at the site of recording. The subject was then sedated and given a lethal dose of anaesthetic. After transcardial perfusion, the brain was removed, and coronal sections (25 μm) were cut, photographed, and stained. The x-ray photographs were aligned with the histological sections to determine the cell locations (accuracy ≈ 1 mm). All cells included in this report were located within the banks of STSa, between 10 and 19 mm anterior to the interaural plane, corresponding to area STPa (Bruce, Desimone, & Gross, 1981) (Figure 6). A histological reconstruction of the locations of cells sensitive to perceptual history and cells not sensitive to perceptual history is shown in Figure 6B for one monkey.

Acknowledgments

This work was supported by grants from the Human Frontier Science Program and the Wellcome Trust. We thank Bruno Wicker for his contribution to some of the experiments, Mary Latimer for her assistance with histology, Robin Edwards for the cell location software, and Mike Oram and Eric Bowman for the helpful comments on the manuscript.

Reprint requests should be sent to Tjeerd Jellema, Department of Experimental Psychology, Helmholtz Research Institute, Utrecht University, Heidelberglaan 2, 3584 CS Utrecht, The Netherlands, or via e-mail: T.Jellema@fss.uu.nl.

REFERENCES

Allison, T., Puce, A., & McCarthy, G. (2000). Social perception from visual cues: Role of the STS region. *Trends in Cognitive Sciences,* 267–278.

Ashbridge, E., Perrett, D. I., Oram, M. W., & Jellema, T. (2000). Effect of image orientation and size on object recognition: Responses of single units in the macaque monkey temporal cortex. *Cognitive Neuropsychology, 17,* 13–34.

Assad, J. A., & Maunsell, J. H. R. (1995). Neuronal correlates of inferred motion in primate posterior parietal cortex. *Nature, 373,* 518–521.

Baker, C. I., Keysers, C., Jellema, T., Wicker, B., & Perrett, D. I. (2001). Neuronal representation of disappearing and hidden objects in temporal cortex of the macaque. *Experimental Brain Research, 140,* 375–381.

Bonda, E., Petrides, M., Ostry, D., & Evans, A. (1996). Specific involvement of human parietal systems and the amygdala in the perception of biological motion. *Journal of Neuroscience, 16,* 3737–3744.

Booth, M., & Rolls, E. T. (1998). View-invariant representations of familiar objects by neurons in the inferior temporal cortex. *Cerebral Cortex, 8,* 510–523.

Bruce, C., Desimone, R., & Gross, C. G. (1981). Visual properties of neurons in a polysensory area in superior temporal sulcus of the macaque. *Journal of Neurophysiology, 46,* 369–384.

Decety, J., Grezes, J., Costes, N., Perani, D., Jeannerod, M., Procyk, E., Grassi, F., & Fazio, F. (1997). Brain activity during observation of actions: Influence of action content and subject's strategy. *Brain, 120,* 1763–1777.

Desimone, R., Albright, T. D., Gross, C. G., & Bruce, C. (1984). Stimulus-selective properties of inferior temporal neurons in the macaque. *Journal of Neuroscience, 4,* 2051–2062.

Dolan, R. J., Fink, G. R., Rolls, E. T., 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.

Földiák, P. (1991). Learning invariance from transformation sequences. *Neural Computations, 3,* 194–200.

Goodale, M. A., & Milner, A. D. (1992). Separate visual pathways for perception and action. *Trends in Neurosciences, 15,* 20–25.

Grafton, S. T., Arbib, M. A., Fadiga, L., & Rizzolatti, G. (1996). Localization of grasp representations in humans by positron emission tomography: 2. Observation compared with imagination. *Experimental Brain Research, 112,* 103–111.

Gross, C. G., Rocha-Miranda, C. E., & Bender, D. B. (1972). Visual properties of neurons in inferotemporal cortex of the macaque. *Journal of Neurophysiology, 35,* 96–111.

Hoffman, E. A., & Haxby, J. V. (2000). Distinct representations of eye gaze and identity in the distributed human neural system for face perception. *Nature Neuroscience, 3,* 80–84.

Jellema, T., Baker, C. I., Oram, M. W., & Perrett, D. I. (2002). Cell populations in the banks of the superior temporal sulcus of the macaque and imitation. In A. N. Meltzoff & W. Prinz (Eds.), *The imitative mind* (pp. 267–290). Cambridge, UK: Cambridge University Press.

Jellema, T., Baker, C. I., Wicker, B., & Perrett, D. I. (2000). Neural representation for the perception of the intentionality of actions. *Brain and Cognition, 44,* 280–302.

Jellema, T., & Perrett, D. I. (2001). Perceptual history and neural responses to body actions and postures. *Journal of Cognitive Neuroscience, 102.*

Jellema, T., & Perrett, D. I. (2002). Coding of visible and hidden objects. In W. Prinz & B. Hommel (Eds.), *Common mechanisms in perception and action. Attention and performance,* vol. XIX (pp. 356–380). Oxford, UK: Oxford University Press.

Kendrick, K. M., & Baldwin, B. A. (1987). Cells in temporal cortex of conscious sheep can respond preferentially to the sight of faces. *Science, 236,* 448–450.

Kobatake, E., & Tanaka, K. (1994). Neuronal selectivities to complex object features in the ventral visual pathway of the macaque cerebral cortex. *Journal of Neurophysiology, 71,* 856–867.

Kobatake, E., Wang, G., & Tanaka, K. (1998). Effects of shape-discrimination training on the selectivity of inferotemporal cells in adult monkeys. *Journal of Neurophysiology, 80,* 324–330.

Koenderink, J. J., & Van Doorn, A. J. (1979). The internal representation of solid shape with respect to vision. *Biological Cybernetics, 32,* 211–216.

Kourtzi, Z., & Kanwisher, N. (2000). Activation in human MT/MST by static images with implied motion. *Journal of Cognitive Neuroscience, 12,* 48–55.

Logothetis, N. K., Pauls, J., & Poggio, T. (1995). Shape representation in the inferior temporal cortex of monkeys. *Current Biology, 5,* 552–563.

Miller, E. K., Li, L., & Desimone, R. (1993). Activity of neurons in anterior inferior temporal cortex during a short-term memory task. *Journal of Neuroscience, 13,* 1460–1478.

Mistlin, A. J., & Perrett, D. I. (1990). Visual and somatosensory processing in the macaque temporal cortex: The role of 'expectation'. *Experimental Brain Research, 82,* 437–450.

Miyashita, Y. (1988). Neuronal correlate of visual associative long-term memory in the primate temporal cortex. *Nature, 335,* 817–820.

Naya, Y., Sakai, K., & Miyashita, Y. (1996). Activity of primate inferotemporal neurons related to a sought target in pair-association task. *Proceedings of the National Academy of Sciences, U.S.A., 93,* 2664–2669.

Oram, M. W., & Perrett, D. I. (1992). Time course of neural responses discriminating different views of the face and head. *Journal of Neurophysiology, 68,* 70–84.

Oram, M. W., & Perrett, D. I. (1996). Integration of form and motion in the anterior superior temporal polysensory area (STPa) of the macaque monkey. *Journal of Neurophysiology, 76,* 109–1297.

Perrett, D. I., Harries, M. H., Bevan, R., Thomas, S., Benson, P. J., Mistlin, A. J., Chitty, A. J., Hietanen, J. K., & Ortega, J. E. (1989). Frameworks of analysis for the neural representation of animate objects and actions. *Journal of Experimental Biology, 146,* 87–113.

Perrett, D. I., Hietanen, J. K., Oram, M. W., & Benson, P. J. (1992). Organization and functions of cells responsive to faces in the temporal cortex. *Philosophical Transactions of the Royal Society of London B, 335,* 23–30.

Perrett, D. I., Mistlin, A. J., & Chitty, A. J. (1987). Visual cells responsive to faces. *Trends in Neurosciences, 10,* 358–364.

Perrett, D. I., & Oram, M. W. (1993). Neurophysiology of shape processing. *Image and Vision Computing, 11,* 317–333.

Perrett, D. I., Smith, P. A. J., Mistlin, A. J., Chitty, A. J., Head, A. S., Potter, D. D., Broennimann, R., Milner, A. D., & Jeeves, M. A. (1985). Visual analysis of body movements by neurons in the temporal cortex of the macaque monkey: A preliminary report. *Behavioural Brain Research, 16,* 153–170.

Perrett, D. I., Smith, P. A. J., Potter, D. D., Mistlin, A. J., Head, A. S., Milner, A. D., & Jeeves, M. A. (1985). Visual cells in the temporal cortex sensitive to face view and gaze direction. *Proceedings of the Royal Society of London B, 223,* 293–317.

Poggio, T., & Edelman, S. (1990). A network that learns to recognize 3D objects. *Nature, 343,* 263–266.

Puce, A., Allison, T., Bentin, S., Gore, J. C., & McCarthy, G. (1998). Temporal cortex activation in humans viewing eye and mouth movements. *Joural of Neuroscience, 18,* 2188–2199.

Rizzolatti, G., Fadiga, L., Matelli, M., Bettinardi, V., Paulesu, E., Perani, D., & Fazio, F. (1996). Localization of grasp representations in human by PET: 1. Observation versus execution. *Experimental Brain Research, 111,* 246–252.

Rolls, E. T. (1992). Neurophysiological mechanisms underlying face processing within and beyond the temporal cortical areas. *Philosophical Transactions of the Royal Society of London B, 335,* 11–21.

Sakai, K., & Miyashita, Y. (1991). Neuronal organization for the long-term memory of paired associates. *Nature, 354,* 152–155.

Stone, J. V. (1998). Object recognition using spatio-temporal signatures. *Vision Research, 38,* 947–951.

Tanji, J., & Shima, K. (1994). Role for supplementary motor area cells in planning several movements ahead. *Nature, 371,* 413–416.

Tinbergen, N. (1951). *The study of instinct.* Oxford, UK: Oxford University Press.

Umilta, M. A., Kohler, E., Gallese, V., Fogassi, L., Fadiga, L., Keysers, C., & Rizzolatti, G. (2001). I know what you are doing: A neurophysiological study. *Neuron, 31,* 155–165.

Wallis, G., & Bülthoff, H. H. (2001). Effects of temporal association on recognition memory. *Proceedings of the National Academy of Sciences, U.S.A., 98,* 4800–4804.