Coordinate Transformations in Object Recognition

Markus Graf
Max Planck Institute for Biological Cybernetics

A basic problem of visual perception is how human beings recognize objects after spatial transformations. Three central classes of findings have to be accounted for: (a) Recognition performance varies systematically with orientation, size, and position; (b) recognition latencies are sequentially additive, suggesting analogue transformation processes; and (c) orientation and size congruency effects indicate that recognition involves the adjustment of a reference frame. All 3 classes of findings can be explained by a transformational framework of recognition: Recognition is achieved by an analogue transformation of a perceptual coordinate system that aligns memory and input representations. Coordinate transformations can be implemented neurocomputationally by gain (amplitude) modulation and may be regarded as a general processing principle of the visual cortex.

Keywords: alignment, coordinate transformations, gain modulation, object recognition, reference frames

How can we recognize objects regardless of spatial transformations such as plane and depth rotation, size scaling, and position changes? This ability is often discussed under the label *object constancy* or *shape constancy*. Even young children recognize objects so immediately and effortlessly that it seems to be a rather ordinary and simple task. However, changes in the spatial relation between observer and object lead to large changes of the image that is projected onto the retina. Hence, to recognize objects regardless of orientation, size, and position is not a trivial problem. No computational system proposed so far can successfully recognize objects over wide ranges of object categories and contexts.

Several different approaches have been proposed over the years (for reviews, see Palmeri & Gauthier, 2004; Ullman, 1996). A number of models rely on abstract object representations, which predict that recognition performance is typically invariant regarding spatial transformations (e.g., structural description models; see Hummel & Biederman, 1992; Marr & Nishihara, 1978). In contrast, image-based or view-based models propose that object representations are close to the format of the perceptual input and therefore depend systematically on image transformations (e.g., Edelman, 1998; Tarr, 2003). More recently, hybrid models have been proposed that aim at integrating both approaches (Edelman & Intrator, 2001; Hummel & Stankiewicz, 1998).

One central issue of research and scientific debate in this area is the question of orientation dependency. A large number of studies have demonstrated that recognition performance depends systematically on the orientation of the stimulus (for reviews, see Jolicoeur & Humphrey, 1998; Lawson, 1999; Tarr, 2003). Even though it is widely accepted that orientation dependency should be interpreted in terms of a pictorial or image-based (or view-based) model of recognition (Jolicoeur & Humphrey, 1998; Tarr & Bülthoff, 1998), there is still no consensus as to which model is best suited to explain the data (e.g., Bar, 2001; Biederman & Bar, 2000; Biederman & Gerhardstein, 1995; Edelman & Intrator, 2001; Foster & Gilson, 2002; Hayward & Tarr, 2000; Tarr & Bülthoff, 1995; Thoma, Hummel, & Davidoff, 2004).

The aim of this article is to lay the groundwork for a new framework of object recognition that accounts for the majority of findings and integrates previously distinct areas of research. When further classes of—previously neglected—data are considered, a new integrative view on recognition emerges that suggests that object recognition relies on *coordinate transformations*, that is, on transformations of a perceptual coordinate system that align input and memory representations. Researchers from computational neuroscience have also proposed that coordinate transformations are crucial for object perception and recognition (e.g., Pouget & Sejnowski, 1997, 2001; Salinas & Abbott, 1997a, 1997b), arriving at this conclusion from an entirely different starting point and providing converging evidence for coordinate transformations in object recognition. Coordinate transformations are fundamental also for visuomotor control and so may be considered as an integrative processing principle for the visual cortex (e.g., Salinas & Abbott, 2001; Salinas & Sejnowski, 2001; Salinas & Thier, 2000). Thus, the transformational framework is integrative at three different levels. First, it accounts for a large number of currently unrelated and neglected studies in the recognition literature. Second, it integrates behavioral findings with approaches from computational neuroscience modeling the behavior of single neurons. Third, the transformational framework suggests common processing principles in object recognition and visuomotor control.

Three central classes of findings are identified that have to be explained by any model of recognition. First, recognition performance deteriorates systematically with increasing changes of orientation, size, and position of the object (addressed in Section 1). Second, transformation processes in recognition (rotations and size scalings) pass through intermediate points along the transformational path, suggesting that compensation processes in recognition are analogue (addressed in Section 2). Third, the recognition of objects that are rotated or size scaled is facilitated when they are presented immediately after a different object shown at the same orientation or size. This indicates that recognition involves the adjustment of a perceptual coordinate system or reference frame. Coordinate transformations can be implemented at the neuronal level by gain modulation, that is, by simple multiplicative interaction (addressed in Section 3). In Section 4, I argue that current object recognition models are not able to explain all three classes of data without introducing new ad hoc assumptions. A transformational framework of recognition (TFR) that can accommodate all three classes of findings in a simple and parsimonious way is proposed in Section 5. According to TFR, recognition is achieved by an analogue alignment transformation of a perceptual coordinate system that specifies correspondences between memory representations and the visual input. In contrast to relatively slow image transformations in mental imagery, recognition is based on relatively fast coordinate transformations (implemented by neural gain modulation).

1. Systematic Relation Between the Amount of Spatial Transformation and Recognition Performance

In traditional models of recognition, the ability to recognize objects after rotations, size scalings, and displacements (translations) has been accounted for by the concept of invariance, that is, on the basis of structures or relations that do not change with spatial transformations (e.g., Biederman, 1987; Cassirer, 1944; Marr & Nishihara, 1978; Pitts & McCulloch, 1947; Selfridge & Neisser, 1963). The concept of invariance is still influential, but is recognition performance actually invariant regarding spatial transformations? Whereas research has often focused narrowly on orientation effects, I also review studies investigating effects of other spatial transformations, like changes of size and position.

Orientation Dependency

The question of how people recognize objects after changes in their spatial orientation has been investigated extensively. Many studies have demonstrated that recognition performance depends on orientation (for reviews, see H. H. Bülthoff, Edelman, & Tarr, 1995; Jolicoeur & Humphrey, 1998; Lawson, 1999; Tarr, 2003; Tarr & Bülthoff, 1998). Most objects can be recognized faster and more accurately from certain perspectives, called *canonical perspectives*. The canonical perspective often corresponds to an upright orientation in between a frontal and side view (approximately a three-quarter view; Blanz, Tarr, & Bülthoff, 1999; Palmer, Rosch, & Chase, 1981). An object can have several canonical perspectives (Edelman & Bülthoff, 1992; Newell & Findlay, 1997). The further an object is misoriented from the canonical orientation, the more time it takes to recognize the object and the more frequently errors are made.[1] Recognition performance de-

pends in a systematic way on orientation, both for rotations in the picture plane (e.g., Jolicoeur, 1985, 1988, 1990b; Lawson & Jolicoeur, 1998, 1999) and rotations in depth (e.g., Lawson & Humphreys, 1998; Lawson, Humphreys, & Jolicoeur, 2000; Palmer et al., 1981; Srinivas, 1993; Tarr, Williams, Hayward, & Gauthier, 1998). This orientation dependency was found even when all major parts or features of an object remained visible after a rotation in depth (Humphrey & Jolicoeur, 1993; Lawson, Humphreys, & Watson, 1994), and thus it is not just a result of self-occlusion (as claimed by Biederman & Gerhardstein, 1993). Moreover, the systematic deterioration of performance with increasing misorientation of the stimulus is not simply due to low-level perceptual processes but rather seems to be caused by high-level object representations (Jolicoeur & Cavanagh, 1992; Lawson & Humphreys, 1998; Verfaillie, 1993).

When objects are presented frequently in specific orientations, these orientations may become canonical perspectives, as demonstrated in two elegant studies (Tarr, 1995; Tarr & Pinker, 1989): Participants had to study novel two-dimensional (2-D) or three-dimensional (3-D) objects from a specific orientation. In naming tasks, reaction times (RTs) increased with increasing departure from the study orientation. With extensive practice, participants recognized the objects almost equally quickly at all familiar orientations. However, when the objects were presented at unfamiliar viewpoints, performance was again viewpoint dependent, now related to the distance from the nearest familiar view. The authors have interpreted these studies as evidence that the recognition of misoriented objects involves both compensation (transformation) processes and the encoding of multiple views (see also B. S. Gibson & Peterson, 1994; Heil, Rösler, Link, & Bajric, 1998; Murray, 1999).

Orientation effects are found not only for novel objects (e.g., H. H. Bülthoff & Edelman, 1992; Edelman & Bülthoff, 1992; Tarr & Pinker, 1989) but also for common, familiar objects (e.g., Hayward & Tarr, 1997; Lawson & Humphreys, 1996, 1998; Murray, 1997, 1999; Newell & Findlay, 1997; Palmer et al., 1981). Orientation-dependent recognition performance is not limited to individual objects, like faces (e.g., Hill, Schyns, & Akamatsu, 1997) or to objects at the subordinate level of categorization (e.g., Edelman & Bülthoff, 1992; Tarr, 1995) but has also been demonstrated for basic level recognition (Hayward & Williams, 2000; Jolicoeur, Corballis, & Lawson, 1998; Lawson & Humphreys, 1998; Murray, 1998; Palmer et al., 1981). Orientation dependency has been observed in the perception of biological motion (Daems & Verfaillie, 1999; Verfaillie, 1993; for a review, see I. Bülthoff &

---

[1] In naming tasks a systematic increase usually holds only for plane rotations from 0° to 120° or 150°, whereas inverted objects are often again recognized faster, leading to an M-shaped response time function (e.g., Jolicoeur, 1985, 1988). For some participants, naming times at 180° still increase relative to naming times at 120° (e.g., Jolicoeur & Milliken, 1989; Murray, 1997). This pattern seems to result from two different compensation processes: The monotonic increase (from 0° to 120° or 150°) appears to be caused by compensating rotations in the picture plane, whereas the fast recognition of inverted objects seems to be due to fast rotations in depth (flipping). There is evidence that participants with monotonically increasing naming latencies make use of only plane transformations, whereas participants with an M-shaped pattern use the fast flipping transformations to recognize inverted objects (Murray, 1997).

Bülthoff, 2003), in scene perception (Diwadkar & McNamara, 1997; Nakatani, Pollatsek, & Johnson, 2002), and in the perception of large, navigable spaces (Shelton & McNamara, 1997).

A dependency on rotations has been demonstrated for sequential picture–picture matching tasks (Lawson & Humphreys, 1996; Murray, 1999; for a review, see Jolicoeur & Humphrey, 1998), picture–name matching tasks (Newell & Findlay, 1997), and naming tasks (e.g., Jolicoeur, 1985, 1988; Lawson & Humphreys, 1998; Palmer et al., 1981; Srinivas, 1993). Picture–picture matching and naming tasks seem to reflect the same basic processes in object recognition (Jolicoeur & Humphrey, 1998; Lamberts, Brockdorff, & Heit, 2002). Moreover, orientation-dependent performance has been found with priming tasks (for reviews, see Jolicoeur & Humphrey, 1998; Lawson, 1999), with a visual search task (Jolicoeur, 1992), and even with figure–ground tasks (B. S. Gibson & Peterson, 1994). Overall, there is convincing evidence that recognition performance depends systematically on the amount of misorientation.

Size and Position Dependency

Recognition performance is also influenced by the size of the stimulus. The pattern of results is quite similar to orientation-dependent recognition: RTs and error rates in (sequential) picture–picture matching tasks depend on the extent of transformation that is necessary to align memory and stimulus representations. RTs increase in a monotonic way with increasing change of perceived size (e.g., Bundesen & Larsen, 1975; Bundesen, Larsen, & Farrell, 1981; K. R. Cave & Kosslyn, 1989; Jolicoeur, 1987; Larsen & Bundesen, 1978; Milliken & Jolicoeur, 1992; for a review, see Ashbridge & Perrett, 1998).

Whereas the dependency of recognition performance on orientation and size scalings is widely accepted in the literature, it is often assumed that recognition performance is invariant regarding position. However, evidence is accumulating that recognition performance is position dependent as well. Several studies have shown a systematic relation between the amount of translation and recognition performance: Increasing displacement between two sequentially presented stimuli can lead to a deterioration of performance, both for novel objects (Dill & Edelman, 2001; Dill & Fahle, 1998; Foster & Kahn, 1985; Nazir & O'Regan, 1990) and familiar objects (K. R. Cave et al., 1994). These results do not appear to merely be due to eye movements or shifts of attention nor a problem of information exchange between the two hemispheres of the brain (K. R. Cave et al., 1994; Dill & Fahle, 1998).

Some priming studies have suggested that recognition performance does not depend on size (Biederman & Cooper, 1992; Cooper, Schacter, Ballesteros, & Moore, 1992; Schacter, Cooper, & Delaney, 1990) or position (Biederman & Cooper, 1991a), because priming effects were independent of size and position changes. However these studies should be considered with caution for several reasons. The logic of these studies depends on accepting the null hypothesis (i.e., the absence of an effect), which is less convincing than the demonstration of view-dependent effects in a large number of studies. In addition, these experiments may not have had the statistical power to measure the small effects that would be expected (for a more detailed discussion, see Jolicoeur & Humphrey, 1998). Moreover, the lack of an effect in these experiments might (at least partially) be due to congruency effects (see

Section 3), as these studies did not control the way in which the perceptual scale or position was set (e.g., by the previous object; see Larsen & Bundesen, 1978, 1998, p. 728). Thus, size and position invariance may have been obtained by transformations of a perceptual reference frame and not by invariant representations. Overall, these priming studies have not provided convincing evidence for invariant recognition performance regarding size and position.

Neurophysiological Evidence for Transformation Dependency

Is the behavioral dependency on spatial transformations consistent with neurophysiological findings? Many laboratories have observed neurons in the inferotemporal (IT) cortex with highly selective responses for particular patterns and objects (for a review, see Farah, 2000, p. 89). Single-cell studies suggest that the responses of the majority of shape-selective cells in IT are orientation dependent, for faces and body parts (Hasselmo, Rolls, Baylis, & Nalwa, 1989; Perrett et al., 1985), and for objects (Logothetis, Pauls, & Poggio, 1995). The typical finding is that cells have a bell-shaped tuning curve, that is, they discharge maximally to one view of an object, and their response declines gradually as the object is rotated away from this preferred view. Few cells respond in a view-invariant manner, and these cells do not seem to be representative for neural processing in object recognition (for a review, see Logothetis & Sheinberg, 1996). Although there is consensus that the responses of IT neurons depend on orientation, there is less agreement regarding size scalings and translations. A number of researchers have claimed that the responses of IT cells depend on size but are invariant regarding translations (e.g., Ashbridge & Perrett, 1998; Perrett, Oram, & Ashbridge, 1998). It has sometimes been argued that position invariance results from the large size of receptive fields of IT neurons, especially in area TE. However, this argument is not convincing for two reasons. First, cell responses in IT depend on the position in the receptive field, with receptive field profiles resembling a two-dimensional Gaussian function (Op de Beeck & Vogels, 2000). Thus, neuronal responses are not invariant regarding the position within the receptive field. Second, receptive field sizes in IT are much smaller than originally assumed (DiCarlo & Maunsell, 2003). A closer inspection of the data reveals that responses of many IT cells vary with both size and position (e.g., DiCarlo & Maunsell, 2003; Lueschow, Miller, & Desimone, 1994; Op de Beeck & Vogels, 2000). These results would not be expected if objects' shapes are encoded with invariant representations, which do not contain spatial information about orientation, size, and position. In contrast, object position and scale information can be read out from small populations of IT neurons (Hung, Kreiman, Poggio, & DiCarlo, 2005). In accordance with these single-cell studies, monkey lesion studies have found that damage to area V4 and posterior IT affects the ability to compensate for spatial transformations more than the ability to recognize nontransformed shapes (Schiller & Lee, 1991; Weiskrantz, 1990; Weiskrantz & Saunders, 1984).

Taken together, the data from behavioral, neurophysiological, and lesion studies indicate that recognition performance depends on the spatial relation between the observer and the object, that is, recognition depends on plane and depth orientation, size, and position. At least three conclusions can be drawn from these

findings. First, object constancy is not equivalent with invariance. Although we are able to recognize objects after spatial transformations, recognition performance depends systematically on the amount of transformation. Therefore I propose that the terms *constancy* and *invariance* should not be used as synonyms. This lack of invariance is in accordance with introspective experience, because the percept is not invariant, even when object constancy is achieved. The percept of a circular disk that is rotated in depth (tilted to the viewer) is not circular but squashed. If the disk is rotated further in depth, the percept changes, although the object may still be perceived as a disk. The second conclusion is that recognition models should not be limited to modeling orientation-dependent performance but also have to account for size- and position-dependent performance. Third, the systematic dependency on spatial transformations (like rotations, size scalings, and translations) suggests that object representations are image-based or imagelike. This dependency is more easily compatible with representations that are in a similar format as the visual input, as compared with abstract representations that should—by definition—be independent of image transformations. The notion of imagelike representations corresponds with the proposal that cognitive functions are embodied, that is, are grounded in sensorimotor mechanisms (e.g., Barsalou, 1999; M. Wilson, 2002).

2. Evidence for Analogue Transformations in Object Recognition

The evidence that recognition performance is not invariant regarding spatial transformations can be accounted for by a number of different approaches. There is, however, a second class of data that provides further constraints to modeling—the finding that recognition seems to imply analogue spatial transformation processes. A transformation (e.g., rotation) is analogue if it proceeds through intermediate points along the transformational path, that is, if it is performed in a continuous or incremental way.

Analogue Transformation Processes in Mental Imagery and in Object Recognition

In a seminal study by Shepard and Metzler (1971), participants had to judge whether two simultaneously presented objects were identical or mirror images. RTs increased linearly with angular disparity between the two objects both in the picture plane and in depth. The authors interpreted these results as evidence for an internal rotation process that they dubbed *mental rotation*. A large number of subsequent studies confirmed the original findings (for reviews, see Finke, 1989; Kosslyn, 1994; Shepard & Cooper, 1982). Similar results were also found for size scalings (e.g., Bundesen & Larsen, 1975) and for mental translations (Bennett, 2002; Larsen & Bundesen, 1998). One of the central questions of the so-called imagery debate was whether linear or monotonic increases of RTs in mental imagery tasks actually reflect an analogue mental rotation, or whether visual representations are propositional and relatively abstract (e.g., Anderson, 1978; Kosslyn, 1981, 1994; Pylyshyn, 1981, 2002). Cooper (1976; see also Cooper & Shepard, 1973, Experiment 2) developed a paradigm that allowed her to test whether rotations in mental imagery were analogue processes or not, that is, whether they passed through intermediate points along the transformational path. Participants

saw random polygons that they had to mentally rotate as soon as the pattern was removed. Some time after stimulus presentation, a test pattern was presented. Participants had to indicate whether the pattern was a normal image or a mirror-image of the starting pattern. The orientation of the test pattern was selected according to participants' individual rate of mental rotation, which was measured in a previous experiment. When the test shape was presented in the expected orientation (based on the participant's normal rate of rotation), RTs were short and constant. RTs increased linearly with increasing departure of the test pattern orientation from where the mentally rotated pattern should have been at that time. This was also true when the test pattern was presented in unfamiliar orientations that had not been shown to the participant before. These results indicate that mental rotations in short-term memory (STM) pass through at least some intermediate points along the rotational path.

This evidence for the analogue nature of imagery transformations does not necessarily transfer to object recognition. The systematic relation between the amount of transformation and recognition performance could be due to any time-consuming and error-prone process (see Perrett et al., 1998). Thus, more direct evidence is necessary to confirm that analogue transformation processes are involved in object recognition. Interestingly, there is evidence for the analogue nature of rotations and size scalings in object recognition, based on the logic of sequential additivity (Bundesen, Larsen, & Farrell, 1981; see also Sternberg, 1998). In each trial of the experiment, two familiar objects (alphanumeric characters) were presented successively and participants had to decide as quickly as possible whether the two stimuli were identical except for size and orientation in the picture plane. If a transformation is analogue, then the time to pass through a given path of transformation can be predicted by the sum of the times that are required to traverse the segments that make up that path: $t_{AC} = t_{AB} + t_{BC}$ (see Figure 1).

In other words, to show that the time for a transformation from A to C is an additive combination of the transformation times from A to B and from B to C provides evidence that the process of transforming from A to C passes through the intermediate point B. Sequential additivity of RTs therefore suggests analogue transformations.[2] The results of Bundesen et al. (1981) confirmed the prediction of additivity for rotations in the picture plane, and also for combinations of rotations and size scalings, providing evidence for analogue rotation and scaling processes in object recognition. The importance of these findings is increased by the fact that sequential additivity for rotations could be demonstrated even though the RT function was nonlinear. In general, it took more time to traverse a sector when the image was farther from upright, and when the direction of rotation was away from upright (see Sternberg, 1998, p. 783). Nonetheless, sequential additivity was found despite these nonlinearities in the RT function.

Kourtzi and Shiffrar (2001) provided further evidence suggesting that analogue transformations are involved in object recognition. They investigated the perception of objects that deform as

---

[2] Notice that an analogue transformation leads to sequential additivity only when the time to traverse any particular sector is the same, regardless of the other sectors with which its traversal is concatenated (Sternberg, 1998, p. 781).

Figure 1. Sequential additivity of transformation times: $t_{AC} = t_{AB} + t_{BC}$. Sequential additivity means that the time that is required to traverse a certain transformational distance (t_{AC}) is equal to the sum of the times that are necessary to traverse its subsections ($t_{AB} + t_{BC}$).

they rotate (i.e., that were bent), using a priming paradigm. Two primes (a normal and a deformed object) were presented successively, and after a short blank interval two target images appeared simultaneously on the screen. The participants' task was to press a key when both targets were physically identical. In Experiments 3 and 4, priming was found for targets at an intermediate orientation and an intermediate level of deformation relative to the two primes. This is consistent with the hypothesis that object perception involves an analogue spatial remapping, even when the objects differ both by a rotation and an elastic deformation.

Studies showing an advantage of ordered versus scrambled sequences in the recognition of rotating objects (Lawson et al., 1994; Vuong & Tarr, 2004) are also suggestive of analogue updating processes in object recognition. Recognition performance was better when several images were sequentially presented in an ordered rotational sequence (which corresponds to a physical rotation) compared with a randomly ordered sequence of frames. Thus, recognition of dynamic objects seems to be facilitated when the visual presentation corresponds to an analogue transformational sequence.

It should be noted that there are also findings suggesting orientation functions that are not monotonic, showing benefits for orientations matching the principal axes (90°, 180°, 270°; Lawson & Jolicoeur, 1999). These results cannot be fully accounted for by analogue transformations but may require additional processes.

In the attention literature, the notion of an attentional spotlight, which can be shifted in location, is an important metaphor. Shulman, Remington, and McLean (1979) and Tsal (1983) have tried to document the analogue nature of the shifting of spatial attention from one location to another. However, a number of criticisms have been raised about these studies (Eriksen & Murphy, 1987; Remington & Pierce, 1984; Sperling & Weichselgartner, 1995; Yantis, 1988; for a review, see K. R. Cave & Bichot, 1999). Thus, the evidence for analogue location shifts of an attentional spotlight is rather weak.

Neurophysiological Findings

Are neurophysiological findings consistent with the psychophysical evidence for analogue transformation processes in object recognition? Electrophysiological evidence for analogue visuomotor transformation processes was found in the motor cortex. Monkeys' mental rotation from the initial to the final direction of movement corresponded to a continuous rotation of a neural population vector that represents the intended direction of movement (Georgopoulos, 2000; Georgopoulos, Lurito, Petrides, Schwartz, & Massey, 1989; Lurito, Georgakopoulos, & Georgopoulos, 1991; Pellizzer & Georgopoulos, 1993). It is not clear whether these results can be transferred to the visual cortex, because the distri-

bution of orientation-tuned neurons is inhomogeneous in the superior temporal sulcus. More cells were optimally tuned to canonical views of the head, like full face or profile, than to other views (Perrett et al., 1991). Statistical methods, which are less susceptible to inhomogeneities in view tuning, may be better suited under these circumstances (e.g., Oram, Földiák, Perrett, & Sengpiel, 1998; Sanger, 1996). However, inhomogenities do not exclude analogue transformation processes in object recognition. An optical imaging study is in accordance with analogue transformation processes in recognition: Wang, Tanifuji, and Tanaka (1998) first determined the critical features for the activation of neurons with single-cell recordings. With subsequent optical imaging techniques, it was demonstrated that these critical features evoked dark spots on the cortex approximately 0.5 mm in diameter. Some spots were specifically activated by faces. The positions of the activation spots changed gradually along the cortical surface as the stimulus face was rotated in depth. This finding was interpreted as evidence that the orientation of objects is continuously mapped and is consistent with analogue transformations occurring in object recognition.

Thus, there is both behavioral and neurophysiological evidence that object recognition involves analogue transformation processes, although the evidence for analogue transformation processes is not as strong as the evidence for orientation dependency. However, the notion of analogue time-consuming transformation processes provides a parsimonious account for the systematic dependency with increasing amount of transformation (see Section 1). In contrast, one-step models of recognition (which do not involve intermediate steps in the recognition process) require additional assumptions to explain this systematic dependency.

3. Congruency Effects in Object Recognition and the Adjustment of Reference Frames

The third relevant class of findings comprises reference frame effects in object recognition. Reference frames, which are a means of specifying locations in space, were investigated in cognitive psychology over a long period of time (e.g., Larsen & Bundesen, 1978; Marr & Nishihara, 1978; Rock, 1973, 1974; Rock & Heimer, 1957; for reviews, see Farah, 2000; Jolicoeur & Humphrey, 1998; Palmer, 1999). A reference frame can be regarded as a specific coordinate system. Reference frames have an origin in a certain location in space, are often conceptualized as orthogonal grids, and usually involve the notion of axes, which correspond to particular directions in space (e.g., Farah, 2000, pp. 71–73, 107–109; Jolicoeur, 1990b; Jolicoeur & Humphrey, 1998; Palmer, 1999, pp. 370–377).

Many different types of reference frames have been proposed that differ along a number of dimensions (for a review, see Jolicoeur & Humphrey, 1998). One important distinction is whether object recognition is based on a viewer-centered or object-centered reference system (Marr & Nishihara, 1978). These frames differ in the location of the origin of the coordinate system: In a viewer-centered system the origin is located on (or in) the viewer, whereas in an object-centered system the origin is located on (or in) the viewed object. A viewer-centered frame may be retinotopic, head-, trunk-, or even hand-centered. Viewer-centered reference frames imply orientation-dependent performance, whereas object-centered reference frames predict that recognition performance is

not influenced by the spatial relation between observer and object, because the reference frame is already centered on the object's intrinsic axes. The findings reviewed in Section 1 indicate that recognition performance deteriorates systematically with increasing amounts of spatial transformation. This is clear evidence for viewer-centered reference frames.

Orientation Congruency Effects in Object Recognition

Evidence for a special role of reference frames in object recognition was supplied by experiments that demonstrated a generic (i.e., not shape-specific) orientation congruency effect, which suggests that recognition involves the adjustment of a perceptual coordinate system. Participants' ability to identify a misoriented stimulus is facilitated if it is preceded by a different stimulus shown at the same orientation. An orientation congruency effect has been found for alphanumeric stimuli (Jolicoeur, 1990b, 1992), for novel objects (Gauthier & Tarr, 1997; Tarr & Gauthier, 1998), and also for common familiar objects (Graf, Kaping, & Bülthoff, 2005). In this latter study, participants had to name two briefly and sequentially displayed objects followed immediately by a pattern mask. The objects were presented either in congruent or incongruent orientations. Recognition accuracy was more than 10 to 15 percentage points higher for congruent orientations, indicating a strong orientation congruency effect. This suggests that recognition involves the adjustment of a perceptual coordinate system. However, there remained a significant effect of orientation in congruent trials, so there was no full compensation for orientation effects.

Previous studies with novel objects have suggested that congruency effects were limited to similar objects (Gauthier & Tarr, 1997; Tarr & Gauthier, 1998) and therefore can be accounted for by class-based processing (Moses & Ullman, 1998), without needing to assume an abstract coordinate system. However, Graf et al. (2005) found congruency effects for dissimilar objects, which would not be predicted by class-based processes. In Experiment 1, Graf et al. (2005) found congruency effects when the two objects were from different basic level categories and even from different superordinate level categories (i.e., a biological and a human-made object; see Figure 2A). The congruency effect was not shape specific, as objects from different superordinate categories (and usually also from different basic level categories) tend to have different shapes (Rosch, Mervis, Gray, Johnson, & Boyes-Braem, 1976). In Experiment 2, congruency effects were found when one object had a horizontal main axis of elongation, while the other object had a vertical main axis (see Figure 2B). Thus, congruency effects were not shape specific and could not be reduced to a priming of the main axis of elongation.

These findings are important in at least two ways. First, they demonstrate congruency effects for common objects and thus suggest that the processes that underlie orientation congruency play a role also in default object recognition. Second, the results provide evidence that a rather abstract (i.e., not shape-specific) frame of reference is adjusted during recognition. This is consistent with the finding that congruency effects transfer from a task that requires the recognition of alphanumeric stimuli to a symmetry detection task for dot pattern stimuli (Pashler, 1990). Overall, the orientation congruency effect is a robust effect. It was found with naming paradigms (Graf et al., 2005; Jolicoeur, 1990b; Gau-
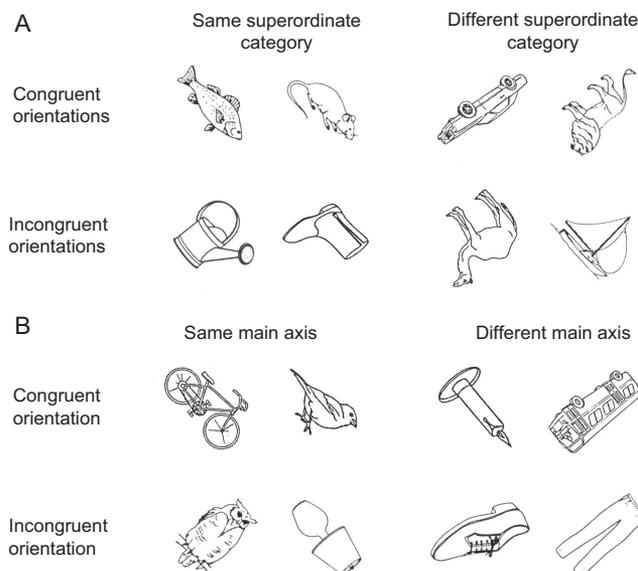
Figure 2. Example displays used in Graf et al. (2005) in order to investigate orientation congruency effects. Objects were presented sequentially in congruent or in incongruent orientations. In both experiments, recognition accuracy was higher when the objects had congruent orientations. A. In Experiment 1, congruency effects were found for objects from the same and from different superordinate level categories. B. In Experiment 2, congruency effects were found when objects had the same and when they had different main axes of elongation. Note. Objects from "A standardized set of 260 pictures: Norms for name agreement, image agreement, familiarity, and visual complexity," by J. G. Snodgrass and M. Vanderwart, 1980, *Journal of Experimental Psychology: Human Learning and Memory*, *6*, 174–215. Copyright, 2000 by Life Science Associates. Adapted with permission.

thier & Tarr, 1997; Tarr & Gauthier, 1998) and with a visual search paradigm (Jolicoeur, 1992), using different dependent measures and a variety of different types of stimuli.

The most parsimonious interpretation of the orientation congruency effect is that in tasks requiring the identification of misoriented patterns, the visual system adjusts the orientation of a perceptual frame of reference by means of a frame rotation process (Graf et al., 2005; Jolicoeur, 1990b). Assuming that the frame can be rotated at a finite rate, a rotation through a larger angle takes more time than a rotation through a smaller angle. In general, the identification of a pattern is achieved by rotating the frame to the orientation of the pattern. A second pattern presented at this orientation can be more readily identified because no further correction for misorientation is necessary. This *frame rotation hypothesis* is also consistent with head-tilt studies that suggest that reference frames can be adjusted or rotated in an analogue way (Corballis, Nagourney, Shetzer, & Stefanatos, 1978; McMullen & Jolicoeur, 1990; see also Jolicoeur & Humphrey, 1998).

The findings of Graf et al. (2005) and Jolicoeur (1990b, 1992) do not rule out two alternative explanations. First, several orientation-dependent frames (with different orientations) may exist in parallel and compete for activation. Perceptual identification may be achieved when one frame becomes dominant over the others (e.g., Hinton, 1981; for a more detailed discussion, see

Jolicoeur, 1990b). This approach again involves reference frames but does not include analogue frame transformations. However, this proposal cannot explain the findings that suggest analogue rotation processes in object recognition (see section 2). Second, identification may be achieved by mentally rotating the stimulus representation until it is upright. In order to account for the orientation congruency effect, it must further be assumed that the rotation process can be facilitated or primed by a prior rotation in the same direction and through the same or a similar angle. This explanation also has to assume that the rotation mechanism is not shape specific, because the orientation congruency effect is not limited to identical stimuli. This second explanation lacks parsimony, and it is not compatible with several studies indicating that recognition does not involve mental rotation (see the subsection *The Relation Between Object Recognition and Mental Rotation* in Section 5).

Size and Position Congruency Effects

Evidence for frame transformations in object recognition was demonstrated also for size-scaling transformations. In their Experiment 2, Larsen and Bundesen (1978) investigated the role of frame transformations relating to long-term memory (LTM) in a recognition task. Uppercase letters were used as stimuli, in four different sizes. In every trial one stimulus was presented. The task was to decide as quickly as possible whether the stimulus was an upright letter or not. The trials were arranged so that the same size format was repeated with a first-order probability of 0.75. Participants were informed about the statistical properties of the stimulus sequence and could build up an expectation about the size of the next stimulus. Both introspective reports and participants' performance indicated that participants perceptually prepared for the expected (cued) size format. RTs were fastest when the stimulus was in the expected size format, increasing monotonically with increasing size divergence. These results cannot be due to shape-specific image transformations, because the same letter was never used in two successive presentations. Larsen and Bundesen (1978) interpreted these size adjustments as coordinate transformation processes in LTM, as the stimulus had to be compared with the representation of a familiar object in LTM. The speed of frame transformations was higher than the rate of image transformations in STM, as measured in simultaneous matching studies (Bundesen & Larsen, 1975) and sequential matching studies with novel objects (Larsen & Bundesen, 1978, Experiment 1). Larsen and Bundesen (1978) argued that two size adjustment processes have to be differentiated. One is a relatively fast frame adjustment process, which refers to representations in LTM. The other is a relatively slow image transformation in STM, which is shape specific.

Larsen and Bundesen's (1978) study confirmed that frame transformations are involved in object recognition. In the object recognition task (Experiment 2), RTs were fastest for the expected size, which suggests that a frame of reference was preadjusted. This finding can only be explained by a frame transformation process, and not by an image transformation. First, only a frame transformation can logically take place before the stimulus has been presented, not an image transformation. Second, only frame transformations are generic in nature, that is, are not stimulus (or shape) specific. Larsen and Bundesen's (1978) results suggest that object

recognition relies on transforming a perceptual reference frame (coordinate transformation) and not on image transformations.

Congruency effects were also found regarding location transformations. There is a large body of experimental evidence on this issue, although it is usually described in terms of shifting the focus of spatial attention rather than transforming the location of a reference frame. Processing one stimulus at a location makes it easier to process another stimulus at that location (K. R. Cave & Pashler, 1995), and there are numerous spatial cuing studies demonstrating that a stimulus can be processed more quickly when its location is known in advance (e.g., Eriksen & Hoffman, 1974; Posner, Snyder, & Davidson, 1980). When a location cue tells a participant to expect a stimulus at a location, the RT for detecting the stimulus generally increases with the distance between the expected and the actual location (Downing & Pinker, 1985), but only under some circumstances (Hughes & Zimba, 1985, 1987; Zimba & Hughes, 1987). The distance effect could be interpreted as the time necessary to shift the location in a reference frame, although LaBerge and Brown (1989) argued for a different explanation based on an attentional gradient (for a review, see K. R. Cave & Bichot, 1999).[3]

Frame Transformations Versus Image Transformations

The distinction between frame transformations and shape-specific image transformations regarding object size was corroborated in a single experimental task. In Larsen and Bundesen's (1978) Experiment 3 also stimulus repetitions could occur. Participants were instructed to decide as rapidly as possible whether the stimulus letter belonged to a set of letters that was defined at the beginning of each block. Performance of both stimulus-repetition and stimulus-nonrepetition trials were fastest for the expected (cued) size and monotonically increased with increasing size disparity. The slopes of the RT functions were different for stimulus repetitions and nonrepetitions. The results of stimulus-nonrepetition trials were similar to Experiment 2, with relatively high speeds of transformation, indicating frame transformations. In contrast, even though in total stimulus-repetition trials were faster than nonrepetition trials (because of a lower *y*-intercept), transformation processes in stimulus-repetition trials were slower than in nonrepetition trials, suggesting image transformations in repetition trials. K. R. Cave and Kosslyn (1989) confirmed these results, using geometrical stimuli. Again, different rates of size transformation were found, with fast size transformation processes for stimulus nonrepetitions and slow transformation processes for stimulus repetitions. The effects of stimulus repetition versus stimulus nonrepetition and the effect of size ratio interacted, suggesting that different size scalings were used in the two conditions. Overall, these studies demonstrated the existence of two different size-scaling processes: a generic (i.e., not shape or stimulus specific) and fast frame scaling process for object recognition—relating to LTM representations—and a stimulus-specific and relatively slow image scaling process for mental imagery size scalings in STM.

Are there also two different adjustment processes regarding orientation? As described earlier, object recognition seems to

---

[3] I thank Kyle Cave for bringing this literature to my attention.

imply frame rotation processes, because orientation congruency effects are not limited to similar shapes (Graf et al., 2005; Jolicoeur, 1990b, 1992). Mental imagery rotations, on the other hand, are typically shape specific, that is, are not frame transformations but image transformations (Cooper & Shepard, 1973; Koriat & Norman, 1984, 1988, 1989; Shepard & Hurwitz, 1984; but see Robertson, Palmer, & Gomez, 1987).[4,5] This distinction is similar to the proposal of two different types of compensation processes regarding orientation that operate at different rates (Simion, Bagnara, Roncato, & Umiltà, 1982): The authors proposed that a slow process operates on mental images (image transformation), while a second—faster but presumably also analogue—process operates directly on visual input (visual code). The latter seems to coincide with frame transformations. Confirming this conception of slow image rotations and fast frame rotations, hypothetical rates of rotation in recognition tasks are typically faster than in mental rotation tasks (e.g., Jolicoeur, 1988; Shepard & Metzler, 1971; Tarr, 1995; Tarr & Pinker, 1989; see also Perrett et al., 1998, pp. 113–114). A simple conclusion is that object recognition involves a relatively fast frame rotation process (coordinate transformations), whereas mental imagery typically relies on a relatively slow and image-specific image rotation process (Larsen & Bundesen, 1998).

It remains an open question whether a similar discrimination between two translation processes can be found concerning position. Kosslyn (1994) postulated two different adjustment processes for position, but there does not seem to be any research that has directly investigated this issue. There is some evidence that the rate of translation in sequential matching tasks is faster for familiar objects (K. R. Cave et al., 1994) than for novel objects (Dill & Fahle, 1998), which is in accordance with the conception of different translation processes in LTM and STM.

The claim that object recognition involves frame transformations does not, however, have to imply that symbolic advance information about orientation (without information about object identity) is sufficient in order to compensate the effects of misorientation. Symbolic cues, like an arrow indicating the expected orientation, have not provided very effective facilitation in naming (recognition) tasks (Gauthier & Tarr, 1997; McMullen, Hamm, & Jolicoeur, 1995; Palmer et al., 1981, Experiment 2) and mental rotation tasks when no additional information about stimulus identity was given (Cooper & Shepard, 1973). It seems that the adjustment of a frame is bound to the presentation of an external stimulus in the corresponding orientation or size (see K. R. Cave & Kosslyn, 1989; Koriat & Norman, 1988, Experiment 4; Larsen & Bundesen, 1978; Robertson et al., 1987) or the previous presentation of a background that provides depth information that may supply a visual reference frame (Humphrey & Jolicoeur, 1993).

Overall, there is converging behavioral evidence for two different compensation processes, both for rotations and size scalings: a relatively fast and generic frame transformation process, related to LTM representations, and a relatively slow and shape-specific image transformation process in STM. Accordingly, generic (i.e., not shape-specific) orientation and size congruency effects have been found in object recognition, whereas mental imagery transformations seem to be shape specific. The faster speed of frame transformations as compared with image transformations fits with performance in recognition and imagery tasks. Orientation and size

adjustment processes in object recognition are faster than in mental imagery transformations (e.g., K. R. Cave & Kosslyn, 1989; Larsen & Bundesen, 1978; Simion et al., 1982).

Neurocomputational and Neurophysiological Evidence

Independent from psychophysical evidence for the adjustment of a perceptual coordinate system in object recognition, researchers in computational neuroscience proposed that object perception and recognition rely on coordinate transformations (Olshausen, Anderson, & Van Essen, 1993, 1995; Salinas & Abbott, 1997a, 1997b, 2001; Salinas & Sejnowski, 2001; Salinas & Thier, 2000). One starting point for this work was research on visuomotor control (e.g., reaching, grasping, and eye movements), for which coordinate transformations are crucial. For instance, if a person wants to grasp an object, a coordinate transformation has to be performed, because eyes and hands rely on different coordinate systems. Visual information coded in retinal coordinates has to be transformed into hand-centered coordinates. Eye movements are another important example. Dynamic updating processes are necessary to cope with the constant changes of eye-centered coordinates relative to head-centered, body-centered, or world-centered coordinates (Duhamel, Colby, & Goldberg, 1992; Mays & Sparks, 1980). In general, visuomotor control requires coordinate transformations (for reviews, see Andersen, Batista, Snyder, Buneo, & Cohen, 2000; Colby, 1998; Salinas & Sejnowski, 2001; Salinas & Thier, 2000; Snyder, 2000). Gain (amplitude) modulation, which is implemented ubiquitously in the visual cortex, provides an efficient solution to the coordinate transformation problem (e.g., Salinas & Abbott, 1995; Zipser & Andersen, 1988; see also Salinas & Thier, 2000). Gain-modulated neurons are ideally suited to perform computations that are fundamental for coordinate transformations (e.g., Pouget, Deneve, & Duhamel, 2002; Pouget & Sejnowski, 1997; Salinas & Abbott, 1995; Salinas & Sejnowski, 2001).

In order to explain how coordinate transformations can be implemented by gain modulation, I use the example of a transformation from eye-centered to head-centered coordinates. Imagine a neuron in the parietal cortex that responds to a spot of light within its visual receptive field and codes the position of the stimulus in retinal coordinates. To factor out the effects of eye movements, extraretinal information about eye position has to be included. Most neurons in the lateral and medial parietal areas and in area 7 respond to the retinal stimulation and are also sensitive to the position of the eyes in the orbit (Andersen, Bracewell, Barash, Gnadt, & Fogassi, 1990; Andersen & Mountcastle, 1983; Andersen, Snyder, Bradley, & Xing, 1997). Eye position modulates the amplitude of visual responses, whereas the shape and position of the receptive field in retinotopic coordinates are unaffected by eye position. Typically, the gain of the sensory response

---

[4] In some experiments on frame effects in mental imagery (that involved LTM representations), frame effects were found (Koriat & Norman, 1988, Experiment 4; Robertson et al., 1987).

[5] Reference frames affected performance in mental imagery experiments in which participants' had their head tilted, but these studies investigated the existence of environmental frames (e.g., Corballis, Zbrodoff, & Roldan, 1976; Corballis, Zbrodoff, Shetzer, & Butler, 1978; McMullen & Jolicoeur, 1990).

increases monotonically as the eye moves along a particular direction in space, corresponding to linear or sigmoidal gain fields (Andersen, Essick, & Siegel, 1985).

The response of gain-modulated neurons can be modeled by a multiplication of the sensory response and the eye position signal (see Figure 3). The interaction between the retinal and the extraretinal (eye position) signals does not have to be exactly multiplicative but simply nonlinear (Pouget & Sejnowski, 1997; Salinas & Abbott, 1997b)—although the responses of gain-modulated neurons can usually be described well by a multiplication (e.g., McAdams & Maunsell, 1999; Treue & Martinez Trujillo, 1999). Several neural mechanisms have been proposed that may underlie this type of nonlinear multiplicative interaction (for a review, see Salinas & Sejnowski, 2001).

The important point for our concerns is that gain-modulated responses at one level of the cortical hierarchy correspond to a coordinate transformation at a higher level. For instance, although gain-modulated responses are still coded in retinotopic coordinates, responses of neurons at the next cortical level can be head centered (Salinas & Abbott, 1995; Zipser & Andersen, 1988; see also Salinas & Abbott, 2001). The following basic principle can be derived: The presence of gain modulation at one stage of a processing pathway suggests that responses at a downstream stage will be in a different coordinate system (Salinas & Abbott, 2001; Salinas & Thier, 2000). In other words, gain-modulated responses
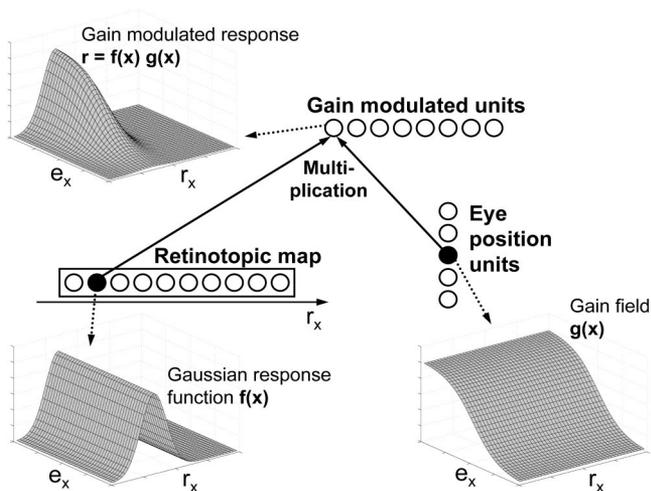
Figure 3. Gain modulation as multiplication. The graph in the lower left shows the Gaussian response function $f(x)$ of a parietal neuron that encodes information in retinotopic coordinates (r_x) and is independent of eye position (e_x). In order to achieve coding in head-centered coordinates, eye position has to be taken into consideration. Eye position information can be described by a gain field $g(x)$, shown in the lower right. The response function of gain-modulated neurons, which are common in the visual cortex, can be described by a multiplicative (nonlinear) interaction. These neurons still code in retinal coordinates (r_x), but the amplitude (gain) of the response depends on eye position (top left graph). Gain-modulated neurons provide the neuronal basis to perform coordinate transformations, such as, in this case, from eye-centered to head-centered coordinates (Pouget & Sejnowski, 1997). *Note.* From "Spatial transformations in the parietal cortex using basis functions," by A. Pouget and T. J. Sejnowski, 1997, *The Journal of Cognitive Neuroscience, 9,* p. 226. Copyright 1997 by MIT Press. Adapted with permission.

at one level of the cortex correspond to coordinate transformations, implemented as transformations of receptive fields at the next cortical level (see Figure 4). A number of studies have provided evidence for dynamic transformations of receptive fields in several brain areas (e.g., Graziano, Hu, & Gross, 1997; Graziano, Yap, & Gross, 1994; Jay & Sparks, 1984; Stricanne, Andersen, & Mazzoni, 1996; Wörgötter et al., 1998; Wöhrgötter & Eysel, 2000), including IT (Rolls, Aggelopoulos, & Zheng, 2003), which fits nicely with this approach.

On the basis of gain-modulated neural responses, several different (e.g., eye-centered and head-centered) coordinate systems can be spanned concurrently in downstream areas (Pouget & Sejnowski, 1997, 2001; Salinas & Abbott, 1995; Salinas & Sejnowski, 2001). Moreover, intermediate coordinate systems can be created, which, for instance, code information in coordinates in between eye- and head-centered frames, corresponding to partially shifting receptive fields of downstream neurons (Pouget et al., 2002). In accordance with intermediate coordinate systems, there is neurophysiological evidence for partially shifting receptive fields (e.g., Cohen & Andersen, 2000; Duhamel, Bremmer, Ben-Hamed, & Graf, 1997; Stricanne et al., 1996).

Gain-modulated neural responses have been found in parietal cortex (Andersen & Mountcastle, 1983; Andersen et al., 1985, 1990, 2000; Batista, Buneo, Snyder, & Andersen, 1999; Cohen & Andersen, 2002) and premotor cortex (Graziano et al., 1994, 1997; Jouffrais & Boussaoud, 1999), in V1 (Trotter & Celebrini, 1999), V3 (Galletti & Battaglini, 1989), and medial superior temporal cortex (Shenoy, Bradley, & Andersen, 1999; Treue & Martinez Trujillo, 1999). Eye-position-dependent gain field modulation has also been demonstrated in the ventral stream in V4 (Bremmer, 2000), which is primarily involved in object recognition and object perception (Milner & Goodale, 1995). Moreover, gaze-dependent gain modulation has significant influences on visual perception. Gaze direction modulates the magnitude of the motion aftereffect, the tilt aftereffect, and the size aftereffect (Nishida, Motoyoshi, Andersen, & Shimojo, 2003).

In two neurocomputational approaches, the notion of coordinate transformations by gain modulation was extended to account for object recognition (Salinas & Abbott, 1997a, 1997b; see also Olshausen et al., 1993) and object perception (Deneve & Pouget, 2003; Pouget & Sejnowski, 1997, 2001). The first approach proposes that differences between the memory representation and the input representation can be compensated for by attentional gain modulation (Salinas & Abbott, 1997a, 1997b), or dynamic routing (Olshausen et al., 1993, 1995). The basic idea is that attentional modulation of neural responses leads to a transformation from retinotopic to attention-based coordinates. For instance, changes in position can be compensated for by shifting the focus of attention to the position of the stimulus. As attention can be shifted independent of the fixation position of the eye, an object can be recognized in different positions, even when it is not fixated. Compensation is achieved by gain-modulated responses, using attention as an extraretinal modulatory signal (instead of eye position as in visuomotor control). This hypothesis has been supported by the demonstration of attentional gain field modulation in the ventral stream. Neurons in V4 have gain fields that are functions of the currently attended location (Connor, Gallant, Preddie, & Van Essen, 1996; Connor, Preddie, Gallant, & Van Essen, 1997). On the basis of these gain-modulated V4 neurons, coordi-
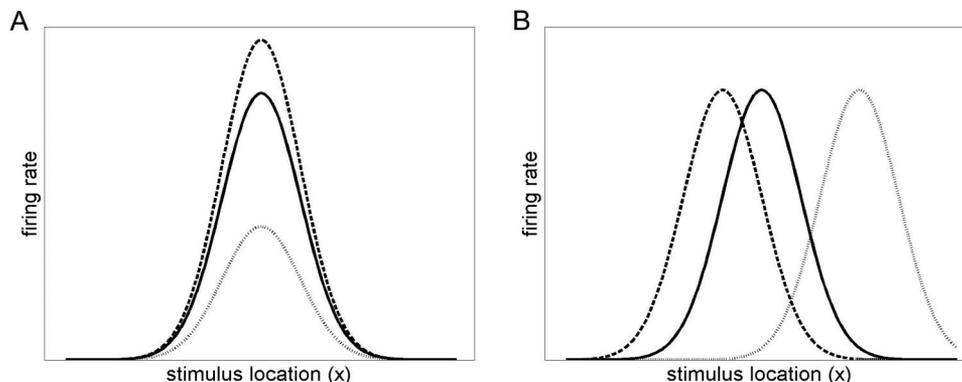
Figure 4. Gain-modulated neurons at one cortical level (A) correspond to transformed (e.g., shifted) receptive fields at a downstream area of processing (B). Thus, gain modulation at one level conforms to a coordinate transformation at the next level (Salinas & Abbott, 2001). The different lines represent Gaussian tuning functions of two exemplary neurons (A and B). In (A), the three tuning functions of the same neuron are gain-modulated, that is, have different amplitudes, while in (B) the tuning functions (and thus the receptive fields) are shifted. This conforms to a shifting of the coordinate system (coordinate transformation) that compensates for a gaze shift, or for the change of an object's position. The gain-modulated Gaussian response functions (in A) correspond to different sections through the gain-modulated response function in Figure 3. Figure 4A and 4B are from "Coordinate transformations in the visual system: How to generate gain fields and what to compute with them," by E. Salinas and L. F. Abbott, 2001. In M.A.L. Nicolelis (Ed.), *Advances in neural population coding: Progress in brain research* (Vol. 130, p. 180), Amsterdam: Elsevier. Copyright 2001 by Elsevier. Adapted with permission.

nate transformations can be performed downstream in IT, by shifting IT receptive fields so that they are centered at the point where attention is directed (Salinas & Abbott, 1997a, 1997b; see also Salinas & Abbott, 2001). As a consequence, differences in position are compensated for by coordinate transformations, and objects can be recognized more or less independent of position. Similarly, coordinate transformations can also account for how we recognize objects at different sizes, depending on viewing distance (Salinas & Abbott, 1997b; Salinas & Sejnowski, 2001). IT receptive fields of variable, attention-controlled spatial scales are obtained when the mechanism is extended to scale dependent attentional gain fields in V4. V4 neurons have been found to be tuned to images of specific sizes and have gain fields that depend on viewing distance (Dobbins, Jeo, Fiser, & Allman, 1998). According to the computational principles of gain modulation, these size-dependent gain fields should correspond with the scaling of receptive fields in IT.

The prominent role of attention in these gain modulation models is consistent with behavioral findings showing that attention plays an important role in object recognition (Mack & Rock, 1998; Thoma et al., 2004). Note that the attentional modulation does not necessarily imply conscious and controlled processes, because the attentional processes in object recognition may be highly automatized (Salinas & Abbott, 2001).

To recapitulate, according to this approach, object recognition is achieved by coordinate transformation processes. Attentional gain modulation in V4 can lead to the transformation of receptive fields of IT neurons, compensating for position and size changes (Salinas & Abbott, 1997a, 1997b; Olshausen et al., 1993, 1995). These models were designed to compensate for translations and size scalings but can be extended to orientation changes, based on orientation-modulated neurons in V4 and corresponding rotating

receptive fields in IT. Accordingly, there is evidence that orientation tuning functions in V4 are gain modulated by attention, and the effects of attention are consistent with a multiplicative scaling (McAdams & Maunsell, 1999; for further evidence for orientation-dependent gain modulation, see Sabes, Breznen, & Andersen, 2002).

The second extension of the gain modulation approach to object perception has been proposed by Alex Pouget and collaborators (Deneve & Pouget, 2003; Pouget et al., 2002; Pouget & Sejnowski, 1997, 2001). Their gain modulation proposal accounts for several physiological and neuropsychological findings related to object orientation, which were previously regarded as evidence for object-based representations. Pouget and colleagues have successfully modeled hemineglect (Pouget & Sejnowski, 2001) and single-cell data related to eye movement control in object perception (Deneve & Pouget, 2003). Instead of using explicit object-centered representations, their approach relies on retinotopic receptive fields modulated by the orientation of the object (Deneve & Pouget, 2003). This type of orientation-dependent modulation has been confirmed by neurophysiological findings (Sabes et al., 2002). In contrast to the approach by Salinas and collaborators, Pouget and coworkers implemented coordinate transformations in parietal and frontal cortex (Deneve & Pouget, 2003), as they investigated object perception in relation to motor control.

In conclusion, there is converging evidence that transformations of perceptual coordinate systems are involved in object recognition from psychophysics, neurophysiology, and computational neuroscience. Several researchers have proposed that object recognition and perception can be modeled neurocomputationally on the basis of coordinate transformations, implemented by gain modulation. These processes may provide the neural basis for the relatively fast

coordinate transformations in object recognition that have been observed in the behavioral studies.

4. Do the Existing Recognition Models Explain the Findings?

A number of approaches have been proposed to explain how the visual system can recognize objects after spatial transformations (for a review, see Ullman, 1996). In the following the most influential models are presented, and their ability to account for the three classes of findings is examined.

Models Based on Abstract (View-Independent) Representations

In view-independent models, the ability to recognize transformed objects is explained on the basis of properties that are invariant to rotations (or to other spatial transformations, like size and position changes). Models of this type differ mainly in the way this spatial invariance is derived. In invariant-property models, formless mathematical properties are defined that are invariant to certain spatial transformations (e.g., Cassirer, 1944; J. J. Gibson, 1950; Pitts & McCulloch, 1947; Van Gool, Moons, Pauwels, & Wagemans, 1994; Wagemans, Van Gool, & Lamote, 1996). Examples for invariant properties are the aspect ratio or the cross ratio. Another influential approach has been to account for recognition on the basis of a decomposition of patterns into a set of transformation-invariant elementary features. The pandemonium model is a well-known example of this model type (Lindsay & Norman, 1972; Selfridge & Neisser, 1963). Most prominently, in structural description models, invariance is derived from a decomposition into object parts and their spatial relations. Objects are represented in terms of their parts (described by geometrical primitives) and the relations between the parts, which are invariant regarding most spatial transformations (e.g., Biederman, 1987; Hummel & Biederman, 1992; Marr & Nishihara, 1978; Sutherland, 1968). As all view-independent models aim at invariance regarding spatial transformations, they cannot account even for the first major class of findings. There are minor exceptions, for instance, the model of Hummel and Biederman may explain effects of rotations in the picture plane. Attempts have been made to reconcile structural description models (Bar, 2001) and invariant-property models (Wagemans et al., 1996) with the lack of invariance in behavioral studies. However, these modified approaches still cannot account for sequential additivity (see Larsen & Bundesen, 1998). Furthermore, none of the view-independent models predict congruency effects, as they rely on invariant representations, or on object-centered reference frames. In general, view-independent models fail to accommodate the second and third classes of findings.

Models With Image-Based (View-Dependent) Representations

Image-based models were developed in order to account for view-dependent recognition performance and thus are well-suited to explain effects of object rotations. However, these models are often limited to orientation effects and do not explain effects of size and position, although they might be extended to encompass

these results. Do these models account for the second and third class of findings?

Alignment models. One of the first models that was developed to explain orientation- and size-dependent recognition performance was the alignment model. The alignment approach proposes that recognition is achieved by spatial transformations that align input and memory representations. As alignment processes are usually assumed to be based on mental rotations (e.g., Jolicoeur, 1985, 1990a; Tarr & Pinker, 1989), these models cannot account for the differences between recognition and mental imagery (see Section 3 and and in Section 5, *The Relation Between Object Recognition and Mental Rotation* subsection). Several computational alignment models of recognition have been developed; Ullman's (1989, 1996) alignment model, which relies on 3-D object representations, is probably the best known example (for a similar model, see Lowe, 1987). Ullman's 3-D alignment model seems compatible with analogue transformations, whereas the 2-D linear combination model (Ullman & Basri, 1991) cannot explain the evidence for analogue transformation processes because a linear combination does not traverse intermediate points on a transformational path. These models cannot account for congruency effects, as they rely on compensation processes that are shape specific.

Interpolation models. In the interpolation approach, recognition is achieved by localization in a multidimensional representational space that is spanned by stored views (Edelman, 1998; Poggio, 1990; Poggio & Edelman, 1990; Poggio & Girosi, 1990). The interpolation model is based on the theory of approximation of multivariate functions and can be implemented with radial basis functions (usually Gaussian classifiers that model neurons tuned to shapes in specific orientations).[6] Object recognition is achieved if the visual input can be approximated by the existing tuned basis functions, that is, if a new data point is localized close to the surface that is spanned by the stored basis functions. The interpolation approach does not require transforming or reconstructing an internal image. It can accommodate view-dependent recognition performance but is not in accordance with the evidence for analogue transformation processes. Edelman (1999) tried to accommodate evidence for analogue transformations in mental rotations (Cooper, 1976) by enhancing the interpolation approach with a binding of views by temporal contiguity. Edelman (1999) argued that temporally contiguous views can be associated to a fixed sequence of snapshot representations, a "footprint" in visual cortex. Images that are frequently seen in close temporal contiguity, for instance when one walks around an object, will tend to be bound together. If, later, the object is activated in memory, the spread of activation through the footprint creates a semblance of mental rotation. After extensive exposure, new connections between the representations of nonneighbors are formed, so that the semblance of mental rotation should disappear with increasing practice. However, the interpolation model cannot account for findings that imply analogue transformations even after extensive practice (Bundesen et al., 1981). Therefore, even this enhanced

---

[6] One of the gain modulation approaches also relies on basis functions (e.g., Pouget & Sejnowski, 1997). In contrast, the interpolation model of recognition does not involve gain modulation processes (but may be extended accordingly).

interpolation model does not explain the evidence for analogue transformations in object recognition. More critically, the interpolation model cannot account for spatial congruency effects. It explains recognition with basis functions that are tuned to both orientation and shape, and therefore congruency effects are predicted only for identical or visually similar objects (see also Gauthier & Tarr, 1997) but not for dissimilar shapes as demonstrated by Graf et al. (2005).

Pooling and threshold models. In pooling and threshold models, recognition is explained on the basis of the behavior of IT cells that are selectively tuned to specific image features in a view-dependent (and size-dependent) way (Perrett & Oram, 1998; Perrett et al., 1998; Riesenhuber & Poggio, 1999, 2002; Wallis & Bülthoff, 1999). A hierarchical pooling of the outputs of view-specific cells provides generalization over viewing conditions. The threshold model (Perrett et al., 1998) accounts for the systematic relation between recognition latencies and the amount of rotation and size scaling: The speed of recognition depends on the rate of accumulation of activity from neurons selective for an object in a specific orientation. A given level of evidence is reached faster for orientations that were perceived more frequently. Pooling and threshold models were partly motivated to avoid the mental rotation account of recognition and its implications (Perrett et al., 1998). Similar to Edelman's (1999) footprint extension, pooling and threshold models were enhanced with the notion that stored views of objects can be associated by temporal contiguity (Perrett & Oram, 1998; Wallis & Bülthoff, 1999). This enhanced model has problems similar to those of Edelman's (1999) footprint extension. Finally, the present pooling and threshold models cannot accommodate congruency effects, as they are based on units that are simultaneously tuned to shape and orientation. Therefore, they do not predict a facilitation effect for the recognition of dissimilar shapes in the same orientation or size. Note that this criticism is not directed against the notion of hierarchical pooling of information in the cortex but points to the necessity of extending present models.

Hybrid Models

Hybrid models combine structural and image-based approaches. Several hybrid models have been proposed to account both for the lack of invariance regarding spatial transformations and for evidence that object representations have a part structure (Biederman & Cooper, 1991b; Goldstone & Medin, 1994; Newell, Sheppard, Edelman, & Shapiro, 2005; but see C. B. Cave & Kosslyn, 1993; for a review, see Graf & Schneider, 2001). A study with novel objects provided evidence that parts-based structured representations and image-based representations operate in parallel (Foster & Gilson, 2002). These findings suggest a hybrid model of recognition, with independent parts-based and image-based processes (see also Hayward, 2003). However, it is not clear yet whether these results transfer to common (and more complex) objects. Some hybrid models have been derived from structural description models (Hummel & Stankiewicz, 1998), whereas others are extensions of view-based models (Edelman & Intrator, 2000, 2001). Hybrid models provide some new interesting predictions, for example, concerning the role of attention in object recognition (Thoma et al., 2004). Present hybrid models usually account for orientation-dependent recognition performance but account for neither analogue transformation processes nor for congruency effects.

To conclude, none of the existing models can account for all three classes of findings without introducing additional ad hoc assumptions. View-independent models typically fail to account for all three classes of findings. Interpolation, pooling, and threshold models have problems in explaining the evidence for analogue transformation processes. Almost all recognition models, including hybrid models, do not address congruency (frame) effects in object recognition. None of the existing models include reference frames or similar structures that are suited to explain these effects—apart from a model by Hinton and Parsons (1981; Hinton, 1981) that would not predict the first two classes of findings. Congruency effects in recognition were mostly ignored in the development of recognition models, even though reference frames seem to play a major role in recognition and shape perception.

5. A Transformational Framework of Recognition

In this section, I lay the foundations for a transformational framework that accommodates all three classes of findings. TFR can be summarized as follows. Object recognition relies on frame (coordinate) transformations related to LTM (implemented by gain modulation), whereas mental imagery seems to involve image transformations in STM. The recognition of objects after spatial transformations is achieved by a compensating transformation of the perceptual coordinate system that defines the correspondence between positions specified in memory and positions in the current visual field. By default, the perceptual coordinate system is aligned with the retinal upright (McMullen & Jolicoeur, 1990), but it can be adjusted to the orientation of the stimulus representation. The adjustment of the perceptual coordinate system is achieved by time-consuming (although relatively fast) and error-prone compensation processes. Object representations are imagelike; both representations and transformation processes are analogue. Objects are represented by one or more canonical views. Recognition is achieved by a transformation of the perceptual coordinate system until the input representation is aligned with the nearest canonical orientation. Coordinate transformations (but not necessarily image transformations) can be neurocomputationally implemented by gain modulation. The transformational framework accounts for recognition performance after rotations, size scalings, and translations, from the level of individual objects up to the basic level of recognition. By allowing for nonlinear transformations, the transformational framework can be extended to account for structural alignment processes in categorization.

Alignment by Analogue Transformations

While the systematic dependency of recognition performance on orientation, size, and position can be explained with a number of models, the evidence for analogue transformation processes imposes stronger limitations on possible models of recognition. The simplest and most intuitive way to explain these two classes of findings is to assume that object recognition involves an alignment of input and memory representations, which is achieved by analogue spatial transformation processes such as rotations and size scalings. When stimulus and memory representations are aligned, a comparison or matching process is relatively straightforward because stimulus representation and object representation are more or less in correspondence (an exact alignment of all shape features

may not be necessary). Thus, the human visual system solves the problem of object constancy by using spatial transformations as compensation processes and not by using spatially invariant memory representations. The first two classes of findings can be easily explained with the plausible assumption that these analogue transformation processes are time-consuming and error prone. First, with increasing transformational distance between stimulus representation and memory representation, more time is required to recognize the object and the probability for errors increases. Second, the evidence for analogue transformation processes can be easily explained within an analogue alignment model. Thus, the analogue model accounts for both classes of findings. Moreover, TFR is in accordance with evidence for a transformational model of similarity (Hahn, Chater, & Richardson, 2003).

The alignment approach implies that memory representation and stimulus representation are brought into correspondence in order to compare them. This correspondence not only helps to determine the identity of an object but also specifies which parts of the image correspond to which parts of the memory representation. When a correspondence between memory and stimulus representation is established, ambiguous parts become more easily recognizable. Moreover, the correspondence helps to direct the attention to selected object parts. Alignment seems to be an integral aspect of the recognition process (Ullman, 1996, pp. 196–197) and was even proposed to underlie language processing (Pickering & Garrod, 2004). Alignment seems to be a general principle of the brain. Alignment is also involved in multisensory and sensorimotor integration; an alignment of neuronal maps can be found in the parietal cortex and even in subcortical areas like the superior colliculus (e.g., Duhamel, Colby, & Goldberg, 1991, 1998; King & Schnupp, 2000; Salinas & Abbott, 1995; Sparks & Nelson, 1987; Stein, Wallace, & Stanford, 2000).

Reference Frames in Object Recognition

The transformational framework, as delineated so far, can explain the first two classes of findings. But what about the third class of findings—the congruency in object recognition? First I want to come back to the question of whether reference frames in object recognition are viewer centered or object centered. The transformational account suggests a viewer-centered reference frame and is in agreement with the finding that recognition performance depends on the spatial relation between observer and object. Moreover, TFR can explain why the impression of an object-centered frame may arise. According to TFR, object recognition is achieved by aligning input and memory representations, so that they are in spatial correspondence. Thus, as a result of the recognition process, the reference frame of the memory representation is centered on the object. It should be noted that this is a result of time-consuming spatial compensation processes (see Morvan & Wexler, 2005) and does not imply inherently object-centered reference frames (for related arguments, see Deneve & Pouget, 2003).

What about congruency effects in object recognition? Previous transformational models of recognition were based on the idea that the alignment between memory representation and stimulus representation is achieved by a process of mental rotation (e.g., Jolicoeur, 1990a; Kosslyn, 1994; Tarr & Pinker, 1989). Initially, this conception appeared reasonable, because performance in ob-

ject recognition tasks was found to be orientation dependent in a similar way as handedness decisions in imagery tasks (e.g., Jolicoeur, 1985, 1988; Tarr & Pinker, 1989, 1990). However, there are also important differences between transformation processes in recognition and mental imagery, which cast doubt on the notion that object recognition relies on imagery transformations (see Section 3 and, in Section 5, *The Relation Between Object Recognition and Mental Rotation* subsection). The orientation or size congruency effect by itself could still be accommodated in a mental rotation model of recognition on the assumption that a frame of reference is activated in addition to the mental rotation process. However, this extended mental rotation account could not explain the fact that the rates of transformation are typically faster in object recognition than in mental imagery rotation. Therefore, it is unlikely that recognition involves transformations of mental images. Instead, TFR proposes that object recognition involves an adjustment of a perceptual coordinate system. Transformation processes in mental imagery, in contrast, are regarded as transformations of mental images, which proceed at a slower rate than frame transformations.

Alignment by Coordinate Transformations

What does the term *frame transformation* really mean; how is it specified in TFR? Perceptual reference frames are coordinate systems, and frame transformations are transformations of a coordinate system. In the alignment process, a perceptual coordinate system is adjusted to the input representation by an analogue coordinate transformation, so that input and memory representations are in correspondence (see Figure 5). Only the reference frame is adjusted, whereas the memory representations (in LTM) remain unchanged. It is likely that the adjusted frame decays quite
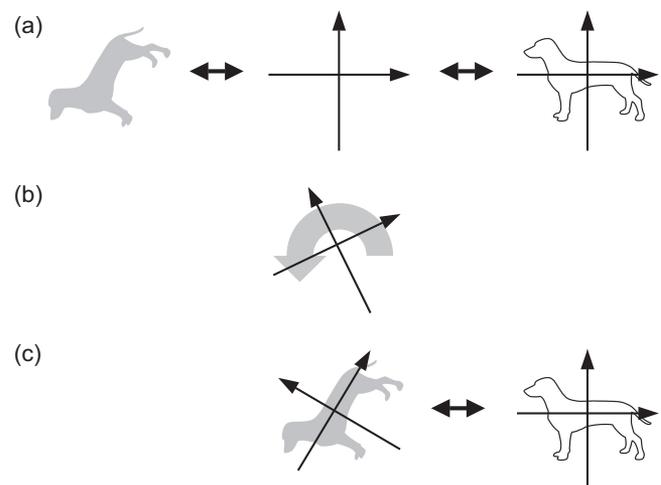
Figure 5. Alignment by analogue coordinate transformations. (a) A misoriented stimulus is presented. The perceptual coordinate system is in its default upright orientation. Memory representations are stored in a canonical orientation. (b) An analogue rotation of the perceptual coordinate system is performed (coordinate transformation), which traverses intermediate points on the rotational path. (c) When the perceptual coordinate system is adjusted to the stimulus representation, it compensates for orientation differences between the input representation and memory representations.

fast, leading to only transient facilitation effects. The notion of frame transformations is in accordance with findings suggesting that memory representations have to be regarded as representations within a frame of reference (for reviews, see Farah, 2000; Jolicoeur & Humphrey, 1998). Frame transformations need to be distinguished from mental imagery, where typically just an image in STM is transformed, not the perceptual coordinate system (Koriat & Norman, 1984, 1988, 1989; but see Robertson et al., 1987). Therefore, frame effects should not be expected in STM mental imagery tasks.

Congruency effects (see Section 3) can be explained in a straightforward way on this theoretical basis. First, the orientation congruency effect (Graf et al., 2005; Jolicoeur, 1990b, 1992) can easily be accounted for with frame rotations. Recognition involves a rotation of the perceptual coordinate system until input and memory representations are aligned. When another stimulus is presented in the same orientation, it can be recognized more easily, because the reference frame is already adjusted (see Figure 6). If the orientation differs from the first presentation, the frame is readjusted until it is in alignment with the stimulus. The orientation adjustment is not stimulus specific, because the perceptual coordinate system is transformed, and not just a mental image. Because the transformation is an analogue, time-consuming, and error-prone process, RTs and error rates increase with increasing amount of transformation.[7]

In a similar way, the size congruency effect in Larsen and Bundesen's (1978, Experiments 2 and 3) experiments can be explained. When a stimulus is presented, the size of the coordinate system is adjusted. Participants were informed that the next stimulus would frequently be in the same size so that the reference frame would be kept active over the intertrial interval. If the next stimulus had the same size, no further size adjustment would be necessary. If the size differed from the expected size, the frame would have to be readjusted until an alignment with the new input representation was achieved. If the same stimulus is repeated, a shape-specific image transformation, which presumably occurs in parallel, may lead to a faster alignment. The image transformation proceeds with a slower rate than a frame transformation, but the RT function has a lower *y*-intercept.

TFR also accounts for the finding that compensation can occur even before the stimulus is presented if adequate prior information is provided (Larsen & Bundesen, 1978). According to TFR, the
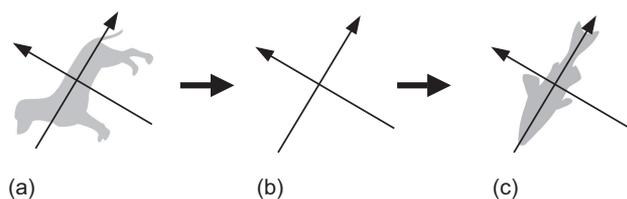
Figure 6. Object recognition by adjusting a perceptual reference frame accounts for orientation congruency effects. (a) Recognition involves the adjustment of a perceptual coordinate system to the orientation of the stimulus. (b) The coordinate system remains active for some time after the initial stimulus has disappeared. (c) A facilitation effect for different objects in the same orientation is expected, as the perceptual coordinate system is already adjusted. This orientation congruency effect is not limited to objects with similar shapes (Graf et al., 2005).

perceptual coordinate system is transformed, not the input representation. When the next stimulus is presented in the expected size, it can be recognized relatively fast, without further adjustment. Moreover, the frame hypothesis fits with evidence that the scene context can provide information that aids object recognition from unfamiliar viewpoints, presumably by providing a consistent reference frame (Christou, Tjan, & Bülthoff, 2003).

The proposed transformational framework has two further (related) advantages. First, it accounts for congruency effects without the need to adjust the memory representations—in contrast to alternative models of recognition, which can only account for congruency effects when all memory representations are adjusted (as congruency effects are not limited to similar shapes and therefore cannot be accounted for by class-specific processes). In TFR it is not necessary to transform every stored shape, but simply the perceptual frame of reference. Thus, the transformational approach based on coordinate transformations avoids the combinatorial explosion that would result if every stored memory representation had to be aligned individually. Second, a framework based on coordinate transformations does not require a mechanism that preselects some memory representations to reduce the computational effort, because the adjustment of a perceptual coordinate system obviates the need to transform every memory representation. Thus, potential problems of a preselection process may be avoided (for a discussion, see Ullman, 1989, pp. 237–238).

TFR is consistent with findings from neurophysiology and computational neuroscience that suggest that recognition relies on coordinate transformations. Changes in the position and size of an object can be compensated for by performing coordinate transformations based on gain modulation (Salinas & Abbott, 1997a, 1997b; Olshausen et al., 1993): IT receptive fields are shifted—because of attentional gain modulation in V4—so that they are centered at the point where attention is directed (for an illustration, see Salinas & Sejnowski, 2001). Similarly, IT receptive fields of variable, attention-controlled spatial scales are obtained when the mechanism is extended to scale dependent attentional gain fields (see also Salinas & Abbott, 2001; Salinas & Sejnowski, 2001; Salinas & Thier, 2000). The recognition of objects in different orientations can be accounted for by orientation-dependent gain modulation (Deneve & Pouget, 2003; Pouget & Sejnowski, 2001). Orientation-dependent gain modulation should correspond to the rotation of receptive fields in IT, which compensates for orientation differences. Moreover, the gain modulation approach also allows partial coordinate transformations (Pouget et al., 2002), so that partial compensation in orientation congruency tasks (Graf et al., 2005; Jolicoeur, 1990b) can be accounted for.

Updating processes involved in transforming the coordinate system can either proceed gradually, as in TFR, or can be performed in one step. The basis function network for coordinate transformations (Deneve & Pouget, 2003; Pouget et al., 2002; Pouget & Sejnowski, 1997) is a one-step model—that is, the transformation does not proceed through intermediate stages (see Pouget & Sejnowski, 2005). The approach proposed by Salinas

---

[7] In addition to the frame that is adjusted to the stimulus orientation, the frame in the retinal upright orientation may also be activated (see Murray, 1999). This might explain why there is still an effect of absolute orientation (Graf et al., 2005; Jolicoeur, 1990b, Experiment 3).

and collaborators (e.g., Salinas & Abbott, 1997a, 1997b) relies on linear combinations and was not explicitly conceptualized as an analogue approach but seems compatible with analogue transformation processes in recognition (Emilio Salinas, personal communication, May 19, 2003). Of interest, gain modulation models can be formalized within a framework of continuous—that is, analogue—field computations (MacLennan, 1997, 1999). Thus, present gain modulation approaches can be translated into a framework that involves analogue coordinate transformations. Several neurocomputational models for coordinate transformations rely on analogue dynamic updating processes (Dominey & Arbib, 1992; Droulez & Berthoz, 1991; Zhang, 1996), consistent with the evidence for analogue transformation processes (see Section 2).

Whatever the exact details of the neuronal implementation of coordinate transformations, the present gain modulation approaches clearly show that object recognition can be modeled on the basis of coordinate transformations, thus demonstrating computational feasibility. Coordinate transformations, based on gain modulation, are also biologically plausible. Gain modulation is implemented ubiquitously in the visual cortex and can be regarded as a general computational principle of the cortex (e.g., Salinas & Sejnowski, 2001; Salinas & Thier, 2000). Coordinate transformations based on gain modulation provide a possible neural implementation of TFR.

Template Matching: Multiple Views Plus Transformations

Even though transformation processes are central to TFR, transformations are not the whole story. It is not necessarily just a single view of an object that is stored; more likely, multiple views are encoded. Consequently, it is possible that several views may serve as canonical perspectives if objects are perceived frequently from specific points of view. Thus, TFR embraces the "multiple views plus transformations" approach (Tarr, 1995; Tarr & Pinker, 1989). Recognition is achieved by transforming the perceptual coordinate system until the input representation is aligned with the nearest canonical orientation.

According to TFR, memory representations in default object recognition are image based (for reviews, see Jolicoeur & Humphrey, 1998; Tarr & Bülthoff, 1998) and can be conceptualized as *templates*. The idea of template matching is still criticized in many introductory textbooks, even though an abundance of evidence has accumulated in favor of a template model of recognition (see Jolicoeur & Humphrey, 1998). An important criticism against template models lies in the difficulties in matching after spatial transformations of the stimulus. In TFR the matching of template and stimulus representation is achieved according to two principles: transformation processes and multiple representations.

This framework, which includes both transformation processes and multiple representations, accommodates the majority of findings in the recognition literature. In TFR, the systematic relation between recognition performance and the transformational distance is weakened with extensive practice, because new canonical perspectives are formed (e.g., Tarr, 1995; Tarr & Pinker, 1989). Evidence for an M-shaped RT function for the naming of rotated objects may be explained as well on the basis of a transformational approach to recognition (Murray, 1997). However, it should be noted that under certain conditions, recognition can be based on the opportunistic use of discriminative features (Murray, Jolicoeur,

McMullen, & Ingleton, 1993). The use of discriminative features, though, does not reflect the default processes in object recognition but is restricted to situations in which a limited set of stimuli is presented repeatedly (Jolicoeur & Humphrey, 1998; Jolicoeur & Milliken, 1989; Lawson, 1999). Moreover, the use of orientation-invariant features is not driven purely by bottom-up stimulus features but requires voluntary strategic top-down processes (K. D. Wilson & Farah, 2003). Thus, view-invariant features may be used under specific conditions and voluntary strategic control but do not reflect default recognition processes.

Now, as the major parts of TFR have been presented, recognition can be described according to the following scheme (see Figure 7). First, when a stimulus is presented, an analogue transformation of the perceptual coordinate system is performed, until the stimulus representation is aligned with the closest memory representation (alignment stage). The perceptual coordinate system specifies correspondences between memory representations and the visual input. Memory representations are stored in the canonical perspective (i.e., within a canonical reference frame). Multiple canonical views may be stored. Second, stimulus and memory representations are compared, and the best matching memory representation is determined (matching stage). These steps do not have to occur strictly sequentially but may be executed in a cascade (e.g., Humphreys & Forde, 2001). At least in mental rotation tasks, the rotation process can start before perceptual discrimination processes are finished (Ruthruff & Miller, 1995), and response preparation can begin before mental rotation is finished (Band & Miller, 1997; Heil, Rauch, & Hennighausen, 1998).

A Process-Based Geometrical Framework

TFR can be regarded as a geometrical approach to object recognition, because the transformations that are necessary to provide a solution to the first main problem of recognition—rotations, size scalings, and translations—correspond to specific transformation

(a)   (b)   (c)

...
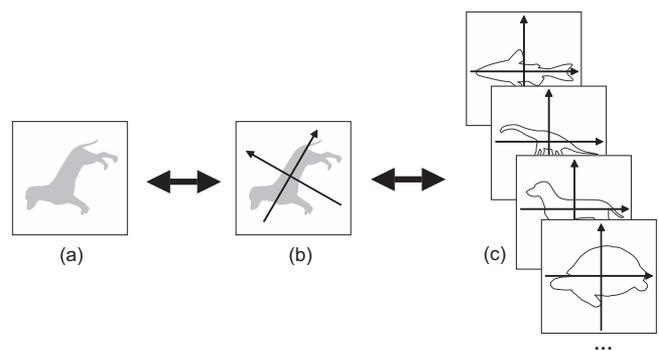
Figure 7. Recognition by alignment via coordinate transformations: (a) A misoriented stimulus is presented, which has to be matched with memory representations in order to be recognized. (b) A perceptual coordinate system is transformed in order to be adjusted to the orientation of the stimulus representation. The coordinate system defines the correspondence between positions specified in memory and positions in the current visual field and thus aligns memory representations and input representation. (c) Object representations in long-term memory are stored in a canonical orientation, typically the retinal upright.

groups of euclidean geometry, as specified in Felix Klein's (1872/1893) *Erlanger Programm*.[8] It seems reasonable to conceptualize object recognition in terms of geometry: If we strive for an understanding of shapes, we have to deal with the branch of science that deals with the description of form and space—that is, with geometry (see Shepard, 1994). However, TFR differs from previous geometrical models, because it is a process-based and dynamic geometrical framework that relies on coordinate transformations. Traditionally, geometrical models of recognition are static models, as they usually postulate the detection of geometrical invariants relative to a specific transformation group (e.g., Cassirer, 1944; Van Gool et al., 1994) or the use of object-centered frames (see Palmer, 1983, 1989, 1999). These models predict that recognition performance is independent of the amount of transformation, as they do not rely on time-consuming compensation processes. Therefore, these static models cannot explain the three classes of findings outlined earlier. An alternative geometrical conception is suggested here: a dynamic system in which object recognition is achieved by a time-consuming and error-prone transformational alignment process.

Structural Alignment and Categorization

Image-based object representations are often confused with holistic representations (for a similar argument, see Barsalou, 1999). However, the question of whether representations are holistic or structured is orthogonal to the issue of whether representations are abstract or image based. The transformational approach does not imply that object parts and relations between parts do not play a role in object recognition. There is evidence that object representations have a part structure (see the *Hybrid Models* subsection of Section 4). However, it seems overstated to associate this evidence for parts-based representations with abstract representations that predict view-invariant recognition performance (e.g., Biederman, 1987), given that most findings speak against invariant recognition performance (see Section 1). It seems more reasonable to integrate parts-based representations into an image-based framework (see Edelman & Intrator, 2000, 2001; Graf, 2002; Graf, Bundesen, & Schneider, 2006; Graf & Schneider, 2001). In a transformational framework, parts may have an important role. The identification of parts can facilitate the alignment process, because corresponding parts indicate possible correspondences between stimulus representation and memory representation. Therefore, knowledge about the hierarchical organization of an object can guide the alignment process (Basri, 1996; Basri, Costa, Geiger, & Jacobs, 1998). Of interest, the notion of structural alignment processes is popular in the categorization and similarity literature. An alignment of structured representations accounts for similarity judgments of objects and scenes (e.g., Goldstone, 1996; Goldstone & Medin, 1994; Medin, Goldstone, & Gentner, 1993) and for object categorization (Markman & Gentner, 1993; Markman & Wisniewski, 1997). Thus, the alignment approach is not in contradiction to the notion of hierarchically structured object representations. By allowing for nonlinear (deforming) transformations (termed *topological transformations* in Klein's, 1872/1893, hierarchy of transformation groups), TFR may be extended to object categorization up to the basic level, that is, the highest level at which category members still have highly similar shapes (Rosch et al., 1976). Shape differences of basic-level category members can be compensated for by

deforming transformation processes—in other words, by deformable template matching (Basri et al., 1998). Consistent with this proposal, categorization performance deteriorates systematically with increasing amount of deforming transformation (Graf, 2002; Graf et al., 2006), reminiscent of orientation dependency in object recognition. These findings can be accounted for by deforming coordinate transformations, that is, by assuming that the brain uses the whole range of transformations specified in Klein's *Erlanger Programm* (cf. Chen, 2005). Deforming transformations seem to be necessary also for the recognition of deformable objects, like plants and animals, and for articulated objects (Ullman, 1989). Moreover, the integration of nonlinear transformations allows rejecting arguments against image-based models of recognition (Hummel, 2000); it provides the basis for a hybrid framework of recognition that is both image based and structural. On the basis of deforming transformations, an image-based alignment of corresponding object parts, that is, a structural alignment, is feasible. Thus, the proposed framework can be regarded as an image-based extension of the structural alignment approach of categorization and similarity (Goldstone, 1996; Goldstone & Medin, 1994; Markman & Gentner, 1993; Markman & Wisniewski, 1997; Medin et al., 1993).

Finally, evidence for a role of nonaccidental properties (NAPs) in recognition (Biederman & Bar, 1999; Vogels, Biederman, Bar, & Lorincz, 2001) is compatible with a transformational framework as well. NAPs are features that are likely to be more or less invariant over large ranges of viewpoints, like instances of connectivity, collinearity, parallelism, and so forth. The conception of NAPs was popularized by Biederman's (1987) recognition-by-components model of recognition, but the use of NAPs was actually proposed first within an alignment model (Lowe, 1985, 1987). NAPs may provide possible constraints in a transformational framework—for example, NAPs may serve as alignment keys that guide the alignment process (Ullman, 1989).

The Relation Between Object Recognition and Mental Rotation

TFR proposes that object recognition relies on coordinate transformations related to LTM, whereas mental imagery typically involves image transformations in STM.[9] This conception accounts for similarities and differences between transformation processes in recognition and imagery. According to TFR, both mental imagery and object recognition rely on time-consuming and error-prone transformation processes in order to achieve an alignment of two representations, which results in a systematic dependency of performance on the amount of transformation. In addition, transformation processes in imagery and recognition seem to be analogue processes (Bundesen et al., 1981; Cooper, 1976; Kourtzi & Shiffrar, 2001).

---

[8] In mathematics, a group is a set that has rules for combining any pair of elements and that obeys four properties: closure, associativity, existence of an identity element, and existence of an inverse element.

[9] The distinction between transformations in mental imagery and object recognition is actually more complex, because tasks with rather different requirements are pooled under the label *mental rotation*, some of them involving LTM (for a more detailed account, see Graf, 2002).

Unlike previous alignment models, TFR does not equate compensation processes in object recognition and mental rotation and, therefore, accounts for several important differences between recognition and imagery. Transformation processes in recognition are fast and shape unspecific, whereas transformation processes are slower and shape specific in typical mental imagery tasks (see Section 3). The distinction between coordinate transformations in recognition and image transformations in mental imagery accounts for several further differences between mental rotation and recognition. First, mental rotations are consciously accessible, whereas compensation processes in recognition are not (see Lawson & Humphreys, 1998). Second, the initial assignment of top and bottom is preserved throughout mental rotation, but not in object recognition. For instance, participants did not recognize the outline of the state of Texas when it was misoriented by 90°, even after mentally rotating it into its canonical orientation. Participants only recognized it when they were implicitly instructed to redefine the top and bottom of the shape (B. S. Gibson & Peterson, 1994; Reisberg & Chambers, 1991). Third, perceived rotary motion influenced speeded naming responses in a recognition task but did not affect responses in an imagery task (Jolicoeur et al., 1998). Fourth, mental rotation performance was facilitated when the axis of rotation corresponded with the main axis of the objects or one of the axes of the environment, whereas recognition performance was not (Willems & Wagemans, 2001). Fifth, studies in which participants had to tilt their head showed that recognition relies on a reference frame more closely aligned with retinal upright, whereas the frame for mental rotation tasks was aligned more closely with environmental upright (McMullen & Jolicoeur, 1990). Sixth, neuropsychological dissociations have been demonstrated between the ability to recognize misoriented objects and to mentally rotate objects. Farah and Hammond (1988) described a patient who failed three different STM mental rotation tasks but was nonetheless able to recognize misoriented letters, numbers, and drawings. Seventh, a functional magnetic resonance imaging study showed that neural activations in the recognition of misoriented objects and mental imagery are not identical, even though there was considerable overlap between both tasks (Gauthier et al., 2002).

All these findings and arguments are directed against the idea that recognition is based on mental imagery transformations, but they are often interpreted as evidence against transformational (or alignment) models in general. However, they actually do not question that recognition involves analogue transformation processes that are not imagery transformations—like analogue coordinate transformations. Instead the findings confirm the distinction between frame and image transformations in TFR. Object recognition seems to be more closely related to visuomotor control, which also relies on coordinate transformations, than to mental rotation.

Unresolved Issues

Although TFR is supported by numerous studies, some issues still remain unresolved. First, what about rotations in depth, which lead to drastic changes in the appearance of the objects due to self-occlusion? TFR proposes that depth rotations are also based on coordinate transformations. Consistent with this view, foreshortened views are recognized faster when a background with strong monocular depth cues is presented whose orientation is congruent with object orientation, supplying a visual reference frame (Humphrey & Jolicoeur, 1993), or when a congruent scene context provides a reference frame (Christou et al., 2003). Initial alignment models were based on the assumption that the visual system uses 3-D representations that are rotated in depth (Ullman, 1989, 1996). A later model proposes that an alignment is achieved by a linear combination of 2-D images (Ullman & Basri, 1991). Also interpolation models, which do not involve alignment processes, rely on 2-D images (Edelman, 1998; Edelman & Bülthoff, 1992). There are good arguments that the underlying memory representations are 2.5-D, that is, include depth information but do not correspond to full 3-D models (for a discussion, see Pinker, 1997, pp. 256–261). But how can coordinate transformations in depth be explained on the basis of 2-D or 2.5-D representations? Coordinate transformations in depth may work more efficiently when memory representations are stored such that different views of an object are bound together to one coherent object representation, conforming to a sequentially ordered rotation in depth. Temporal contiguity is clearly one important principle for associating views to an integrated representation: When one walks around objects or interacts with them, one typically observes sequences of continuous depth rotations. Recent experiments using human faces have shown that views that were initially presented in a temporal sequence were bound together to form a coherent identity (e.g., Wallis, 2002; Wallis & Bülthoff, 2001; for a computational model, see Wallraven & Bülthoff, 2001). Accordingly, it has been shown that several 2-D projections can be associated to a 3-D percept by visual experience (Sinha & Poggio, 1996). Such 2.5-D representations, consisting of associated views stored in sequential order, may provide the representational basis for coordinate transformations in depth. This principle alone, however, does not account for all findings, because generalization across views is possible with a single view of a novel object (Biederman & Bar, 1999). It seems possible that unfamiliar views are approximated by views of similar objects from the relevant perspective, especially from members of the same basic- or subordinate-level category (Vetter, Hurlbert, & Poggio, 1995).

A second unresolved issue is related to animal versus nonanimal categorization tasks with natural images that showed ultrarapid categorization decisions (e.g., Thorpe, Fize, & Marlot, 1996; Rousselet, Fabre-Thorpe, & Thorpe, 2002). It has been argued that recognition is so fast in these tasks that it is essentially a feedforward process. There remains little or no time for recurrent processes or compensation processes like alignment. However, the results on ultrarapid categorization do not speak against alignment processes in recognition, for three reasons. First, an event-related potential study confirmed that there is enough time for compensation processes in recognition (Johnson & Olshausen, 2003). The authors identified two signals related to object recognition. The early signal at around 135 ms represents a presentation-locked component that did not covary with recognition latencies. This signal is present when there are low-level feature differences between images, which appeared sufficient to do the animal versus nonanimal categorization task (see also Torralba & Oliva, 2003). The other component arises between 150 and 300 ms, and its latency covaries with subsequent RTs for identification. Thus, the neural signatures of recognition have a substantially later and variable time of onset, leaving enough time for alignment pro-

cesses in object recognition. The second reason is that the bottom-up processes reflected in the results of Thorpe and collaborators may simply correspond to a fast feedforward sweep in visual processing that does not lead to a conscious percept (e.g., Lamme, 2003). And actually, participants in animal detection tasks often cannot report basic-level category membership. Consequently, conscious object perception may require recurrent processes (Lamme, 2003; Lamme & Roelfsema, 2000), including coordinate transformations. Third, in accordance with TFR, orientation congruency effects have been found recently even in an ultrarapid categorization task with natural images (Rieger, Köchy, Schalk, & Heinze, 2006). In Experiment 2, Rieger et al. varied picture plane orientation of objects and scene background independently. They presented either upright images, 90° rotated full images, 90° rotated objects (with upright background), or 90° rotated backgrounds (with upright objects). RT was fastest for upright full images, followed by rotated full images, and was slowest when the orientation of the object and background was incongruent. Thus, recognition was facilitated when the background provided a frame of reference that was congruent with the object orientation.

A third unresolved issue is that in the alignment approach the stimulus apparently has to be recognized first before the compensating alignment transformation (along the shortest path) can be determined, leading to an apparent paradox (e.g., Corballis, 1988). In an elegant model, Ullman (1989) demonstrated that an alignment can be achieved on the basis of information that is available before object identification. A number of heuristics may be used to determine correspondences (Palmer, 1999, pp. 375–376). Further solutions to this correspondence problem have been proposed in the computer vision literature (e.g., Belongie, Malik, & Puzicha, 2002; Sclaroff, 1997; Sclaroff & Liu, 2001; Witkin, Terzopoulos, & Kass, 1987; see also Ullman, 1996). It should be noted that the correspondence problem is not specific for the alignment approach to recognition but also arises in (apparent) motion perception and stereoscopic vision (for reviews, see Chen, 2001; Palmer, 1999). A possible solution to this problem is that the shortest direction of alignment is determined by early perceptual processes that occur in a fast and unconscious feedforward sweep (Lamme, 2003), whereas conscious recognition requires an alignment of input and memory representations and, therefore, time-consuming (but still fast) recurrent alignment processes.

6. Conclusions and Outlook

Three important classes of findings regarding the recognition of objects after spatial transformations were identified, including independent behavioral and neurocomputational evidence that object recognition relies on coordinate transformations. All three classes of findings can be explained in a consistent and parsimonious way with a transformational account of recognition, whereas existing recognition models cannot accommodate all findings. The main difference from previous alignment models is that compensation processes in recognition are conceptualized in TFR as coordinate transformations and not as mental imagery transformations. TFR covers a broad range of data on the basis of few processing principles—such as alignment and the use of analogue transformation processes. TFR is supported by an impressive number of studies, the weakest point probably being the evidence for

analogue transformations. For instance, analogue transformations and coordinate transformations have not yet been demonstrated within one single experiment. However, given the current state of research, the assumption of analogue transformations seems to provide the most parsimonious account.

Several extensions of TFR seem possible. A number of findings indicate that object categories contain knowledge about possible transformations (Landau, 1994; Stone, 1998; Zaki & Homa, 1999). The transformational account offers a possibility of explaining how this transformational knowledge may be represented in the visual system. In addition, TFR is in accordance with evidence for massive top-down processes in the visual cortex (e.g., Bar, 2003; Mumford, 1994), because the alignment approach is easily compatible with interactions between top-down and bottom-up processing (Salinas, 2004; Ullman, 1995).

As described earlier, at present some issues remain unresolved within TFR. However, new questions arise that seem to be fruitful for further research. First, even though TFR is compatible with many neurophysiological and neurocomputational findings, there are open questions regarding the neuronal implementation of TFR. There is still little evidence for predicted dynamic transformations of IT receptive fields due to spatial compensation processes in object recognition, although context-dependent changes of receptive fields in IT cortex have been demonstrated (Rolls et al., 2003). Another important issue relates to the question of in which parts of the visual cortex the compensation processes are performed. The notion that recognition relies on coordinate transformations is compatible with the view that recognition is achieved exclusively in the ventral pathway (Salinas & Abbott, 1997b). However, it seems possible that spatial transformation processes in recognition also involve the dorsal pathway (as implemented by Pouget and collaborators), which is traditionally associated with spatial processing and coordinate transformations. There is suggestive evidence that the dorsal stream is involved in the recognition of objects that are rotated or size scaled (Eacott & Gaffan, 1991; Faillenot, Decety, & Jeannerod, 1999; Faillenot, Toni, Decety, Grégoire, & Jeannerod, 1997; Gauthier et al., 2002; Kosslyn et al., 1994; Sugio et al., 1999; Vuilleumier, Henson, Driver, & Dolan, 2002; Warrington & Taylor, 1973, 1978).[10] These findings challenge the assumption of a strict functional and anatomical separation of the visual cortex into two distinct pathways (Milner & Goodale, 1995; Ungerleider & Haxby, 1994; Ungerleider & Mishkin, 1982). However, it still must be shown whether these dorsal processes are related to transformation processes in recognition.

Second, the role of attention in object recognition requires further elaboration. Different gain modulation models differ in the role they assign to attentional mechanisms. Attentional processes are central in the proposals of Salinas and collaborators (e.g., Salinas & Abbott, 1997b; Salinas & Sejnowski, 2001) and Olshausen and collaborators (e.g., Olshausen, Anderson, & Van Essen, 1993). In contrast, gain modulation is simply a function of stimulus orientation in Pouget's approach and does not depend on attention (e.g., Pouget & Sejnowski, 1997, 2001). Of interest, gain modulation has been proposed as one of two basic neural mechanisms of attention (Bundesen, Habekost, & Kyllingsbæk, 2005).

---

[10] For a review discussing parietal activations in mental imagery, see Jagaroo (2004).

These open questions indicate that the study of coordinate transformations in recognition is still in an early phase. Clearly, much further work is necessary to establish these claims. Not only is further behavioral and neural evidence required but so are computational implementations of the behavioral data. Nevertheless, the transformational framework seems computationally feasible and biologically plausible, as exemplified by neurocomputational implementations based on gain modulation (e.g., Pouget & Sejnowski, 2001; Salinas & Abbott, 2001). Although a lot of work still has to be done, these neurocomputational models indicate that in principle, the gain modulation approach can be extended to object recognition.

TFR proposes that recognition and visuomotor control involve similar processing principles. This is in contrast to the claims by Milner and Goodale (1995) but fits well with the proposal that perception and action planning are coded in a common representational medium (e.g., Hommel, Müsseler, Aschersleben, & Prinz, 2001; Prinz, 1990, 1997). Accordingly, TFR is compatible with evidence for a close coupling between object recognition and perception for action (e.g., Chao & Martin, 2000; Creem & Proffitt, 2001; Helbig, Graf, & Kiefer, 2006; Tucker & Ellis, 2001).

Coordinate transformations (based on gain modulation) provide a unifying computational principle for diverse tasks such as eye and limb movements, spatial perception, navigation, attention, and object recognition (Andersen et al., 2000; Bizzi & Mussa-Ivaldi, 2000; Salinas & Sejnowski, 2001; Salinas & Thier, 2000). The brain may solve the problem of object constancy by principles similar to those used to solve the problem of spatial constancy, that is, the question of how the perceived outer world is stable despite body and eye movements (e.g., Nishida et al., 2003). This seems reasonable because when the observer moves the object perception system has to compensate for changes induced by self-motion.

In conclusion, the transformational framework is fruitful for further research, and it is highly parsimonious, because it allows the integration of previously distinct literatures on several levels. It explains a range of findings that were mostly neglected in the recognition literature. TFR accounts for similarities and differences between object recognition and mental imagery without needing to equate processes in recognition and imagery. TFR relies on processing principles that are already established for the visual cortex. And last but not least, TFR allows for an integrative framework of object recognition and action.

References

Andersen, R. A., Batista, A. P., Snyder, L. H., Buneo, C. A., & Cohen, Y. E. (2000). Programming to look and reach in the posterior parietal cortex. In M. S. Gazzaniga (Ed.), *The new cognitive neurosciences* (pp. 515–524). Cambridge, MA: MIT Press.

Andersen, R. A., Bracewell, R. M., Barash, S., Gnadt, J. W., & Fogassi, L. (1990). Eye position effects on visual, memory, and saccade-related activity in areas LIP and 7a of macaque. *Journal of Neuroscience, 10,* 1176–1196.

Andersen, R. A., Essick, G. K., & Siegel, R. M. (1985, October 25). Encoding of spatial location by posterior parietal neurons. *Science, 230,* 456–458.

Andersen, R. A., & Mountcastle, V. B. (1983). The influence of the angle of gaze upon the excitability of the light-sensitive neurons of the posterior parietal cortex. *Journal of Neuroscience, 3,* 532–548.

Andersen, R. A., Snyder, L. H., Bradley, D. C., & Xing, J. (1997).

Encoding of intention and spatial location in the posterior parietal cortex. *Annual Review of Neuroscience, 20,* 303–330.

Anderson, J. R. (1978). Arguments concerning representations for mental imagery. *Psychological Review, 85,* 249–277.

Ashbridge, E., & Perrett, D. I. (1998). Generalizing across object orientation and size. In V. Walsh & J. Kulikowski (Eds.), *Perceptual constancy. Why things look as they do* (pp. 192–209). Cambridge, England: Cambridge University Press.

Band, G. P. H., & Miller, J. (1997). Mental rotation interferes with response preparation. *Journal of Experimental Psychology: Human Perception and Performance, 23,* 319–338.

Bar, M. (2001). Viewpoint dependency in visual object recognition does not necessarily imply viewer-centered representation. *Journal of Cognitive Neuroscience, 13,* 793–799.

Bar, M. (2003). A cortical mechanism for triggering top-down facilitation in visual object recognition. *Journal of Cognitive Neuroscience, 15,* 600–609.

Barsalou, L. W. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences, 22,* 577–660.

Basri, R. (1996). Recognition by prototypes. *International Journal of Computer Vision, 19,* 147–167.

Basri, R., Costa, L., Geiger, D., & Jacobs, D. (1998). Determining the similarity of deformable shapes. *Vision Research, 38,* 2365–2385.

Batista, A. P., Buneo, C. A., Snyder, L. H., & Andersen, R. A. (1999, July 9). Reach plans in eye-centered coordinates. *Science, 285,* 257–260.

Belongie, S., Malik, J., & Puzicha, J. (2002). Shape matching and object recognition using shape contexts. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 24,* 509–522.

Bennett, D. (2002, May). *Evidence for a pre-match "mental translation" on a form-matching task.* Paper presented at the 2nd annual meeting of the Vision Sciences Society, Sarasota, FL.

Biederman, I. (1987). Recognition-by-components: A theory of human image understanding. *Psychological Review, 94,* 115–147.

Biederman, I., & Bar, M. (1999). One-shot viewpoint invariance in matching novel objects. *Vision Research, 39,* 2885–2899.

Biederman, I., & Bar, M. (2000). Differing views on views: Response to Hayward and Tarr (2000). *Vision Research, 28,* 3901–3905.

Biederman, I., & Cooper, E. E. (1991a). Evidence for complete translational and reflectional invariance in visual object priming. *Perception, 20,* 585–593.

Biederman, I., & Cooper, E. E. (1991b). Priming contour-deleted images: Evidence for intermediate representations in visual object recognition. *Cognitive Psychology, 23,* 393–419.

Biederman, I., & Cooper, E. E. (1992). Size invariance in visual object priming. *Journal of Experimental Psychology: Human Perception and Performance, 18,* 121–133.

Biederman, I., & Gerhardstein, P. C. (1993). Recognizing depth-rotated objects: Evidence and conditions for three-dimensional viewpoint invariance. *Journal of Experimental Psychology: Human Perception and Performance, 19,* 1162–1182.

Biederman, I., & Gerhardstein, P. C. (1995). Viewpoint-dependent mechanisms in visual object recognition: Reply to Tarr and Bülthoff (1995). *Journal of Experimental Psychology: Human Perception and Performance, 21,* 1506–1514.

Bizzi, E., & Mussa-Ivaldi, F. A. (2000). Toward a neurobiology of coordinate transformations. In M. S. Gazzaniga (Ed.), *The new cognitive neurosciences* (pp. 489–500). Cambridge, MA: MIT Press.

Blanz, V., Tarr, M. J., & Bülthoff, H. H. (1999). What object attributes determine canonical views? *Perception, 28,* 575–599.

Bremmer, F. (2000). Eye position effects in macaque area V4. *NeuroReport, 11,* 1277–1283.

Bülthoff, H. H., & Edelman, S. (1992). Psychophysical support for a two-dimensional view interpolation theory of object recognition. *Proceedings of the National Academy of Sciences, USA, 89,* 60–64.

Bülthoff, H. H., Edelman, S. Y., & Tarr, M. J. (1995). How are three-dimensional objects represented in the brain? *Cerebral Cortex, 3,* 247–260.

Bülthoff, I., & Bülthoff, H. H. (2003). Image-based recognition of biological motion, scenes and objects. In M. A. Peterson & G. Rhodes (Eds.), *Analytic and holistic processes in the perception of faces, objects, and scenes* (pp. 146–176). New York: Oxford University Press.

Bundesen, C., Habekost, T., & Kyllingsbæk, S. (2005). A neural theory of visual attention: Bridging cognition and neurophysiology. *Psychological Review, 112,* 291–328.

Bundesen, C., & Larsen, A. (1975). Visual transformation of size. *Journal of Experimental Psychology: Human Perception and Performance, 1,* 214–220.

Bundesen, C., Larsen, A., & Farrell, J. E. (1981). Mental transformations of size and orientation. In J. Long & A. Baddeley (Eds.), *Attention and performance* (Vol. 9, pp. 279–294). Hillsdale, NJ: Erlbaum.

Cassirer, E. (1944). The concept of group and the theory of perception. *Philosophy and Phenomenological Research, 5,* 1–35.

Cave, C. B., & Kosslyn, S. M. (1993). The role of parts and spatial relations in object identification. *Perception, 22,* 229–248.

Cave, K. R., & Bichot, N. P. (1999). Visuo-spatial attention: Beyond a spotlight model. *Psychonomic Bulletin & Review, 6,* 204–223.

Cave, K. R., & Kosslyn, S. M. (1989). Varieties of size-specific visual selection. *Journal of Experimental Psychology: General, 118,* 148–164.

Cave, K. R., & Pashler, H. (1995). Visual selection mediated by location: Selecting successive visual objects. *Perception & Psychophysics, 57,* 421–432.

Cave, K. R., Pinker, S., Giorgi, L., Thomas, C. E., Heller, L. M., Wolfe, J. M., et al. (1994). The representation of location in visual images. *Cognitive Psychology, 26,* 1–32.

Chao, L. L., & Martin, A. (2000). Representation of manipulable man-made objects in the dorsal stream. *NeuroImage, 12,* 478–484.

Chen, L. (2001). Perceptual organization: To reverse back the inverted (upside-down) question of feature binding. *Visual Cognition, 8,* 287–303.

Chen, L. (2005). The topological approach to perceptual organization. *Visual Cognition, 12,* 553–637.

Christou, C. G., Tjan, B. S., & Bülthoff, H. H. (2003). Extrinsic cues aid shape recognition from novel viewpoints. *Journal of Vision, 3,* 183–198. Retrieved May 5, 2003, from http://journalofvision.org/3/3/1/

Cohen, Y., & Andersen, R. (2000). Reaches to sounds encoded in an eye-centered reference frame. *Neuron, 27,* 647–652.

Cohen, Y., & Andersen, R. (2002). A common reference frame for movement plans in the posterior parietal cortex. *Nature Reviews Neuroscience, 3,* 553–562.

Colby, C. L. (1998). Action-oriented spatial reference frames in cortex. *Neuron, 20,* 15–24.

Connor, C. E., Gallant, J. L., Preddie, D. C., & Van Essen, D. C. (1996). Responses in area V4 depend on the spatial relationship between stimulus and attention. *Journal of Neurophysiology, 75,* 1306–1308.

Connor, C. E., Preddie, D. C., Gallant, J. L., & Van Essen, D. C. (1997). Spatial attention effects in macaque area V4. *Journal of Neuroscience, 17,* 3201–3214.

Cooper, L. A. (1976). Demonstration of a mental analog of an external rotation. *Perception & Psychophysics, 19,* 296–302.

Cooper, L. A., Schacter, D. L., Ballesteros, S., & Moore, C. (1992). Priming and recognition of transformed three-dimensional objects: Effects of size and reflection. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18,* 43–57.

Cooper, L. A., & Shepard, R. N. (1973). Chronometric studies of the rotation of mental images. In W. G. Chase (Ed.), *Visual information processing* (pp. 75–176). New York: Academic Press.

Corballis, M. C. (1988). Recognition of disoriented shapes. *Psychological Review, 95,* 115–123.

Corballis, M. C., Nagourney, B. A., Shetzer, L. I., & Stefanatos, G. (1978). Mental rotation under head tilt: Factors influencing the location of the subjective reference frame. *Perception & Psychophysics, 24,* 263–273.

Corballis, M. C., Zbrodoff, N. J., & Roldan, C. E. (1976). What's up in mental rotation? *Perception & Psychophysics, 19,* 525–530.

Corballis, M. C., Zbrodoff, N. J., Shetzer, L. I., & Butler, P. B. (1978). Decisions about identity and orientation of rotated letters and digits. *Memory & Cognition, 6,* 98–107.

Creem, S. H., & Proffitt, D. R. (2001). Grasping objects by their handles: A necessary interaction between cognition and action. *Journal of Experimental Psychology: Human Perception and Performance, 27,* 218–228.

Daems, A., & Verfaillie, K. (1999). Viewpoint-dependent priming effects in the perception of human actions and body postures. *Visual Cognition, 6,* 665–693.

Deneve, S., & Pouget, A. (2003). Basis functions for object-centered representations. *Neuron, 37,* 347–359.

DiCarlo, J. J., & Maunsell, J. H. R. (2003). Anterior inferotemporal neurons of monkeys engaged in object recognition can be highly sensitive to object retinal position. *Journal of Neurophysiology, 89,* 3264–3278.

Dill, M., & Edelman, S. (2001). Imperfect invariance to object translation in the discrimination of complex shapes. *Perception, 30,* 707–724.

Dill, M., & Fahle, M. (1998). Limited translation invariance of human visual pattern recognition. *Perception & Psychophysics, 60,* 65–81.

Diwadkar, V. A., & McNamara, T. P. (1997). Viewpoint dependence in scene recognition. *Psychological Science, 8,* 302–307.

Dobbins, A. C., Jeo, R. M., Fiser, J., & Allman, J. M. (1998, July 24). Distance modulation of neural activity in the visual cortex. *Science, 281,* 552–555.

Dominey, P. F., & Arbib, M. A. (1992). A cortico-subcortical model for generation of spatially accurate sequential saccades. *Cerebral Cortex, 2,* 153–175.

Downing, C. J., & Pinker, S. (1985). The spatial structure of visual attention. In M. I. Posner & O. S. M. Marin (Eds.), *Attention and performance: Vol. 11. Mechanisms of attention* (pp. 171–187). Hillsdale, NJ: Erlbaum.

Droulez, J., & Berthoz, A. (1991). A neural network model of sensoritopic maps with predictive short-term memory properties. *Proceedings of the National Academy of Sciences, USA, 88,* 9653–9657.

Duhamel, J. R., Bremmer, F., BenHamed, S., & Graf, W. (1997, October 23). Spatial invariance of visual receptive fields in parietal cortex neurons. *Nature, 389,* 845–848.

Duhamel, J. R., Colby, C. L., & Goldberg, M. E. (1991). Congruent representations of visual and somatosensory space in single neurons of monkey ventral intra-parietal cortex area (area VIP). In J. Paillard (Ed.), *Brain and space* (pp. 223–236). Oxford, England: Oxford University Press.

Duhamel, J. R., Colby, C. L., & Goldberg, M. E. (1992, January 3). The updating of the representation of visual space in parietal cortex by intended eye movements. *Science, 255,* 90–92.

Duhamel, J. R., Colby, C. L., & Goldberg, M. E. (1998). Ventral intraparietal area of the macaque: Congruent visual and somatic response properties. *Journal of Neurophysiology, 79,* 126–136.

Eacott, M. J., & Gaffan, D. (1991). The role of monkey inferior parietal cortex in visual discrimination of identity and orientation of shapes. *Behavioural Brain Research, 46,* 95–98.

Edelman, S. (1998). Representation is representation of similarities. *Behavioral and Brain Sciences, 21,* 449–498.

Edelman, S. (1999). *Representation and recognition in vision.* Cambridge, MA: MIT Press.

Edelman, S., & Bülthoff, H. H. (1992). Orientation dependence in the recognition of familiar and novel views of three-dimensional objects. *Vision Research, 32,* 2385–2400.

Edelman, S., & Intrator, N. (2000). (Coarse coding of shape fragments) + (retinotopy) ≈ representation of structure. *Spatial Vision, 13,* 255–264.

Edelman, S., & Intrator, N. (2001). A productive, systematic framework for the representation of visual structure. In T. K. Lean, T. G. Dietterich, & V. Tresp (Eds.), *Advances in neural information processing systems* (Vol. 13, pp. 10–16). Cambridge, MA: MIT Press.

Eriksen, C. W., & Hoffman, J. E. (1974). Selective attention: Noise suppression or signal enhancement? *Bulletin of the Psychonomic Society, 4,* 587–589.

Eriksen, C. W., & Murphy, T. D. (1987). Movement of attentional focus across the visual field: A critical look at the evidence. *Perception & Psychophysics, 42,* 299–305.

Faillenot, I., Decety, J., & Jeannerod, M. (1999). Human brain activity related to the perception of spatial features of objects. *NeuroImage, 10,* 114–124.

Faillenot, I., Toni, I., Decety, J., Grégoire, M.-C., & Jeannerod, M. (1997). Visual pathways for object-oriented action and object recognition: Functional anatomy with PET. *Cerebral Cortex, 7,* 77–85.

Farah, M. J. (2000). *The cognitive neuroscience of vision.* Oxford, England: Blackwell Publishers.

Farah, M. J., & Hammond, K. M. (1988). Mental rotation and orientation-invariant object recognition: Dissociable processes. *Cognition, 29,* 29–46.

Finke, R. A. (1989). *Principles of mental imagery.* Cambridge, MA: MIT Press.

Foster, D. H., & Gilson, S. J. (2002). Recognizing novel three-dimensional objects by summing signals from parts and views. *Proceedings of the Royal Society, London, B, 269,* 1939–1947.

Foster, D. H., & Kahn, J. I. (1985). Internal representations and operations in visual comparison of transformed patterns: Effects of pattern point-inversion, positional symmetry, and separation. *Biological Cybernetics, 51,* 305–312.

Galletti, C., & Battaglini, P. P. (1989). Gaze-dependent visual neurons in area V3A of monkey prestriate cortex. *Journal of Neuroscience, 9,* 1112–1125.

Gauthier, I., Hayward, W. G., Tarr, M. J., Anderson, A. W., Skudlarski, P., & Gore, J. C. (2002). BOLD activity during mental rotation and viewpoint-dependent object recognition. *Neuron, 34,* 161–171.

Gauthier, I., & Tarr, M. J. (1997). Orientation priming of novel shapes in the context of viewpoint-dependent recognition. *Perception, 26,* 51–73.

Georgopoulos, A. P. (2000). Neural mechanisms of motor cognitive processes: Functional MRI and neurophysiological studies. In M. S. Gazzaniga (Ed.), *The new cognitive neurosciences* (pp. 525–538). Cambridge, MA: MIT Press.

Georgopoulos, A. P., Lurito, J. T., Petrides, M., Schwartz, A. B., & Massey, J. T. (1989, January 13). Mental rotation of the neuronal population vector. *Science, 243,* 234–236.

Gibson, B. S., & Peterson, M. A. (1994). Does orientation-independent object recognition precede orientation-dependent recognition? Evidence from a cuing paradigm. *Journal of Experimental Psychology: Human Perception and Performance, 20,* 299–316.

Gibson, J. J. (1950). *The perception of the visual world.* Boston: Houghton Mifflin.

Goldstone, R. L. (1996). Alignment-based nonmonotonicities in similarity. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22,* 988–1001.

Goldstone, R. L., & Medin, D. L. (1994). Time course of comparison. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20,* 29–50.

Graf, M. (2002). *Form, space and object. Geometrical transformations in object recognition and categorization.* Berlin, Germany: Wissenschaftlicher Verlag Berlin.

Graf, M., Bundesen, C., & Schneider, W. X. (2006). *Topological transformations in basic level object categorization.* Manuscript submitted for publication.

Graf, M., Kaping, D., & Bülthoff, H. H. (2005). Orientation congruency effects for familiar objects: Coordinate transformations in object recognition. *Psychological Science, 16,* 214–221.

Graf, M., & Schneider, W. X. (2001). Structural descriptions in HIT—A problematic commitment. *Behavioral and Brain Sciences, 24,* 483–484.

Graziano, M. S. A., Hu, T. X., & Gross, C. G. (1997). Visuospatial properties of ventral premotor cortex. *Journal of Neurophysiology, 77,* 2268–2292.

Graziano, M. S. A., Yap, G. S., & Gross, C. G. (1994, November 11). Coding of visual space by premotor neurons. *Science, 266,* 1054–1057.

Hahn, U., Chater, N., & Richardson, L. B. (2003). Similarity as transformation. *Cognition, 87,* 1–32.

Hasselmo, M. E., Rolls, E. T., Baylis, G. C., & Nalwa, V. (1989). Object-centered encoding of face-selective neurons in the cortex in the superior temporal sulcus of the monkey. *Experimental Brain Research, 75,* 417–429.

Hayward, W. G. (2003). After the viewpoint debate: Where next in object recognition? *Trends in Cognitive Sciences, 7,* 425–427.

Hayward, W. G., & Tarr, M. J. (1997). Testing conditions for viewpoint invariance in object recognition. *Journal of Experimental Psychology: Human Perception and Performance, 23,* 1511–1521.

Hayward, W. G., & Tarr, M. J. (2000). Differing views on views: Comments on Biederman & Bar (1999). *Vision Research, 28,* 3895–3899.

Hayward, W. G., & Williams, P. (2000). Viewpoint dependence and object discriminability. *Psychological Science, 11,* 7–12.

Heil, M., Rauch, M., & Hennighausen, E. (1998). Response preparation begins before mental rotation is finished: Evidence from event-related brain potentials. *Acta Psychologica, 99,* 217–232.

Heil, M., Rösler, F., Link, M., & Bajric, J. (1998). What is improved if a mental rotation task is repeated—The efficiency of memory access, or the speed of a transformation routine? *Psychological Research, 61,* 99–106.

Helbig, H. B., Graf, M., & Kiefer, M. (2006). The role of action representations in visual object recognition. *Experimental Brain Research, 174,* 221–224.

Hill, H., Schyns, P. G., & Akamatsu, S. (1997). Information and viewpoint dependence in face recognition. *Cognition, 62,* 201–222.

Hinton, G. E. (1981). A parallel computation that assigns canonical object-based frames of reference. *Proceedings of the seventh international joint conference on artificial intelligence,* 683–685.

Hinton, G. E., & Parsons, L. M. (1981). Frames of reference and mental imagery. In J. Long & A. Baddeley (Eds.), *Attention and performance* (Vol. 9, pp. 261–277). Hillsdale, NJ: Erlbaum.

Hommel, B., Müsseler, J., Aschersleben, G., & Prinz, W. (2001). The theory of event coding (TEC): A framework for perception and action planning. *Behavioral and Brain Sciences, 24,* 849–937.

Hughes, H. C., & Zimba, L. D. (1985). Spatial maps of directed visual attention. *Journal of Experimental Psychology: Human Perception and Performance, 11,* 409–430.

Hughes, H. C., & Zimba, L. D. (1987). Natural boundaries for the spatial spread of directed visual attention. *Neuropsychologia, 25,* 5–18.

Hummel, J. E. (2000). Where view-based theories of human object recognition break down: The role of structure in human shape perception. In E. Dietrich & A. B. Markman (Eds.), *Cognitive dynamics: Conceptual change in humans and machines* (pp. 157–185). Hillsdale, NJ: Erlbaum.

Hummel, J. E., & Biederman, I. (1992). Dynamic binding in a neural network for shape recognition. *Psychological Review, 99,* 480–517.

Hummel, J. E., & Stankiewicz, B. J. (1998). Two roles for attention in shape perception: A structural description model of visual scrutiny. *Visual Cognition, 5,* 49–79.

Humphrey, G. K., & Jolicoeur, P. (1993). An examination of the effects of axis foreshortening, monocular depth cues, and visual field on object

identification. *Quarterly Journal of Experimental Psychology, 46A,* 137–159.

Humphreys, G. W., & Forde, E. M. E. (2001). Hierarchies, similarity, and interactivity in object recognition: "Category-specific" neuropsychological deficits. *Behavioral and Brain Sciences, 24,* 453–509.

Hung, C. P., Kreiman, G., Poggio, T., & DiCarlo, J. J. (2005, November 4). Fast readout of object identity from macaque inferior temporal cortex. *Science, 310,* 863–866.

Jagaroo, V. (2004). Mental rotation and the parietal question in functional neuroimaging: A discussion of two views. *European Journal of Cognitive Psychology, 16,* 717–728.

Jay, M. F., & Sparks, D. L. (1984, May 24). Auditory receptive fields in primate superior colliculus shift with changes in eye position. *Nature, 309,* 345–347.

Johnson, J. S., & Olshausen, B. A. (2003). Timecourse of neural signatures of object recognition. *Journal of Vision, 3,* 499–512. Retrieved September 14, 2003, from http://journalofvision.org/3/7/4/

Jolicoeur, P. (1985). The time to name disoriented natural objects. *Memory & Cognition, 13,* 289–303.

Jolicoeur, P. (1987). A size-congruency effect in memory for visual shape. *Memory & Cognition, 15,* 531–543.

Jolicoeur, P. (1988). Mental rotation and the identification of disoriented objects. *Canadian Journal of Psychology, 42,* 461–478.

Jolicoeur, P. (1990a). Identification of disoriented objects: A dual-systems theory. *Mind & Language, 5,* 387–410.

Jolicoeur, P. (1990b). Orientation congruency effects on the identification of disoriented shapes. *Journal of Experimental Psychology: Human Perception and Performance, 16,* 351–364.

Jolicoeur, P. (1992). Orientation congruency effects in visual search. *Canadian Journal of Psychology, 46,* 280–305.

Jolicoeur, P., & Cavanagh, P. (1992). Mental rotation, physical rotation, and surface media. *Journal of Experimental Psychology: Human Perception and Performance, 18,* 371–384.

Jolicoeur, P., Corballis, M. C., & Lawson, R. (1998). The influence of perceived rotary motion on the recognition of rotated objects. *Psychonomic Bulletin & Review, 5,* 140–146.

Jolicoeur, P., & Humphrey, G. K. (1998). Perception of rotated two-dimensional and three-dimensional objects and visual shapes. In V. Walsh & J. Kulikowski (Eds.), *Perceptual constancy. Why things look as they do* (pp. 69–123). Cambridge, England: Cambridge University Press.

Jolicoeur, P., & Milliken, B. (1989). Identification of disoriented objects: Effects of context of prior presentation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15,* 200–210.

Jouffrais, C., & Boussaoud, D. (1999). Neuronal activity related to eye–hand coordination in the primate premotor cortex. *Experimental Brain Research, 128,* 205–209.

King, A. J., & Schnupp, J. W. H. (2000). Sensory convergence in neural function and development. In M. S. Gazzaniga (Ed.), *The new cognitive neurosciences* (pp. 437–450). Cambridge, MA: MIT Press.

Klein, F. (1893). Vergleichende Betrachtungen über neuere geometrische Forschungen [A comparative review of recent researches in geometry]. *Mathematische Annalen, 43,* 63–100. (Original work published 1872)

Koriat, A., & Norman, J. (1984). What is rotated in mental rotation? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 10,* 421–434.

Koriat, A., & Norman, J. (1988). Frames and images: Sequential effects in mental rotation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14,* 93–111.

Koriat, A., & Norman, J. (1989). Establishing global and local correspondence between successive stimuli: The holistic nature of backward alignment. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15,* 480–494.

Kosslyn, S. M. (1981). The medium and the message in mental imagery: A theory. *Psychological Review, 88,* 46–66.

Kosslyn, S. M. (1994). *Image and brain.* Cambridge, MA: MIT Press.

Kosslyn, S. M., Alpert, N. M., Thompson, W. L., Chabris, C. F., Rauch, S. L., & Anderson, A. K. (1994). Identifying objects seen from different viewpoints. A PET investigation. *Brain, 117,* 1055–1071.

Kourtzi, Z., & Shiffrar, M. (2001). Visual representation of malleable and rigid objects that deform as they rotate. *Journal of Experimental Psychology: Human Perception and Performance, 27,* 335–355.

LaBerge, D., & Brown, V. (1989). Theory of attentional operations in shape identification. *Psychological Review, 96,* 101–124.

Lamberts, K., Brockdorff, N., & Heit, E. (2002). Perceptual processes in matching and recognition of complex pictures. *Journal of Experimental Psychology: Human Perception and Performance, 28,* 1176–1191.

Lamme, V. A. F. (2003). Why visual attention and awareness are different. *Trends in Cognitive Sciences, 7,* 12–18.

Lamme, V. A. F., & Roelfsema, P. R. (2000). The distinct modes of vision offered by feedforward and recurrent processing. *Trends in Neurosciences, 23,* 571–579.

Landau, B. (1994). Object shape, object name, and object kind: Representation and development. In D. L. Medin (Ed.), *The psychology of learning and motivation* (Vol. 31, pp. 253–304). New York: Academic Press.

Larsen, A., & Bundesen, C. (1978). Size scaling in visual pattern recognition. *Journal of Experimental Psychology: Human Perception and Performance, 4,* 1–20.

Larsen, A., & Bundesen, C. (1998). Effects of spatial separation in visual pattern matching: Evidence on the role of mental translation. *Journal of Experimental Psychology: Human Perception and Performance, 24,* 719–731.

Lawson, R. (1999). Achieving visual object constancy across plane rotation and depth rotation. *Acta Psychologica, 102,* 221–245.

Lawson, R., & Humphreys, G. W. (1996). View-specificity in object processing: Evidence from picture matching. *Journal of Experimental Psychology: Human Perception and Performance, 22,* 395–416.

Lawson, R., & Humphreys, G. W. (1998). View-specific effects of depth rotation and foreshortening on the initial recognition and priming of familiar objects. *Perception & Psychophysics, 60,* 1052–1066.

Lawson, R., Humphreys, G. W., & Jolicoeur, P. (2000). The combined effects of plane disorientation and foreshortening on picture naming: One manipulation or two? *Journal of Experimental Psychology: Human Perception and Performance, 26,* 568–581.

Lawson, R., Humphreys, G. W., & Watson, D. G. (1994). Object recognition under sequential viewing conditions: Evidence for viewpoint-specific recognition procedures. *Perception, 23,* 595–614.

Lawson, R., & Jolicoeur, P. (1998). The effects of plane rotation on the recognition of brief masked pictures of familiar objects. *Memory & Cognition, 26,* 791–803.

Lawson, R., & Jolicoeur, P. (1999). The effect of prior experience on recognition thresholds for plane-disoriented pictures of familiar objects. *Memory & Cognition, 27,* 751–758.

Lindsay, P. H., & Norman, D. A. (1972). *An introduction to psychology.* New York: Academic Press.

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

Logothetis, N. K., & Sheinberg, D. L. (1996). Visual object recognition. *Annual Review of Neuroscience, 19,* 577–621.

Lowe, D. G. (1985). *Perceptual organization and visual recognition.* Boston: Kluwer Academic.

Lowe, D. G. (1987). Three-dimensional object recognition from single two-dimensional images. *Artificial Intelligence, 31,* 355–395.

Lueschow, A., Miller, E. K., & Desimone, R. (1994). Inferior temporal mechanisms for invariant object recognition. *Cerebral Cortex, 5,* 523–531.

Lurito, T., Georgakopoulos, T., & Georgopoulos, A. P. (1991). Cognitive spatial–motor processes: VII. The making of movements at an angle from a stimulus direction: Studies of motor cortical activity at the single cell and population levels. *Experimental Brain Research, 87,* 562–580.

Mack, A., & Rock, I. (1998). *Inattentional blindness.* Cambridge, MA: MIT Press.

MacLennan, B. J. (1997). Field computation in motor control. In P. G. Morasso & V. Sanguineti (Eds.), *Self-organization, computational maps and motor control* (pp. 37–74). North-Holland: Elsevier.

MacLennan, B. J. (1999). Field computation in natural and artificial intelligence. *Information Sciences, 119,* 73–89. Retrieved March 5, 2004 from http://www.cs.utk.edu/~mclennan/fieldcomp.html

Markman, A. B., & Gentner, D. (1993). Splitting the differences: A structural alignment view of similarity. *Journal of Memory and Language, 32,* 517–535.

Markman, A. B., & Wisniewski, E. J. (1997). Similar and different: The differentiation of basic-level categories. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23,* 54–70.

Marr, D., & Nishihara, H. K. (1978). Representation and recognition of the spatial organization of three-dimensional shapes. *Proceedings of the Royal Society, London, B, 200,* 269–294.

Mays, L. E., & Sparks, D. L. (1980). Dissociation of visual and saccade-related responses in superior colliculus neurons. *Journal of Neurophysiology, 43,* 207–232.

McAdams, C. J., & Maunsell, J. H. R. (1999). Effects of attention on orientation-tuning functions of single neurons in macaque cortical area V4. *Journal of Neuroscience, 19,* 431–441.

McMullen, P., Hamm, J., & Jolicoeur, P. (1995). Rotated object identification with and without orientation cues. *Canadian Journal of Experimental Psychology, 49,* 133–149.

McMullen, P. A., & Jolicoeur, P. (1990). The spatial frame of reference in object naming and discrimination of left–right reflections. *Memory & Cognition, 18,* 99–115.

Medin, D. L., Goldstone, R. L., & Gentner, D. (1993). Respects for similarity. *Psychological Review, 100,* 254–278.

Milliken, B., & Jolicoeur, P. (1992). Size effects in visual recognition memory are determined by perceived size. *Memory & Cognition, 20,* 83–95.

Milner, A. D., & Goodale, M. A. (1995). *The visual brain in action.* Oxford, England: Oxford University Press.

Morvan, C., & Wexler, M. (2005). Reference frames in early motion detection. *Journal of Vision,* 5, 131–138. Retrieved March 1, 2005 from http://jornalofvision.org/5/2/4/

Moses, Y., & Ullman, S. (1998). Generalization to novel views: Universal, class-based, and model-based processing. *International Journal on Computer Vision, 29,* 233–253.

Mumford, D. (1994). Neuronal architectures for pattern-theoretic problems. In C. Koch & J. L. Davis (Eds.), *Large-scale neuronal theories of the brain* (pp. 125–152). Cambridge, MA: MIT Press.

Murray, J. E. (1997). Flipping and spinning: Spatial transformation procedures in the identification of rotated natural objects. *Memory & Cognition, 25,* 96–105.

Murray, J. E. (1998). Is entry-level recognition viewpoint invariant or viewpoint dependent? *Psychonomic Bulletin & Review, 5,* 300–304.

Murray, J. E. (1999). Orientation-specific effects in picture matching and naming. *Memory & Cognition, 27,* 878–889.

Murray, J. E., Jolicoeur, P., McMullen, P. A., & Ingleton, M. (1993). Orientation-invariant transfer of training in the identification of rotated objects. *Memory & Cognition, 21,* 604–610.

Nakatani, C., Pollatsek, A., & Johnson, S. H. (2002). Viewpoint-dependent recognition of scenes. *The Quarterly Journal of Experimental Psychology, 55A,* 115–139.

Nazir, T. A., & O'Regan, J. K. (1990). Some results on translation invariance in the human visual system. *Spatial Vision, 5,* 81–100.

Newell, F. N., & Findlay, J. M. (1997). The effect of depth rotation on object identification. *Perception, 26,* 1231–1257.

Newell, F. N., Sheppard, D. M., Edelman, S., & Shapiro, K. L. (2005). The interaction of shape- and location-based priming in object categorization: Evidence for a hybrid "what + where" representation stage. *Vision Research, 45,* 2065–2080.

Nishida, S., Motoyoshi, I., Andersen, R. A., & Shimojo, S. (2003). Gaze modulation of visual aftereffects. *Vision Research, 43,* 639–649.

Olshausen, B. A., Anderson, C. H., & Van Essen, D. C. (1993). A neurobiological model of visual attention and invariant pattern recognition based on dynamic routing of information. *Journal of Neuroscience, 13,* 4700–4719.

Olshausen, B. A., Anderson, C. H., & Van Essen, D. C. (1995). A multiscale routing circuit for forming size- and position-invariant object representations. *The Journal of Computational Neuroscience, 2,* 45–62.

Op de Beeck, H., & Vogels, R. (2000). Spatial sensitivity of macaque inferior temporal neurons. *The Journal of Comparative Neurology, 426,* 505–518.

Oram, M. W., Földiák, P., Perrett, D. I., & Sengpiel, F. (1998). The "ideal homunculus": Decoding neural population signals. *Trends in Neurosciences, 21,* 259–265.

Palmer, S. E. (1983). The psychology of perceptual organization: A transformational approach. In J. Beck, B. Hope, & A. Rosenfeld (Eds.), *Human and machine vision* (pp. 269–339). New York: Academic Press.

Palmer, S. E. (1989). Reference frames in the perception of shape and orientation. In B. E. Shepp & S. Ballesteros (Eds.), *Object perception: Structure and process* (pp. 121–163). Hillsdale, NJ: Erlbaum.

Palmer, S. E. (1999). *Vision science. Photons to phenomenology.* Cambridge, MA: MIT Press.

Palmer, S. E., Rosch, E., & Chase, P. (1981). Canonical perspective and the perception of objects. In J. Long & A. Baddeley (Eds.), *Attention and performance* (Vol. 9, 135–151). Hillsdale, NJ: Erlbaum.

Palmeri, T. J., & Gauthier, I. (2004). Visual object understanding. *Nature Reviews Neuroscience, 5,* 1–13.

Pashler, H. (1990). Coordinate frame for symmetry detection and object recognition. *Journal of Experimental Psychology: Human Perception and Performance, 16,* 150–163.

Pellizzer, G., & Georgopoulos, A. P. (1993). Mental rotation of the intended direction of movement. *Current Directions in Psychological Science, 2,* 12–17.

Perrett, D. I., & Oram, M. W. (1998). Visual recognition based on temporal cortex cells: Viewer-centred processing of pattern configurations. *Zeitschrift für Naturforschung, C, 53,* 518–541.

Perrett, D. I., Oram, W. M., & Ashbridge, E. (1998). Evidence accumulation in cell populations responsive to faces: An account of generalization of recognition without mental transformations. *Cognition, 67,* 111–145.

Perrett, D. I., Oram, W. M., Harries, M. H., Bevan, R., Hietanen, J. K., Benson, P. J., & Thomas, S. (1991). Viewer-centred and object-centred coding of heads in the macaque temporal cortex. *Experimental Brain Research, 86,* 159–173.

Perrett, D. I., Smith, P. A., 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, London, B, 223,* 293–317.

Pickering, M. J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences, 27,* 169–225.

Pinker, S. (1997). *How the mind works.* London: Penguin Press.

Pitts, W., & McCulloch, W. S. (1947). How we know universals: The perception of auditory and visual forms. *Bulletin of Mathematical Biophysics, 9,* 127–147.

Poggio, T. (1990). A theory of how the brain might work. *The Brain: Cold Spring Harbor Symposia on Quantitative Biology* (pp. 899–910). New York: CSH Laboratory Press.

Poggio, T., & Edelman, S. (1990, January 18). A network that learns to recognize three-dimensional objects. *Nature, 343,* 263–266.

Poggio, T., & Girosi, F. (1990). Regularization algorithms for learning that are equivalent to multilayer networks. *Science, 247,* 978–982.

Posner, M. I., Snyder, C. R. R., & Davidson, B. J. (1980). Attention and the detection of signals. *Journal of Experimental Psychology: General, 109,* 160–174.

Pouget, A., Deneve, S., & Duhamel, J.-R. (2002). A computational perspective on the neural basis of multisensory spatial representations. *Nature Reviews Neuroscience, 3,* 741–747.

Pouget, A., & Sejnowski, T. J. (1997). Spatial transformations in the parietal cortex using basis functions. *The Journal of Cognitive Neuroscience, 9,* 222–237.

Pouget, A., & Sejnowski, T. J. (2001). Simulating a lesion in a basis function model of spatial representations: Comparison with hemineglect. *Psychological Review, 108,* 653–673.

Pouget, A., & Sejnowski, T. J. (2005). Dynamic remapping. In M. A. Arbib (Ed.), *The handbook of brain theory and neural networks* (2nd ed., pp. 335–338). Cambridge, MA: MIT Press.

Prinz, W. (1990). A common coding approach to perception and action. In O. Neumann & W. Prinz (Eds.), *Relationships between perception and action: Current approaches* (pp. 167–201). Berlin, Germany: Springer-Verlag.

Prinz, W. (1997). Perception and action planning. *European Journal of Cognitive Psychology, 9,* 129–154.

Pylyshyn, Z. W. (1981). The imagery debate: Analogue media vs. tacit knowledge. *Psychological Review, 88,* 16–45.

Pylyshyn, Z. W. (2002). Mental imagery: In search of a theory. *Behavioral and Brain Sciences, 25,* 157–238.

Reisberg, D., & Chambers, D. (1991). Neither pictures nor propositions: What can we learn from a mental image? *Canadian Journal of Psychology, 45,* 336–352.

Remington, R., & Pierce, L. (1984). Moving attention: Evidence for time-invariant shifts of visual selective attention. *Perception & Psychophysics, 35,* 393–399.

Rieger, J. W., Köchy, N., Schalk, F., & Heinze, H. J. (2006). *Speed limits: Orientation and semantic context interactions constrain natural scene discrimination.* Manuscript submitted for publication.

Riesenhuber, M., & Poggio, T. (1999). Hierarchical models of object recognition in cortex. *Nature Neuroscience, 2,* 1019–1025.

Riesenhuber, M., & Poggio, T. (2002). Neural mechanisms of object recognition. *Current Opinion in Neurobiology, 12,* 162–168.

Robertson, L. C., Palmer, S. E., & Gomez, L. M. (1987). Reference frames in mental rotation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 13,* 368–379.

Rock, I. (1973). *Orientation and form.* New York: Academic Press.

Rock, I. (1974, January). The perception of disoriented figures. *Scientific American, 230,* 78–85.

Rock, I., & Heimer, W. (1957). The effect of retinal and phenomenal orientation on the perception of form. *American Journal of Psychology, 70,* 493–511.

Rolls, E. T., Aggelopoulos, N., & Zheng, F. (2003). The receptive fields of inferior temporal cortex neurons in natural scenes. *Journal of Neuroscience, 23,* 339–348.

Rosch, E., Mervis, C. B., Gray, W. D., Johnson, D. M., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology, 8,* 382–439.

Rousselet, G. A., Fabre-Thorpe, M., & Thorpe, S. J. (2002). Parallel processing in high-level categorization of natural images. *Nature Neuroscience, 5,* 629–630.

Ruthruff, E., & Miller, J. (1995). Can mental rotation begin before perception finishes? *Memory & Cognition, 23,* 408–424.

Sabes, P. N., Breznen, B., & Andersen, R. A. (2002). Parietal representation of object-based saccades. *Journal of Neurophysiology, 88,* 1815–1829.

Salinas, E. (2004). Fast remapping of sensory stimuli onto motor actions on the basis of contextual modulation. *Journal of Neuroscience, 24,* 1113–1118.

Salinas, E., & Abbott, L. F. (1995). Transfer of coded information from sensory to motor networks. *Journal of Neuroscience, 15,* 6461–6474.

Salinas, E., & Abbott, L. F. (1997a). Attentional gain modulation as a basis for translation invariance. In J. Bower (Ed.), *Computational neuroscience: Trends in research* (pp. 807–812). New York: Plenum Press.

Salinas, E., & Abbott, L. F. (1997b). Invariant visual responses from attentional gain fields. *Journal of Neurophysiology, 77,* 3267–3272.

Salinas, E., & Abbott, L. F. (2001). Coordinate transformations in the visual system: How to generate gain fields and what to compute with them. In M. A. L. Nicolelis (Ed.), *Advances in neural population coding: Progress in brain research* (Vol. 130, pp. 175–190). Amsterdam: Elsevier.

Salinas, E., & Sejnowski, T. J. (2001). Gain modulation in the central nervous system: Where behavior, neurophysiology, and computation meet. *The Neuroscientist, 7,* 430–440.

Salinas, E., & Thier, P. (2000). Gain modulation: A major computational principle of the central nervous system. *Neuron, 27,* 15–21.

Sanger, T. D. (1996). Probability density estimation for the interpretation of neural population codes. *Journal of Neurophysiology, 76,* 2790–2793.

Schacter, D. L., Cooper, L. A., & Delaney, S. M. (1990). Implicit memory for unfamiliar objects depends on access to structural descriptions. *Journal of Experimental Psychology: General, 119,* 5–24.

Schiller, P. H., & Lee, K. (1991, March 8). The role of the primate extrastriate area V4 in vision. *Science, 251,* 1251–1253.

Sclaroff, S. (1997). Deformable prototypes for encoding shape categories in image databases. *Pattern Recognition, 30,* 627–642.

Sclaroff, S., & Liu, L. (2001). Deformable shape detection and description via model-based region grouping. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 23,* 475–489.

Selfridge, O. G., & Neisser, U. (1963). Pattern recognition by machine. In E. A. Feigenbaum & J. Feldman (Eds.), *Computers and thought* (pp. 237–250). New York: McGraw-Hill.

Shelton, A. L., & McNamara, T. P. (1997). Multiple views of spatial memory. *Psychonomic Bulletin & Review, 4,* 102–106.

Shenoy, K. V., Bradley, D. C., & Andersen, R. A. (1999). Influence of gaze rotation on the visual response of primate MSTd neurons. *Journal of Neurophysiology, 81,* 2764–2786.

Shepard, R. N. (1994). Perceptual–cognitive universals as reflections of the world. *Psychonomic Bulletin & Review, 1,* 2–28.

Shepard, R. N., & Cooper, L. A. (1982). *Mental images and their transformations.* Cambridge, MA: MIT Press.

Shepard, R. N., & Hurwitz, S. (1984). Upward direction, mental rotation, and discrimination of left and right turns in maps. *Cognition, 18,* 161–193.

Shepard, R. N., & Metzler, J. (1971, February 19). Mental rotation of three-dimensional objects. *Science, 171,* 701–703.

Shulman, G. L., Remington, R. W., & McLean, J. P. (1979). Moving attention through physical space. *Journal of Experimental Psychology: Human Perception and Performance, 5,* 522–526.

Simion, F., Bagnara, S., Roncato, S., & Umiltà, C. (1982). Transformation processes upon the visual code. *Perception & Psychophysics, 31,* 13–25.

Sinha, P., & Poggio, T. (1996, December 5). Role of learning in three-dimensional form perception. *Nature, 384,* 460–463.

Snodgrass, J. G., & Vanderwart, M. (1980). A standardized set of 260 pictures: Norms for name agreement, image agreement, familiarity, and visual complexity. *Journal of Experimental Psychology: Human Learning and Memory, 6,* 174–215.

Snyder, L. H. (2000). Coordinate transformations for eye and arm movements in the brain. *Current Opinion in Neurobiology, 10,* 747–754.

Sparks, D. L., & Nelson, J. S. (1987). Sensory and motor maps in the mammalian superior colliculus. *Trends in Neuroscience, 10,* 312–317.

Sperling, G., & Weichselgartner, E. (1995). Episodic theory of the dynamics of spatial attention. *Psychological Review, 102,* 503–532.

Srinivas, K. (1993). Perceptual specificity in nonverbal priming. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19,* 582–602.

Stein, B. E., Wallace, M. T., & Stanford, T. R. (2000). Merging sensory signals in the brain: The development of multisensory integration in the superior colliculus. In M. S. Gazzaniga (Ed.), *The new cognitive neurosciences* (pp. 55–71). Cambridge, MA: MIT Press.

Sternberg, S. (1998). Discovering mental processing stages: The method of additive factors. In D. Scarborough & S. Sternberg (Eds.), *An invitation to cognitive science: Vol. 4. Methods, models, and conceptual issues* (pp. 703–863). Cambridge, MA: MIT Press.

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

Stricanne, B., Andersen, R. A., & Mazzoni, P. (1996). Eye-centered, head-centered, and intermediate coding of remembered sound locations in area LIP. *Journal of Neurophysiology, 76,* 2071–2076.

Sugio, T., Inui, T., Matsuo, K., Matsuzawa, M., Glover, G. H., & Nakai, T. (1999). The role of the posterior parietal cortex in human object recognition: A functional magnetic resonance imaging study. *Neuroscience Letters, 276,* 45–48.

Sutherland, N. S. (1968). Outlines of a theory of visual pattern recognition in animals and man. *Proceedings of the Royal Society, London, B, 171,* 297–317.

Tarr, M. J. (1995). Rotating objects to recognize them: A case study on the role of viewpoint dependency in the recognition of three-dimensional objects. *Psychonomic Bulletin & Review, 2,* 55–82.

Tarr, M. J. (2003). Visual object recognition: Can a single mechanism suffice? In M. A. Peterson & G. Rhodes (Eds.), *Perception of faces, objects, and scenes: Analytic and holistic processes* (pp. 177–211). Oxford, England: Oxford University Press.

Tarr, M. J., & Bülthoff, H. H. (1995). Is human object recognition better described by geon structural descriptions or by multiple views? Comment on Biederman and Gerhardstein (1993). *Journal of Experimental Psychology: Human Perception and Performance, 21,* 1494–1505.

Tarr, M. J., & Bülthoff, H. H. (1998). Image-based object recognition in man, monkey and machine. In M. J. Tarr & H. H. Bülthoff (Eds.), *Object recognition in man, monkey, and machine* (pp. 1–20). Cambridge, MA: MIT Press.

Tarr, M. J., & Gauthier, I. (1998). Do viewpoint-dependent mechanisms generalize across members of a class? *Cognition, 67,* 71–109.

Tarr, M. J., & Pinker, S. (1989). Mental orientation and orientation-dependence in shape recognition. *Cognitive Psychology, 21,* 233–282.

Tarr, M. J., & Pinker, S. (1990). When does human object recognition use a viewer-centered reference frame? *Psychological Science, 1,* 253–256.

Tarr, M. J., Williams, P., Hayward, W. G., & Gauthier, I. (1998). Three-dimensional object recognition is viewpoint-dependent. *Nature Neuroscience, 1,* 275–277.

Thoma, V., Hummel, J. E., & Davidoff, J. (2004). Evidence for holistic representations of ignored images and analytic representations of attended images. *Journal of Experimental Psychology: Human Perception and Performance, 30,* 257–267.

Thorpe, S., Fize, D., & Marlot, C. (1996, June 6). Speed of processing in the human visual system. *Nature, 381,* 520–522.

Torralba, A., & Oliva, A. (2003). Statistics of natural image categories. *Computation in Neural Systems, 14,* 391–412.

Treue, S., & Martinez Trujillo, J. C. (1999, June 10). Feature-based attention influences motion processing gain in macaque visual cortex. *Nature, 399,* 575–579.

Trotter, Y., & Celebrini, S. (1999, March 18). Gaze direction controls response gain in primary visual-cortex neurons. *Nature, 398,* 239–242.

Tsal, Y. (1983). Movements of attention across the visual field. *Journal of Experimental Psychology: Human Perception and Performance, 9,* 523–530.

Tucker, M., & Ellis, R. (2001). The potentiation of grasp types during visual object categorization. *Visual Cognition, 8,* 769–800.

Ullman, S. (1989). Aligning pictorial descriptions: An approach to object recognition. *Cognition, 32,* 193–254.

Ullman, S. (1995). Sequence-seeking and counter streams: A computational model for bi-directional information flow in the visual cortex. *Cerebral Cortex, 5,* 1–11.

Ullman, S. (1996). *High-level vision. Object recognition and visual cognition.* Cambridge, MA: MIT Press.

Ullman, S., & Basri, R. (1991). Recognition by linear combinations of models. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 13,* 992–1006.

Ungerleider, L. G., & Haxby, J. V. (1994). "What" and "where" in the human brain. *Current Opinion in Neurobiology, 4,* 157–165.

Ungerleider, L. G., & Mishkin, M. (1982). Two cortical visual systems. In D. J. Ingle, M. A. Goodale, & R. J. W. Mansfield (Eds.), *Analysis of visual behavior* (pp. 549–586). Cambridge, MA: MIT Press.

Van Gool, L. J., Moons, T., Pauwels, E., & Wagemans, J. (1994). Invariance from the Euclidean geometer's perspective. *Perception, 23,* 547–561.

Verfaillie, K. (1993). Orientation-dependent priming effects in the perception of biological motion. *Journal of Experimental Psychology: Human Perception and Performance, 19,* 992–1013.

Vetter, T., Hurlbert, A., & Poggio, T. (1995). View-based models of 3D object recognition: Invariance to imaging transformations. *Cerebral Cortex, 3,* 261–269.

Vogels, R., Biederman, I., Bar, M., & Lorincz, A. (2001). Inferior temporal neurons show greater sensitivity to nonaccidental than to metric shape differences. *Journal of Cognitive Neuroscience, 13,* 444–453.

Vuilleumier, P., Henson, R. N., Driver, J., & Dolan, R. J. (2002). Multiple levels of visual object constancy revealed by event-related fMRI of repetition priming. *Nature Neuroscience, 5,* 491–499.

Vuong, Q. C., & Tarr, M. (2004). Rotation direction affects object recognition. *Vision Research, 44,* 1717–1730.

Wagemans, J., Van Gool, L., & Lamote, C. (1996). The visual system's measurement of invariants need not itself be invariant. *Psychological Science, 7,* 232–236.

Wallis, G. (2002). The role of object motion in forging long-term representations of objects. *Visual Cognition, 9,* 233–247.

Wallis, G., & Bülthoff, H. (1999). Learning to recognize objects. *Trends in Cognitive Sciences, 3,* 22–31.

Wallis, G. M., & Bülthoff, H. H. (2001). Effect of temporal association on recognition memory. *Proceedings of the National Academy of Sciences, USA, 98,* 4800–4804.

Wallraven, C., & Bülthoff, H. H. (2001). Acquiring robust representations for recognition from image sequences. In B. Radig & S. Florczyk (Eds.), *Pattern recognition. Lecture Notes in Computer Science 2191* (pp. 216–222). Berlin, Germany: Springer.

Wang, G., Tanifuji, M., & Tanaka, K. (1998). Funcitional architecture in monkey inferotemporal cortex revealed by in vivo optical imaging. *Neuroscience Research, 32,* 33–46.

Warrington, E. K., & Taylor, A. M. (1973). The contribution of the right parietal lobe to object recognition. *Cortex, 9,* 152–164.

Warrington, E. K., & Taylor, A. M. (1978). Two categorical stages of object recognition. *Perception, 7,* 695–705.

Weiskrantz, L. (1990). Visual prototypes, memory, and the inferotemporal lobe. In E. Iwai & M. Mishkin (Eds.), *Vision, memory and the temporal lobe* (pp. 13–28). New York: Elsevier.

Weiskrantz, L., & Saunders, R. C. (1984). Impairments of visual object transforms in monkeys. *Brain, 107,* 1033–1072.

Willems, B., & Wagemans, J. (2001). Matching multicomponent objects

from different viewpoints: Mental rotation as normalization? *Journal of Experimental Psychology: Human Perception and Performance, 27,* 1090–1115.

Wilson, K. D., & Farah, M. J. (2003). When does the visual system use viewpoint-invariant representations during recognition? *Cognitive Brain Research, 16,* 399–415.

Wilson, M. (2002). Six views of embodied cognition. *Psychonomic Bulletin & Review, 9,* 625–636.

Witkin, A., Terzopoulos, D., & Kass, M. (1987). Signal matching through scale space. *International Journal of Computer Vision, 2,* 133–144.

Wöhrgötter, F., & Eysel, U. T. (2000). Context, state and the receptive fields of striatal cortex cells. *Trends in Neurosciences, 23,* 497–503.

Wörgötter, F., Suder, K., Zhao, Y., Kerscher, N., Eysel, U. T., & Funke, K. (1998, November 12). State-dependent receptive field restructuring in the visual cortex. *Nature, 396,* 165–168.

Yantis, S. (1988). On analog movements of visual attention. *Perception & Psychophysics, 43,* 203–206.

Zaki, S. R., & Homa, D. (1999). Concepts and transformational knowledge. *Cognitive Psychology, 39,* 69–115.

Zhang, K. (1996). Representation of spatial orientation by the intrinsic dynamics of the head-direction cell ensemble: A theory. *Journal of Neuroscience, 16,* 2112–2126.

Zimba, L. D., & Hughes, H. C. (1987). Distractor–target interactions during directed visual attention. *Spatial Vision, 2,* 117–149.

Zipser, D., & Andersen, R. A. (1988, February 25). A back-propagation programmed network that simulates response properties of a subset of posterior parietal neurons. *Nature, 331,* 679–684.