Journal of Experimental Psychology:
Human Perception and Performance
1996, Vol. 22, No. 2, 395–416

0096-1523/96/$3.00

View Specificity in Object Processing: Evidence From Picture Matching

Rebecca Lawson and Glyn W. Humphreys
University of Birmingham

Four experiments investigated the types of representations mediating sequential visual matching of objects depicted at different depth rotations. Matching performance was affected by the similarity between depicted views of the objects. Effects of view similarity were not influenced by the presence of a meaningless mask in the interstimulus interval (ISI), but they were reduced by long ISIs and by familiarity with the stimuli. It is suggested that with longer ISIs or increased stimulus familiarity, a number of object representations are activated that, although abstracted from some image characteristics, remain view specific. Under these conditions, matching is less reliant on representations closely tied to the view of the initial stimulus presented. The results are consistent with both the derivation and the long-term representation of view-specific rather than view-invariant descriptions of objects.

Models of human visual object recognition must account for the generally rapid and accurate achievement of object constancy: that is, the relative immunity of object recognition to changes in the retinal image projected by an object, dependent on the position of the object with respect to the viewer. Transformations of position, scale, and distance and rotations in plane and in depth can profoundly alter the retinal image projected by the same object from one viewing occasion to another. Despite this, human visual object recognition generally proceeds efficiently and shows remarkably little sensitivity to changes in the viewpoint of the observer.

The precise procedures by which object constancy is achieved remain poorly understood. Several accounts assume that a series of different representations are computed en route to a relatively abstract, view-invariant representation. The derivation of a single, abstract, view-invariant representation is held to be necessary to minimize long-term storage requirements and to allow semantic and associative information specific to a given object to be retrieved from a range of retinal projections (e.g., Hinton, 1981; Lowe, 1987; Marr, 1982). Contrasting approaches suggest the existence of a small number of view-specific representations of each object, perhaps invariant over certain properties such as retinal position and scale but not invariant over rotation in depth (Edelman & Weinshall, 1991; Perrett, Benson, Oram, & Hietanen, 1994; Tarr & Pinker, 1989) or even the existence of numerous highly view-specific templates for each object that necessitate only a minimal degree of abstraction to match an image to a stored representation.

In the current studies, we examined the effects of depth rotation on object recognition. The manipulation of rotation in depth is likely to provide the most stringent test of the ability of the human recognition system to achieve object constancy. Two depth-rotated images of the same object will often reveal different features and parts and will frequently have very different global shapes, whereas position, scaling, or plane rotation transformations generally have less catastrophic effects on the image subtended by an object. The task of sequential picture–picture matching was used to investigate the effects of depth rotation. This task provides good control over the depth relationship between two to-be-compared views without requiring a complex, difficult, or highly variable response (cf. picture naming; see also Tarr & Bülthoff, 1995), and thus the task is suited to detecting small effects resulting from depth rotation transformations.

Empirical evidence from studies of matching relates to the issue of the types of representation constructed during object processing. Initial matching studies by Posner and colleagues (Posner, 1969; Posner & Keele, 1967) used letters with either the same or different names, presented either simultaneously or sequentially. These authors reported an advantage for the matching of identical letters in comparison with the matching of letters with the same name but differing case. Furthermore, this "identity benefit" (the advantage for matching identical relative to nonidentical stimuli) occurred only with a short interstimulus interval (ISI) between the to-be-matched stimuli. Posner suggested that two types of representation were necessary to account for the results: a visual representation mediating the rapid matching of identical stimuli and a name representation enabling matches to be made between letters differing in case.

Shepard and Metzler (1971) considered the simultaneous matching of novel, three-dimensional objects made up of

Rebecca Lawson, Cognitive Science Research Centre, School of Psychology, University of Birmingham, Birmingham, United Kingdom; Glyn W. Humphreys, Cognitive Science Research Centre, School of Psychology, University of Birmingham, Birmingham, United Kingdom.

This work was supported by a grant from the Science and Engineering Research Council of Great Britain (Grant 9031212X). We are grateful to Karl Verfaillie for comments on an earlier version of this article.

Correspondence concerning this article should be addressed to Rebecca Lawson, who is now at Department of Psychology, University of Waterloo, Waterloo, Ontario, Canada N2L 3G1. Electronic mail may be sent via Internet to rlawson@cgl.uwaterloo.ca.

blocks joined at right angles. On a given trial, the views were depth or plane rotations of the same object (match trials) or of two different objects that were mirror-images of each other (mismatch trials). For match trials, Shepard and Metzler found a linear relation between the speed of response and the angle of rotation necessary to align the two views. This relationship held irrespective of whether the transformation necessary to align the two views was a plane or a depth rotation. Shepard and Metzler suggested that matching involved the rotation of a mental representation of an object and that the time taken for rotation was directly proportional to the angle through which the representation had to be rotated.

However, experiments using simple alphanumeric stimuli or novel objects may not provide valid ecological conditions under which to examine the processes involved in human visual recognition of familiar objects. For example, alphanumeric stimuli are two-dimensional, and therefore only plane rotations (not rotations in depth) can be examined; also, processing may differ for familiar relative to unfamiliar stimuli (e.g., stored representations are available only to mediate the matching of familiar stimuli).

Later researchers have alleviated these problems by examining the matching of more naturalistic, familiar objects (Bartram, 1976; Ellis & Allport, 1986; Ellis, Allport, Humphreys, & Collis, 1989; Humphrey & Lupker, 1993; Kelter et al., 1984; Klatzky & Stoy, 1974). Bartram (1976) conducted sequential matching experiments, presenting line drawings and photographs of familiar objects. He compared the matching of (a) identical views of the same object (identical view matches), (b) different views of the same object (different view matches), and (c) different exemplars of the same object category that had the same name (same name matches). On mismatch trials, objects from two different object categories were presented. Bartram found that identical view matches were more rapid than different view matches, which were, in turn, more rapid than same name matches. From these results, Bartram argued that three different levels of representation were involved: (a) a two-dimensional, view-specific, perceptual representation mediating identical view matches; (b) a visual but more abstract representation mediating different view matches; and (c) an abstract representation mediating same name matches.

Ellis and colleagues (Ellis & Allport, 1986; Ellis et al., 1989) argued that Bartram's (1976) results could be accounted for simply by assuming that a visual similarity measure is computed and a "same" response is made when a criterion level of similarity is reached. Because, in general, views were increasingly dissimilar across identical view, different view, and same name matches, the pattern of results reported by Bartram would follow. Ellis and colleagues conducted further sequential matching experiments with photographs of familiar objects to try to evaluate the three representations proposed by Bartram. They examined matching across a range of blank ISIs between stimuli and when an intervening, meaningless pattern mask was presented in the ISI. Participants responded "same" to objects with the same name, irrespective of differences in view. Ellis and colleagues found an advantage for identical view matches over different view matches that was present at short ISIs (100 ms and 500 ms) but not at a long ISI (2,000 ms). This advantage was also lost if the ISI was filled with a visual mask. Furthermore, there was an advantage for different view matches over matches between objects with the same name that was maintained at the long ISI and with an intervening visual mask.

From this pattern of results, Ellis and colleagues (Ellis & Allport, 1986; Ellis et al., 1989) argued in support of Bartram's (1976) distinction between three types of representation in object processing. The first type is a view-specific code that can mediate identical view matches. It is rapidly derived but dissipates quickly and is disrupted by a visual mask. The second is a more abstract code that can mediate different view matches. It is derived more slowly but is robust to the effects of masking. The third type is a semantic or name representation involved in matching different objects with the same name (cf. Posner & Keele, 1967; Warrington & Taylor, 1978). Neither Bartram (1976) nor Ellis and colleagues (Ellis & Allport, 1986; Ellis et al., 1989) attempted to determine whether the abstract representation apparently involved in different view matches was object-centered and view-invariant or whether it was view-sensitive, although perhaps less precisely tied to image features and more robust to masking than the representation mediating identical view matches. We term this latter type of representation an *abstract, view-specific* representation to highlight the fact that the depicted view of the object may be encoded, although the representation may not be tied to image-based properties such as the exact spatial locations of edges in the image (e.g., it may be robust across lateral shifts of images).

Humphrey and Lupker (1993) conducted a study similar to those of Ellis and colleagues, but they presented line drawings rather than photographs and manipulated within-plane rotation rather than depth rotation. They too found an identity benefit for matching objects shown in the same rather than different views; in contrast to prior research, however, they reported that this benefit was maintained at long ISIs. Their results suggest that durable object representations are view-sensitive, and there was no evidence for a distinct and transient representation that mediated only short ISI identical view matches.

Even for a theory proposing that object recognition is based on single, stored, object-centered representations of familiar objects, to which all views of an object are matched with equal efficiency, there are at least two reasons why effects of view might occur in matching. First, the matching of identical views, for which image descriptions are the same at every level of representation, might be mediated by a transient, image-based, view-specific representation that would be available after very little processing. This form of low-level matching would probably be more rapid than that mediated by an abstract representation, resulting in a specific benefit for identical view matches (e.g., Bartram, 1976; Ellis & Allport, 1986; Ellis et al., 1989; Kelter et al., 1984; Klatzky & Stoy, 1974). However, matching image descriptions at such an unabstracted, image-based, low level of representation would not be useful for normal recognition, in which matching to stored object representations must take place.

Second, some views of an object might be processed more efficiently than other views and, hence, might activate a stored, abstract representation faster. For instance, "good" views might optimally reveal important, diagnostic features of the object, whereas "bad" views might lead to an ambiguity in the three-dimensional interpretation of the image as a result of an accident of viewpoint. Palmer, Rosch, and Chase (1981) operationally defined the canonicality of a view of an object in terms of converging results from ratings of the goodness of a view, the judged perspective when imagining objects, and the best view from which to photograph a previously imagined object. Palmer et al. then demonstrated that more canonical views of objects were named faster. They concluded that canonical views maximize the salient information about an object within the image. Thus, it is possible that matching to an object-centered representation may be influenced by the view at which an object is depicted, with privileged, canonical views enabling object-centered representations to be accessed more efficiently.

From these arguments, it is clear that mere effects of viewpoint change on object matching do not rule out the involvement of object-centered, view-invariant representations in object matching and recognition. What is needed is a more focused attempt to assess whether there are systematic effects of the degree of viewpoint change on matching and whether effects occur even with canonical views of objects. For example, effects of the degree of viewpoint change, even for canonical views, would be consistent with the involvement of view-specific representations and would be difficult to account for in terms of view-invariant representations.

The present study assessed whether there are systematic effects of viewpoint change on the matching of canonical as well as noncanonical views of objects. One problem with matching studies is that they may not reflect important stages in object recognition; instead, matching efficiency may primarily reflect episodic memories of objects or early stages of processing before object recognition (e.g., see Cooper, Biederman, & Hummel, 1992). In particular, matching could be achieved without the use of stored knowledge about the specific objects presented and could be based solely on image-derived information. Previous researchers who have presented familiar objects have assumed that stored representations of the depicted objects were involved in matching, but they have not demonstrated this. We investigated whether stored knowledge influenced matching efficiency by assessing performance as a function of the familiarity of the orientation of the stimuli (Experiment 2). We return to consider the relations between object matching and object recognition in the General Discussion section.

We used a sequential picture matching task in a series of four experiments that investigated the nature of the representations involved in matching identical and different, depth-transformed views of objects and the factors determining the effects of view on matching efficiency. The representation mediating different view matches could be a single object-centered, view-invariant, three-dimensional representation (e.g., Lowe, 1987; Marr, 1982) or one of a set of view-specific representations (e.g., Tarr & Pinker, 1989). If the representation is view-invariant, then the same representation should be accessed by any view of the object presented. Hence, matching involving such a representation should be independent of viewpoint. However, as discussed earlier, view-independence could be compromised by either an advantage for identical view matches mediated by low-level, image-based representations or variation in the ease of encoding image descriptions or of matching them to abstract representations, with canonical views activating stored representations more efficiently than noncanonical views. In contrast, if only view-specific representations are involved, then matching should be more efficient for similar views than for different views, and the efficiency of matching should increase with increased similarity between the views. This should occur with both canonical and noncanonical views.

Line drawings of familiar objects were used, and the view in depth of the depicted object was varied systematically (see Figure 1). Unlike photographs, line drawings possess no surface features that might reduce the effects of depth rotation by permitting low-level matching of surface characteristics (see Bartram, 1976; Humphrey & Lupker, 1993). The use of a range of different views allowed the depth relation between the to-be-matched views to be manipulated quantitatively. In all of the experiments, the first picture was presented for 100 ms, followed after a fixed ISI by the second picture, which was presented until a response was made. Participants decided whether or not the two pictures depicted the same object. Throughout this article, the first stimulus and the second stimulus presented on a trial are referred to as the reference and target, respectively. In identical view matches, the reference and target were identical views of the same object; in mirror-image view matches, the reference and target were mirror-image views

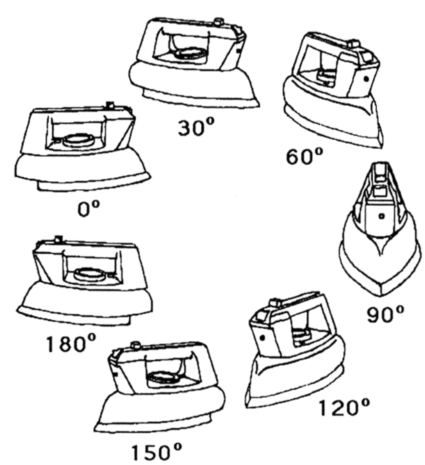

Figure 1. Examples of different views of an iron, which was one of the 36 objects used in all of the experiments reported in this article. The depicted angle of view of the iron is labeled beneath each line drawing. In Experiments 1 and 2 only, the 30° and 60° views were replaced by mirror-image reflections of the 150° and 120° views, respectively, to ensure that the 30° and 150° views and the 60° and 120° views, respectively, were exact mirror-images.

of the same object; and in different view matches, the reference and target were different views of the same object. On mismatch trials, the reference and target depicted different objects. Finally, an identity benefit indicates that identical view matches were made more rapidly than mirror-image and different view matches.

Experiment 1 manipulated both target view and ISI to investigate the view specificity and longevity of the representations involved in identical, mirror-image, and different view matches. Experiment 2 investigated whether stored knowledge was involved in matching and whether the results of Experiment 1 were due to within-experiment frequency effects. Experiments 1 and 2 showed that participants had a specific difficulty in matching foreshortened targets; Experiment 3 considered whether this disadvantage could be overcome by presenting foreshortened reference views. Finally, Experiment 4 examined the view specificity of the representations mediating different view matches by comparing directly the effects of two different reference views.

Experiment 1

In Experiment 1, participants had to match a reference depicted at a depth rotation of 150° to one of five targets at different depth rotations: 150°, 120°, 90°, 60°, and 30°. The targets were thus rotated 0°, 30°, 60°, 90°, and 120°, respectively, from the reference. In the 0° view, the main axis of the object was depicted as perpendicular to the line of sight of the viewer (see Figures 1 and 2).

The 150° reference was selected as being a good, but not the best (most canonical), view of an object. The choice of the best view was based on data from a rating study described subsequently. Prior studies suggest that at longer ISIs, matching is mediated by representations that are less view-specific and more robust to masking, whereas more transient and highly view-specific representations may be involved at shorter ISIs (e.g., Ellis et al., 1989). At longer ISIs, it is possible that participants activate a "best-view," canonical representation from the reference. This might result in matches of the 150° reference being faster to a canonical target than to an identical but noncanonical 150° target. Such a result could not have been observed in previous matching studies because the "goodness" of the target view was not manipulated systematically.

The views selected here were chosen on the basis of a study in which 20 independent participants rated which was the best, most revealing view from a full set of 12 views of each object, each differing by a 30° rotation. Participants consistently preferred the 60° and 120° views (together accounting for 48.1% of choices, 70.7% if the corresponding 210° and 330° rear views are also included). The 30° and 150° views were chosen less frequently but still accounted for a proportion of preferences (15.6% of choices,

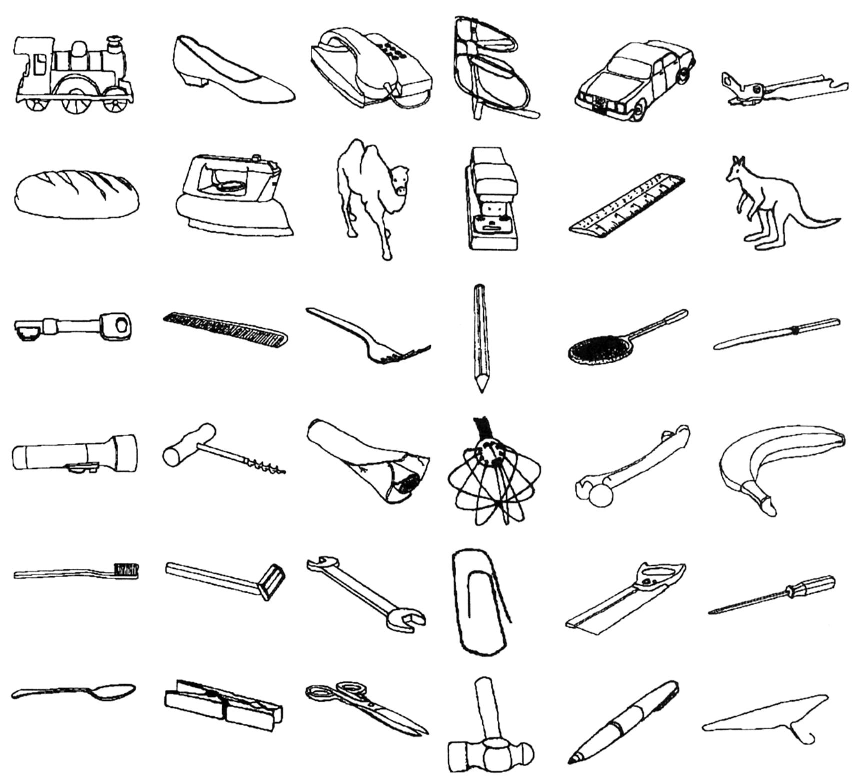

Figure 2. One view of each of the 36 experimental objects presented. From left to right, the columns show views at 0°, 30°, 60°, 90°, 120°, and 150°. The top two rows of objects are the inverting items (see Experiment 2); the remaining 24 objects are the noninverting items.

20.5% if the corresponding 240° and 300° rear views are included). Very few participants selected views that were not rotated 30° or 60° from the 0° view (only 4.0% chose foreshortened views, either 90° or 270°, in which the main axis of the object pointed directly toward the viewer; 4.7% chose 0° or 180° views, in which the main axis of the object was fully revealed). Converging evidence for the choice of views came from picture-naming and word–picture verification tasks conducted with the same set of stimuli (Lawson, 1994). These studies revealed that the 60° and 120° views were named and verified rapidly; responses to the 30° and 150° views tended to be less efficient, and responses to 90° stimuli were much slower and less accurate. Taken together, the results indicate that the 60° and 120° views of the object are more canonical than the 30° and 150° views and that the 0° and 90° views are particularly disadvantaged, both in preference judgments and in recognition performance.

Experiment 1 examined three different strategies that participants may use to perform different view matches. The first strategy is *mental rotation*: Participants may rotate in depth the image descriptions of the target and reference to align the main axes of the descriptions. A comparison of the two descriptions could then be made. This hypothesis predicts that reaction times (RTs) should increase with increasing disparity between the views of the reference and the target. The 150° reference would be matched fastest to the 150° target, followed in turn by the 120°, 90°, and 60° targets, with the 30° target's being matched slowest. A strong version of this hypothesis would take the rate of mental rotation to be constant and, hence, would predict a linear relation between RT and the view disparity between the reference and target (Jolicoeur, 1985; Shepard & Metzler, 1971).

The second strategy is *rotation and reflection*: Participants both rotate and reflect image descriptions to achieve matching. This hypothesis extends the first and posits that, when beneficial, participants perform rapid mirror-image reflections of image descriptions in addition to mentally rotating descriptions. This account would explain the results of Klatzky and Stoy (1974, Experiment 1). They conducted a picture-matching experiment and found that identical view and mirror-image view matches were equally rapid and faster than same name matches. According to this second hypothesis, RTs should increase as a function of the disparity in the views of the reference and the target, where the disparity is calculated from either the target view or its mirror image, whichever value is smaller. In Experiment 1, the 30° and 150° views and the 60° and 120° views, respectively, were mirror-images of each other. Accordingly, with a 150° reference, matches to 30° and 150° targets should be faster than matches to 60° and 120° targets, and matches to 90° targets should be the slowest. If reflection takes some time to achieve, then 30° mirror-image targets may be matched more slowly than 150° targets (which are identical to the reference); also, 60° targets may be matched slower than 120° targets, because a reflection transformation would be required for 60° but not 120° targets.

The final strategy is *matching to abstract, durable representations*: Matching is contingent on both the reference and target's accessing a common, relatively abstract representation. Different predictions can be made here, depending on the type of representation posited. If the representation is object-centered and view-invariant, matches to all target views should be equally efficient (with the proviso that all views are encoded equally rapidly and do not differ in canonicality; see earlier discussion). In contrast, if the representation remains tuned to the view of the reference, although perhaps in a more abstract and durable form than low-level, image-based descriptions, targets similar in view to the reference may be matched more efficiently than targets dissimilar in view.

Following from the results of Bartram (1976) and Ellis and colleagues (Ellis & Allport, 1986; Ellis et al., 1989), identical view matches are predicted to be faster than matches between different views, at least at the short ISI. This could be because identical view matches require only a temporary, image-based representation to achieve matching, as discussed earlier. This representation might be the usual precursor to a more durable and abstract representation used in normal object recognition. The predicted identity benefit is outside the scope of the three accounts of matching between different views of objects just described.

To provide conditions that tapped both transient and durable representations, we manipulated the ISI between the reference and target. Ellis and Allport (1986) found that, at a long ISI, the benefit for identical view relative to different view matches found at short ISIs was lost. By manipulating the ISI, we attempted to replicate this interaction and to motivate the distinction between the representations mediating identical view and different view matches proposed by Bartram (1976) and by Ellis and colleagues (Ellis & Allport, 1986; Ellis et al., 1989).

Method

Participants. There were 16 participants (8 men and 8 women). In this and all of the following experiments reported here, participants were from the University of Birmingham and were paid to take part. Participants were between 18 and 35 years of age and had normal or corrected-to-normal vision.

Materials. A set of 12 views of each of 36 familiar objects was produced (see Figure 2; the objects are listed in Appendix A). Each view was separated by a 30° horizontal rotation in depth. The 12 views ranged over a full 360° horizontal depth rotation, although only a subset of views was used in each experiment. All of the objects possessed an unambiguous main axis of elongation, and the objects were rotated about the vertical axis running through their center point. The angle of view was defined with respect to the line of sight of the viewer relative to the main axis of the object. The 0° view revealed the main axis perpendicular to the line of sight of the viewer. In the 90° foreshortened view, the main axis of elongation pointed directly toward the viewer and revealed either the most familiar view (the front of the object) or the view with the most important feature to the fore (e.g., the 90° view of a camel was depicted facing head on rather than tail on, and the fork was depicted with the prongs closest to the viewer and the handle farthest away). In Experiments 1 and 2 only, the 30° and 60° views were replaced by mirror-image reflections of the 150° and 120°

views, respectively, to ensure that the 30° and 150° views and the 60° and 120° views were mirror-images of each other.

All stimuli were line drawings produced by tracing photographs of either the objects or scale models of the objects. Photographs were taken from a slightly elevated angle of between 15° and 30° above the horizontal plane on which the object rested. This angle was maintained as a constant during the depth rotation of each object. An elevated angle was used to ensure that the effects of foreshortening the main axis of the object were not too severe and that all views were recognizable. Each picture was scaled to occupy a square of 6 × 6 cm. In Experiment 1, the reference on each trial was a 150° view, and the targets were 30°, 60°, 90°, 120°, and 150° views.

Design. Each trial began with the presentation of a 150° reference of one of the 36 objects. This was followed with equal probability by a target view of that object at 30°, 60°, 90°, 120°, or 150° (a match trial) or by a view of a different object at 30°, 60°, 90°, 120°, or 150° (a mismatch trial). For mismatch trials, 5 different target objects, each at a different view, were paired with each reference. Two sets of 360 reference and target picture pairs were produced. Different object pairings were used for the mismatch trials in the second set; otherwise, the two picture pair sets were identical. In each of the two picture pair sets, all five target views of all 36 objects were shown twice, once on a match trial and once on a mismatch trial. There were 720 trials in total, with each target view being shown four times in all. Each set of 360 picture pairs was divided into two blocks of 180 pairs. This reduced the time taken to complete the long ISI blocks to approximately 20 min. Each block of 180 trials contained two or three match and two or three mismatch trials for each reference object. The order of presentation of trials within a block was random and different for each participant. The two blocks derived from the same picture pair set were always shown consecutively. The order of presentation of blocks within a picture pair set was balanced across participants, as was the order of presentation of the two picture pair sets. For each participant, one picture pair set was presented with a short ISI of 585 ms between the reference and the target; the other picture pair set had a long ISI of 2,510 ms. The assignment of picture pair sets to ISI condition and the order of presentation of ISI were balanced across participants.

Apparatus and procedure. A Macintosh IIci computer with the Psychlab Version 8.5 presentation package was used to display the stimuli. The experiment lasted about 1 hr. The procedure for each trial was as follows: A fixation cross appeared on the screen for 200 ms, immediately followed by the 150° reference, which was presented for 100 ms. There was then a blank ISI of either 585 ms or 2,510 ms, followed by the target, which was displayed until the participant responded. Participants responded with their preferred hand to "same" match trials, and responded with their nonpreferred hand to "different" mismatch trials, by hitting either the *M* or *Z* key of the keyboard. Before the first block of trials, participants were given practice trials randomly selected from the first block of trials. Participants decided whether the target and reference both represented the same object. Participants were instructed that, if the same object was depicted at different views, it was still a same match trial. Participants were encouraged to respond as rapidly and as accurately as possible.

Results

In this experiment, and in all of the following experiments reported, RTs of less than 300 ms or exceeding 1,200 ms were discarded as errors. All participants scoring an average of more than 10% errors in the experiment were replaced. Three participants were replaced in Experiment 1 as a result of these criteria. Mean correct same RTs across participants are shown in Figure 3; errors are given in Table 1.

An analysis of variance (ANOVA) was carried out on the mean correct RTs for same responses. In this and in all of the following experiments, the results across participants (F_1) and across items (F_2) are reported. For across-participants analyses, the data for each participant were pooled over the 36 reference objects presented in each View × ISI × ISI Order condition. For across-items analyses, the data for each object were pooled over the 8 participants in each View × ISI × ISI Order condition. There were two within-subject variables, view of the target (30°, 60°, 90°, 120°, or 150°) and ISI (short [585 ms] or long [2,510 ms]), and one between-subjects variable, ISI order (order of presentation of ISI blocks [short ISI first or long ISI first]).

The main effect of view was significant, $F_1(4, 56) = 13.450, p < .001, MSE = 12{,}238, F_2(4, 140) = 20.227, p < .001, MSE = 52{,}285$, as was that of ISI, $F_1(1, 14) = 68.961, p < .001, MSE = 288{,}758, F_2(1, 35) = 650.941, p < .001, MSE = 1{,}297{,}528$. Short ISI matches were faster than long ISI matches. The main effect of ISI order was significant across items only, $F_1(1, 14) = 0.101, p > .7, MSE = 5{,}221, F_2(1, 35) = 7.765, p < .009, MSE = 15{,}233$. Matches tended to be faster for participants who had the long ISI block first.

There were three significant interactions: ISI × View, $F_1(4, 56) = 3.894, p < .008, MSE = 1{,}472, F_2(4, 140) = 3.126, p < .02, MSE = 7{,}735$; ISI × ISI Order (across items only), $F_1(1, 14) = 2.895, p > .1, MSE = 12{,}120, F_2(1, 35) = 21.625, p < .001, MSE = 56{,}903$; and ISI × View × ISI Order (across participants only), $F_1(4, 56) = 3.386, p < .02, MSE = 1{,}280, F_2(4, 140) = 1.799, p < .1, MSE = 4{,}331$.

The three-way interaction was decomposed by separate two-way analyses for each ISI order. Data for participants who had the short ISI block first and for participants who had the long ISI block first are shown separately in Figure 3. For participants who had the short ISI block first, there were main effects of view, $F_1(4, 28) = 7.129, p < .001, MSE = 5{,}911, F_2(4, 140) = 11.915, p < .001, MSE = 27{,}293$, and ISI, $F_1(1, 7) = 27.068, p < .002, MSE = 91{,}280, F_2(1, 35) = 234.632, p < .001, MSE = 405{,}493$. Short ISI matches were faster than long ISI matches. The View × ISI interaction was also significant, $F_1(4, 28) = 6.627, p < .001, MSE = 2{,}519, F_2(4, 140) = 5.956, p < .001, MSE = 10{,}916$. At the short ISI, 90° targets were matched slower than all other targets ($p < .01$; Newman–Keuls analysis), and 150° targets were matched faster than all other targets ($p < .05$). At the long ISI, there were no differences between any target view matches.

For participants who had the long ISI block first, there were main effects of view, $F_1(4, 28) = 6.444, p < .001, MSE = 6{,}383, F_2(4, 140) = 8.442, p < .001, MSE = 25{,}712$, and ISI, $F_1(1, 7) = 41.901, p < .001, MSE = 209{,}599, F_2(1, 35) = 327.626, p < .001, MSE = 948{,}938$. Short ISI matches were faster than long ISI matches. The 90° targets were matched slower than all other targets ($p < .01$; Newman–Keuls analysis). The ISI × View interaction

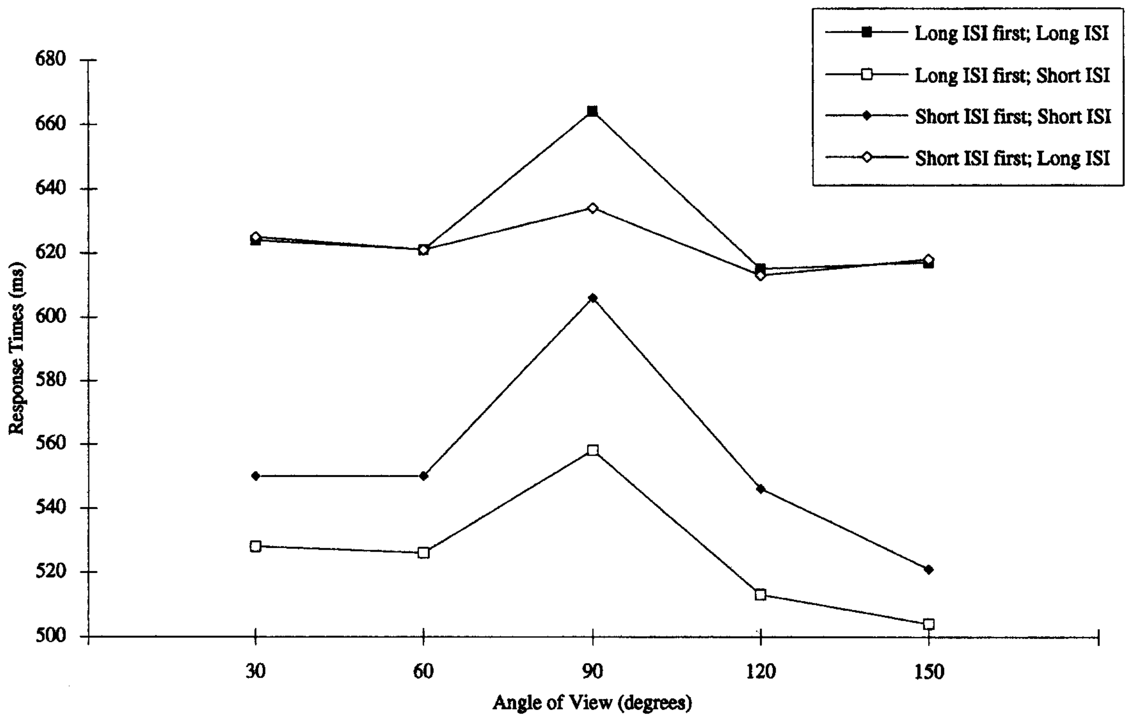

Figure 3. Mean correct response times for match trials at short and long ISIs, for participants presented with the short ISI first or the long ISI first, as a function of target view: Experiment 1. ISI = interstimulus interval.

was not significant, $F_1(4, 28) = 0.621, p > .6, MSE = 233$, $F_2(4, 140) = 0.377, p > .8, MSE = 1{,}150$.

In all of the experiments reported here, a log-linear transformation was performed on the mean error scores for each cell, followed by ANOVAs across participants and across items, for both same and different trials. There was no evidence of a speed–accuracy trade-off in the error analyses, and the results were similar to those provided by the RT analyses. Thus, error analyses are not reported in the remainder of the article.

The mean RTs for correct different trials were 597 ms at the short ISI (4.0% errors) and 676 ms at the long ISI (6.1% errors). The main effect of view was marginally significant, $F_1(4, 56) = 2.178, p > .08, MSE = 1{,}482$, $F_2(4, 140) = 2.243, p > .06, MSE = 6{,}668$ (see Table 2), whereas ISI was significant, $F_1(1, 14) = 38.820, p < .001, MSE = 251{,}895$, $F_2(1, 35) = 439.171, p < .001, MSE = 1{,}114{,}473$. Short ISI matches were faster than long ISI matches. The main effect of ISI order was significant across items only, $F_1(1, 14) = 0.351, p > .5, MSE = 18{,}085$, $F_2(1, 35) = 28.436, p < .001, MSE = 75{,}750$. Matches tended to be faster for participants who had the long ISI block first. The ISI × ISI Order interaction was significant across items only, $F_1(1, 14) = 2.385, p > .1, MSE = 15{,}474$, $F_2(1, 35) = 33.230, p < .001, MSE = 69{,}984$. This interaction took the same form as that found for same trials; short ISI trials for participants who had the long ISI block first were faster than short ISI trials for participants who had the short ISI block first ($p < .01$; Newman–Keuls analysis for items only). These were, in turn, faster than long ISI trials for both participants who had the short ISI block first and those who had the long ISI block first. There was no difference between the last two conditions. No other interactions were significant.

Discussion

A number of separate effects are apparent from these results. First, matches to foreshortened targets were much slower than matches to targets depicting other views; we term this the foreshortened view disadvantage. This result was consistent over both ISI orders and concurs with previous studies revealing a similar disadvantage. For instance,

Table 1
Mean Error Rates (%) for Match Trials as a Function of Target View: Experiment 1

Block order and ISI	Target view 30°	60°	90°	120°	150°
Short ISI first					
Short	5.6	3.8	4.9	4.5	3.5
Long	7.3	6.3	5.9	5.6	8.0
Long ISI first					
Short	4.2	2.4	9.7	4.5	5.6
Long	5.9	9.4	12.5	8.3	4.5

Note. ISI = interstimulus interval.

Table 2
Mean Correct Response Times (ms) for Mismatch Trials at Short and Long ISIs as a Function of Target View: Experiment 1

	Target view				
ISI	30°	60°	90°	120°	150°
Short	597	596	607	595	589
Long	678	682	685	668	669

Note. ISI = interstimulus interval.

Humphrey and Jolicoeur (1993) reported that partly foreshortened views were named less efficiently than nonforeshortened views, despite the same major parts and features of the object being visible in both views. Furthermore, a number of studies have found that patients with right hemisphere lesions have specific difficulties in matching foreshortened views of objects to more prototypical views (Humphreys & Riddoch, 1984; Warrington & Taylor, 1973, 1978; see also Marr, 1982). Second, for participants who had the short ISI block first, there was an identity benefit, with 150°–150° matches of identical views being faster than different view matches. This replicates the identity benefit found by Bartram (1976) and by Ellis and colleagues (Ellis & Allport, 1986; Ellis et al., 1989). The identity benefit did not extend to 30° targets, although the 30° view was an exact mirror-image of the 150° reference. Third, all matches were slower at the long ISI.

These effects of viewpoint on matching provide clear evidence against two of the possible matching strategies discussed earlier. First, the mental rotation hypothesis predicts that participants should be increasingly slow to match targets rotated away from the 150° reference. This was refuted by the results. Foreshortened targets were matched slower than 30° and 60° targets, although the 30° and 60° views were rotated further in depth from the 150° reference.

As mentioned in the *Method* section, in Experiments 1 and 2 only, the 30° and 60° views of objects were replaced by mirror-image reflections of the 150° and 120° views, respectively. Half of the objects used were symmetrical about their main axis of elongation, and so this manipulation had no effect on the views presented for these objects because their 30° and 150° views and their 60° and 120° views were identical anyway (barring a mirror-image reflection). However, for the remaining asymmetrical objects, the manipulation did alter the view of the object presented. The manipulation was used to ensure that the second reflection plus rotation strategy could be optimal, because reflections of either 120° or 150° views always mapped exactly onto 60° and 30° views, respectively. Because this strategy would always be successful in Experiments 1 and 2, a failure to find consistent evidence for the strategy makes it highly unlikely that the strategy is used under normal viewing circumstances, when it would be much less likely to be successful. Nevertheless, the manipulation may have discouraged participants from using the rotation strategy in Experiments 1 and 2 because, for some trials, mental rotation would not map the image descriptions of the target and reference onto each other exactly (because the manipulation involved a reflection rather than a rotation transformation). However, in Experiment 4, the rotation strategy would have invariably been successful because the true 30° and 60° views of objects were always presented. The results from Experiment 4 indicated that participants still did not use a rotation strategy to achieve matching. Indeed, the pattern of results was identical across Experiments 1 and 4; with a 150° reference, 60° targets were matched faster than 90° targets, whereas the rotation strategy would predict the opposite.

Second, the reflection plus rotation hypothesis predicts equal differences in RTs for matching 30° compared to 60° targets, 90° to 120° targets, and 120° to 150° targets because the difference between each pair involves a 30° depth rotation. However, 90° targets were consistently slower to be matched than all other target views, which did not differ from each other apart from a weak identity benefit for 150° targets. Thus, at least for familiar objects depicted at different views in depth, mental rotation does not appear to play a major role in matching.

The results further suggest that the two effects of view observed (the strong foreshortening disadvantage and the relatively weak identity benefit) are modulated by two factors. First, there is an effect of ISI, with stronger effects of view at the shorter ISI. Second, there is an effect of block order; familiarity with the stimuli in the first block reduces the view effects for the same stimuli when they are presented in the second block. This result is consistent with previous findings of Jolicoeur (1985) on the effects of plane disorientation on picture naming. He found that plane disorientation initially strongly disrupts picture naming but that the effects of disorientation rapidly reduce with repeated presentation of the same pictures. The result of the combination of the ISI and familiarity effects here is that the strongest effects of view occurred in the short ISI block, when it was presented first, and the weakest effects occurred in the long ISI block, when it was presented second.

These viewpoint effects can be accommodated by the third hypothesis outlined in the introduction: namely, that for different view matches, image descriptions are matched to abstract, durable representations. The results from Experiment 1 suggest that the representations are accessed relatively slowly, both after the reference is presented (hence the reduced view effects with a long ISI) and over the course of the experiment (hence the reduced view effects when stimuli are familiar, in the second block of trials). It is not clear whether the representations involved are constructed solely from image-based information or whether the reference may activate a stored representation that is then used to match to the target. These two alternatives were examined in Experiment 2. In addition, either (a) a number of different representations might be available for a given object, each tuned to a different view (with foreshortened views not being represented), ensuring that all nonforeshortened views can be recognized equally rapidly if all of the representations can be activated equally efficiently, or (b) a single, abstract, relatively view-insensitive representation may be available, such that matching is not reliant on specific image-based properties. These two possibilities were tested in Experiment 4.

The low error rates indicate that almost all pairs of views could be matched accurately, although both RTs and errors indicate that matches to foreshortened targets were less efficient than matches to other targets. The foreshortened view disadvantage could be due to various factors. One possible factor is that it is harder to derive three-dimensional structure from a foreshortened view, slowing encoding of the image description. A second is that foreshortened views reveal fewer discriminating features than other views, and hence the image description is a poor match to a stored object representation. This could be the case if the stored representation is either view-invariant or view-specific and (here) activated by a 150° reference.

The identity benefit at the short ISI was probably due to matching based on rapidly available image-based and highly view-specific representations of the stimuli. Relatively low-level matching of identical stimuli can, at most, be only one part of the full process necessary for normal object recognition because such matching does not require access to stored knowledge about object structure and cannot overcome the problem of object constancy. It is interesting to note that matches to 30° targets were not as rapid as 150°–150° identical view matches, although the 30° and 150° views were mirror-images of each other. In contrast, in a picture matching experiment with ISIs ranging between 300 ms and 4,000 ms, Klatzky and Stoy (1974) found that identical view and mirror-image view matches were equally rapid, even at the shortest ISI. However, in Klatzky and Stoy's experiment there were only 12 objects, all of the experimental pictures were presented for participants to name before the experiment began, pictures were presented for a relatively long time (500 ms), and the data were collapsed across the three experimental blocks before analysis. These four points ensured that for all or most trials, participants were familiar with the stimuli. In Experiment 1, we showed that the identity benefit can be overridden, even at the short ISI, when participants are familiar with stimuli (e.g., for participants who received the long ISI block first), perhaps because a canonical representation can then be derived rapidly. Hence, Klatzky and Stoy may not have found any benefit for identical relative to mirror-image view matches because of the high within-experiment familiarity of stimuli in their study. In contrast, Experiment 1 here shows that, at short ISIs, there was no reflectional invariance for the matching of unfamiliar pictures of familiar objects.

Experiment 2

Experiment 1 showed that matches between different views of familiar objects could not be accounted for by strategies relying on mental rotation or rotation plus reflection. We have suggested instead that matching was based on accessing relatively abstract, durable representations from the reference and target. However, Experiment 1 does not allow us to specify whether matching involved access to stored knowledge about objects or whether it was based solely on information derived from the image, without recourse to stored knowledge. Experiment 2 was conducted to distinguish between these possibilities.

In one block of trials, participants matched upright pictures of objects, as in Experiment 1. In the other block, participants matched the same pictures, but both reference and target were inverted and, hence, were in an unfamiliar orientation. This manipulation allowed a controlled comparison of two conditions, one in which stored knowledge could readily be accessed (with upright stimuli) and one in which stored knowledge could be used but was more difficult to access (with inverted stimuli; see, e.g., Boucart & Humphreys, 1992; Jolicoeur, 1985). Image-based matching should be equal for upright and for inverted stimuli; if matches are faster for upright stimuli than for inverted stimuli, this would suggest that stored knowledge influences matching.

Experiment 2 used the same 36 objects as Experiment 1. Of these, 12 had a normal, familiar orientation and no horizontal axis of symmetry. These stimuli appeared unfamiliar and disoriented when inverted and were termed the *inverting items* (Appendix B lists the 12 inverting items; see also Figure 2). The remaining 24 noninverting items either had a horizontal axis of symmetry or did not have a familiar, upright orientation. If stored knowledge is involved in matching, the inverting items should show larger effects of inversion than the noninverting items (which should show little or no effect). The noninverting items were included as controls, because any increase in RTs in the inverted block may have been due to image-based matching being disadvantaged for inverted stimuli. For instance, when inverted, all stimuli appeared to be suspended on a surface angled like a ceiling (see Figure 2), whereas upright stimuli appeared to lie on a surface angled like a floor, which is more ecologically familiar. This could have resulted in slower matching of all inverted stimuli, irrespective of whether they were familiar or unfamiliar when inverted (for noninverting and inverting items, respectively).

Experiment 2 also investigated a possible confound in Experiment 1. It might be argued that the 90° target was at a disadvantage in Experiment 1 because, if participants do rapidly reflect image descriptions as part of their matching strategy, then target images of the 60° and 120° and the 30° and 150° mirror-image views would have been activated twice as often as image descriptions of the 90° target view because all five targets were shown equally often. Differential effects of familiarity within Experiment 1 could thus have disadvantaged the foreshortened view. It is unlikely that within-experiment familiarity can explain all of the results of Experiment 1; for example, the 150° view was seen six times more often than the other views because it was always the reference, but the mirror image of the 150° reference, the 30° target, was not matched any faster than the 60° target. Nevertheless, to equate within-experiment familiarity in Experiment 2, a 150° reference was presented as in Experiment 1, and 30°, 60°, and 90° targets were shown equally often. The mirror image of the 60° view, the 120° view, was never shown.

Only the short ISI condition was used in Experiment 2 because this had produced the strongest effects of view in Experiment 1. Also, any effects of image-based matching should be most pronounced at a short ISI because representations need not then be coded in a durable form.

Method

Participants. There were 16 participants (7 men and 9 women).

Materials. The stimuli were a subset of those used in Experiment 1, and all stimuli were presented inverted as well as upright. The references were 150° views; the targets were 30°, 60°, and 90° views. As in Experiment 1, the 30° and 60° views were mirror-image reflections of the 120° and 150° views, respectively.

Design. Every trial began with the presentation of a 150° reference of one of the 36 objects. This was followed with equal probability by a target view of that object at 30°, 60°, or 90° (a match trial) or by a view of a different object at 30°, 60°, or 90° (a mismatch trial). For mismatch trials, 3 different objects, at different views, were paired with each reference. Two sets of 216 picture pairs were produced; they were identical, apart from having different object pairings in the mismatch trials. All of the picture pairs in each set were shown in a single block. Participants received two blocks of trials. In one, both the reference and the target were upright; in the other, all stimuli were inverted. In each block, all three target views of all 36 objects were shown twice, once on a match trial and once on a mismatch trial. There were 432 trials in total, with each target being shown four times in all. The order of presentation of trials within a block was random and different for each participant. The assignment of sets to the upright or the inverted block and the order of presentation of the upright and the inverted blocks were counterbalanced across participants. The ISI was 585 ms on all trials.

Apparatus and procedure. The apparatus and procedure were identical to those of Experiment 1, except that participants completed just two blocks of trials, and the experiment lasted about 30 min. Before the block of inverted pictures, participants were warned that all of the pictures in the block would be upside down. However, it was emphasized that it was not necessary to recognize an object to match the two views on a trial and that simple visual cues were sufficient to allow accurate and rapid matching.

Results

Two participants were replaced because they each had an overall error rate greater than the maximum of 10%. Mean correct same RTs over participants, for both inverting and noninverting items, are shown in Figure 4; errors are given in Table 3.

Two ANOVAs were carried out on mean correct RTs for same responses. The inverting and noninverting items were analyzed separately because there were different numbers of items in each group. For each set of stimuli, there were two within-subject variables, view of the target (30°, 60°, or 90°) and inversion of the target and reference (upright or inverted), and one between-subjects variable, inversion order (order of presentation of blocks [upright first or inverted first]).

For the 12 inverting items, there was a significant main effect of view, $F_1(2, 28) = 22.060$, $p < .001$, $MSE = 16{,}331$, $F_2(2, 22) = 11.907$, $p < .001$, $MSE = 36{,}589$. The

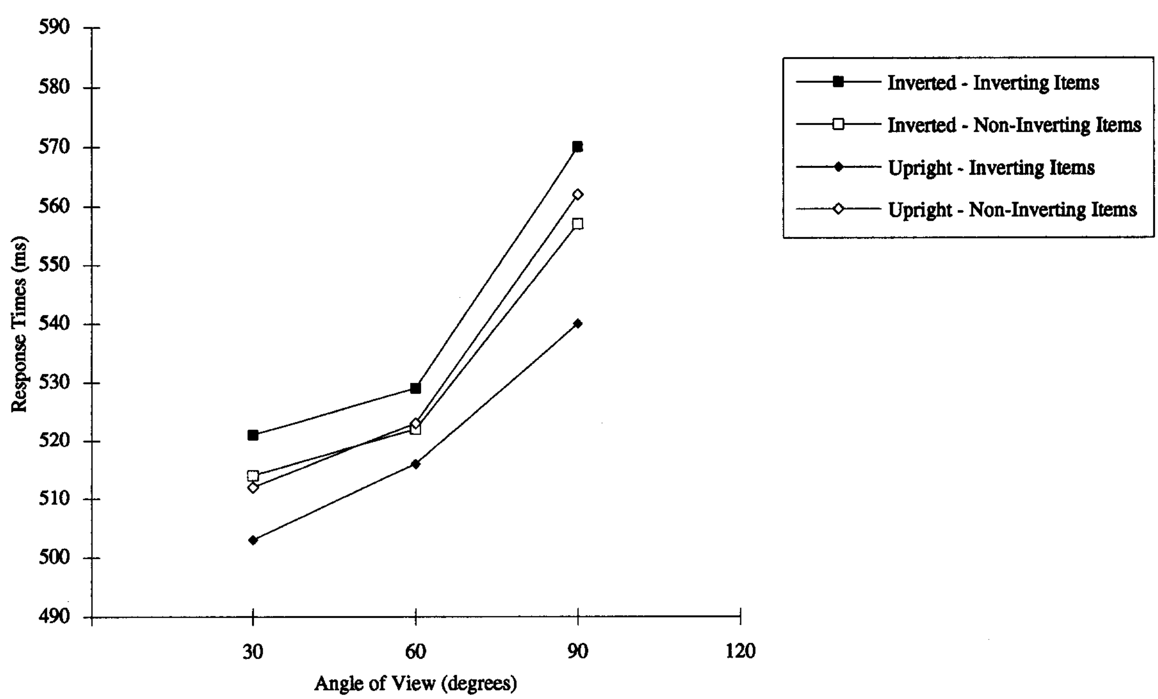

Figure 4. Mean correct response times for match trials, for the upright and inverted presentation of inverting and noninverting stimuli, as a function of target view: Experiment 2.

Table 3
Mean Error Rates (%) for Match Trials as a Function of Target View: Experiment 2

Block order and orientation	Item set	Target view 30°	60°	90°
Upright first				
Upright	Inverting	1.0	4.2	11.5
	Noninverting	2.6	6.3	6.8
Inverted	Inverting	2.1	4.2	10.4
	Noninverting	3.1	2.1	5.7
Inverted first				
Upright	Inverting	4.2	6.3	13.5
	Noninverting	3.6	4.2	5.2
Inverted	Inverting	2.1	11.5	17.7
	Noninverting	5.7	3.6	8.9

Note. See Appendix B for a list of the 12 inverting items; the noninverting items are the remaining 24 items listed in Appendix A.

90° targets were matched slower than the 30° and 60° targets, which did not differ ($p < .01$; Newman–Keuls analysis). The effect of inversion was also significant, $F_1(1, 14) = 5.309$, $p < .04$, $MSE = 10{,}091$, $F_2(1, 11) = 8.677$, $p < .02$, $MSE = 15{,}147$. Upright matches were faster than inverted matches. Inversion order was not significant, $F_1(1, 14) = 0.035$, $p > .8$, $MSE = 2{,}140$, $F_2(1, 11) = 2.049$, $p > .1$, $MSE = 2{,}879$. Only the Inversion × Inversion Order interaction was significant, $F_1(1, 14) = 25.061$, $p < .001$, $MSE = 47{,}632$, $F_2(1, 11) = 21.491$, $p < .001$, $MSE = 80{,}830$. Inverted stimuli were matched slower than upright stimuli for the participants who had the inverted block first ($p < .01$; Newman–Keuls analysis), but there was no significant difference between upright and inverted matches for the participants who had the upright block first.

For the 24 noninverting items, the main effect of view was significant, $F_1(2, 28) = 39.196$, $p < .001$, $MSE = 18{,}969$, $F_2(2, 46) = 22.341$, $p < .001$, $MSE = 60.980$. Inversion was not significant, $F_1(1, 14) = 0.018$, $p > .8$, $MSE = 42$, $F_2(1, 23) = 0.017$, $p > .8$, $MSE = 31$. Inversion order was significant across items only, $F_1(1, 14) = 0.055$, $p > .8$, $MSE = 2{,}880$, $F_2(1, 23) = 7.429$, $p < .02$, $MSE = 7{,}986$. Matches tended to be faster for participants who had the upright block first. There were two significant interactions: Inversion × Inversion order, $F_1(1, 14) = 6.085$, $p < .03$, $MSE = 14{,}106$, $F_2(1, 23) = 35.250$, $p < .001$, $MSE = 45{,}998$, and Inversion × Inversion Order × View, $F_1(2, 28) = 6.043$, $p < .007$, $MSE = 2{,}388$, $F_2(2, 46) = 4.755$, $p < .02$, $MSE = 5{,}673$.

The three-way interaction for the noninverting items only was decomposed by separate two-way analyses for each ISI order. For participants presented with the upright block first, the main effect of view was significant, $F_1(2, 14) = 14.636$, $p < .001$, $MSE = 7{,}009$, $F_2(2, 46) = 13.551$, $p < .001$, $MSE = 24{,}291$. The 90° targets were matched slower than the 30° and 60° targets, which did not differ ($p < .01$ for items; Newman–Keuls analysis). Inversion was significant across items only, $F_1(1, 7) = 2.455$, $p > .1$, $MSE = 7{,}841$, $F_2(1, 23) = 12.468$, $p < .002$, $MSE = 24{,}211$. Inverted matches tended to be faster than upright matches. The View × Inversion interaction was significant across participants only, $F_1(2, 14) = 6.783$, $p < .009$, $MSE = 1{,}702$, $F_2(2, 46) = 2.445$, $p > .09$, $MSE = 4{,}091$, with upright 90° targets being matched slower than all other targets ($p < .01$ for participants; Newman–Keuls analysis).

For participants presented with the inverted block first, the main effect of view was significant, $F_1(2, 14) = 25.813$, $p < .001$, $MSE = 12{,}623$, $F_2(2, 46) = 18.460$, $p < .001$, $MSE = 37{,}717$. The 90° targets were matched slower than 30° and 60° targets, which did not differ ($p < .01$; Newman–Keuls analysis). Inversion was significant across items only, $F_1(1, 7) = 4.373$, $p > .07$, $MSE = 6{,}307$, $F_2(1, 23) = 19.032$, $p < .001$, $MSE = 21{,}818$. Upright matches tended to be faster than inverted matches. The View × Inversion interaction was not significant, $F_1(2, 14) = 1.439$, $p > .2$, $MSE = 776$, $F_2(2, 46) = 1.216$, $p > .3$, $MSE = 1{,}827$. There was no evidence of a speed–accuracy trade-off.

The mean RTs for correct different responses for the 12 inverting items were 559 ms (4.3% errors) for upright stimuli and 577 ms (5.4% errors) for inverted stimuli. The main effect of view was not significant, $F_1(2, 28) = 0.579$, $p > .5$, $MSE = 528$, $F_2(2, 22) = 0.213$, $p > .8$, $MSE = 531$, but inversion was significant across participants and marginally significant across items, $F_1(1, 14) = 6.513$, $p < .03$, $MSE = 8{,}574$, $F_2(1, 11) = 4.291$, $p < .07$, $MSE = 10{,}219$. Upright matches tended to be faster than inverted matches. Inversion order was significant across items only, $F_1(1, 14) = 0.243$, $p > .6$, $MSE = 14{,}549$, $F_2(1, 11) = 12.761$, $p < .001$, $MSE = 27{,}333$. Matches tended to be faster for participants who had the upright block first. Only the Inversion × Inversion order interaction was significant, $F_1(1, 14) = 19.883$, $p < .001$, $MSE = 26{,}176$, $F_2(1, 11) = 24.551$, $p < .001$, $MSE = 41{,}230$. This interaction took the same form as that found for same trials; that is, inverted stimuli were matched slower than upright stimuli by the participants who had the inverted block first ($p < .01$; Newman–Keuls analysis), but there was no significant difference between upright and inverted matches for the participants who had the upright block first.

The mean RTs for correct different responses for the 24 noninverting items were 574 ms (4.4% errors) for upright stimuli and 568 ms (4.4% errors) for inverted stimuli. The main effect of view was significant across participants only, $F_1(2, 28) = 5.027$, $p < .02$, $MSE = 1{,}459$, $F_2(2, 46) = 1{,}808$, $p > .1$, $MSE = 5{,}264$, with 30° targets being matched faster than 90° targets ($p < .05$ for participants only; Newman–Keuls analysis). Inversion was not significant, $F_1(1, 14) = 0.706$, $p > .4$, $MSE = 794$, $F_2(1, 23) = 0.293$, $p > .5$, $MSE = 555$, and neither was inversion order, $F_1(1, 14) = 0.002$, $p > .9$, $MSE = 137$, $F_2(1, 23) = 0.737$, $p > .3$, $MSE = 1{,}997$. The only significant interaction was that between inversion and inversion order, $F_1(1, 14) = 12.388$, $p < .004$, $MSE = 13{,}929$, $F_2(1, 23) = 22.175$, $p < .001$, $MSE = 31{,}085$. This interaction took the same form as that found for same trials and occurred because matches were generally faster on the second block of trials. Inverted matches for participants who had the inverted stimuli first and upright matches for participants who had the upright stimuli first were slower than both inverted matches for participants who had the upright stimuli first ($p < .05$ for participants, $p < .01$ for items) and upright matches for participants who had the inverted stimuli first (not significant for participants, $p < .05$ for items).

Discussion

First, there was an effect of inversion that was specific to the set of inverting items, which had a familiar, upright orientation. This suggests that matching was not achieved solely on the basis of image-based representations but that stored knowledge was also important. The noninverting (control) items were matched equally rapidly when inverted as when upright (531 ms and 532 ms, respectively), whereas the inverting items were matched, on average, a reliable 20 ms slower when inverted than when upright (540 ms and 520 ms, respectively). It follows that the effect of inversion on the inverting items was not a consequence of participants' having to match inverted pictures per se; rather, the data implicate the involvement of stored knowledge in matching.

It might be argued that the effects observed were due to structural differences between the inverting and noninverting items; for instance, the inverting items may have possessed more symmetry or may have been more complex than the noninverting items. However, it seems highly unlikely that such geometric differences would result in differential effects on matching dependent on the depicted orientation of the object, as was observed in Experiment 2, rather than simply influencing the overall efficiency of matching.

Second, the overall effect of view was similar to that found in Experiment 1. There was a substantial foreshortened view disadvantage for both inverting and noninverting stimuli in both the upright and the inverted blocks. This counters the proposal that the foreshortened view disadvantage in Experiment 1 was due to relative intraexperimental familiarity with the different views, because this variable was controlled in Experiment 2. The view effect cannot be accounted for in terms of a simple depth rotation of the image description because, given a 150° reference, 90° targets require less rotation in depth to match than do 60° targets. Neither can the effect be explained in terms of participants both reflecting and rotating image descriptions. Matches of the 30° target could be achieved after a simple reflection of the 150° reference; matching the 60° view would require a 90° depth rotation, yet performance was not reliably different for the 30° and 60° targets. In addition, the 150° reference was seen three times more often than the target views, yet its mirror-reflected image, the 30° target, was not matched significantly faster than the 60° target.

Third, there was an additive relationship between the target's orientation in depth and the effect of inversion on inverting stimuli (Figure 4). If inverted stimuli were matched without recourse to stored knowledge, one might conclude that the disadvantage for foreshortened views was independent of stored knowledge. However, given the involvement of stored knowledge, inverted stimuli may be disadvantaged by taking longer to access stored knowledge. Inverted stimuli were not necessarily matched on image features alone.

Experiment 3

The results from Experiments 1 and 2 disconfirm accounts of viewpoint effects in the matching of different views that propose (a) mental rotation, (b) reflection plus rotation, and (c) relative within-experiment familiarity. Matches were particularly slow to foreshortened targets. The reason may have been that foreshortened views are difficult to match to stored representations of objects, for instance, because such views obscure components present in stored object representations (e.g., see Biederman & Gerhardstein, 1993). Alternatively, the reason may have been that foreshortened views are difficult to encode: for instance, because the main axis of the image description is difficult to derive for foreshortened views, as Marr (1982) proposed, or because there are fewer orthogonal angles in foreshortened views, making it more difficult to construct an image-based description. Many objects, especially artifacts, have right-angled junctions that may constitute important distinguishing features. In good views of objects, there is normally sufficient information to derive these angles. In foreshortened views, angles are often distorted, so the information available to derive a junction can be misleading or inadequate, which may make such views difficult to match, irrespective of the nature of the stored representation to which the views are matched (e.g., whether it is object-centered or view-specific).

Experiment 3 investigated whether foreshortened views would always be matched more slowly than other views by using 90° references. If matching remained slow for 90° targets, this would indicate that foreshortened views are difficult to encode, irrespective of the view of the reference. Three targets were used: 60° (rated the best, preferred view), 150° (the good view used as the reference in Experiments 1 and 2), and 90°. Performance was examined at two ISIs. There may be a foreshortened view identity benefit at the short ISI mediated by a highly view-specific, image-based representation. However, this representation appears to be transient (e.g., the identity benefit dissipated at the long ISI in Experiment 1). At a long ISI, matches may necessarily involve stored representations (accessed by the reference) that, if optimally activated by canonical views, may penalize 90° targets while benefiting 60° targets.

Method

Participants. There were 16 participants (8 men and 8 women).

Materials. The stimuli were a subset of those used in Experiment 1. The references were 90° views; the targets were 60°, 90°, and 150° views.

Design. Every trial began with the presentation of a 90° reference of one of the 36 objects. This was followed with equal probability by a target view of that object at 60°, 90°, or 150° (a match trial) or by a view of a different object at 60°, 90°, or 150° (a mismatch trial). For mismatch trials, 3 different target objects, at different views, were paired with each reference. Two sets of 216 pairings of stimuli were produced. Different object pairings were used for the mismatch trials for the second picture pair set; otherwise, the two sets were identical. In each of the two picture pair sets, all three target views of all 36 objects were shown twice, once on a match trial and once on a mismatch trial. There were 432 trials in total, with each target being shown four times in all. The order of presentation of picture pairs within each set was random and different for each participant. The order of presentation of the two picture pair sets was balanced across participants. For each participant, one picture pair set was presented with a short ISI of 585 ms, and the other set was presented with a long ISI of 2,510 ms. The assignment of picture pair sets to ISI condition and the order of presentation of ISI were balanced across participants.

Apparatus and procedure. The apparatus and procedure were identical to those of Experiment 2.

Results

Six participants were replaced because they had an overall error rate greater than 10%. An ANOVA was carried out on mean correct RTs to same responses. There were two within-subject variables, view of the target (60°, 90°, or 150°) and ISI (short [585 ms] or long [2,510 ms]), and one between-subjects variable, ISI order (order of presentation of ISI blocks [short ISI first or long ISI first]). Mean correct same RTs across participants are shown in Figure 5; errors are given in Table 4.

The main effect of view was significant, $F_1(2, 28) = 20.556$, $p < .001$, $MSE = 18{,}232$, $F_2(2, 70) = 33.881$, $p < .001$, $MSE = 99{,}458$; ISI, $F_1(1, 14) = 1.418$, $p < .2$, $MSE = 8{,}092$, $F_2(1, 35) = 9.719$, $p < .004$, $MSE = 22{,}685$, and ISI order, $F_1(1, 14) = 0.763$, $p < .3$, $MSE = 41{,}435$, $F_2(1, 35) = 94.173$, $p < .001$, $MSE = 182{,}245$, were significant across items only. Short ISI matches tended to be faster than long ISI matches, and matches tended to be faster for participants who had the long ISI block first.

The View × ISI interaction was significant, $F_1(2, 28) = 12.522$, $p < .001$, $MSE = 6{,}873$, $F_2(2, 70) = 16.877$, $p < .001$, $MSE = 32{,}902$, whereas the ISI × ISI Order interaction was significant across items only, $F_1(1, 14) = 1.410$, $p > .2$, $MSE = 8{,}049$, $F_2(1, 35) = 13.823$, $p < .001$, $MSE = 30{,}065$. These interactions were qualified by a reliable three-way View × ISI × ISI Order interaction, $F_1(2, 28) = 4.896$, $p < .02$, $MSE = 2{,}687$, $F_2(2, 70) = 5.043$, $p < .01$, $MSE = 10{,}859$.

The three-way interaction was decomposed by separate two-way analyses for each ISI order. Data for participants who had the short ISI block first and for participants who had the long ISI block first are shown separately in Figure 5. For participants who had the short ISI block first, the main effect of view was significant, $F_1(2, 14) = 6.448$, $p < .02$, $MSE = 7{,}617$, $F_2(2, 70) = 15.533$, $p < .001$, $MSE = 45{,}216$, whereas the main effect of ISI was not, $F_1(1, 7) = 0.000$, $p > .9$, $MSE = 0.0$, $F_2(1, 35) = 0.097$, $p > .7$, $MSE = 259$. The View × ISI interaction was significant, $F_1(2, 14) = 10{,}480$, $p < .002$, $MSE = 8{,}018$, $F_2(2, 70) = 19.641$, $p < .001$, $MSE = 37{,}832$. At the short ISI, 90°

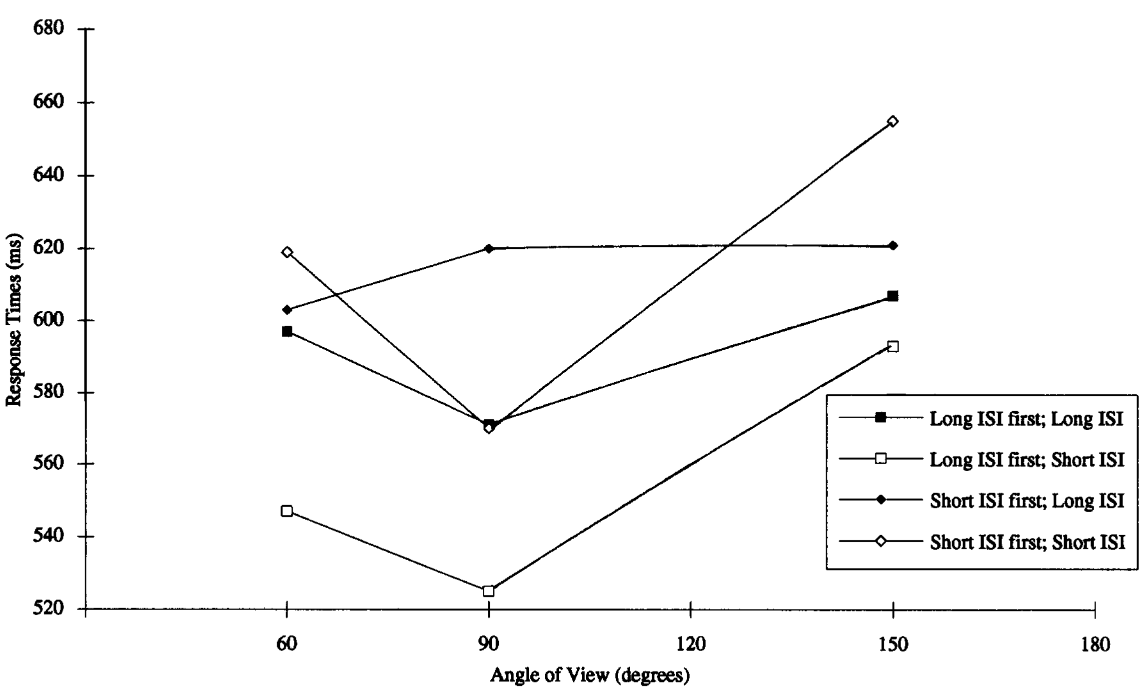

Figure 5. Mean correct response times for match trials at short and long ISIs, for participants presented with the short ISI first or the long ISI first, as a function of target view: Experiment 3. ISI = interstimulus interval.

Table 4
Mean Error Rates (%) for Match Trials as a Function of Target View: Experiment 3

Block order and ISI	Target view		
	60°	90°	150°
Short ISI first			
Short	3.8	1.0	9.4
Long	7.3	6.9	8.7
Long ISI first			
Short	6.9	4.5	11.1
Long	10.1	6.3	11.8

Note. ISI = interstimulus interval.

targets were matched faster than 60° targets ($p < .05$ for participants, $p < .01$ for items; Newman–Keuls analysis), which were, in turn, matched faster than 150° targets (not significant for participants, $p < .01$ for items). At the long ISI, there were no significant differences between 60°, 90°, and 150° target matches.

For participants who had the long ISI block first, the main effect of view was significant, $F_1(2, 14) = 18.253$, $p < .001$, $MSE = 10{,}816$, $F_2(2, 70) = 23.089$, $p < .001$, $MSE = 54{,}544$. For items only, 90° targets were matched faster than 60° targets ($p < .01$; Newman–Keuls analysis), which were, in turn, matched faster than 150° targets ($p < .01$). ISI was significant across items only, $F_1(1, 7) = 2.490$, $p > .1$, $MSE = 16{,}141$, $F_2(1, 35) = 28.745$, $p < .001$, $MSE = 52{,}491$. Short ISI matches tended to be faster than long ISI matches. The View × ISI interaction was significant across participants only, $F_1(2, 14) = 4.636$, $p < .03$, $MSE = 1{,}542$, $F_2(2, 70) = 2.724$, $p > .07$, $MSE = 5{,}929$. For participants only, at the short ISI, 90° targets were matched faster than 60° targets ($p < .05$; Newman–Keuls analysis), which were, in turn, matched faster than 150° targets ($p < .01$). At the long ISI, 90° targets were matched faster than 60° targets ($p < .05$) and 150° targets ($p < .01$). There were no significant differences between 60° and 150° target matches. There was no evidence of a speed–accuracy trade-off.

The mean RTs for correct different results were 618 ms (4.9% errors) at the short ISI and 636 ms (6.1% errors) at the long ISI. The main effect of view was not significant, $F_1(2, 28) = 0.531$, $p > .5$, $MSE = 362$, $F_2(2, 70) = 1.172$, $p > .3$, $MSE = 2{,}773$; ISI, $F_1(1, 14) = 1.295$, $p > .2$, $MSE = 7{,}390$, $F_2(1, 35) = 16{,}171$, $p < .001$, $MSE = 34{,}330$, and ISI order, $F_1(1, 14) = 0.384$, $p > .5$, $MSE = 20{,}194$, $F_2(1, 35) = 39{,}961$, $p < .001$, $MSE = 87{,}926$, were significant across items only. Short ISI matches tended to be faster than long ISI matches, and matches tended to be faster for participants who had the long ISI block first. The ISI × ISI Order interaction was significant across items only, $F_1(1, 14) = 1.143$, $p > .3$, $MSE = 6{,}523$, $F_2(1, 35) = 9.590$, $p < .004$, $MSE = 27{,}779$. This interaction took the same form as that found for same trials. Short ISI matches for participants who had the long ISI block first were faster than all other types of match ($p < .01$; Newman–Keuls analysis for items only). No other interactions were significant.

Discussion

There were three main findings. First, the results clearly demonstrate the identity benefit, which was also found in Experiment 1. However, in Experiment 3, the facilitation due to matching identical views was associated with the 90° view. Thus, foreshortened targets can be matched reliably faster than other targets under certain conditions (identical view matches at a short ISI in this case). Second, participants matched 150° targets slower than 60° targets at the short ISI. In comparison with the 150° view, the 60° view was both more similar in view to the 90° reference and rated as more canonical; either or both factors could have caused the 60° target advantage relative to the 150° target. Third, as in both Experiment 1 and Ellis and Allport (1986), matches at a short ISI were generally faster than those at a long ISI.

The identity benefit in Experiment 3 actually appeared to be stronger than that found in Experiment 1; it was significant at the short ISI block for both ISI orders and at the long ISI for participants receiving the long ISI first. At the short ISI, we suggest that a view-specific, image-based representation allows rapid matches to take place between identical stimuli, resulting in an identity benefit for foreshortened views here. As in Experiment 1, the effects of view were attenuated at a long ISI and on the second block of trials.

These results present difficulties for an encoding account of the foreshortened view disadvantage found in Experiments 1 and 2. Such an account would predict that matches to 90° targets should be slower than all other matches, irrespective of the view of the reference. Clearly, rapid identical view matches can be made with foreshortened stimuli. However, transient, image-based representations may mediate efficient identical view matching, and this may mask difficulties in accessing a more abstract, durable representation from foreshortened views. It may remain the case that this more abstract representation is harder to encode from foreshortened views. Against this is the present finding of an advantage for foreshortened views at the long ISI for participants who had the long ISI first. The data from Experiment 1 suggested that any transient, image-based representation of the reference would have dissipated at the long ISI, so the present advantage for foreshortened targets cannot be linked to such a representation. Instead, durable, abstract, view-specific representations are implicated in both the present advantage for foreshortened targets and their disadvantage in Experiment 1 at long ISIs. Such view-specific representations may be less precisely tied to image features than are the transient, image-based representations involved at short ISIs. One remaining problem is that foreshortened views were seen four times more frequently than other views in Experiment 3. An effect due to the frequency of presentation could have helped participants encode foreshortened views. Experiment 4 controlled for this by directly comparing performance with 90° and 150° references, which were balanced for frequency of presentation.

In Experiment 3, 60° targets were matched more efficiently than 150° targets. Experiments 1 and 2 eliminated simple reflection and rotation accounts of the matching process, so the advantage of 60° over 150° targets must have a different cause. It could be due to either of two factors. First, the 60° view was independently rated the most preferred view and hence is probably more canonical than the 150° view. Canonical (60°) views of objects may match abstract, stored representations more efficiently than noncanonical (150°) views, for instance, because canonical views contain more diagnostic image features. Second, as suggested earlier, different view matches may be mediated by an abstract representation that remains view-specific. In this latter case, 90° references would match 60° targets better than 150° targets because 90° and 60° stimuli are more similar in view than 90° and 150° views. Because the 150° view is neither canonical nor similar in view to the 90° reference, the two factors of canonicality and view similarity were confounded in Experiment 3. However, canonicality seems less likely to be important than the similarity in view between the reference and target in Experiment 3. First, there was a benefit for noncanonical foreshortened targets, even at the long ISI, for participants presented with the long ISI block first. Second, the 60° benefit over 150° targets disappeared at the long ISI. If 60° targets benefit because the 60° view is canonical, then 60° targets should enjoy an advantage at the long ISI over both the less canonical 150° view and the noncanonical 90° view (because transient, image-based representations would no longer be available to mediate identical view matches for foreshortened targets). Neither of these effects occurred. In contrast, view similarity effects between the reference and target can explain the results from Experiment 3 if view similarity effects are assumed to be present but to weaken across the ISI.

Although the present evidence suggests that temporary, highly view-specific, image-based representations are derived fastest (given the identity benefit at the short ISI), the more durable, abstract, view-specific representations must also be accessed rapidly to mediate different view matches at short ISIs. At the long ISI, the transient, image-based representation appears to exert little or no influence, and view effects are qualified by the effects of ISI order. At the long ISI, foreshortened targets were matched slower than all other targets in Experiment 1 and faster than all other targets in Experiment 3 for participants shown the long ISI block first. In contrast, there were no differences between long ISI matches for the different target views for participants shown the short ISI block first in both Experiments 1 and 3. This ISI order effect suggests that experience with the stimuli was an important factor in both experiments. View effects occur at the long ISI only if participants are unfamiliar with the stimuli. This suggests that with practice, there is rapid access to a number of different abstract, view-specific representations of an object during the long ISI, enabling all targets to be matched about equally efficiently. We return to this in the General Discussion section.

Experiment 4

Experiment 4 investigated whether different view matches were sensitive to the view of the reference by comparing 90° and 150° references in a single study. Targets could be seen at 60°, 90°, 150°, or 180° views. This allowed view similarity effects between the reference and target to be systematically examined. View similarity effects should favor 180° targets following 150° references but 60° targets following 90° references (each being 30° rotations in depth). This would indicate that the representations mediating different view matches are view sensitive. Matching based on a single, stored representation should favor the canonical 60° target, irrespective of the view of the reference.

In addition to incorporating two references into the study, Experiment 4 also varied the nature of the ISI. Ellis and Allport (1986) showed that the transient, highly view-specific representation mediating identical view matches was disrupted by a meaningless mask presented in the ISI, but the abstract, durable representation mediating different view matches was unaffected by a mask (the advantage at the short ISI for identical view matches over different view matches, but not the advantage for different view matches over same name matches, was eliminated by a pattern mask). We compared matching with and without an irrelevant pattern mask. The identity benefit was predicted to be disrupted by the mask. Only the short 585 ms ISI was used.

Method

Participants. There were 16 participants (8 men and 8 women).

Materials. The stimuli were a subset of those used in Experiment 1. The references were 90° and 150° views; the targets were 60°, 90°, 150°, and 180° views. The pattern mask used was designed to maximize disruption to low-level, image-based representations without providing any meaningful part or object-level information that might disrupt high-level, abstract representations. The mask was composed of small, meaningless, overlapping line fragments taken from pictures of the objects presented in Experiment 4 but derived from different views to those shown (see Figure 6). The mask occupied a square of 8 × 8 cm, which was greater than the 6 × 6 cm area occupied by the pictures.

Design. Each trial began with the presentation of either a 90° or a 150° reference of one of the 36 objects. The reference was followed with equal probability by a target view of that object at 60°, 90°, 150°, or 180° (a match trial) or a view of a different object at 60°, 90°, 150°, or 180° (a mismatch trial). For mismatch trials, 8 different target objects, 2 at each of the four target views, were paired with the 90° and 150° references of each object. Two sets of 576 pairings of pictures were produced. Different object pairings were used for the mismatch trials for each picture pair set; otherwise, the two sets were identical. In each of the sets, all four targets of all 36 objects were shown four times as the target on a trial, on two trials (one match and one mismatch trial), they were preceded by a 90° reference, and, on two other trials, they were preceded by a 150° reference. There were 1,152 trials in total, with each target being shown eight times in all.

Each set of 576 picture pairs was divided into two blocks of 288 pairs, so that each block took approximately 10 min to complete. Each block of 288 trials contained four match and four mismatch trials for each object. The 90° and 150° references each occurred on half of the trials within each block. The order of presentation of trials within a block was random and different for each participant.

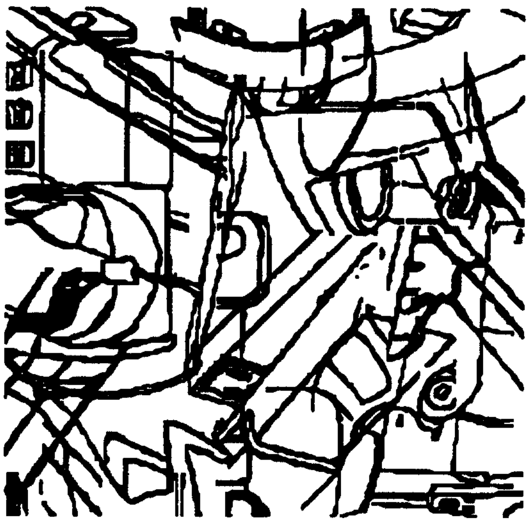

Figure 6. Pattern mask presented in Experiment 4.

The two blocks derived from the same picture pair set were always shown consecutively. The order of presentation of blocks within a picture pair set was balanced across participants, as was the order of presentation of the two sets. For each participant, one picture pair set was presented with a short blank ISI of 585 ms between the reference and target, and the other was presented with a short masked ISI of 585 ms. The pattern mask was the same on every trial. The mask was presented immediately after the reference and was immediately followed by the target. The assignment of picture pair sets to ISI type (blank or masked ISI) and the order of presentation of ISI type were balanced across participants.

The 90° and 150° views and the 60° and 180° views were thus presented equally often. This permitted a direct comparison between these pairs of views and avoided the frequency of presentation confound of Experiments 1 and 3.

Apparatus and procedure. The apparatus and procedure were identical to those of Experiment 1. The experiment lasted about 40 min.

Results

All participants had an overall error rate of less than 10%. An ANOVA was carried out on mean correct RTs to same responses. There were three within-subject variables: target (60°, 90°, 150°, or 180°), reference (90° or 150°), and ISI type (blank or masked ISI). There was one between-subjects variable: ISI order (order of presentation of blocks [blank ISI first or masked ISI first]). Mean correct same RTs across participants are shown in Figure 7; errors are given in Table 5.

The main effects of both target, $F_1(3, 42) = 8.294$, $p < .001$, $MSE = 2{,}033$, $F_2(3, 105) = 5.477$, $p < .002$, $MSE = 10{,}608$, and reference, $F_1(1, 14) = 21.871$, $p < .001$, $MSE = 7{,}188$, $F_2(1, 35) = 17.926$, $p < .001$, $MSE = 35{,}794$, were significant. Matches were faster for 150° references than for 90° references. The main effect of ISI type was significant across items only, $F_1(1, 14) = 0.990$, $p > .3$, $MSE = 5{,}498$, $F_2(1, 35) = 18.566$, $p < .001$, $MSE =$

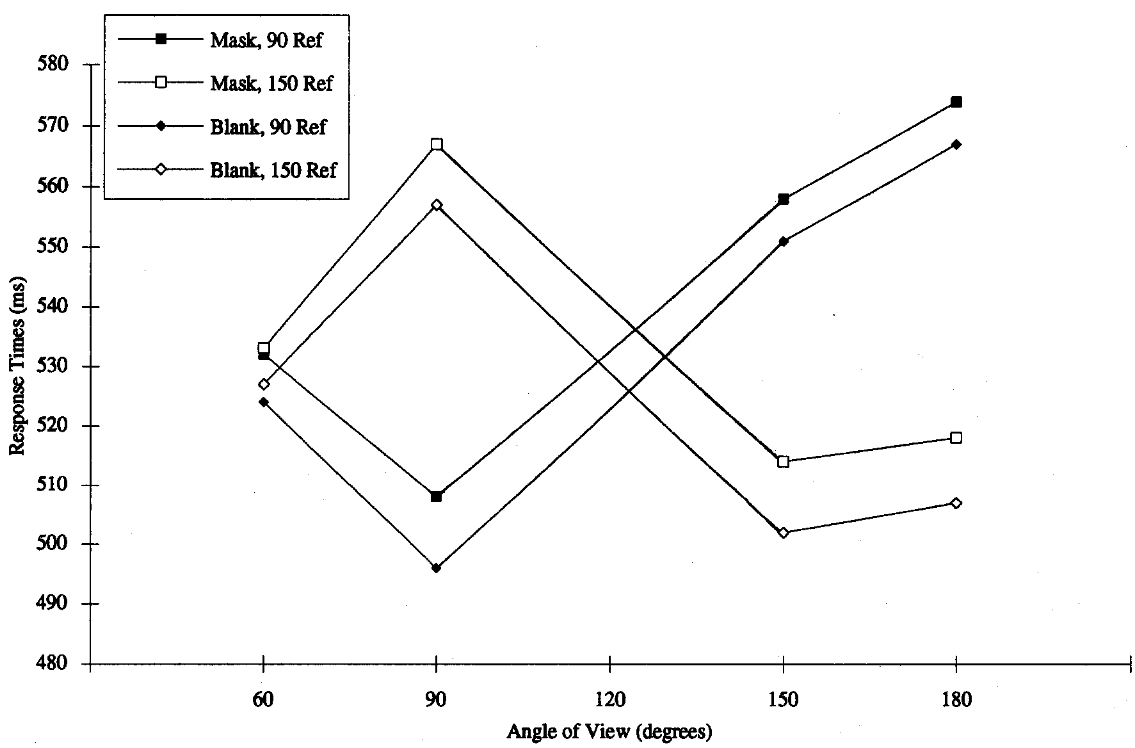

Figure 7. Mean correct response times for match trials with blank and masked interstimulus intervals for 90° and 150° references, as a function of target view: Experiment 4. Ref = reference.

Table 5
Mean Error Rates (%) for Match Trials as a Function of Target View: Experiment 4

Reference view and ISI type	Target view 60°	90°	150°	180°
90°				
Blank	4.2	2.8	8.7	13.5
Mask	5.7	3.3	8.2	10.0
150°				
Blank	4.0	8.2	3.8	4.2
Mask	5.7	8.2	4.3	4.9

Note. ISI = interstimulus interval.

23,484. Matches with a blank ISI tended to be faster than those with a masked ISI. The main effect of ISI order was not significant, $F_1(1, 14) = 0.017$, $p > .8$, $MSE = 1{,}460$, $F_2(1, 35) = 2.688$, $p > .1$, $MSE = 5{,}162$. There were two significant two-way interactions: Target × Reference, $F_1(3, 42) = 77.052$, $p < .001$, $MSE = 46{,}437$, $F_2(3, 105) = 112.883$, $p < .001$, $MSE = 225{,}119$, and ISI Type × ISI order, $F_1(1, 14) = 11.095$, $p < .005$, $MSE = 61{,}594$, $F_2(1, 35) = 168.919$, $p < .001$, $MSE = 285{,}448$. The ISI Type × ISI Order interaction occurred because matches were faster on the second block of trials. If the masked ISI block was presented first, matches were faster for blank ISIs than for masked ISIs ($p < .05$ for participants, $p < .01$ for items; Newman–Keuls analysis). Conversely, if the blank ISI block was presented first, matches tended to be faster for masked ISIs than for blank ISIs (not significant for participants, $p < .01$ for items).

For 90° references, 90° targets were matched faster than 60° targets ($p < .01$; Newman–Keuls analysis), which were, in turn, matched faster than 150° targets ($p < .01$), and 150° targets were matched faster than 180° targets ($p < .05$ for participants, $p < .01$ for items). For 150° references, there were no significant differences between 150° and 180° targets, but both were matched faster than 60° targets ($p < .05$), which were, in turn, matched faster than 90° targets ($p < .01$). Thus, with 90° references, 90° and 60° targets were matched faster than 150° and 180° targets; with 150° references, the opposite results held. With respect to target view, the following occurred: For the 60° target, there was no difference between matches with 90° and 150° references; for the 90° target, 90° references were matched faster than 150° references ($p < .01$); and for the 150° and 180° targets, 90° references were matched slower than 150° references ($p < .01$). There was no evidence of a speed–accuracy trade-off.

The mean RTs for correct different results were 566 ms (4.1% errors) with a blank ISI and 561 ms (4.3% errors) with a masked ISI; the mean RTs were 558 ms (3.8% errors) for 90° references and 569 ms (4.6% errors) for 150° references. The main effect of reference was significant, $F_1(1, 14) = 27.155$, $p < .001$, $MSE = 7{,}890$, $F_2(1, 35) = 31.941$, $p < .001$, $MSE = 39{,}987$, but the main effect of target was not significant, $F_1(3, 42) = 0.176$, $p > .9$, $MSE = 61.555$, $F_2(3, 105) = 0.245$, $p > .8$, $MSE = 555$. Matches were faster for 150° references than for 90° references. The main effect of ISI type was not significant, $F_1(1, 14) = 0.145$, $p > .7$, $MSE = 1{,}300$, $F_2(1, 35) = 3.361$, $p > 0.7$, $MSE = 3{,}840$, and ISI order was significant across items only, $F_1(1, 14) = 0.370$, $p > .5$, $MSE = 31{,}551$, $F_2(1, 35) = 81.208$, $p < .001$, $MSE = 126{,}247$. Matches tended to be faster for participants who had the masked ISI block first. The only significant interaction was that between ISI type and ISI order, $F_1(1, 14) = 6.822$, $p < .03$, $MSE = 61{,}068$, $F_2(1, 35) = 147.011$, $p < .001$, $MSE = 262{,}389$. This interaction took the same form as that found for same trials and occurred because matches tended to be faster on the second block of trials. If the masked ISI block was presented first, matches tended to be faster for blank ISIs than for masked ISIs (not significant for participants, $p < .01$ for items; Newman–Keuls analysis). Conversely, if the blank ISI block was presented first, matches tended to be faster for masked than for blank ISIs (not significant for participants, $p < .01$ for items).

Discussion

These results have important implications for understanding object encoding and matching. In particular, RTs were dependent not only on the view of the target but also on the similarity in view of the reference and target. There was a crossover interaction reflecting effects of both reference and target viewpoint: 150° and 180° targets were slow to be matched to 90° references and fast to be matched to 150° references, whereas the reverse held for 60° and 90° targets (see Figure 7). This indicates that effects of target view are not simply a result of the time taken to encode targets or the canonicality of targets. This crossover would not occur if different view matches were mediated by a single, object-centered, view-invariant representation derived from the reference. Such a representation should not favor any view (or it should favor only canonical views, such as the 60° view), and any effects of view on matching should reflect the target view only and should be independent of the view of the reference. However, the reference view was clearly important for both identical view matches and different view matches. Such effects of the similarity of reference and target views suggest that a number of different abstract, view-specific representations of each object mediate different view matches.

These effects of the similarity in view of the reference and target for different view matches are unlikely to be due to the involvement of transient, image-based representations, even though only short ISIs were used here. In Experiment 4, the 30° depth rotations that separated different views of the same object were close to the rotations shown in other studies to require abstract, durable representations to mediate different view matches, rather than the transient image-based representations assumed to mediate identical view matches (Ellis & Allport, 1986; Ellis et al., 1989). Also, matches between references and targets rotated 30° in depth were found to be slower than identical view matches in Experiments 1 and 3. Thus, 30° depth rotations are sufficient to disrupt matches mediated by transient image-based representations. Finally, in Experiment 4 we found no effect of an intervening mask, which ought to disrupt any transient image-based representations (Ellis & Allport, 1986). We conclude that abstract, durable representations mediated performance for different view matches and that these representations are view-specific.

This argument is supported by comparisons between the results of Experiments 1 and 3, at the long ISI, for participants who had the long ISI block first. Relative to 150° targets, there was a strong disadvantage for 90° targets in Experiment 1 (in which 150° references were used) but an advantage in Experiment 3 (in which 90° references were used). Because the same 90° targets were encoded in Experiments 1 and 3, this implies that the efficiency of matching is crucially determined by the relationship between the reference and target and not just the efficiency of encoding targets. The 90° view was seen more frequently in Experiment 3 than in Experiment 1, but frequency of presentation cannot account for performance with 90° targets in Experiment 3. In Experiment 1, the 150° view was seen many more times than any other view, yet there was only a weak identity benefit for 150° targets relative to other targets at the short ISI and none at the long ISI. This implies that in Experiment 3, at the long ISI, there should have been no identity benefit for 90° targets, and any encoding difficulty for 90° views should have been revealed. This was not the case.

Finally, although there was a trend for the mask to slow RTs relative to a blank ISI, we failed to find any interaction between the mask and the view of either the target or the reference (see Figure 7). This result runs counter to the results of Ellis and Allport (1986), who found that a pattern mask during a 100 ms ISI eliminated the identity benefit. Ellis and Allport used photographs of objects, which were more complex than the line drawings used in the current experiments. They also used a shorter ISI (100 ms, in comparison with the 585-ms ISI used in our Experiment 4). It is possible either that masking effects are more pronounced at shorter ISIs (although Ellis et al., 1989, found that the size of the identity benefit was not greatly affected by ISI, being 16 ms with a 100 ms ISI and 19 ms with a 600 ms ISI) or that the image description formed from line drawings is less prone to disruption by a pattern mask than the image description of photographs (e.g., because coding is simpler and perhaps more rapid for line drawings).

General Discussion

A series of four experiments investigated the representations mediating the matching of objects depicted from different views. First, Experiments 1 and 2 indicated that participants were not using simple mental rotation or mirror-image reflection plus rotation strategies to mediate matches between different views of an object. Instead, the data suggest that abstract, durable representations mediate the matching of different views.

Second, Experiments 1, 3, and 4 each demonstrated a benefit for matching identical views of objects; this effect was shortlived, however, only being reliable at the short ISI (Experiments 1 and 3). This implicates the involvement of a transient image-based, view-specific representation of the reference in mediating identical view matches. In Experiment 1 we found that, contrary to previous results (Klatzky & Stoy, 1974), mirror-image view matches were not as rapid as identical view matches. This suggests that the transient image-based representations assumed to underlie the identity benefit are not reflectionally invariant, and therefore their use is strictly limited to mediating identical view matches at short ISIs.

Third, different view matches were strongly influenced by the similarity in view of the reference and the target. Following a 90° reference, 60° targets were matched faster than 150° and 180° targets (Experiments 3 and 4). Conversely, following a 150° reference, 180° targets were matched faster than 60° and 90° targets (Experiment 4). Matching was more efficient if the reference and target were similar in view, independent of the canonicality of the views presented. Such "view similarity" effects suggest that the abstract representations mediating different view matches are view specific; the representations are not object centered because they cannot be matched to all views of an object equally efficiently (cf. Marr, 1982).

Throughout this article, we have measured visual similarity between target and reference views of an object in terms of the degree of rotation between the depicted viewpoints. It should be noted that the physical difference in view between the reference and target is unlikely to be the best psychological measure of the perceived visual similarity of the two views. Nevertheless, view similarity is important because it is likely to correlate strongly with factors that are important in determining perceived visual similarity but that are difficult to measure either objectively or quantitatively, such as the number of common features and parts across two views. Also, even if the factors determining perceived visual similarity could be measured, their nature, number, and relative importance would still be unspecified. Here we have shown that at least factors correlated with the angle of rotation between depicted views influence matching.

Finally, we reported evidence that matching was affected by stored knowledge derived from the reference. In Experiment 2, the reference and target were inverted. Items that were unfamiliar and disoriented when inverted were matched more slowly when inverted than when upright. Effects of stored knowledge were also apparent in Experiments 1 and 3, in which effects of view were eliminated at the long ISI for participants given the short ISI block first (Figures 3 and 5) so view effects disappeared once participants were familiar with the stimuli (see also Jolicoeur, 1985). One reason why prior experience with the stimuli might reduce view effects is that stored representations are established for objects in previously seen views. When a target is then represented at that view, matches can be made directly to the stored view rather than being mediated by a representation tied to the view of the reference. This proposal for the involvement of view-specific stored representations is similar to that of Tarr and Pinker (1989). According to such a view, object constancy is achieved by storing a number of different representations of each object, each tuned to a different view.

Following Ellis and Allport (1986) and Ellis et al. (1989), we could interpret our evidence in terms of the involvement of at least two types of representation in object matching: (a) transient, image-based, highly view-specific representations, which can mediate only identical view matches at short ISIs, and (b) abstract, durable representations, which mediate all different view matches. However, some of the distinctions made by Ellis and colleagues between these two types of representation were not supported by our data. First, Ellis and colleagues assumed that the latter representation was view invariant and object centered, but this was disconfirmed by our results, which suggest that it is view specific. Second, in Experiment 4, the presence of an intervening mask did not eliminate the identity benefit, which is assumed to be mediated by the first type of representation (cf. Ellis & Allport, 1986); in fact, the mask did not interact with view effects whatsoever. Third, the identity benefit was not always present at short ISIs and was not always absent at long ISIs. In Experiment 1, for participants shown the long ISI block first, there was no identity benefit at the short ISI, and, in Experiment 4, following a 150° reference, matches to 150° identical targets and 180° nonidentical targets were equally rapid. In Experiment 3, for participants shown the long ISI block first, the identity benefit was still present at the long ISI (see also Humphrey & Lupker, 1993). In Ellis and Allport (1986), ISI was a between-subjects variable, so ISI order effects could not be examined; in Ellis et al. (1989), ISI order effects were not reported.

As Humphrey and Lupker (1993) have argued, it is possible that some or all of the identity benefit reported by Ellis and colleagues (Ellis & Allport, 1986; Ellis et al., 1989) was due to the presentation of photographs rather than line drawings in their studies. Unlike line drawings, photographs possess surface characteristics such as texture that might permit the use of matching strategies based on low-level features. Such strategies could mediate only identical view matches, because surface characteristics are highly dependent on viewing conditions and would usually change significantly across different depth-rotated views of an object. If information about surface characteristics decays rapidly and is disrupted by masking and by changes in stimulus size, then the results reported by Ellis and colleagues can be accounted for.

A more parsimonious account of our results than that just provided can be postulated. If a number of different, view-specific representations of each object are stored (cf. Tarr & Pinker, 1989), then the target would be matched most efficiently if it accessed the same view-specific representation as the reference (i.e., an identity benefit would be observed). However, over time, stored representations of different views of the object may also be activated. The time course of activating these representations may be dependent on the view of the reference. In particular, stored representations similar in view to the reference may be activated before those of views rotated farther from the reference view. Thus, over time, the initial highly view-specific identity benefit would weaken, and eventually all views of the object that could access a stored representation would be matched efficiently because all stored representations would be activated. Such an account postulates only one type of representation rather than specifying distinct transient, image-based and abstract, durable representations. However, because the different view-specific representations of a given object are not all assumed to be activated simultaneously, view-specific effects on matching may still be observed, with initially more efficient matching of more similar views because such stimuli are more likely to activate either the same stored representation or two stored representations similar in view.

Other workers, notably Biederman and colleagues, have come to rather different conclusions concerning the representations mediating object processing based on studies of long-term priming in object naming tasks (Biederman & Cooper, 1991a, 1991b, 1991c, 1992; Biederman & Gerhardstein, 1993). Biederman and colleagues used a long-term priming task to investigate the limits of object constancy. Typically, participants named a set of prime pictures of objects in one block of trials and then named the same set of objects as targets in a second block. Some target pictures were transformed from their initial presentation. In general, naming latencies in the second block were facilitated relative to those in the first block, and the magnitude of any facilitatory priming effects was equally large when the pictures were identical and when they were transformed in size, lateral position, mirror-image reflection, and orientation in depth. Facilitatory priming was also as large when the pictures were complementary contour-deleted versions of each other as when they were identical. Moreover, this facilitation for transformed depictions of the same object was greater than that produced when two different exemplars from the same object class were named. From these results, Biederman and colleagues inferred that the additional facilitation in the identical view and different view conditions, in comparison with the different exemplar condition, is visual rather than conceptual in nature because the prime view picture, the different view targets, and the different exemplar targets all belong to the same object category. The results suggest that the same visual representation of an object is accessed, despite changes in image size, position, and orientation in depth, for different views of an object. Access to this common representation enables object recognition to be robust to changes in view, thus achieving object constancy.

There are several reasons why the results reported by Biederman and colleagues (Biederman & Cooper, 1991a, 1991b, 1991c; Biederman & Gerhardstein, 1993) might differ from ours; for instance, the time intervals separating the first and second presentations of the stimuli were much shorter here, and we used a matching rather than a naming task. It might be argued that the representations mediating matching and those mediating naming are different, and hence contrasting results would be expected from studies using the different tasks. Specifically, matching might involve episodic memories, because the task requires the target to be recognized as depicting the same object as the previously presented reference. Biederman and colleagues (Biederman & Cooper, 1992; Cooper et al., 1992) have argued that object identification tasks that tap episodic recognition memory involve a separate object processing system in addition to the system mediating object identification. They have suggested that object naming is mediated solely by the object identification system, probably represented in occipitotemporal pathways in the brain, whereas object matching, for instance, may also tap a motor-action system, probably represented in occipitoparietal pathways (Ungerleider & Mishkin, 1982). Viewpoint is important for the object action system (because the view of the object will determine the precise action required) but not for the object identification system. It is possible that both sets of results, those from long-term priming of picture naming and those from picture matching, reflect different aspects of object processing. We provided evidence that matching did tap stored representations of objects, because the matching of inverted stimuli was affected by whether the objects were normally seen in an upright orientation. Such stored representations are normally associated with the object identification system rather than with the episodic system involved in the on-line generation of actions to objects, and therefore it appears that the object identification system is involved in matching. Nevertheless, the view-sensitive effects reported here may reflect the involvement of view-sensitive representations that are used only in explicit recognition memory tasks and are not implicated in pure object identification tasks such as naming.

However, we suggest that although it is a possibility that all of the view effects we have reported are the result of processing mediated via the motor interaction system and that this system is not involved in object recognition, it seems most likely that there is a systematic effect of depth rotation on object recognition. Object naming studies using the same set of stimuli presented in our experiments produced effects of viewpoint on priming very similar to those reported here (Lawson, 1994; see also Tarr, Hayward, Gauthier, & Williams, 1994). Furthermore, at present, there is no consensus as to what is the most appropriate task with which to study object recognition (Tarr & Bülthoff, 1995), and Biederman and Gerhardstein (1993) themselves used a sequential picture matching task to examine the effects of depth rotation on object recognition.

The effects of depth rotation on performance might have emerged because different parts were visible across different views of an object. Biederman and Gerhardstein (1993) claimed that even in object naming tasks, priming is predicted to be viewpoint invariant only if the same parts are visible in the same spatial relations across different views of an object. When different object parts are visible, contrasting object representations are constructed, because these representations explicitly code the parts present in the image. The difficulty with this proposal, however, is in specifying the nature and scale of the visible parts. If object parts are defined sufficiently small, the proposal is unfalsifiable. Biederman and Gerhardstein (1993) claimed that small parts are important, but they put no lower limits on the size of parts that are coded in object representations, and they provided no other criteria for determining the importance of a part. Also, when visible parts change and priming reduces, one cannot assume that it is the change in the visible parts that is important relative to the correlated change in some other stimulus property (e.g., the global shape of the object). In the present study, effects of viewpoint change occurred even when there were relatively small (30°) rotations between images, when it was unlikely that there were changes in the main visible parts. Also, in object identification studies involving the same set of stimuli presented in the current studies, we found systematic effects of depth rotation on performance, which did not differ between objects rated to have major changes in their visible parts across different views and objects rated to have minimal changes (Lawson, Humphreys, & Watson, 1994). Furthermore, results from both sequential picture matching and naming tasks using novel objects in which quantitative changes were compared with qualitative changes in the visibility of parts across different views (Tarr et al., 1994) provided no support for the claim that the occlusion of parts across different views is a major factor determining whether view-invariant recognition is observed. The lack of an effect of major part changes, coupled with the sensitivity to viewpoint, suggests that familiarity and overall change in view are primary determinants of both identification and matching performance.

References

Bartram, D. J. (1976). Levels of coding in picture–picture comparison tasks. *Memory & Cognition, 4,* 593–602.

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). Object recognition and laterality: Null effects. *Neuropsychologia, 29,* 685–694.

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

Biederman, I., & Cooper, E. E. (1992). Size invariance in visual object priming. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 13,* 640–645.

Biederman, I., & Gerhardstein, P. C. (1993). Recognizing depth-rotated objects: Evidence for 3-D viewpoint invariance. *Journal of Experimental Psychology: Human Perception and Performance, 19,* 1162–1182.

Boucart, M., & Humphreys, G. W. (1992). Global shape cannot be attended without object identification. *Journal of Experimental Psychology: Human Perception and Performance, 18,* 785–806.

Cooper, E. E., Biederman, I., & Hummel, J. (1992). Metric invariance in object recognition: A review and further evidence. *Canadian Journal of Psychology, 46,* 191–214.

Edelman, S., & Weinshall, D. (1991). A self-organizing multiple-view representation of 3D objects. *Biological Cybernetics, 64,* 209–219.

Ellis, R., & Allport, D. A. (1986). Multiple levels of representation for visual objects: A behavioral study. In A. G. Cohn & J. R. Thomas (Eds.), *Artificial intelligence and its applications* (pp. 245–247). Chichester, England: Wiley.

Ellis, R., Allport, D. A., Humphreys, G. W., & Collis, J. (1989). Varieties of object constancy. *Quarterly Journal of Experimental Psychology, 41A,* 775–796.

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, 2,* 683–685.

Humphrey, G. K., & Jolicoeur, P. (1993). Visual object identification: Some effects of image foreshortening, monocular depth cues, and visual field on object identification. *Quarterly Journal of Experimental Psychology, 46A,* 137–159.

Humphrey, G. K., & Lupker, S. J. (1993). Codes and operations in picture matching. *Psychological Research, 55,* 237–247.

Humphreys, G. W., & Riddoch, M. J. (1984). Route to object constancy: Implications from neurological impairments of object constancy. *Quarterly Journal of Experimental Psychology, 36A,* 385–415.

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

Kelter, S., Grötzbach, H., Freiheit, R., Höhle, B., Wutzig, S., & Diesch, E. (1984). Object identification: The mental representation of physical and conceptual attributes. *Memory & Cognition, 12,* 123–133.

Klatzky, R. L., & Stoy, A. M. (1974). Using visual codes for comparisons of pictures. *Memory & Cognition, 2,* 727–736.

Lawson, R. (1994). *The effects of viewpoint on object recognition.* Unpublished doctoral dissertation, Birmingham University, Birmingham, United Kingdom.

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.

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

Marr, D. (1982). *Vision.* San Francisco: Freeman.

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

Perrett, D. I., Benson, P. J., Oram, M. W., & Hietanen, J. K. (1994). Issues of representation in object vision. In M. Farah & G. Ratcliff (Eds.), *The neuropsychology of high-level vision: Collected tutorial essays* (pp. 33–62). Hillsdale, NJ: Erlbaum.

Posner, M. I. (1969). Abstraction and the process of recognition. In J. T. Spence & G. Bower (Eds.), *The psychology of learning and motivation* (pp. 43–100). New York: Academic Press.

Posner, M. I., & Keele, S. W. (1967). Decay of visual information from a single letter. *Science, 158,* 137–139.

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

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

Tarr, M. J., Hayward, W. G., Gauthier, I., & Williams, P. (1994, November). *Geon recognition is viewpoint dependent.* Paper presented at the 35th Annual Meeting of the Psychonomic Society, St. Louis, MO.

Tarr, M. J., & Pinker, S. (1989). Mental rotation and orientation dependence in shape recognition. *Cognitive Psychology, 21,* 233–283.

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 behaviour* (pp. 549–586). Cambridge, MA: MIT Press.

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.

(Appendixes follow on next page)

Appendix A

The 36 Items Presented in Experiments 1–4

Banana	Paper clip
Bone	Pen
Camel	Pencil
Can opener	Tennis racquet
Car	Razor
Clothes peg	Ruler
Coat hanger	Saw
Comb	Scissors
Corkscrew	Screwdriver
Fork	Shoe
Glasses	Spanner
Hammer	Spoon
Iron	Stapler
Kangaroo	Telephone
Key	Toothbrush
Knife	Torch
Leek	Train
Loaf	Whisk

Appendix B

The 12 Inverting Items Presented in Experiment 2

Camel
Can opener
Car
Glasses
Iron
Kangaroo
Loaf
Ruler
Shoe
Stapler
Telephone
Train

Received July 19, 1993
Revision received December 6, 1994
Accepted February 21, 1995 ■