

The role of characteristic motion in object categorization

Fiona N. Newell

Department of Psychology, Trinity College,
Dublin, Ireland



Christian Wallraven

Max Planck Institute for Biological Cybernetics,
Tübingen, Germany



Susanne Huber

Friedrich Miescher Laboratory of the Max Planck Society,
Tübingen, Germany



We report three experiments where we investigated the role of movement in object recognition. Previous studies have suggested a distinct and separate mechanism for object motion encoding, related to the action or motor-based system. To date, however, the role of an object's motion in long-term memory representations has not been explicitly tested. Here we were specifically interested in whether an object's characteristic motion patterns are integrated with static properties in an object's representation in memory. To that end, we used a simple categorization task where novel objects were categorized on the basis of two static (color and shape) and two dynamic (action and path) properties. The "action" of an object referred to its intrinsic motion pattern, whereas "path" referred to an object's extrinsic motion pattern (i.e., the route an object took). In Experiment 1, we found that all properties were relevant for categorization with the exception of path. This result was not due to path being less salient than other properties (Experiment 2). In Experiment 3, we found that when the action property was redundant that path was now used for categorization, suggesting that path was not used with action in Experiment 1 because of temporal order effects. Our findings argue for a cue-integrated model of object representation in memory.

Keywords: object categorization, motion and shape, characteristic motion

Introduction

One of the most fundamental tasks for the human visual system is to recognize the objects that surround us in our world. In order to do this task it is assumed that we can make use of all an object's properties for optimal recognition performance. For example, an object's shape or part structure, color, or texture may reveal information unique to this object that can be used for recognition. However, many objects in our world are not stationary and the manner in which an object moves can often act as a unique signature for the identity of that object. Thus, it seems clear that our visual system should also make use of the way in which an object moves for the purposes of recognition. What is not obvious, however, is whether motion information is an alternative way of recognizing an object when static information is reduced or unavailable, or whether motion is integrated with static information into an object's representation in memory.

To date, most theories of object representation have concentrated on static cues for recognition (by static, we mean all object features that can be extracted from a single frame). Indeed, the two main current approaches to object representation are based on static properties. According to the structuralist approach, for example, objects are represented as unique relations between an object's parts or primitives (Marr, 1982; Biederman, 1987; Hummel, 2001). The holistic or image-based account, on the other hand,

proposes that objects are represented as a collection of views or snapshots in memory (Tarr & Pinker, 1989; Bühlhoff & Edelman, 1992; Tarr & Bühlhoff, 1995) and that object images are recognized based on their similarity to other stored views (Edelman & Bühlhoff, 1992; Lawson & Humphreys, 1996; Newell & Findlay, 1997) or other objects (see Edelman, 1999). Yet none of these approaches provides a detailed description of how movement information can be used for recognition.

It is long established that movement information in itself is an important cue for recognition. For example, motion information alone from point light displays can be enough to perceive a person walking (Johansson, 1973), to identify the person, their gender, and the weight a person is lifting (Kozlowski & Cutting, 1977; Cutting, Proffitt, & Kozlowski, 1978) and even whether the person is angry or not (Pollick, Lestou, Ryu, & Cho, 2002). Using a morphing procedure on point light displays, Giese and Lappe (2002) recently demonstrated that observers are remarkably sensitive to changes between similar movement patterns (e.g., walking to running) but not to changes between dissimilar movements, such as walking to boxing. Furthermore, Wallach and O'Connell (1953) have demonstrated in the kinetic depth effect that motion can reveal information about three-dimensional (3D) form in an otherwise ambiguous, static random dot pattern. Although studies on structure-from-motion and the kinetic depth effect provide evidence that the shape of an object can be determined for

recognition when only temporal information is available, the discrimination of even very simple objects from motion alone can be an effortful and attentionally demanding task (Cavanagh, Labianca, & Thornton, 2001). Furthermore, studies on biological motion or point-light displays do not explain whether motion is necessary for recognizing objects when static information is fully available.

Suggestions that motion is integrated with shape information for recognition have recently come from the literature on face perception. For example, Hill and Johnston (2001) found that movement information in a statically noninformative face can help determine the identity and gender of that face. In their studies they superimposed two different types of facial movement, rigid and non-rigid motion, from individual actors onto an average head. They reported that rigid motion (i.e., movement of the entire head) was a better indicator of the identity of the person, whereas non-rigid motion (i.e., the relative movement of the facial features) was better for sex classification. Facial movement has also been shown to improve the recognition of famous faces (Lander, Christie, & Bruce, 1999) when static information is reduced or at threshold. However, rigid movement has no effect on the recognition of unfamiliar faces relative to static images of those faces (Christie & Bruce, 1998), unless the unfamiliar faces are shown from novel views during test (e.g., Pike, Kemp, Towell, & Phillips, 1997). These findings demonstrate that static information alone is sufficient for the recognition of unfamiliar faces but that rigid motion can help derive a 3D representation leading to better generalization across novel face views. A recent study on face recognition has suggested a more important role for motion than was previously thought. Knappmeyer, Thornton, and Bülthoff (2003) morphed the identity of one individual's face into another face and superimposed the facial, non-rigid movement of one or the other individuals onto the face morphs. They found that the superimposed motion information caused a bias in the identification of the shape of a face.

We would argue that although faces can be identified based on motion characteristics, this occurred mainly when the shape of the face was not available (Bassili, 1978) from a novel viewpoint (Pike et al., 1997), degraded (Lander & Bruce, 2000), not easily discriminable (Knappmeyer et al., 2003), or redundant (Hill & Johnston, 2001), suggesting two alternative routes to recognition where motion information compensates for impoverished static information. In fact, O'Toole, Roark, and Abdi (2002) propose such a "two-route" model of face recognition arguing that the moving aspects of a face are encoded and represented separately from the static-based aspects of a face. Their model is based on evidence from neuroimaging studies suggesting functional separation of the motion and structural aspects of face perception in humans (e.g., Haxby, Hoffman, & Gobbini, 2002). Haxby et al. found that facial movement on the one hand activates the superior temporal sulcus (STS) area, whereas the more shape-based or invariant aspects of a face activate the fusiform gyrus.

We might also argue that the general benefit of movement on face perception per se is not surprising given that motion conveys information about a face necessary for social interaction (O'Toole et al., 2002; Haxby et al., 2002), such as speech (Massaro & Cohen, 1995) and expression recognition (Bassili, 1978; Kamachi, Bruce, Mukaida, Gyoba, Yoshikawa, & Akamatsu, 2001). Therefore, motion integration in faces may be an overlearned associated cue for recognizing faces only and may not generalize to other classes of objects. The recent findings of Mak and Vera (1999) provide evidence for a role of class familiarity in motion and shape integration. They reported that for older children, motion information was important in tasks with shapes that are commonly seen as moving in the real world, but not in static geometric shapes. For younger children, on the other hand, motion information was used for both shape types. Consequently, we suggest that it is only through the use of novel objects that the exact role of motion for general object recognition can be determined.

Nevertheless, some studies in the literature have suggested that motion patterns can affect the recognition of other types of objects apart from faces. For example, the direction of movement has been shown to influence the interpretation of ambiguous figures (Tinbergen, 1951; Bernstein & Cooper, 1997). More pertinently however, Stone (1998, 1999) has argued that objects are represented in visual memory as unique spatiotemporal signatures, where both information about form and movement are integrated. He investigated this notion in a task where participants were first required to learn unfamiliar 3D-amoeboid objects shown rotating in a constant direction. In a subsequent recognition task, he found that performance was reduced when the familiar direction of rotation of the target object was reversed (Stone, 1998). Furthermore, Stone demonstrated that the manner in which an object moves during learning can also bias the type of views represented for object recognition (Stone, 1999). Despite these findings, previous models of object recognition have ignored the role of motion and instead have concentrated on how objects are represented on the basis of their static information alone (Biederman, 1987; Biederman & Gerhardstein, 1993; Edelman & Bülthoff, 1992; Cutzu & Edelman, 1994; Tarr & Pinker, 1989). This may be because motion is often assumed to be an alternative route to recognition as suggested by studies in neurophysiology (Ungerleider & Mishkin, 1982; Felleman & Van Essen, 1991), neuropsychology (Goodale & Milner, 1992), and psychophysics (Kourtzi & Shiffrar, 2001; Kourtzi & Nakayama, 2002). Until now, however, the role of explicit, characteristic movement in an object's representation in memory has not been investigated.

Although the studies reported by Stone and others indicate the importance of motion on object recognition, they suggest (as do the studies on face recognition) that motion can be used as a direct cue for recognition when static properties are impoverished. In Stone's experiments specifically, the object stimuli used were novel amoeboid

shapes that are very difficult to recognize from a single stationary image. Recognition of this type of object class may well benefit from additional cues to its identity, such as exposure to other viewpoints or perceiving the manner in which it moves. In the real world, however, where objects are made of myriad shapes, which are readily identifiable, we rarely have to perform such a difficult task. Our question here was whether motion is used when static cues are available. Specifically, we asked whether the recognition of an object is impaired if that object is shown moving in a pattern inconsistent with its familiar motion pattern or whether only shape information prevails in object recognition.

For the purposes of our investigation, we chose a task that was indicative of the type of object recognition task we conduct in our daily lives, namely an object categorization task (Biederman, 1987). All objects in our experiments comprise four different properties: two static (form and color) and two dynamic. The two different types of dynamic patterns used were “path” and “action.” Path refers to the manner in which an object moves relative to some external reference, such as another object or its environment. This type of motion has also been referred to as extrinsic motion (Kersten 1998a, 1998b). The action motion refers to the movement of the object with respect to an internal reference frame, also known as intrinsic motion.

The aim of our study was to investigate the importance of motion cues in object categorization using novel object-shapes. In particular, we were interested in whether static and dynamic cues are used to the same extent in object categorization. Our reasoning behind the investigation was as follows. If an object is represented in terms of both static and motion information, then a change to any of these cues should cause a decrease in categorization performance. If, on the other hand, motion information is not integrated into an object’s representation but is instead an alternative way of recognizing an object, then any changes in motion should have no effect on performance, provided the static cues are unchanged.

Experiment 1

To study this research question, we used 3D objects that differed in shape, color, and intrinsic and extrinsic motion. The task for the participants was to first learn the prototype objects and then to categorize new exemplars in a test phase. We measured the degree of response bias when each of the four object properties differed from the prototype objects.

We predicted that if static properties alone define the category to which an object belongs, then only changes to shape or color should cause a decision change in the categorization of that object. In this case, a change in motion cues should not change the category to which the object belongs. Alternatively, if motion is equally important for

object recognition, then motion changes should also cause a change in categorization response.

Methods

Participants

Twenty-four persons from the Max-Planck Institute for Biological Cybernetics, Tübingen, and 16 undergraduate students from the department of psychology, Trinity College, Dublin, participated in this experiment for pay (€8.00/hr). Twenty-nine of the participants were female, aged from 18 to 39 years, with a mean age of 24.7 years. All participants gave written consent to partake in the study, and all had normal or corrected-to-normal vision.

Stimuli

Stimuli were created using 3D Studio Max 3.0 and rendered as 320×160 pixel avi sequences (Indeo-Codec) consisting of 300 frames with 30 frames/s. The shape of each object was defined by either a discontinuous or a continuous curve in two dimensions, and this contour was then rotated around the upright axis to create 3D objects. The four colors used were pure red, green, blue, and yellow. The motion properties of the objects were defined as follows: The four types of actions were created by a combination of either a swinging or a continuous rotation of an object around its upright or horizontal axis. Each action was completed four times during the movie sequence. For the path features, all paths had equal length and were determined by a combination of a rectangular or sinusoidal wave pattern and by a smooth or sharp loop, yielding four different types of path (see Figure 1).

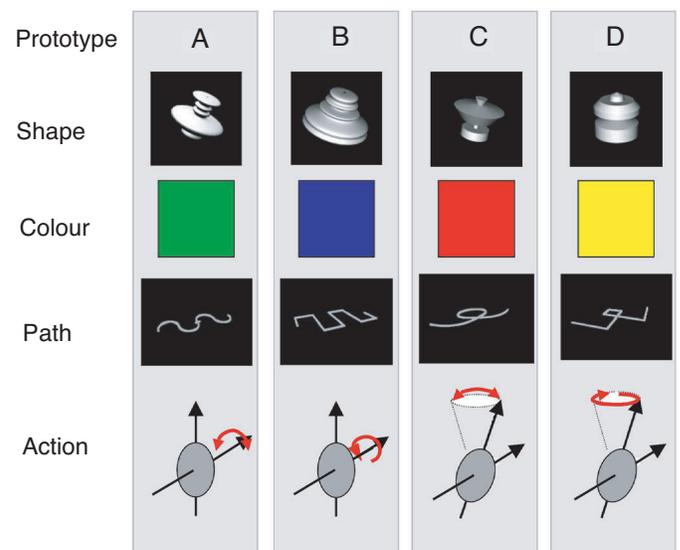
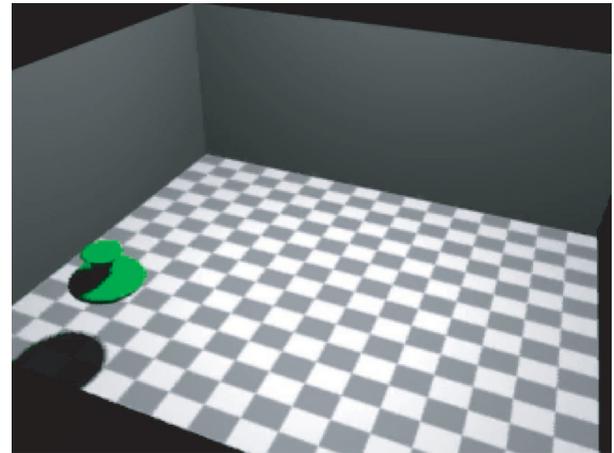


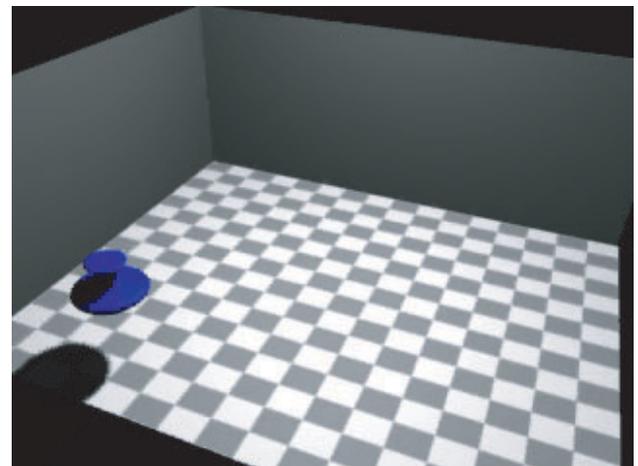
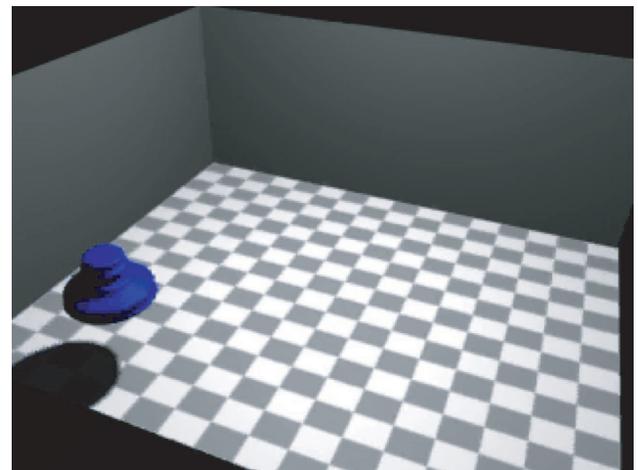
Figure 1. Plot showing the features that defined the four prototype objects.

Objects were placed in a “room” consisting of a checkboard-pattern floor and two grey walls with the start point of the sequence in one corner of the room and the end point in the opposite corner. A spotlight illuminated the scene from above to create a shadow of the object on the floor in order to facilitate the perception of depth and object motion. Prototypes were defined by selecting four different features, each from the set of shape, color, action, and path features. Thus, each prototype (A,B,C, and D) was uniquely defined along all dimensions. Exemplars were created by exchanging one or more features between prototypes, which yielded the whole set of stimuli for the experiments, creating a total set of 40 exemplar objects. The display size of each stimulus subtended a visual angle of 4.5° in the horizontal axis and 2.8° in the vertical axis.



Design

The experiment was based on a two-way mixed design with one between-subjects factor (paired prototypes learned) and one within-subjects factor (feature changes from prototype). The between-group factor had two levels (AB,CD prototype pairings or AC,BD prototype pairings). The within-group factor had five levels indicating the feature differences between the exemplar and the prototype (shape, color, path, action, and shape+color/path+action). The experiment consisted of two phases: a learning phase followed by a test phase. Feedback was provided during the learning trials and there was no indication of performance given during the test trials. To help the reader, the design of the experiment is illustrated in Table 1.



Movie 1-3. Prototype A (top panel), prototype B (middle panel), and exemplar with color change (AsBcApAa) (lower panel).

	Static cues		Dynamic cues	
	Shape	Color	Path	Action
Prototype A	A	A	A	A
Exemplar objects	B	A	A	A
	A	B	A	A
	A	A	B	A
	A	A	A	B
	A	A	B	B
	B	B	A	A
	A	B	B	B
	B	A	B	B
	B	B	A	B
	B	B	B	A
Prototype B	B	B	B	B

Table 1. An illustration of the design of Experiment 1.

Participants first learned the prototype objects in pairs. After prototype learning, participants were presented with the learning phase where exemplar objects that differed by one feature only from a prototype were presented. The feature change always belonged to the other prototype. Thus, for prototype pairs AB, a color change to prototype A (AsAcApAa) would be the color of prototype B (AsBcApAa) (see also Movie 1-3).

For each exemplar of a prototype, a feature change was either a motion feature (action or path) or a static feature (color or shape). Feature changes were counterbalanced across all participants. For each participant in group AB,CD prototype pairings, three of the four possible feature changes were learned and one feature was always tested without being learned. For example, if exemplars to the AB pair involved a path and color change, then exemplars to the CD pair would involve a path and shape change. Thus, in this case, action was never learned. The non-learned features were counterbalanced across participants in this group. The result of this design was that for each participant, one feature was not tested (e.g., in this case, path).

We changed the design slightly for the participants in the AC,BD group, such that features learned in one prototype pairing were always different from features learned in the other prototype pairing. Thus, for each participant, feature learning was counterbalanced across prototypes. Similarly, the test features were not presented during the learning phase. For example, if color and path changes were learned for the AC pair, then shape and action were learned for the BD pair. Furthermore, shape and action were tested for the AC pair, and color and path were tested for the BD pair. The learned features were randomly assigned across participants with the constraint that both a motion and static feature had to be learned for each prototype pair.

We measured the categorization decisions in our experiment. Reaction times were not collected due to the nature of the stimuli, which required viewing before responding.

Procedure

The prototype objects were first learned using a trial and error procedure. Each participant was shown a movie file and instructed to first learn the prototype object and then choose, by pressing one of two keys, the category to which the prototype belonged. Throughout the experiment, the participants were instructed to view the movies in their entirety and to respond as fast and as accurately thereafter to which category the object belonged. Six trials were presented and participants received “correct” or “incorrect response” feedback after each trial.

After prototype learning, participants were tested on their categorization of a subset of exemplars from each of the two categories. Each exemplar matched a prototype on three of four of the features. Feature changes to the exemplars were counterbalanced according to the constraints described above. The task for the participant was to correctly categorize each exemplar according to the category to which the nearest prototypes belonged. Participants received feedback throughout the learning phase. Following learning, participants were tested on new exemplars involving changes to features of prototypes that were not learned during the learning phase. There were 16 test trials per prototype pair (i.e., 32 total), which included exemplars where one feature differed from the prototype. A further 8 trials

per prototype pair included exemplars where the two static features matched one prototype and the two motion features matched the other prototype. There was no feedback given during the test trials.

The experiment consisted of two blocks and participants could take a self-timed break between blocks. Each block consisted of one pair of prototypes. The order of the blocks was counterbalanced across participants. The experimenter remained in the testing laboratory while the participant learned to categorize the prototype objects. Any questions were answered before the start of the learning phase of the experiment. Participants were told to categorize each object as accurately as possible and to consider all information present in the stimulus as relevant for categorization, without explicitly mentioning the features. Each participant completed the experiment in approximately 25 min.

For our data analyses, we used response bias as our dependent variable because a relative measure of accuracy was the most appropriate for our dataset. Each exemplar presented in our experiment differed from one prototype, say prototype A, by 1, 2, or 3 features (or 25%, 50%, and 75% feature changes). Our measure of response bias was then calculated by first measuring the mean frequency “prototype A” categorization responses as a function of the number of feature differences between the exemplar and the prototype. We then subtracted the actual number of feature changes from participants’ responses. We would expect, for example, that if all features were used for categorization, then the distribution of responses would reflect exactly the distribution of feature changes and subtraction would yield a score of zero. Otherwise, if one feature was used more often for categorization, for example, then we would expect a response bias or deviation in categorization responses away from the actual number of feature changes. Finally, in order to investigate whether some features were used more often for categorization than others, we calculated the response bias as a function of the type of feature changes (i.e., shape, color, path, and action).

Results

We found a total of 2.78% errors for the prototype objects. There were 5.89% errors made during the learning trials. The error rates across participants for all trials were then calculated as a bias from the actual percentage difference between the exemplar and the prototype. The mean percentage bias for each feature change is presented in [Figure 2](#). A positive bias means that the participants were sensitive to this feature and tended to overestimate changes from the prototype with changes to that feature. A negative bias, on the other hand, meant that the participants underestimated changes to this feature in their categorization decisions. The feature changes included shape, color, path, action, and the combined feature changes of shape and color or path and action. We included this last feature combination because participants were not explicitly

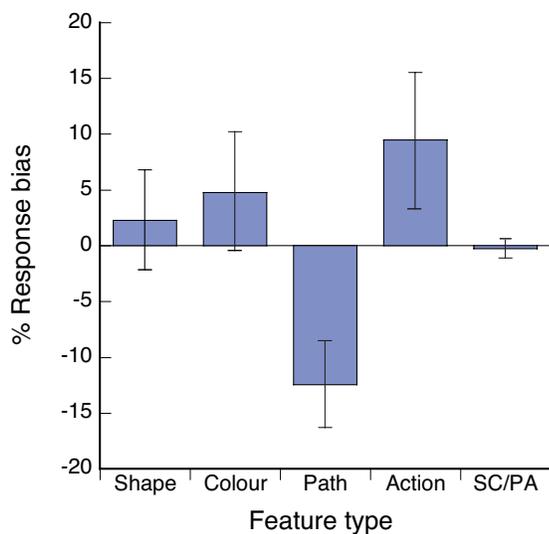


Figure 2. Plot showing percentage bias to each of the individual feature differences in exemplar objects from the prototype objects, and the combined static or motion feature differences in Experiment 1. A positive response bias indicates that participants were overusing the feature in their category judgments. A negative bias indicates that the feature was effectively underestimated for the purposes of categorization. Here we found that the feature “path” was not used for categorization. Error bars are SEM.

trained on two feature changes. In this case, if both motion features or both static features were ignored, then a response bias would be found. Otherwise, responses to either motion or static features would cancel each other out.

We conducted a two-way ANOVA using one between-factor (paired prototypes learned) and one within-factor (feature changes from prototype) on the percent bias responses across all trials (i.e., learning and test trials). We found no effect of paired prototype learned, $F(1, 38) = 1.49$, *ns*]. A main effect of feature was found, $F(4, 152) = 2.96$, $p < .05$]. A post hoc Newman-Keuls test revealed that the “path” feature was significantly different from all other features ($p < .05$) except the combined sc/ap feature. There were no other differences between the features. We found no interaction between the factors, $F(4, 152) = 0.62$, *ns*].

Using a nonparametric Sign test, we compared the extent of the response bias to each feature change against no bias (essentially against a bias of zero). None of the shape, color, and action features showed any significant difference from zero, indicating no evidence of a bias to these features (shape, $Z = 1.11$; color, $Z = 0.64$; and action, $Z = 0.81$) Also the exchange of two static or two dynamic features did not result in any bias (sc/ap, $Z = 1.31$). On the other hand, responses to the path feature were significantly different from zero, indicating a response bias to this feature ($Z = 3.95$, $p = .0001$).

In a further analysis of the data, we separated responses to the test trials only and again compared the bias to each

feature change against no bias. We conducted this analysis to ensure that our findings were not due to the feedback given in the learning phase of the experiment. Using Sign test we found no evidence of a bias to any of the feature changes except the path feature that was, again, significantly different from zero [$Z = 2.60$, $p = .009$].

Discussion

In this experiment, we found that observers used most information available, including intrinsic motion, in their category decisions. Extrinsic motion or path information, on the other hand, was not used for categorization. Instead, we found a negative response bias to the path feature, indicating that it was effectively ignored during categorization; participants used a path change in their categorization decisions much less frequently than any other feature change.

Several explanations are possible why path was less likely to be used for categorization than other features. First, path may simply be a perceptually less salient feature in our stimuli than color, shape, and action, and was, therefore, not sufficiently available for the purpose of categorization. We examined this possibility in Experiment 2. Alternatively, path might have been obscured or overshadowed by the second motion, action. When overshadowing occurs, a feature’s usefulness for discrimination is less likely to be attended to when a second similar feature is present (Gluck & Bower, 1988). As a result, action may have overshadowed path when action was sufficient to differentiate between the two categories. In Experiment 3, we made action redundant to test whether path, in this case, is used for categorization. Third, the time it takes for the different features to be revealed may have been a factor influencing the bias against path information. For example, color and shape can be instantly perceived by the observers, whereas action information is fully revealed after about 2.5 s, but path is only sufficiently revealed after about 5 s. Finally, path may not have been encoded for categorization due to the low familiarity of such a property in common objects. For example, path information in the real world is rarely diagnostic of object identity. Therefore, path may never be used for categorization because it is neither an ecologically valid nor familiar feature (see Mak & Vera, 1999). The following experiments were designed to elucidate reasons why path was not used for categorization in Experiment 1.

Experiment 2

In this second experiment, we tested for any potential differences in perceptual saliency across the four features that might explain why path was not used for categorization. If path was found to be less salient than the other features, this may explain our findings in Experiment 1.

The rationale behind the design of this experiment is based on a study described by Schwarzer (2000). The participant’s task was to rate the similarity of a pair of objects that always consisted of a prototype and an exemplar object

differing in one or three features. First, exemplars that had only one feature in common with the prototype (i.e., three feature differences) should always be judged as less similar than exemplars that shared three features with the prototype (i.e., one feature change). Furthermore, when the pair of objects differed in one feature only, the similarity ratings across these pairs should be the same irrespective of the type of feature type, provided these features are equally salient. For example, if a nonsalient feature defines the difference between a pair of objects, then this pair will be judged as more similar than if a salient feature defines the difference. On the other hand, if all features are equally salient then similarity ratings should be the same. For the path feature in particular, if this feature is not as salient as other shape, color, or action features, then two objects differing in path only will look more similar than two objects differing in, say, color. If path is equally salient, then similarity ratings to path differences should be the same as all other feature differences. The same logic applies to the alternative situation where, for example, path is the only feature in common to the pair of objects.

Method

Participants

Twelve members of the Max Planck Institute, Tübingen, participant list took part in this experiment for pay (about €4.00). Four of the participants were female, aged from 20 to 39 years, with a mean age of 29.3 years. All participants had normal or corrected-to-normal vision. None of these persons participated in the previous experiment, and all gave written consent to partake in the study.

Stimuli

See [Experiment 1](#) for a description of the stimuli. As in [Experiment 1](#), we again used 4 prototype pairings in our task: AB, CD, AC, and BD. A stimulus consisted of two moving objects (as described in [Experiment 1](#)) presented left and right of the center of the computer monitor. The two movies were simultaneously presented.

Design

The experiment was based on a two-way repeated measures design using number of feature differences (one or three) and feature type (shape, color, action, and path) as factors. In any one trial, two objects had to be rated for their perceived similarity. One of the objects was always a prototype object and the other an exemplar object. The position of the prototype (left or right of center screen) was counterbalanced across participants. The exemplar was either one feature or three features different from the prototype.

Procedure

Participants were instructed to rate the perceptual similarity of two objects using a Likert scale from 1 to 7, where a rating of 1 indicated a high degree of similarity. They

were encouraged to use the entire scale in their ratings. In each trial, one of the objects presented was a prototype object, the other object was an exemplar with either one feature or three features different from the prototype. The two objects were presented next to each other and started moving at the same time. Participants conducted four test blocks and the blocks differed in the pair of prototypes used (AB, AC, BD, or CD). The order of the test blocks was counterbalanced across participants. In each block, participants conducted two similarity ratings for each one- and three-feature change and each prototype resulting in 32 trials per block. The experiment took approximately 15 min to complete.

Results

The mean ratings per feature are shown in [Figure 3](#). Similarity ratings for one-feature difference were significantly higher than similarity ratings for object pairs with three-feature differences, $t(11) = 31.203, p < .001$.

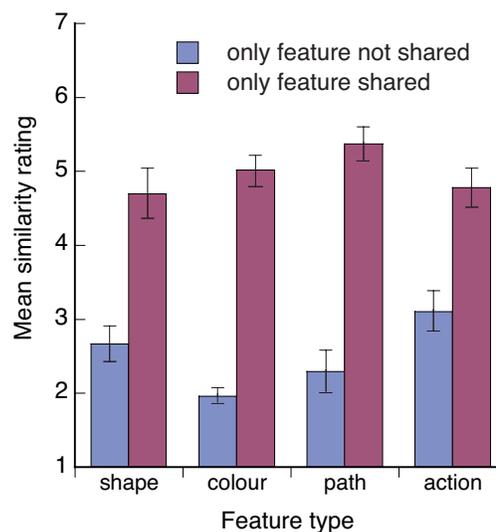


Figure 3. Plot showing similarity ratings to feature differences between an exemplar object and a prototype. A score of 1 indicates high similarity and 7 indicates low similarity. Feature differences were either one (“only feature not shared”) or three (“only feature shared”) between the prototype and the exemplar. Error bars are SEM.

We conducted one-way ANOVA on the ratings for pairs with one-feature difference and three-feature differences separately. There was a significant difference between the one-feature changes, $F(3, 33) = 3.8896, p < .05$. Post hoc Newman-Keuls analysis revealed that objects with only a color change were rated as significantly more similar than objects with an action change ($p < .05$). An ANOVA on the three-feature differences revealed no differences, $F(3, 33) = 1.3989, ns$. The ratings to each single feature change were compared to a perfect similarity rating of 1 using nonparametric analyses. Each of the shape, action, and path feature

changes was rated as significantly greater than 1 ($\chi^2 = 41.089$, $p < .0001$; $\chi^2 = 63.315$, $p < .0001$; and $\chi^2 = 30.6649$, $p < .002$, respectively). There was no difference found between the ratings to color changes and 1 ($\chi^2 = 14.215$, ns). A separate comparison between the three-feature changes and a dissimilarity rating of 7 revealed no significant differences (if the only feature shared is shape, $\chi^2 = 11.446$, ns ; color, $\chi^2 = 7.547$, ns ; action, $\chi^2 = 9.439$, ns ; and path, $\chi^2 = 5.8378$, ns). Therefore, not only was there no difference between each of the three feature changes (as revealed by ANOVA), but ratings to each of these object changes were not significantly different from the most dissimilar rating of 7.

Discussion

In terms of our aim in this experiment, the main result was that path was as perceptually salient as the other features. Exemplars were judged to be as dissimilar to the prototype if path was the only feature change than if any other features were changed.

Thus, we can say that path can be used for categorization just like any of the other features. A color change, however, seemed to have no effect on similarity judgments, in that a change of color only did not make object pairs look as dissimilar as object pairs with other feature changes. We are unsure why the effect occurred. Our observers commented that they did notice color changes, as well as other features, but for some reason decided that color changes were not as relevant to the similarity judgments as other shape or motion changes. Whatever the reason behind this finding, the important result is that path changes had an equal effect on similarity judgments as action and shape, indicating that path is at least as salient as these other features.

Experiment 3

The possibility still remains that when path is used together with action, the resulting motion reduces the likelihood that path information is used. In this experiment, all objects shared the same action pattern, leaving three features, color, shape and path, relevant for categorization. Therefore, action information is still present but all objects have the same action pattern. If path information is not used for object categorization, then responses should be related to changes in the static characteristics only. On the other hand, if all motion information is important for categorizing objects, then here path should be as important for category decisions as color or shape.

Method

Participants

Sixteen individuals from the Max Planck Institute, Tübingen, took part in this experiment for pay (€8.00/hr). Thirteen of the participants were female, aged from 18 to

30 years, with a mean age of 22.3 years. All participants were screened and had normal or corrected-to-normal vision. None of these persons participated in any of the previous experiments, and all gave written consent to partake in the study.

Stimuli

See [Experiment 1](#) for a description of the stimuli. Unlike in [Experiment 1](#), in this experiment, we used only 2 prototype pairings; AB and CD. The action information was the same for all prototypes, therefore only shape, color, and path were indicators of the category of the object.

Design

The design followed that outlined in [Experiment 1](#) for the AC,BD group of participants only. Feature changes were therefore counterbalanced across prototype pairs for each participant. The experiment was based on a repeated measures design with feature type as the main factor (shape, color, and path). Trials were randomly presented across participants according to the constraints of feature allocation to the learning and test trials described in [Experiment 1](#) above.

Procedure

The procedure followed that of [Experiment 1](#). The experiment lasted approximately 25 min for each participant.

Results

We found a total of 1.19% errors to the prototype objects and 15.63% errors during the learning trials. As in [Experiment 1](#), the error rates across participants for all trials were then calculated as a bias from the actual percentage difference between the exemplar and the prototype. The mean percentage bias for each feature change is presented in [Figure 4](#).

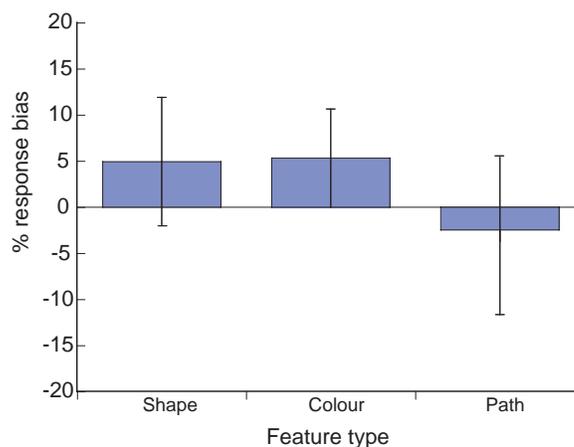


Figure 4. Plot showing percentage response bias to each of the individual feature differences in exemplar objects from the prototype objects in [Experiment 3](#). We found no evidence of a response bias to any of the feature differences, indicating that all features were relevant for categorization.

A one-way repeated measures ANOVA was conducted on the bias scores using feature type as the factor. There were three levels to the feature type factor, indicating the types of feature differences between the exemplar and prototype (shape, color, and path). We found no effect of feature type [$F(1, 3) < 1$]. We then compared the level of bias in each feature type to 0. We found no evidence of a bias to any of the features (shape, $Z = 0.25$, *ns*; color, $Z = 1.25$, *ns*; and path, $Z = .25$, *ns*). We separated the test trials from the data and conducted a one-way ANOVA using feature type as the factor. Here we found a main effect of feature type [$F(1, 3) = 2.99$, $p < .05$]. A post hoc Newman-Keuls test revealed that the biases for path and color were significantly different ($p < .05$) (see Figure 4). Sign tests on the test trials data revealed that responses to any of the features were not biased away from zero (shape, $Z = -0.25$, *ns*; color, $Z = 1.75$, *ns*; and path, $Z = 0.25$, *ns*).

Discussion

When the action feature was irrelevant for categorization, path was used as readily as shape and color to categorize the objects. Therefore, when path is a useful motion feature for categorization, it is used accordingly. During the test trials, we found that participants seemed to be more sensitive to changes in color of the objects than changes in path information, when making category decisions. This finding may have been due to our task instructions. We asked participants to view the movie and then categorize the object as fast and as accurately as possible. Because color information is rapidly perceived (relative to shape or motion), then the instructions may have biased participants to using color as a cue for categorization more often than path. Nevertheless, the responses to each of these features were not found to be significantly biased, indicating that despite differences between features, overall each feature was used for categorization.

In order to ensure that our participants did not adopt a simple response strategy to do the task, we tested participants' explicit knowledge of the individual properties of the objects after they had finished the experiment. The reasoning here was that if these named features correlated with the categorization responses, then participants were using some explicit, conscious "feature-listing" in order to perform the task. The percentage of features mentioned across participants was 34% "shape", 28% "color," and 38% "movement." This distribution of responses did not concur with the correct responses made during the experiment itself (Kendall's coefficient of concordance = 0.11, $p > .5$), suggesting that it is unlikely that participants adopted an explicit feature-listing approach to do the task.

General discussion

In the three experiments reported above, we found that an object's characteristic motion is as useful for categorization as any of its static properties. Unlike in previous stud-

ies in object perception where motion was found to be used when static information was unavailable (Johansson, 1973, 1988; Cutting et al., 1978), impoverished (Wallach & O'Connell, 1953; Lander et al., 1999), or redundant (Hill & Johnston, 2001), here we found that motion is used even when color and shape information is fully available. Each of the features used in our experiments was fully discriminable in the absence of other features (see Figure 1); therefore, motion was not merely useful for revealing shape information, for example.

In Experiment 1, we found some evidence for a response bias against extrinsic over intrinsic motion cues in object categorization. We established that this bias was not due to the path feature being less perceptually salient than other features (Experiment 2) or that it was obscured by the action feature (Experiment 3). Instead, we suspect that bias against path in Experiment 1 was due to a temporal order effect: All other information, such as color, shape, and action, is revealed earlier in the stimulus presentation than path information, and this may have influenced the response decision in favor of these "earlier" features. Furthermore, perceiving information about an object's path may be more effortful than perceiving action information because path is revealed only by integrating the object's position over a relatively long time (i.e., between 5 to 10 s). Action, on the other hand, is revealed in a quarter of the time of path and, moreover, it is presented 4 times more often than path information. Consequently, we suggest that temporal order, or even repetition effects, may have affected a bias in response toward action information relative to path information. Despite the temporal differences, however, path information is a perceptually salient property like other action, color, and shape properties (Experiment 2) and can be useful for categorization when other motion information is noninformative (Experiment 3).

The idea that motion and shape information is integrated is also supported by recent neuroimaging and unit recording studies. For example, it is known that more than 90% of neurons in middle temporal visual area (MT) and medial superior temporal visual area (MST) of monkeys are selective for direction of motion (Dubner & Zeki, 1971; Tanaka & Saito, 1989). However, a recent fMRI imaging study has shown that motion sensitive areas (MT+/V5) in humans are activated by static images of objects if movement is implied in these images (Kourtzi & Kanwisher, 2000). For example, Kourtzi and Kanwisher found that the MT/MST area was activated to a static image of a basketball player about to throw the ball, but not to the same basketball player shown standing. Kourtzi and Kanwisher concede that activation may be due to motion imagery; however, they also suggest that some object categories may activate cortical areas highly associated with those categories. Therefore, static images of objects that move in the real world may activate motion areas through learning.

Other studies from neurophysiology reveal a candidate cortical area likely to be directly involved in shape and motion integration. Tanaka, Koyama, and Mikami (2002), for

example, recorded activation in neurons in the STS of monkeys when presented with moving objects. The STS has been proposed as the candidate area for neuronal convergence from the dorsal (“where” or “how”) and ventral (“what”) streams (see Oram & Perrett, 1996). Tanaka and colleagues investigated whether neurons in STS are sensitive to shape or motion or both. Prior to testing, they trained monkeys on a set of moving object stimuli. As in our study, targets were defined as having both shape and motion characteristics. Thus an object had a particular contour shape and rotated in either a clockwise or counterclockwise direction. When Tanaka et al. tested the tuning properties of neurons in STS, they found that at least 50% of the neurons were not selective to the particular shape or motion used in the experiment. Of the remaining neurons, 34% were selective to shape characteristics only, 3% were selective to motion characteristics only, and 13% were selective to motion and shape characteristics together. Other neurophysiological studies have also found neurons selective to both movement and shape of more familiar stimuli, such as social signals from the face such as eye gaze (Allison, Puce, & McCarthy, 2000), and body shape and movement (Oram & Perrett, 1996). For example, Oram and Perrett (1996) reported cells in anterior STS that were selective to both the view of a body shape and the direction of motion of the body shape. Interestingly, of these cells responsive to both shape and motion, most were responsive to compatible motion and relatively fewer were selective to incompatible motion, suggesting a role of familiarity in cell response selectivity to integrated cues.

Although the STS is a candidate cortical area for motion and shape recognition of objects in monkeys, or face recognition in humans, it would be interesting to investigate if the same cortical area is involved in the integration of motion and shape for object recognition in humans. Furthermore, as O’Toole et al. (2002) have argued, familiarity may play a role in the recognition of face from motion information represented in the STS cortical area. The role of STS in the recognition of familiar objects from motion, however, has yet to be established.

What do our findings suggest about how objects are represented in memory? Given that our findings provide evidence for a role of motion in object categorization, we would suggest that motion information should be incorporated into any theory of how objects are recognized. As previously discussed, other studies have also provided evidence for motion integration in object representation (Stone, 1998, 1999; Wallis & Bühlhoff, 2001). The evidence is clearly weighted in terms of a higher-level spatiotemporal representation of objects. Yet current theories provide no explanation of how motion might be integrated.

In a recent exception, two studies have suggested that temporal proximity, or sequencing, may very well be the mechanism that integrates separate static “images” of an object into a single object representation (Wallis, 2002; Kourtzi & Nakayama, 2002). Thus, Wallis found that different multiple views of an object encoded in close tempo-

ral proximity are likely to be integrated into an object representation, even if the images can be markedly different. Similarly, Kourtzi and Nakayama (2002) argued that motion information may be useful for encoding and updating object properties from one moment to the next. These proposals indeed provide an interesting qualification of the multiple views model espoused by, for example, Tarr and Pinker (1989) and Tarr and Bühlhoff (1995), among others, by providing an explanation of how views are encoded into memory. Kourtzi and Nakayama argue that if motion is indeed represented, then it is a rapidly decaying representation and useful for moment-to-moment updating for mediating visual guidance and action, but not recognition per se. They argue that motion-based and shape-based mechanisms are distinct and used for action and recognition, respectively. However, none of these studies directly investigated the role of motion in long-term object representations.

Although motion may indeed be the underlying mechanism for the action system, we would argue that motion might also be involved in long-term representations for object recognition. Our findings, along with those of Stone (1998, 1999), suggest a more explicit representation of an object’s characteristic motion beyond that of a mechanism for integrating object views or for action responses. Motion information seems to be incorporated into visual long-term memory along with other object information, such as shape and color. As a result, category decisions are based on all object properties available in the objects representation through a process of cue integration. Our task, however, measured categorization performance of a small number of objects; therefore, we concede that it is not entirely clear from this current study how motion is integrated with spatial information in long-term representations. Our study does not allow us to determine, for example, whether motion is integrated into a holistic representation of all object information or as a part-based representation incorporating generic motion information. The precise nature of how motion information is integrated in long-term memory representations for objects is the focus of our next investigations. Nevertheless, our present study was designed to test if motion was useful for categorization even when spatial information was fully available. As such, our findings suggest that the manner in which an object moves, that is, its characteristic motion pattern, is clearly a source of information that the perceiver uses in classifying an object. Any future developments in object recognition theory would, therefore, need to account for such findings.

Acknowledgments

Commercial relationships: none.

Corresponding author: Fiona N. Newell

Address: Department of Psychology, Trinity College,
Dublin, Ireland.

Email: Fiona.Newell@tcd.ie.

References

- Allison, T., Puce, A., & McCarthy, G. (2000). Social perception from visual cues: Role of the STS region. *Trends in Cognitive Science*, 4(7), 267-278. [PubMed]
- Bassili, J. N. (1978). Facial motion in the perception of faces and of emotional expression. *Journal of Experimental Psychology: Human Perception and Performance*, 4(3), 373-379. [PubMed]
- Biederman, I. (1987). Recognition-by-components: A theory of human image understanding. *Psychological Review*, 94, 115-147. [PubMed]
- 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. [PubMed]
- Bernstein, L. J., & Cooper, L. A. (1997). Direction of motion influences perceptual identification of ambiguous figures. *Journal of Experimental Psychology: Human Perception and Performance*, 23(3), 721-737. [PubMed]
- Bülthoff, H. H., & Edelman, S. (1992). Psychophysical support for a 2-dimensional view interpolation theory of object recognition. *Proceedings of the National Academy of Sciences of the U. S. A.*, 89(1), 60-64. [Article]
- Cavanagh, P., Labianca, A. T., & Thornton, I. M. (2001). Attention-based visual routines: Sprites. *Cognition*, 80, 47-60. [PubMed]
- Christie, F., & Bruce, V. (1998). The role of dynamic information in the recognition of unfamiliar faces. *Memory and Cognition*, 26, 780-790. [PubMed]
- Cutting, J. E., Proffitt, D. R., & Kozlowski, L. T. (1978). Biomechanical invariant for gait perception. *Journal of Experimental Psychology: Human Perception and Performance*, 4(3), 357-372. [PubMed]
- Cutzu, F., & Edelman, S. (1994). Canonical views in object representation and recognition. *Vision Research*, 34(22), 3037-3056. [PubMed]
- Dubner, R., & Zeki, S. M. (1971). Response properties and receptive fields of cells in an anatomically defined region of the superior temporal sulcus in the monkey. *Brain Research*, 35, 528-532. [PubMed]
- Edelman, S. (1999). *Representation and recognition in vision*. Cambridge: 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. [PubMed]
- Felleman, D. J., & Van Essen, D. C. (1991). Distributed hierarchical processing in the primate cerebral cortex. *Cerebral Cortex*, 1, 1-47. [PubMed]
- Giese, M. A., & Lappe, M. (2002). Measurement of generalisation fields for the recognition of biological motion. *Vision Research*, 42, 1847-1858. [PubMed]
- Gluck, M. A., & Bower, G. H. (1988). From conditioning to category learning: An adaptive network model. *Journal of Experimental Psychology: General*, 117, 227-247. [PubMed]
- Goodale, M. A., & Milner, A. D. (1992). Separate visual pathways for perception and action. *Trends in Neuroscience*, 15, 20-25. [PubMed]
- Haxby, J. V., Hoffman, E. A., & Gobbini, M. I. (2002). Human neural systems for face recognition and social communication. *Biological Psychiatry*, 51, 59-67. [PubMed]
- Hill, H., & Johnston, A. (2001). Categorizing sex and identity from the biological motion of faces. *Current Biology*, 11(11), 880-885. [PubMed]
- Hummel, J. E. (2001). Complementary solutions to the binding problem in vision: Implications for shape perception and object recognition. *Visual Cognition*, 8, 489-517.
- Johansson, G. (1973). Visual perception of biological motion and a model for its analysis. *Perception and Psychophysics*, 14, 201-211.
- Kamachi, M., Bruce, V., Mukaida, S., Gyoba, J., Yoshikawa, S., & Akamatsu, S. (2001). Dynamic properties influence the perception of facial expressions. *Perception*, 30, 875-887. [PubMed]
- Kersten A. W. (1998a). An examination of the distinction between nouns and verbs: Associations with two different kinds of motion. *Memory and Cognition*, 26(6), 1214-1232. [PubMed]
- Kersten A. W. (1998b). A division of labor between nouns and verbs in the representation of motion. *Journal of Experimental Psychology: General*, 127(1), 34-54.
- Knappmeyer, B., Thornton, I. M., & Bülthoff, H. H. (2003). The use of facial motion and facial form during the processing of identity. *Vision Research*, 43(18), 1921-1936. [PubMed]
- Kourtzi, Z., & Kanwisher, N. (2000). Activation in human MT/MST by static images with implied motion. *Journal of Cognitive Neuroscience*, 12(1), 48-55. [PubMed]
- Kourtzi, Z., & Nakayama, K. (2002). Distinct mechanisms for the representation of moving and static objects. *Special Issue of Visual Cognition*, 9(1/2), 248-264.
- Kourtzi, Z., & Shiffrar, M. (2001). The visual representation of malleable and rigid objects that deform as they rotate. *Journal of Experimental Psychology: Human Perception and Performance*, 27, 335-355. [PubMed]
- Kozlowski, L. T., & Cutting, J. E. (1977). Recognizing sex of a walker from a dynamic point-light display. *Perception and Psychophysics*, 21(6), 575-580.

- Lander, K., & Bruce, V. (2000). Recognizing famous faces: Exploring the benefits of facial motion. *Ecological Psychology*, 12(4), 259-272.
- Lander, K., Christie, F., & Bruce, V. (1999). The role of movement in the recognition of famous faces. *Memory and Cognition*, 27, 974-985. [PubMed]
- 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. [PubMed]
- Mak, B. S. K., & Vera, A. H. (1999). The role of motion in children's categorization of objects. *Cognition*, 71, B11-B21. [PubMed]
- Marr, D. (1982). *Vision*. New York: W.H. Freeman & Company.
- Massaro, D.W., & Cohen, M. M. (1995). Perceiving talking faces. *Current Directions in Psychological Science*, 4(4), 104-109.
- Newell, F. N., & Findlay, J. M. (1997). The effect of depth rotation on object identification. *Perception*, 26(10), 1231-1257. [PubMed]
- O'Toole, A. J., Roark, D. A., & Abdi, H. (2002). Recognizing moving faces: A psychological and neural synthesis. *Trends in Cognitive Sciences*, 6(6), 261-266. [PubMed]
- Oram, M. W., & Perrett, D. I. (1996). Integration of form and motion in the anterior superior temporal polysensory area (STPa) of the macaque monkey. *Journal of Neurophysiology*, 76, 109-129. [PubMed]
- Pike, G. E., Kemp, R. I., Towell, N. A., & Phillips, K. C. (1997). Recognizing moving faces: The relative contribution of motion and perspective view information. *Visual Cognition*, 4(4), 409-437.
- Pollick, F. E., Lestou, V., Ryu, J., & Cho, S.-B. (2002). Estimating the efficiency of recognizing gender and affect from biological motion. *Vision Research*, 42, 2345-2355. [PubMed]
- Stone J. V. (1998). Object recognition using spatiotemporal signatures. *Vision Research*, 38, 947-951. [PubMed]
- Stone J. V. (1999). Object recognition: View-specificity and motion-specificity. *Vision Research*, 39, 4032-4044. [PubMed]
- Schwarzer, G. (2000). Development of face processing: The effect of face inversion. *Child Development*, 71, 391-401. [PubMed]
- Tanaka, K., & Saito, H. (1989). Analysis of motion of the visual field by direction, expansion/contraction, and rotation cells in the dorsal part of the medial superior temporal area of the macaque monkey. *Journal of Neurophysiology*, 62, 626-641. [PubMed]
- Tanaka, Y. Z., Koyama, T., & Mikami, A. (2002). Visual responses in the temporal cortex to moving objects with invariant contours. *Experimental Brain Research*, 146, 248-256. [PubMed]
- 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(6), 1494-1505. [PubMed]
- Tarr, M. J., & Pinker, S. (1989). Mental rotation and orientation-dependence in shape recognition. *Cognitive Psychology*, 21, 233-282. [PubMed]
- Tinbergen, N. (1951). *The Study of Instinct*. Oxford: Clarendon Press.
- Ungerleider, L., & Mishkin, M. (1982). Two cortical visual systems. In D. Ingle, M. Goodale, and R. Mansfield (Eds.), *Analysis of visual behaviour*. Cambridge: MIT Press.
- Wallach, H., & O'Connell, D. N. (1953). The kinetic depth effect. *Journal of Experimental Psychology*, 45, 205-217.
- Wallis, G. (2002). The role of object motion in forging long-term representations of objects. *Visual Cognition*, 9(1-2), 233-247.
- Wallis, G., & Bülthoff, H. H. (2001). Effects of temporal association on recognition memory. *Proceedings of the National Academy of Sciences of the U. S. A.*, 98, 4800-4804. [PubMed]