T.N. DAVIES and D.D. HOFFMAN

FACIAL ATTENTION AND SPACETIME FRAGMENTS ⋆

ABSTRACT. Inverting a face impairs perception of its features and recognition of its identity. Whether faces are special in this regard is a current topic of research and debate. Kanizsa studied the role of facial features and environmental context in perceiving the emotion and identity of upright and inverted faces. He found that observers are biased to interpret faces in a retinal coordinate frame, and that this bias is readily overruled by increased realism of facial features, but not easily overruled by environmental context. An additional factor contributing to a retinal coordinate-frame interpretation may be the ambiguous nature of the face stimuli. Since his facial expressions are interpretable both upright and inverted, they may in both orientations activate an endogenous attentional process for faces. We present visual search and change-blindness experiments that explore how inversion, negation, and facial emotion affect visual attention to static faces. We find that attention to faces is impaired by inversion and negation. We also find that the parts of the face that receive greater attention can be influenced by the emotional expression of the face. We propose to extend these experiments to dynamic faces. To this end, we develop a theory of the visual representation of dynamic faces, in which faces are represented by classes of 'spacetime fragments'-moving regions of the face with high informational content. We then present ideas for future experiments which are motivated by the spacetime fragment theory, and which should serve to constrain its further development.

1. KANIZSA'S STUDY OF FACES

Among the many topics to which Kanizsa applied his talents as an artist and experimental psychologist was the perception of faces (Kanizsa and Tampieri 1976). He was interested in whether the perceptual orientation of visual objects depends on a retinal or an environmental frame of reference. Some research had suggested the importance of the environmental frame (James 1890; Miyakawa 1950), some the importance of the retinal frame (von Helmholtz 1867; Thouless 1947), and some the importance of viewer expectations (Rock 1956, 1973).

Kanizsa noted that these differing results were obtained with different kinds of visual objects, and proposed that study of the specific properties of these visual objects could explain the different results. To this end he

⋆ Paper contributed to the Mitteleuropa Foundation's conference: "The Legacy of Kanizsa in Cognitive Science", June 10–12, 2002, Bolzano, Italy.

Axiomathes **13:** 303–327, 2003.

created a variety of schematic faces, carefully varying their visual properties, and studied the effects of these properties on the perceived orientation of the faces. His faces were bistable, in that they could be seen as one face when shown upright and another when inverted.

He found that subjects are strongly biased to interpret a face in a retinal frame of reference, so that the top of the face is oriented towards the top of the retinal frame. Adding a potent environmental clue, such as attaching an inverted body to the face, did little to influence this bias. However, he found that the retinal bias could be overcome if he drew individual parts of the face unambiguously. A clearly-drawn inverted mouth, for instance, could force subjects to see a face as inverted in the retinal frame.

These results have an interesting interpretation in terms of visual attention to faces. Kanizsa's finding of a retinal bias suggests that observers have a default attentional strategy for searching faces, a strategy which assumes that the face is oriented upright with respect to the retinal frame of reference. In this attentional strategy, if a visual feature is identified as, say, an eye, then to find the nose or mouth, attention is directed downward in the retinal frame of reference. If, instead, a mouth is found first, then attention is directed upward in the retinal frame to find an eye or nose. However this default attentional strategy can be overcome, at least in part, if the details of a facial feature clearly indicate that the strategy is incorrect. In the default strategy, if one first finds a nose, then one looks for the eyes above the nose in the retinal frame. If, however, the nose one finds is clearly inverted in the retinal frame, then the default strategy is aborted, and eyes are sought for below the nose in the retinal frame.

Given these interpretations of Kanizsa's results, one might expect that attention to a face is more efficient if the face is upright in the retinal frame, and less efficient if the face is inverted. The reason is that the default attentional strategy assumes that the face is upright in the retinal frame. If this assumption is false, it will take time for the system to discover this fact and to switch attentional strategies. Moreover, the reason for the default strategy is that it is often correct: faces usually do appear upright in the retinal frame. Because of this empirical bias in visual experience, the visual system might not have sufficient experience with inverted faces to develop efficient attentional strategies for processing them.

It is these predictions from Kanizsa's work that we explore in the experiments discussed here. We use change-detection tasks to study attention to upright and inverted faces, and also to faces shown in photographic negative. We find that attention is more efficient for upright faces than for inverted faces, and for faces shown in photographic positive than photographic negative. Our stimuli, like Kanizsa's, are static images of faces.

It is, of course, of great interest to see if our attentional results extend to the case of moving faces, since these are the ecologically more natural stimuli. To this end, we propose a 'spacetime fragments' theory for the representation of moving faces and objects, and use this theory to motivate proposals for future experiments that explore how we perceive and attend to dynamic faces.

2. ATTENTION TO FACES

Humans have special skills developed for the perception of faces. These skills can be evidenced even in infancy. By the early age of 30 minutes infants will seek out and follow a moving face stimulus over another moving object of equal visual complexity (Johnson et al. 1991). Moreover, if the cortical areas specialized for faces are damaged in infancy, then other brain areas are not capable of taking over the task of face perception and recognition, suggesting that cortical face processing areas are specialized even at birth (Farah et al. 2001). In particular, bilateral or right unilateral damage to the lateral temporal brain areas can result in a disorder known as prosopagnosia. Prosopagnosia selectively impairs an individual's ability to recognize face identity while often leaving other forms of object perception, and even perception of emotional face expressions, intact.

Although there is a general right hemisphere specialization for faces, specific cortical regions and their purported areas of expertise for faces include the superior temporal sulcus involved in emotion perception, the lateral fusiform gyrus involved in identity perception, and the fusiform face area involved in face detection (Haxby et al. 2000; Tong et al. 2000). In addition to neurophysiological evidence from fMRI and PET studies, there is also evidence of specialization for faces from studies of EEG. For human observers, the initial processing of faces triggers an obligatory event-related potential (ERP), the N170, at the lateral posterior temporal electrodes, and the processing of identity triggers an N400 and P600 (Bentin et al. 1996; Bentin and Deouell 2000; Cauquil et al. 2000; Eimer 1998, 2000a; Eimer and McCarthy 1999; George et al. 1996; Jemel et al. 1999). Interestingly, the N170 is evoked by face stimuli even if they are not attended, but it is enhanced by centrally focused attention (Eimer 2000).

Although this suggests that attention is not required by human vision to perceive faces, it does not rule out the possibility that attention is required to build a detailed description of faces. One study by Gosselin and Schyns (2001) incorporates a technique known as 'bubbles' to show how human vision uses attentional strategies to perceive faces. In the bubbles method, observers are shown a set of individual face images and asked to

judge the faces by a certain property (e.g., gender, emotion). The faces are completely occluded by a gray field except for small portions seen through a gaussian-shaped window (the bubbles) which varies randomly from trial to trial in size and location.

In analyzing which faces, with which random placement and size of bubbles, resulted in the highest accuracy rates, Gosselin and Schyns were able to deduce which areas of the face received the most attention. They found that attention varies based on the size of the window and the nature of the categorization task. This suggests that there may be top-down, or endogenous, influences for the allocation of attention to faces that are spatial-frequency specific. The bubbles technique links attention and looking because it allows the eye to look only at restricted portions of the image in its assessment of attention.

Studies of eye-movements also suggest that attention to faces depends on the nature of the information extracted. Recent studies show that observers look more at the upper half of the face when making judgments about speech intonation in comparison to judgments about speech segments (Lansing and McConkie 1999), and that more fixations are made to the left visual field (right side of the face) when making judgments of emotion (Borod et al. 1988). There are even differences among individual emotions, with happy faces receiving more fixations to the corners of the eyes in comparison to other emotions (Williams et al. 2001). Eye movements to unfamiliar faces are more prevalent in the left visual field, and are overall more systematic and constrained in movement in comparison to famous faces (Althoff and Cohen 1998). Observers fixate the eyes more often than the mouth, and this difference is more pronounced for familiar than for unfamiliar faces.

3. CHANGE DETECTION AND CHANGE BLINDNESS

Because an observer's direction of gaze can be strongly influenced by where the observer allocates attention (Deubel and Schneider 1996), studies of bubbles and eye movements provide an indirect means of studying attention. However, there are cases in which where one looks is not where one attends. This has been confirmed by studies of covert attention incorporating techniques such as change blindness and inattentional blindness (Klein et al. 1992; Posner 1980; Posner et al. 1980). In one study, observers viewed images of outdoor scenes and were instructed to press a button each time they saw something change. Each time the observer blinked a change was made, although observers were not told this. Interestingly, observers failed to detect 40% of the changes made to locations they fixated just

before and after a blink, suggesting that looking does not entail seeing or attending (O'Regan et al. 2000). In a similar study, observers were instructed to copy blocks displayed on a computer screen. Even when observers looked directly at the blocks, they often failed to notice any changes (Ballard et al. 1995; Hayhoe et al. 1998).

In order to more directly measure the way human vision allocates attention to faces, it is useful to complement techniques such as bubbles and measurement of eye movements, with techniques such as the flicker paradigm from the field of change blindness (Rensink et al. 1995, 1997). In a flicker task, observers see a brief presentation of one image, then a brief blank screen, and then presentation of the original image, with or without a change. Observers are instructed to press a button if they detect any changes. The types of changes that can be made are addition, deletion, and change in location or other property of an object in the scene. The three-image sequence continues to cycle until the observer decides if a change has or has not been made.

Surprisingly, the task is very difficult. Although we feel like we simultaneously see everything in our field of vision, we can in fact only attend to and store a limited number of items in our visual short-term memory. Since the blank screen blocks human vision from detecting low-level motion signals which would reveal the location of the change, observers must attend and build object descriptions one-by-one, store these in visual short-term memory, and then compare them with descriptions built in the second image (Rensink 2000a, b, c). Therefore, changes made to items which human vision preferentially attends to have higher likelihoods of being successfully detected. Possible factors that influence the allocation of attention are the low-level visual salience of objects (exogenous capture) or the task goals and interest of the observer (endogenous control) (O'Regan et al. 2000; Shore and Klein 2000).

Although change blindness has traditionally been used to study attention to outdoor and indoor scenes, recently its application has been expanding. Rensink (2000b) used it to study detection of change in arrays of rectangles, Williams and Simons (2000) used it to study detection of changes to a single multi-part object, and Simons and Levins (1998) used it to study attention to people in a real-world interaction. To help better understand the attention processes that human vision uses in the perception of faces, we applied the flicker task of change blindness to faces, where the face serves as the entire scene and changes can be made to the individual face features and relationships between features.

4. THE PERCEPTION OF INVERTED FACES

Previously we discussed different ways in which human vision is specialized for perceiving faces. However, evidence suggests that this specialization comes with certain limitations. In particular, there is a well known *face-inversion effect*, which was first discovered by Yin (1969). In his original study, Yin conducted a test of recognition memory for faces and other stimuli usually seen upright. He found that when all the stimuli were presented upright, faces were recognized most accurately. However, when all the stimuli were turned upside down, faces had the lowest recognition rate, showing that recognition memory for faces is disproportionately impaired by inversion in comparison to other objects.

The face-inversion effect has also been found in studies of perceptual matching tasks, where observers simultaneously compare two inverted faces presented side-by-side (Searcy and Bartlett 1996; Valentine 1988). This suggests that inverting a face does not just interfere with memory retrieval, but also with perceptual encoding of the face.

There is also evidence that inversion has differential effects of impairment based upon the type of face information under investigation. A study by Searcy and Bartlett (1996) shows that perceptual judgments related to local feature information are impacted much less by inversion than perceptual judgments related to relationships between features, or configural information. Diamond and Carey (1996) further breakdown the category of configural information by dividing it into first order and second order relations. First order relations refer to the general spatial arrangement of the face (such as eyes above nose, nose above mouth), while second order relations refer to the spatial locations of parts relative to the prototypical arrangement of parts (Bill's eyes are more wide-set than Tracey's). In their study, they found that the second order relational information is impacted more by inversion than first order.

In light of this evidence, it is natural to ask if, in addition to recognition memory and perceptual encoding, face inversion also impairs the normal strategies for attending to facial features and the relationships between features. Our experiments studied both of these questions.

5. THE PERCEPTION OF FACES IN PHOTOGRAPHIC NEGATIVE

In addition to the face-inversion effect, there is another transformation which impairs human perception of faces: contrast negation. Contrast negation involves reversing the brightness levels of an image so that the face image looks like it is in photographic negative. Early studies showed that

negating a face impairs perception of emotional expression (Galper 1970; Galper and Hochberg 1971), but more recent studies show that negation also affects perceptual processing of faces more generally. This is evidenced in tasks of face-matching (Lewis and Johnston 1997) and studies of brain-imaging (George et al. 1996).

Reversing the brightness levels can affect a face image in several ways. First, it changes the apparent direction of lighting. Although light sources are typically seen as being overhead (e.g., Ramachandran 1988), negation of a positive face changes the apparent lighting direction so that the face has the appearance of being bottom-lit (Liu et al. 1999). Additionally, negation impairs shading information. This triggers our visual system to construct 3D representations of structures that are physically impossible. Lastly, since negation leaves lines and edges intact, the impairment may result from a disruption to information in low spatial frequencies (Kemp et al. 1996). Since configural information is found in low spatial frequencies, this may indicate a greater disruption to perception of configural rather than local information of faces.

Although both negation and inversion impair face perception, studies suggest that they have independent causes (Kemp et al. 1990; Bruce and Langton 1994; Lewis and Johnston 1997). For this reason, we investigated how human vision attends to faces in contrast negation, and if the results differ based upon attention to local or configural information.

6. CHANGE BLINDNESS EXPERIMENTS

To answer these questions, we conducted two main experiments (Davies and Hoffman 2002). Experiment 1a investigated how human vision attends to configural features in upright and inverted face images, and experiment 1b investigated attention to configural features in positive and negative contrast face images. Experiment 2a investigated how human vision attends to local face features in upright and inverted images, and experiment 2b investigated attention to local features in positive and negative contrast images.

To study attention we used the flicker task described above, and the two dependent variables measured were time to detect changes, and accuracy rate. Changes were made to all areas of the face, but the two face areas under investigation for this study were the eyes and mouth. Configural changes were made by moving face features up or down by 10 pixels. Local changes were made by rotating features in place by 180 degrees. All face expressions were neutral, which minimized any configural change occurring as a result of the local manipulation.

For experiment 1a we found that attention to configural features was more efficient for upright than inverted faces. Specifically, observers were both faster and more accurate when faces were seen in an upright orientation. This reveals a new aspect of the face-inversion effect. If attention is necessary, or at least beneficial, in building descriptions of faces, then perhaps an attentional impairment also contributes to the perceptual impairments found when faces are inverted. Moreover, this result also shows that human vision uses endogenous mechanisms of attention for faces. Turning a face upside down eliminates the meaning, or any previously learned strategy, since features are no longer where we expect them to be and, as a result of this, attentional efficiency is lost. This is an interesting result not only for the field of face perception but also for the field of change blindness, since previous results suggested that the flicker task cannot be used to study endogenous attention (Shore and Klein 2000).

For experiment 1b we found that attention to configural face features was more efficient for positive than negative contrast images, with observers having higher accuracy rates for positive face images. Although there was no main effect of detection times, a between-subjects ANOVA on detection times from experiments 1a and 1b reveal a significant difference between experiments for performance on the upright/positive contrast images, even though the exact same upright/positive images were used in both experiments. It is possible that because negated images depict impossible shading patterns, or because they promote a low-level reliance on color as a cue, this altered attentional strategies for both positive and negative contrast images. This suggests that although visual attentional strategies for faces are highly specialized, perhaps these strategies can be altered or impaired by visual experience.

For experiment 2a we found that attention to local face features is faster and more accurate when face images are upright rather than inverted. Moreover, the accuracy levels were higher for the eye than for the mouth. This difference is not readily explained by low-level image differences since the mean number of pixels involved in the changes, and the mean squared difference in images for the changes, were both less for the eye than for the mouth. In change-detection tasks those elements of a scene that receive the most attention have the highest rate of change detection (Rensink 1997), suggesting that in a change-detection task human vision attends more to the eye than to the mouth. This comports well with Althoff and Cohen's (1999) findings that observers spend more time looking at the eyes when judging the fame or emotion of a face.

For experiment 2b we found that attention to local face features is more accurate and faster when face images are in positive rather than

negative contrast. It is interesting to note that, although stimuli for the uninverted/positive contrast images were exactly the same, there was no advantage of the eye over the mouth as in experiment 2a. This again could be related to the difference cited above for negated images in experiment 1b: negation could alter attentional strategies for both positive and negative contrast images.

There has been much discussion concerning the similarities and differences between the two transformations of inversion and negation. Perhaps one difference is the attentional strategies human vision employs. For an inverted image, individual features, although in the wrong orientation, may still be recognized as characteristic of a face and thus prompt our normal face attention mechanisms. This would result in a slower (since features may require mental rotation) but still similar attention pattern as used for the viewing of upright faces. However, for negated images, we may employ normal attentional strategies but because the 3D representations are skewed and illogical, features may not be recognized as part of the face 'class'. This may prompt human vision to adopt a new mechanism or pattern of attention that is more efficient for extracting information from negated images (e.g., one that relies on color cues). In summary, inversion may slow down special face mechanisms, and negation may not engage special face mechanisms. Focused investigation of these possible differences is necessary and may reveal interesting information about how human vision deals with inconsistencies among face representations. This may also provide valuable assistance in building computer network models whose job is to recognize faces under a variety of conditions.

Because it is a well-known result that inversion and negation impair perceptual processing of faces, we conducted a series of control experiments to ensure that our data reflect attentional, and not simply perceptual, impairments. Our control experiments were identical to our original experiments, except that before the presentation of each initial face image, a cue word was presented to the observer (e.g., EYE, CHIN, MOUTH) which directed their attention to a part of the face. If there was change at all, it occurred at the location indicated by the cue word; but changes only occurred on about half of the trials, so observers still had a nontrivial detection task to perform. We reasoned that if perceptual difficulties were responsible for the observed impairments in negated or inverted faces, than an observer's performance should not be improved by the use of attentional cues. However, if attentional impairments did contribute to the results, then performance should improve once the role of attention was eliminated, or at least significantly reduced.

For all experiments we found that cueing resulted in a significant improvement in both accuracy and timing, suggesting that inversion and negation affect not only perception, but also attention to faces. Moreover, we found that the addition of cues significantly improved performance more for the mouth than for the eye across all experiment versions. This comports well with data from experiment 2a that the eye received more attention than the mouth. When attention was directed to the mouth, it was perceptually easier for observers' to see the change than for the eye, but this was not reflected in the original data since the eye was the recipient of more focused attention.

The notion of endogenous attentional mechanisms for face perception has some interesting ramifications concerning the work done by Kanizsa (1976). In his study he found that observers were more likely to detect the emotional expression in an image corresponding to the retinal frame of reference, even when this required that the head appeared to be connected to the body by the forehead and not the chin. This is a surprising result since the image of a forehead connected to a body does not make ecological sense. However, given that there are endogenous processes for looking at faces, because the stimuli that Kanizsa used displayed valid emotions whether they were upright or inverted, attention may have been automatically allocated to the most easily extracted expression. In this case, the local facial features may have automatically been processed in the retinal frame of reference and engaged attention mechanisms, thus influencing observer's perceptions, and even overriding the external cue of the attached body.

This poses some interesting questions concerning the nature of attention to faces, its influence on perception, and the importance of surrounding context. From the above experiments, we know that inversion and negation can interfere with normal attention mechanisms. But are there any external factors or attributes that can reverse or interfere with mechanisms of face attention? In addition to using the whole body as a cue, future experiments may investigate use of head hair, facial hair, skin texture (e.g., the direction of wrinkles), and dynamic features, such as face movement. Below we explore current theories of attention and how adding a dynamic element is an informative and natural way to learn more about the attention mechanisms human vision employs for faces.

7. RECENT THEORIES OF ATTENTION

How does human visual attention navigate about an image of a scene? Perhaps attention acts like a 'spotlight', moving across the scene and selecting

any visual features falling within the confines of a spatial region. According to this proposal, empty space, primitive visual features, and meaningful objects or their parts can all fall within the focus of attention: attention is deployed only to spatial regions. Studies show that the speed of detecting targets outside a cued area falls off monotonically as the distance between the target and cue is increased (Posner et al. 1980; Downing and Pinker 1985). This suggests the idea of a spatial gradient of attention, with areas on the perimeter of the spotlight receiving less attention than the center.

However, some results from recent studies of attention are not easily explained by the spotlight theory. In particular, the technique of 'selective looking' reveals that when two scenes are superimposed on each other, observers attending to one scene can remain unaware of events occurring in the other scene (Neisser 1967, 1979; Neisser and Becklen 1975; Simons and Chabris 1999). This suggests that attention is not always spatially governed since a spatial focus in the attended scene would also encompass events in the unattended scene.

Studies of multiple object tracking (MOT) reveal that subjects are at least 85% accurate in tracking and identifying up to five randomly and independently moving targets among a field of identical distractors (Pylyshyn and Storm 1988). A single-spotlight theory would require that the focus of attention monitors each moving object by visiting them cyclically, but such a mechanism is too slow to account for subjects' performance (Pylyshyn and Storm 1988). This suggests that attention is not just a single spotlight, but that it can be divided, and it can be divided equally.

Moreover, the division of attention can be governed by and applied to the tracking of individual objects (Intriligator 1997; Sears and Pylyshyn 2000). More evidence that attention can be directed towards discrete visual objects comes from studies of a 'same-object advantage'. In divided-attention tasks, subjects are more accurate in judging two properties of a single object than judging two properties of two different objects (Duncan 1984; Watt 1988). If subjects must shift their attention by a certain distance, they complete the shift more quickly if they need not cross an object boundary during the shift (Egly et al. 1994).

Attention can be directed not just to a single object, but also to a group of objects or to part of an object. The same-object advantage still holds if the 'object' is a collection of objects that form a perceptual group (Egly et al. 1994). This suggests that some parsing of the visual scene into units occurs preattentively (Davis and Driver 1994; Enns and Rensink 1998) and that attention can be directed to these units (Driver and Baylis 1998; Driver et al. 2001).

Attention can also be directed to parts of objects defined by the minima and short-cut rules (Hoffman and Richards 1984; Hoffman and Singh 1997; Singh et al. 1999). Shifts of attention that cross such part boundaries take longer than equal-length shifts that do not, with a larger effect for more salient boundaries (Singh and Scholl 2000; Scholl 2001).

As mentioned earlier, perceptual units created by the human visual system may be grouped together preattentively, and attention can be directed towards those units. This idea is incorporated in two theories of object segmentation. One is Rensink's coherence theory (Rensink 2000a, c), which defines three different levels of object-based attention. First, early visual processes segment the scene into 'proto-objects'. These can be complex, but are volatile and constantly regenerated. Second, a 'setting system' computes the gist of the scene and its spatial layout, and directs attention to proto-objects in the scene. Finally, attention grabs up to six of the volatile proto-objects (Pashler 1988; Rensink 2000b), requiring about 100 m s to grab each proto-object, and holds them in coherence over space and time. This permits more detailed visual descriptions and more stable storage in visual working memory. Once attention is released, the proto-objects lose coherence and return to the volatile field of proto-objects.

'Object-file' theory differs from coherence theory by positing that object descriptions persist after the release of attention (Kahneman and Treisman 1984; Kahneman et al. 1992; Treisman 1988, 1993). One consequence is that coherence theory requires focused attention to detect change in an object, whereas object-file theory does not.

Object files are updated using spatiotemporal data: an object seen at two times with matching properties is represented as a single object instead of two distinct objects. To accomplish this, object-file theory, like coherence theory, posits preattentional processes that track and represent objects prior to encoding. In object-file theory this tracking is known as 'visual indexing' (Pylyshyn 1989, 1994, 2001).

Once attention is focused upon an object, how does encoding proceed? Are all properties of the object encoded automatically or are some given priority over others? Although some theories propose that attending to an object entails encoding all of its properties in visual working memory (Kahneman and Henik 1981; Duncan 1993a, b; Duncan and Nimmo-Smith 1996; O'Craven et al. 1999), there is evidence that, at least under high attentional loads, object properties may be encoded based upon their relevance to the task at hand. For instance, change-blindness studies have shown that an observer's search speed is influenced by the shapes of the items if the task is to detect changes of orientation, but not if the task is to detect changes of contrast polarity (Rensink 2000b). And in a MOT task,

spatiotemporal properties are more easily encoded than featural properties, such as color and shape (Scholl et al. 2001).

There is a key difference between visual indexing and the setting system of coherence theory, although both are pre-attentive processes for object-based attention. The setting system uses endogenous meaning and exogenous salience to guide attention, whereas visual indexing uses only exogenous salience (Pylyshyn 2001). The setting system allows experience to influence attentional selection of objects: the more experience with a particular scene, the more influence a top-down procedure may have in selecting the most meaningful objects for encoding in that scene. Indeed, change-blindness studies show better detection for objects of central interest in a scene than for objects of marginal interest (Rensink et al. 1997),

8. SPACETIME FRAGMENTS

With this survey of attention as background, we now propose a *spacetime fragment* (STF) theory of object perception and attention, to understand our change-detection results and to generate new empirical predictions. The STF theory extends the fragment theory of Ullman et al. (2001) by incorporating temporal change as an integral part of the fragments, and by using spacetime fragments in an account of visual attention. We first briefly review the fragment theory of Ullman et al. and then discuss the extensions introduced by STF theory.

Identification and classification are different tasks, and pose different recognition problems. To identify Bill's face, one can use a specific model of Bill's face, and this model facilitates the identification process. If, for instance, Bill has widely-set eyes and a long nose, then one can look for matches to these specific features. However, to classify a face as a face, one cannot rely simply on a model of a specific person's face, since faces can vary in many different dimensions.

Several methods have been proposed for identification, including interpolation (Poggio and Edelman 1990), linear combinations of views (Ullman and Basri 1991), and recognition polynomials (Bennett et al. 1993). Among the methods proposed for classification are recognition by components (Biederman 1985), eigenspaces (Turk and Pentland 1990), and nearest-neighbor grouping in feature spaces (Murase and Nayar 1995).

The fragment theory of Ullman et al. classifies visual objects by analyzing the fragments of which they are composed. A fragment is simply a connected region of an image. Fragments can vary in size, position, and resolution. A fragment of a face, for instance, could be a rectangular region

of the face that contains just an eye, or a nose, or both eyes, or the whole face.

To learn which fragments are most useful in recognizing members of a class, fragments of various sizes and resolutions are examined, and those fragments are chosen which have the highest mutual information for the class. The mutual information between a class C and collection of fragments F is denoted by $I(C, F)$ and defined as $I(C, F) = H(C) - H(C/F)$, where H denotes entropy (Haykin 1999, p. 493). A fragment with the highest mutual information for the class resolves the most uncertainty about whether a member of that class is present in an image. This is similar to the 'functionality principle' of Schyns and Murphy (1994), which says that 'if a fragment of a stimulus categorizes objects (distinguishes members from non-members), the fragment is instantiated as a unit in the representational code of object concepts' (Schyns 1998, p. 155).

Once these fragments have been learned for each class, the fragments are organized into types. With the class of faces, for instance, there might be several types, including right eye, left eye, mouth, and whole face. Several fragments might be of the right-eye type, each showing a different instance of a right eye. One such fragment might depict a front view of a right eye with an averted gaze, another a three-quarters view with a forward gaze, and so on.

After the fragments have been learned and organized into types, they can be used for classification. Each time a new image is encountered, it is searched for matches to the set of fragments that have been learned. The matching process is tolerant, and does not require exact identity of all pixels between the stored fragment and the subset of the image. Once all the fragment matches have been found, that class is chosen which maximizes the likelihood ratio $P(F|C)/P(F/-C)$, where F denotes the set of matching fragments and C denotes a class. This method has proven to be effective in classifying faces and cars (Ullman et al. 2001).

The spacetime-fragment theory extends the fragment theory of Ullman et al. (2001) in two ways. First, it extends the definition of fragments. In fragment theory a fragment is a continuous region of an image; in STF theory a spacetime fragment is a region of an image as it changes through a continuous period of time. Fragments are two-dimensional; spacetime fragments are three-dimensional. A fragment might depict a static view of an eye; a spacetime fragment might depict that same eye blinking once, or averting its gaze to the right.

Second, spacetime-fragment theory proposes a new account of visual attention, in which attention is directed to the most informative spacetime

fragments. Which spacetime fragments are most informative depends on the tasks and goals of the observer. If the task is to decide if a person is happy or sad, then spacetime fragments depicting actions of the corners of the mouth might be most informative and attract the greatest attention. If the task is to decide where a person is attending, then spacetime fragments depicting movements of the eyes might be most informative. The precise definition of 'most informative' is 'highest mutual information' with the class of interest. With each task one can associate classes of objects. With the task of deciding if a person is happy or sad, one can associate a class of happy faces and a class of sad faces, and one can order the possible spacetime fragments by their mutual information with each of these classes.

The distinction between fragments and spacetime fragments is nicely illustrated by the success of comic impersonators. A comic who successfully impersonates, say, Bush or Blair, need not have the same facial features or facial configuration as Bush or Blair, which entails that they need not have the same facial fragments. Their success depends, in large part, on mimicking the characteristic facial and bodily movements of Bush or Blair, perhaps a pursing of the lips or gesture of the hand. That is, it depends on mimicking the temporal aspects of key spacetime fragments of the face and body. Even if the spatial aspects of a spacetime fragment are not matched, a close enough matching of its temporal aspects can be enough to activate the entire spacetime fragment. Indeed, this very incongruity may contribute to the humor of the successful comic. Hill and Johnston (2001), using a neutral 3D face animated with motion-capture data, have found that gender and identity can be determined from facial motion alone.

This dual aspect of spacetime fragments, viz., their integration of spatial and temporal patterns, leads to the prediction of some interesting visual illusions. It should be possible to match the spatial pattern of a spacetime fragment with sufficient precision that the entire spacetime fragment is activated, leading to the illusory perception of the (possibly complex) movements encoded in the temporal pattern of the spacetime fragment: Complex-motion illusions should be possible, triggered by the right spatial pattern. Conversely, it should be possible to match the temporal pattern of a spacetime fragment with sufficient precision that the entire spacetime fragment is activated, leading to the illusory perception of the (possibly complex) spatial patterns encoded in the spacetime fragment: Complex-shape illusions should be possible, triggered by the right temporal pattern.

Anecdotal support for these predictions is reported by Bellefeuille and Faubert (1998), in studies of biological motion of animals. Some of the

stimuli depicted silhouette contours of animals but not their biological motions, while others depicted the biological motions by moving dots, but not their silhouette contours. Subjects occasionally reported that contour-defined animals actually appeared to be moving in a biologically correct fashion, or that animals defined by dots moving in biological motion appeared to have a contour (Bellefeuille and Faubert 1998, p. 235). These are precisely the kinds of complex illusions of form or motion that one would expect based on the spacetime-fragment theory. Activation of a spacetime fragment by either its spatial or temporal aspect should lead to the illusory perception of its other aspect. A visually-simpler example is the spoke-brightness illusion: when a white disk revolves around a fixation point on a gray background, observers perceive a dark gray region that connects the disk to the fixation point (Holcombe et al. 2000).

The dual aspect of spacetime fragments also suggests that there should be many cells in the ventral visual pathway with neural receptive fields that are selective for both form and motion. These should be found from the earliest stages of processing in V1 to the most advanced stages in TE. In the early stages there should be cells selective not just for simple forms, such as oriented edges, but also for simple motions of these edges. In intermediate and later stages, such as in V4 through TE and the STS, there should be cells selective for motions of forms of intermediate and greater complexity. Some support for these predictions comes from studies that find cortical areas that process both biological motion and contour information (Bruce et al. 1981; Desimone et al. 1984; Oram and Perret 1994; Perrett et al. 1985).

9. SPACETIME FRAGMENTS AND SPRITES

A sprite is a set of visual routines, of the type proposed by Ullman (1984), 'that is responsible for detecting the presence of a specific characteristic motion in the input array, for modeling or animating the object's changing configuration as it makes this stereotypical motion, and for filling in the predictable details of the motion over time and in the face of noisy or absent image details' (Cavanagh et al. 2001, p. 48). Sprites have been proposed by Cavanagh et al. to govern the perception of all kinds of dynamic stimuli, from simple motions of a single dot, to complex biological motions of the type studied by Johansson (1973). Given that sprites and spacetime fragments both claim to account for our perception of dynamic stimuli, it is natural to ask how the two accounts differ.

One difference between spacetime fragments and sprites is that spacetime fragments are chosen to maximize the mutual information between

the fragments and certain object classes; sprites are not posited to maximize mutual information with any class. But the key difference is that sprites, being based on visual routines, require attention for their construction. By contrast, spacetime fragments need not require attention for their construction, although attention might assist in the construction of complex spacetime fragments.

The differences between sprites and spacetime fragments have consequences for theories of eye movements and attention. Since attention is required to construct a sprite, sprites cannot be the foundation of a theory of eye movements and attention in which attention is directed to preexisting, preattentively constructed, sprites. However, since spacetime fragments don't require attention for their construction, attention and eye movements can be directed to preexisting, preattentively constructed, spacetime fragments that maximize mutual information with a class of interest. Thus spacetime fragments can motivate a theory of eyemovements and attention in a way that sprites cannot. And this difference is critical to our account of face processing.

10. SPACETIME FRAGMENTS AND FACE PROCESSING

Spacetime fragments can account for a range of well-known phenomena in face perception, such as the difficulty with interpreting or recognizing inverted, negated, or foreign faces. In each of these cases the account follows from the statistical nature of spacetime fragments. They are acquired by experience, and prioritized by their mutual information to classes of interest. Since we have little experience with inverted, negated, or foreign faces, we have few spacetime fragments that will match features of these faces. Therefore it should not be possible, or it should take longer, for enough matches to be found so that such faces can be properly interpreted. This was the result in the experiments discussed here: changes to inverted or negated faces took longer to detect, and were detected with less accuracy. The subjects simply had fewer appropriate spacetime fragments at their disposal for the inverted and negated faces, and their performance suffered accordingly. In particular, their attentional strategies suffered because, according to STF theory, attention is directed to the spacetime fragments that maximize mutual information for the class of interest, and there simply were few such fragments available. As a result, subjects had to abandon an endogenous strategy for the control of attention, which requires the proper spacetime fragments to drive it, and to fall back on an exogenous strategy, governed by the low-level salience of image features.

STF theory accounts well, as we have just seen, for the experimental results discussed here. Indeed, it has more resources than are necessary to explain these results, since spacetime fragments incorporate temporal dynamics into the structure of the fragments, whereas our experimental stimuli were simply static views of faces. These extra resources provide a theoretical framework, and theoretical motivation, to extend the experiments discussed here to the dynamic case.

The experiments discussed here had observers search for changes between two static images of a face. But faces, when not asleep, are rarely static in real life. They continuously change as a person talks, blinks, coughs, sneezes, and displays emotion. A natural extension of our experiments will have subjects search for changes between two video sequences of a moving face. This will allow one to explore how attention to the face is affected by facial motions such as talking, crying, laughing, or glancing to one side. The interpretation of such experiments will require the full temporal facilities of spacetime fragments.

For instance, if the two video sequences display a face laughing, then specific motion fragments of specific parts of the face will be preattentively discovered by the visual system of the observer, and these spacetime fragments will activate the object class for which they maximize the mutual information, namely the class 'laughing faces'. Activation of this class will then direct attention towards certain features of the face and away from others. This difference in attention may lead to differences in speed and accuracy of change detection for these features, differences which can be explored experimentally.

11. FUTURE EXPERIMENTS

Recent studies of emotion processing reveal several interesting findings. In particular, some expressions may demand more attention than others. A study of attentional interference found that presentation of an irrelevant angry face delayed response to an unrelated cognitive task more than presentation of an irrelevant happy face (White, 1996). Additionally, a study of visual search found that search slopes are shallower, meaning that detection is faster, for sad faces in comparison to happy faces (Eastwood et al. 2001). Given these findings of differential attention levels for different emotions, it would be interesting to study the different patterns of attention for different emotions.

For instance, previous studies of eye movements report more fixations to the corners of the eyes when the face is smiling than when portraying other emotions (Williams et al. 2001). Given these results, we might

predict that in a visual search task, if one face differs from all the other faces only by its eyes, observers might be more successful if all the faces (both target and distracters) displayed happy expressions. Conversely, if the target face differed only by its nose, perhaps a field of faces displaying the disgust expression (which usually involves a wrinkling of the nose) may result in a shallower search slope. In exploring the subconscious guidance of emotions in triggering different systematic ways of looking and attending, we may learn valuable information about which features and aspects of the face human vision gives preference to in making judgments of emotion.

To incorporate the theory of space-time fragments these studies can also be carried out dynamically. For instance, we can see the real-time switch of emotion patterns of attention if we present a field of faces in a neutral face but then, simultaneously, animate them all so that they display a certain expression. If all the faces move into a smiling face, will the face that differs by the eyes be detected more readily than if all the faces moved to a frown? We can also investigate the robustness of the space-time fragment. For example, how sensitive are we to a space-time fragment that has slightly different movement? If all faces move into a smile, but one face has a slightly different smile, how quickly is that difference detected? Are there certain space-time fragments that are more robust than others? And, if so, are they all related to displaying the same class of expression, such as sadness or anger?

In addition to using space-time fragments to learn more about perception of emotional expressions, we can also investigate perception of identity. For identifying individuals, which space-time fragments are most informative? One possible study could include both a learning and a test phase in which observers view a set of dynamic faces which they are later asked to identify. These faces will not reveal any static visual information, but will be completely neutral and recognized solely on their patterns of movement. During the test phase some of these faces can be presented with conflicting information. For example, the eye movement patterns of one face could be crossed with the mouth movement patterns of another face. To control for possible gender differences (Hill and Johnston 2001), only motions from same-sex faces would be swapped. The results would then be analyzed to see which space-time fragments were given more weight in the identification process.

Whether expanding emotional expression models, such as the facial action coding system (FACS) to include the dynamic movements of space-time fragments, or incorporating movement patterns into computer

programs built to model human vision's approach to face recognition, spacetime fragments are a rich source of new research.

12. CONCLUSION

Previous studies have shown that inversion and negation impair perception of configural features of static faces. The experiments discussed here extend these results by showing that inversion and negation impair attention to local and configural features of static faces. These impairments of attention can be explained by STF theory, which posits that preattentively-constructed spacetime fragments activate that object class for which they maximize mutual information, and this activated object class in turn guides deployment of attention. STF theory motivates extension of the experiments presented here to the case of dynamic faces. Further developments of the STF theory might incorporate dynamic gabor-filter representations of images (see, e.g., Simoncelli and Olshausen, 2001) rather than using simply the raw image pixels.

ACKNOWLEDGEMENTS

This material is based upon work supported by the National Science Foundation under Grant No. 0090833. We thank Shimon Ullman and Michel Vidal-Naquet for helpful discussions.

REFERENCES

Althoff, R.R. and N.J. Cohen: 1999, 'Eye-Movement-Based Memory Effect: A Reprocessing Effect in Face Perception', *Journal of Experimental Psychology: Learning, Memory, and Cognition* **25**, 997–1010.

Ballard, D.H., M.M. Hayhoe, and J.B. Pelz: 1995, 'Memory Representations in Natural Tasks', *Journal of Cognitive Neuroscience* **B**, 66–80.

Bellefeuille, A. and J. Faubert: 1998, 'Independence of Contour and Biological-Motion Cues for Motion-Defined Animal Shapes', *Perception* **27**, 225–235.

Bennett, B.M., D.D. Hoffman, and C. Prakash: 1993, 'Recognition polynomials', *Journal of the Optical Society of America* **A10**, 759–764.

Bentin, S., T. Allison, A. Puce, E. Perez, and G. McCarthy: 1996, 'Electrophysiological Studies of Face Perception in Humans', *Journal of Cognitive Neuroscience* **8**, 551–565.

Bentin, S. and L.Y. Deouell: 2000, 'Structural Encoding and Identification in Face Processing: ERP Evidence for Separate Processes', *Cognitive Neuropsychology* **17**, 35–54.

Biederman I.: 1985, 'Human Image Understanding: Recent Research and Theory', *Computer Vision, Graphics and Image Processing* **32**, 29–73.

Borod, J.C., W. Vingiano, and F. Cytryn: 1988, 'The Effects of Emotion and Ocular Dominance on Lateral Eye Movement', *Neuropsychologia* **26**, 213–220.

Bruce, C.J., R. Desimone, and C.G. Gross: 1981, 'Visual Properties of Neurons in the Polysensory Area in Superior Temporal Sulcus of the Macaque', *Journal of Neurophysiology* **46**, 369–384.

Bruce, V. and S. Langton: 1994, 'The Use of Pigmentation and Shading Information in Recognising the Sex and Identities of Faces', *Perception* **23**, 803–822.

Cauquil, A.S., G.E. Edmonds, and M.J. Taylor: 2000, 'Is the Face-Sensitive N170 the Only ERP not Affected by Selective Attention?', *NeuroReport* **11**, 2167–2171.

Cavanagh, P., A. Labianca, and I. Thornton: 2001, 'Attention-Based Visual Routines: Sprites',*Cognition* **80**, 47–60.

Davies, T.N. and D.D. Hoffman: 2002, 'Attention to Faces: A Change-Blindness Study', *Perception* **31**, 1123–1146.

Davis, G. and J. Driver: 1994, 'Parallel Detection of Kanizsa Subjective Figures in the Human Visual System', *Nature* **371**, 791–793.

Desimone, R., T.D. Albright, C.G. Gross, and C. Bruce: 1984, 'Stimulus-Selective Properties of Inferior Temporal Neurons in the Macaque', *Journal of Neuroscience* **8**, 2051–2062.

Deubel, H. and W.X. Schneider: 1996, 'Saccade Target Selection and Object Recognition: Evidence for a Common Attentional Mechanism', *Vision Research* **36**, 1827–1837.

Diamond, R. and S. Carey: 1986, 'Why Faces are and are Not Special: An Effect of Expertise', *Journal of Experimental Psychology: General* **115**, 107–117.

Downing, C. and S. Pinker: 1985, 'The Spatial Structure of Visual Attention', in M. Posner and O.S.M. Marin (eds.), *Attention and Performance*, Vol. XI, London: Erlbaum) pp. 171–187.

Driver, J. and G.C. Baylis: 1998, 'Attention and Visual Object Segmentation', in R. Parasuraman (ed.), *The Attentive Brain*, Cambridge, MA: MIT Press, pp. 299–325.

Driver, J., G. Davis, C. Russell, M. Turatto, and E. Freeman: 2001, 'Segmentation, Attention, and Phenomenal Visual Objects', *Cognition* **80**, 61–95.

Duncan, J.: 1984, 'Selective Attention and the Organization of Visual Information', *Journal of Experimental Psychology: General* **113**, 501–517.

Duncan, J.: 1993a, 'Coordination of What and Where in Visual Attention', *Perception* **22**, 1261–1270.

Duncan, J.: 1993b, 'Similarity between Concurrent Visual Discriminations: Dimensions and Objects', *Perception & Psychophysics* **54**, 425–430.

Duncan, J. and I. Nimmo-Smith: 1996, 'Objects and Attributes in Divided Attention: Surface and Boundary Systems', *Perception & Psychophysics* **58**, 1076–1084.

Eastwood, J.D., D. Smilek, and P.M. Merikle: 2001, 'Differential Attentional Guidance by Unattended Faces Expressing Positive and Negative Emotion', *Perception and Psychophysics* **63**(6), 1004–1013.

Egly, R., J. Driver, and R. Rafal: 1994, 'Shifting Visual Attention between Objects and Locations: Evidence for Normal and Parietal Lesion Subjects', *Journal of Experimental Psychology: General* **123**, 161–177.

Eimer, M.: 1998, 'Does the Face-Specific N170 Component Reflect the Activity of a Specialized Eye Detector?', *NeuroReport* **9**, 2945–2948.

Eimer, M.: 2000a, 'Event-Related Brain Potentials Distinguish Processing Stages Involved in Face Perception and Recognition', *Clinical Neurophysiology* **111**, 694–705.

Eimer, M.: 2000b, 'Attentional Modulations of Event-Related Brain Potentials Sensitive to Faces', *Cognitive Neuropsychology* **17**, 103–116.
Eimer, M.: 2000c, 'Effects of Face Inversion on the Structural Encoding and Recognition of Faces: Evidence from Event-Related Brain Potentials', *Cognitive Brain Research* **10**, 145–158.
Eimer, M., and R.A. McCarthy: 1999, 'Prosopagnosia and Structural Encoding of Faces: Evidence from Event-Related Potentials', *NeuroReport* **10**, 255–259.
Enns, J.T. and R. Rensink: 1998, 'Early Completion of Occluded Objects', *Vision Research* **38**, 2489–2505.
Farah, M.J., C. Rabinowitz, G.E. Quinn, and G.T. Liu: 2001, 'Early Commitment of Neural Substrates for Face Recognition', *Cognitive Neuropsychology* **17**(1/2/3), 117–123
Galper, R.E: 1970, 'Recognition of Faces in Photographic Negative', *Psychonomic Science* **19**, 207–208.
Galper, R.E. and J. Hochberg: 1971, 'Recognition Memory for Photographs of Faces', *American Journal of Psychology* **84**, 351–354.
George, N., J. Evans, N. Fiori, J. Davidoff, and B. Renault: 1996, 'Brain Events Related to Normal and Moderately Scrambled Faces', *Cognitive Brain Research* **4**, 65–76.
Gosselin, F. and P.G. Schyns: 2001, 'Bubbles: A Technique to Reveal the Use of Information in Recognition Tasks', *Vision Research* **41**, 2261–2271.
Haxby, J.V., E.A. Hoffman, and M.I. Gobbini: 2000, 'The Distributed Human Neural System for Face Perception', *Trends in Cognitive Sciences* **4**, 223–233.
Hayhoe, M.M., D. Bensinger, and D.H. Ballard: 1998, 'Task Constraints in Visual Working Memory', *Vision Research* **38**, 125–137.
Helmholtz, H.: 1867, *Handbuch der physiologischen optik*, Leipzig: Voss.
Hill, H. and A. Johnston: 2001, 'Categorizing Sex and Identity from the Biological Motion of Faces', *Current Biology* **11**, 880–885.
Hoffman, D.D. and W.A. Richards: 1984, 'Parts of Recognition', *Cognition* **18**, 65–96.
Hoffman, D.D. and M. Singh: 1997, 'Salience of Visual Parts', *Cognition* **69**, 29–78.
Holcombe, A.O., J. Intriligator, and P.U. Tse: 2000, 'The Spoke Brightness Illusion Originates at an Early Motion Processing Stage', *Perception & Psychophysics* **62**, 1619–1624.
Intriligator, J.M: 1997, 'The Spatial Resolution of Visual Attention', Unpublished doctoral dissertation, Harvard University, Cambridge, MA.
James, W.: 1890, *Principles of Psychology*, New York: Henry Holt.
Jemel, B., N. George, L. Chaby, N. Fiori, and B. Renault: 1999, 'Differential Processing of Part-to-Whole and Part-to-Part Face Priming: An ERP Study', *NeuroReport* **10**, 1069–1075.
Johansson, G: 1973, 'Visual Perception of Biological Motion and a Model for its Analysis', *Perception & Psychophysics* **14**, 201–211.
Johnson, M.H., S. Dziurawiec, H. Ellis, and J. Morton: 1991, 'Newborns' Preferential Tracking of Face-Like Stimuli and its Subsequent Decline', *Cognition* **40**, 1–19.
Kahneman, D. and A. Henik: 1981, 'Perceptual Organization and Attention', in M. Kubovy and J. Pomerantz (eds.), *Perceptual Organization*, Hillsdale, NJ: Erlbaum, pp. 181–211.
Kahneman, D. and A. Treisman: 1984, 'Changing Views of Attention and Automaticity' in R. Parasuraman & D.R. Davies (eds.), *Varieties of Attention*, New York: Academic Press, pp. 29–61.
Kahneman, D., A. Treisman, and B.J. Gibbs: 1992, 'The Reviewing of Object Files: Object-Specific Integration of Information', *Cognitive Psychology* **24**, 174–219.

Kanizsa, G. and G. Tampieri: 1976, 'Environmental and Retinal Frames of Reference in Visual Perception', *Italian Journal of Psychology* **3**, 317–332.

Kemp, R., C. McManus, and T. Pigott: 1990, 'Sensitivity to the Displacement of Facial Features in Negative and Inverted Images', *Perception* **19**, 531–543.

Kemp, R., G. Pike, P. White, and A. Musselman: 1996, 'Perception and Recognition of Normal and Negative Faces: The Role of Shape from Shading and Pigmentation Cues', *Perception* **25**, 37–52.

Klein, R., A. Kingstone, and A. Pontefract: 1992, 'Orienting of Visual Attention', in K. Rayner (ed.), *Eye Movements and Visual Cognition: Scene Perception and Reading*, New York: Springer, pp. 46–65.

Lansing, C.R. and G.W. McConkie: 1999, 'Attention to Facial Regions in Segmental and Prosodic Visual Speech Perception Tasks', *Journal of Speech, Language, and Hearing Research* **42**, 526–539.

Lewis, M.B. and R.A. Johnston: 1997, 'Familiarity, Target Set and False Positives in Face Recognition', *European Journal of Cognitive Psychology* **9**, 437–459.

Liu, C.H., C.A. Collin, A.M. Burton, and A. Chaudhuri: 1999, 'Lighting Direction Affects Recognition of Untextured Faces in Photographic Positive and Negative', *Vision Research* **39**, 4003–4009.

Miyakawa, T.: 1950, 'Experimental Research on the Structure of Visual Space When We Bend Forward and Look Backward between the Spread Legs', *Japan Journal of Psychology* **20**, 14–23.

Murase, H. and S.K. Nayar: 1995, 'Visual Learning and Recognition of 3-D Objects from Appearance', *International Journal of Computer Vision* **14**, 5–24.

Neisser, U.: 1967, *Cognitive Psychology*, New York: Appleton-Century-Crofts.

Neisser, U.: 1979, 'The Control of Information Pickup in Selective Looking', in A. Pick (ed.), *Perception and its Development*, Hillsdale, NJ: Erlbaum, pp. 201–219.

Neisser, U. and R. Becklen: 1975, 'Selective Looking: Attending to Visually Specified Events', *Cognitive Psychology* **7**, 480–494.

O'Craven, K., P. Downing, and N. Kanwisher: 1999, 'fMRI Evidence for Objects as the Units of Attentional Selection', *Nature* **401**, 584–587.

Oram, M.W. and D.I. Perrett: 1994, 'Responses of Anterior Superior Temporal Polysensory (STPa) Neurons to 'Biological Motion' Stimuli', *Journal of Cognitive Neuroscience* **6**, 99–116.

O'Regan, J.K., H. Deubel, J.J. Clark, and R.A. Rensink: 2000, 'Picture Changes during Blinks: Looking without Seeing and Seeing without Looking', *Visual Cognition* **7**, 191–211.

Pashler, H.: 1988, 'Familiarity and Visual Change Detection', *Perception & Psychophysics* **44**, 369–378.

Perrett, D.I., A. Chitty, A. Mistlin, and H. Harries: 1985, 'Visual Cells Sensitive to Biological Motion', *Behavioral Brain Research* **16**, 153–170.

Poggio, T. and S. Edelman: 1990, 'A Network that Learns to Recognize Three-Dimensional Objects', *Nature* **343**, 263–266.

Posner, M.I.: 1980, 'Orienting of Attention', *Quarterly Journal of Experimental Psychology* **32**, 3–26.

Posner, M.I., C.R.R. Snyder, and B.J. Davidson: 1980, 'Attention and the Detection of Signals', *Journal of Experimental Psychology: General* **109**, 160–174.

Pylyshyn, Z.W.: 1989, 'The Role of Location Indexes in Spatial Perception: A Sketch of the FINST Spatial Index Model', *Cognition* **32**, 65–97.

Pylyshyn, Z.W.: 1994, 'Some Primitive Mechanisms of Spatial Attention', *Cognition* **50**, 363–384.

Pylyshyn, Z.W.: 2001, 'Visual Indexes, Preconceptual Objects, and Situated Vision', *Cognition* **80**, 127–158.

Pylyshyn, Z.W. and R.W. Storm: 1988, 'Tracking Multiple Independent Targets: Evidence for a Parallel Tracking Mechanism', *Spatial Vision* **3**, 179–197.

Ramachandran, V.S.: 1988, 'Perception of Shape from Shading', *Nature* **331**, 163–166.

Rensink, R.A.: 2000a, 'The Dynamic Representation of Scenes', *Visual Cognition* **7**, 17–42.

Rensink, R.A.: 2000b, 'Visual Search for Change: A Probe into the Nature of Attentional Processing', *Visual Cognition* **7**, 345–376.

Rensink, R.A.: 2000c, 'Seeing, Sensing, and Scrutinizing', *Vision Research* **40**, 1469–1487.

Rensink, R.A., J.K. O'Regan, and J.J. Clark: 1995, 'Image Flicker is as Good as Saccades in Making Large Scene Changes Invisible', *Perception* **24**(Suppl.), 26–27.

Rensink, R.A., J.K. O'Regan, and J.J. Clark: 1997, 'To See or Not to See: The Need for Attention to Perceive Changes in Scenes', *Psychological Science* **8**, 368–373.

Rock, I.: 1956, 'The Orientation of Forms on the Retina and in the Environment', *American Journal of Psychology* **69**, 513–528.

Rock, I.: 1973, *Orientation and Form*, New York: Academic Press.

Scholl, B.J., Z.W. Pylyshyn, and S.L. Franconeri: 2001, 'The Relationship between Property-Encoding and Object-Based Attention: Evidence from Multiple Object Tracking', submitted.

Searcy, J.H. and J.C. Bartlett: 1996, 'Inversion and Processing of Component and Spatial-Relational Information in Faces', *Journal of Experimental Psychology: Human Perception and Performance* **22**, 904–915.

Sears, C.R. and Z.W. Pylyshyn: 2000, 'Multiple Object Tracking and Attentional Processing', *Canadian Journal of Experimental Psychology* **54**, 1–14.

Shore, D.I. and R.M. Klein: 2000, 'The Effects of Scene Inversion on Change Blindness', *The Journal of General Psychology* **127**, 27–43.

Simoncelli, E.P. and B.A. Olshausen: 2001, 'Natural Image Statistics and Neural Representation', *Annual Review of Neuroscience* **24**, 1193–1216.

Simons, D.J. and C.F. Chabris: 1999, 'Gorillas in Our Midst: Sustained Inattentional Blindness for Dynamic Events', *Perception* **28**, 1059–1074.

Simons, D.J. and D.T. Levin: 1998, 'Failure to Detect Changes to People during a Real-World Interaction', *Psychnomic Bulletin and Review* **5**(4), 644–649.

Singh, M. and B.J. Scholl: 2000, 'Using Attentional Cueing to Explore Part Structure', Poster presented at the 2000 *Pre-Psychonomics Object Perception and Memory Meeting*, New Orleans, LA.

Singh, M., G. Seyranian, and D.D. Hoffman: 1999, 'Parsing Silhouettes: The Short-Cut Rule', *Perception & Psychophysics* **61**, 636–660.

Thouless, R.H.: 1947, 'The Experience of 'Upright' and 'Upside-Down' in Looking at Pictures', In I. Nuttih (ed.), *Miscellanea Psychologica A. Michotte*, Louvain: Institut Superieur de Philosophie, pp. 130–137.

Tong, F., K. Nakayama, M. Moscovitch, O. Weinrib, and N. Kanwisher: 2000, 'Response Properties of the Human Fusiform Face Area', *Cognitive Neuropsychology* **17**(1/2/3), 257–279.

Treisman, A.: 1988, 'Features and Objects: The Fourteenth Bartlett Memorial Lectures', *Quarterly Journal of Experimental Psychology* **40**, 201–237.

Treisman, A.: 1993, 'The Perception of Features and Objects', in A. Baddeley and L. Weiskrantz (eds.), *Attention: Selection, Awareness, and Control*, Oxford: Clarendon Press, pp. 5–35.

Ullman, S. and R. Basri: 1991, 'Recognition by Linear Combination of Models', *IEEE Transactions on Pattern Analysis and Machine Intelligence* **13**, 992–1006.

Ullman, S., E. Sali, and M. Vidal-Naquet: 2001, 'A Fragment-Based Approach to Object Representation and Classification', *Proceedings of the 4th International Workshop on Visual Form*, pp. 85–100.

Valentine, T.: 1988, 'Upside-Down Faces: A Review of the Effects of Inversion upon Face Recognition', *British Journal of Psychology 79*, 471–491.

Watt, R.J.: 1988, *Visual Processing: Computational, Psychophysical, and Cognitive Research*, Hillsdale, NJ: Erlbaum.

White, M.: 1996, 'Anger Recognition is Independent of Spatial Attention', *New Zealand Journal of Psychology* **25**, 30–35.

Williams, L.M., C. Senior, A.S. David, C.M. Loughland, and E. Gordon: 2001, 'In Search of the Duchenne Smile: Evidence from Eye Movements', *Journal of Psychopathology* **15**, 122–127.

Williams, P. and Simons, D.J.: 2000, 'Detecting Changes in Novel, Complex Three-Dimensional Objects', *Visual Cognition* **7**, 297–322.

Yin, R.K.: 1969, 'Looking at Upside-Down Faces', *Journal of Experimental Psychology* **81**, 141–145.