Recognizing Objects Seen From Novel Viewpoints: Effects of View Similarity and Time

John Schwoebel
Moss Rehabilitation Research Institute
and Temple University School of Medicine

Kavitha Srinivas
Boston College

View combination refers to a process that allows generalization to novel views of an object by the combination of object views represented in memory. The present experiments examined whether the perceived similarity between views or the temporal separation of views poses constraints on the view combination process. The results indicated that, although similar views are mapped onto the same object more easily than dissimilar views, dissimilar views, when experienced contingously in time, produce greater generalization after view combination than similar views. These findings suggest that although the mapping of dissimilar views to the same object is difficult, exposure to temporally contiguous dissimilar views produces better generalization than does exposure to similar views, presumably because they provide more structural information about the object. However, view combination appears to be constrained to views experienced continguously in time.

One of the fundamental problems in object recognition research concerns the issue of how successful recognition of objects occurs despite the fact that objects are often perceived from different viewpoints in depth. Research suggests that recognition of objects occurs by access to representations that specify viewpoint in depth, with costs in reaction time or accuracy when views are changed between study and test. Thus, sensitivity to changes in viewpoint has been consistently observed on a number of long-term memory tasks, including recognition, naming, object decision, and symmetry decision (Bartram, 1974; Bülthoff & Edelman, 1990, 1992; Bülthoff, Edelman, & Sklar, 1991; Edelman & Bülthoff, 1992; Humphrey & Khan, 1992; Rock & DiVita, 1987; Srinivas, 1993, 1995; Srinivas & Schwoebel, 1998; Tarr, 1995; Tarr, Williams, Hayward, & Gauthier, 1998; but see Biederman & Gerhardstein, 1993, for an exception). Viewpoint sensitivity has also been observed on many short-term memory tasks that require participants to determine whether sequentially presented objects are the same or different (Bartram, 1976; Ellis & Allport, 1986; Ellis, Allport, Humphreys, & Collis, 1989; Hayward, 1998; Kelter et al., 1984; Klatzky & Stoy, 1974; Lawson & Humphreys, 1996; but see Biederman & Gerhardstein, 1993, and Srinivas, 1995, for exceptions). The majority of these findings suggest that the representations formed are sensitive to the viewpoint that an object is initially seen from and that there are computational costs involved in identifying or recognizing novel views of objects. How are view-sensitive representations of objects used to allow generalization to novel views? The focus in this article is on a specific mechanism known as view combination that allows transfer to novel views from the combination of two or more representations of views of an object in memory.

Empirical support for view combination was first reported by Bülthoff and Edelman (1992). In their study, participants were presented with two different training views of each wire-frame object, which were separated by a 75° rotation in depth. Each training view was composed of a motion sequence of views rotating through 15° in depth. Thus, for instance, the participant was presented with a 0° and a 75° view, each of which was seen rotating through 15° in depth. Following training with the two views, participants saw novel views of the object that were either interpolated views (between 0° and 75°) or extrapolated views (between 75° and 360°). These novel views of studied objects were presented with a set of distractor objects, and participants were asked to determine whether each object was previously studied or nonstudied. Recognition accuracy on this task was found to be better for interpolated views than for extrapolated views. Interestingly, the advantage for interpolated views was observed even when the angle of rotation between these novel views and the nearest training view was held constant. Bülthoff and Edelman (1992) attributed this advantage to the fact that when views of objects are combined through a process of linear interpolation, interpolated views show better generalization than do extrapolated views. Subsequently, Tarr (1995) also reported an effect of view combination in the naming of novel objects composed of several bricks. Specifically, he found that naming times were relatively
constant for interpolated views between two training views of an object but increased linearly as a function of rotation from the nearest training view for extrapolated views that were beyond the training views. This failure to observe linear effects of rotation in depth for interpolated views was interpreted as being consistent with the view combination mechanism proposed by Bülthoff and Edelman (1992).

In a recent study, we demonstrated three additional facets of the view combination process (Srinivas & Schwoebel, 1998). First, we found that view combination can allow generalization to extrapolated views when the trained set is composed of objects with distinct part structures. Thus, view combination appears to operate not only for within-category discrimination (i.e., the discrimination of objects with highly similar structures, as in the Bülthoff & Edelman, 1992, and Tarr, 1995, studies), but for basic-level discrimination as well (i.e., the discrimination of objects with distinct parts, as in our study). Second, the advantage we obtained for view combination was over and above generalization from each of the training views when they were presented in isolation. Thus, exposure to 80° and 110° training views of an object produced a generalization advantage to a 0° test view, over and above exposure to the training views in isolation. Third, and most surprising, generalization from view combination was observed even when the views were temporally separated by an average of 26 intervening views of other objects. Thus, view combination does not appear to be dependent on temporal contiguity between views. This set of findings suggests that two views of an object can be combined to improve recognition of subsequently experienced novel views. If this is the case, the two views must first be mapped to the same object before combination can take place. How might this occur? We hypothesized that the views are mapped on the basis of the perceptual similarity between the two views of the object. Specifically, whenever an object is perceived, features visible in that particular viewpoint are represented in memory. Upon exposure to a second view of the object, the visible features from this new viewpoint are compared with the features of objects already represented in memory. If there is a sufficient amount of similarity between the features of a previously represented view of an object and the currently perceived novel view, then the two views are considered to result from the same object. Accordingly, similar views of an object should be more easily mapped to the same object for view combination than should dissimilar views.

However, it is difficult to map dissimilar views to the same object; once this mapping occurs, the less redundant sets of features in these views should allow for better generalization to new views. In other words, if the mapping process is somehow facilitated across dissimilar views, greater generalization should occur from the combination of dissimilar views than of similar views. From prior reports, the mapping process across views seems to be dependent on the similarity between the features of a previously represented view of an object and the currently perceived novel view, then the two views are considered to result from the same object. Accordingly, similar views of an object should be more easily mapped to the same object for view combination than should dissimilar views.

In Experiment 1, ratings of visual similarity between different views of each of the 21 different objects were collected and submitted to MDS analyses. This provided us with an empirically derived measure of the similarity between different views of each object and allowed us to examine the underlying dimensions that govern similarity between views.

Method

Participants. Forty-two Boston College undergraduates participated to partially fulfill a course requirement.

Materials. Twenty-one novel computer-generated 3-D objects were constructed for this experiment. Each object was bilaterally symmetrical and was composed of a main part with several smaller parts attached to the main part. Thus, each object was composed of distinct volumetric parts arranged in different spatial relations (see Figure 1). Twelve different views of each object were obtained by rotating the object around its y-axis in 30° increments. The initial view of each object was determined by aligning the object so that its 2-D image on the computer screen was perfectly symmetrical and then rotating the object 10° to the left. This was done to ensure that none of the rotations would produce views that were mirror reflections of one another. Subsequent views were achieved by rotating the object to the right in 30° increments. The initial view will be referred to as the 0° view, with subsequent views being 30°, 60°, 90°, 120°, 150°, 180°, 210°, 240°, 270°, 300°, and 330° rotations from the 0° view. All objects were rendered under even illumination as 32-bit gray-scale images against a white background.

Design and procedure. Each participant was tested individually and was initially exposed to all 12 views of an object, which were each printed on a separate sheet of paper and distributed on a table. While viewing all 12 views of an object, participants were asked to identify the 2 most similar and 2 least similar views of the object. This allowed participants to see the range of different views and to establish anchors or reference points for the subsequent similarity-rating task.

For the similarity-rating task, participants were instructed to press a number on the keyboard between 1 (representing a least similar pair) and 7 (representing a most similar pair) for each pair of views appearing on the computer screen in a self-paced procedure. Participants viewed 76 pairs of
views of the same object seen previously during the anchoring task. All 12 views of an object were paired with each other once, resulting in 66 different pairings. Ten of these pairings were presented a second time during the similarity-rating task so that a measure of reliability could be established for each participant. Thus, participants were required to rate the similarity of a total of 76 pairs of views of an object. After a 5-min break, each participant repeated the anchoring and similarity-rating tasks for a second object. The entire session lasted approximately 45 min.

The pairs of views were presented in a different random order for each participant, with the constraint that none of the 10 identical pairs were presented contiguously in time. The objects that each participant viewed were randomly determined, with the constraints that 4 different participants rated the view similarities of each object and that each object was seen first by 2 participants and second by 2 participants. Similarity ratings were collected from 4 different participants for each of the 21 objects so that a measure of interparticipant reliability could be obtained for each object.

Figure 1. Examples of objects used in Experiments 1–5 and the views used in Experiment 2 (left column) and in Experiments 3–5 (right column).
Results and Discussion

To determine the consistency of each individual participant’s similarity ratings for the pairs of views of each object, we performed Spearman rank-order correlations between the first and second presentation of the 10 repeated pairs of views. There were 8 participants with correlations below .70 for an object. The participants’ ratings for these objects were deleted and replaced.

Correlations between the 4 different participants’ similarity ratings for each object were also performed to assess the degree of agreement across participants. All of the pairwise correlations between participants’ ratings for each of the 21 objects were positive, and all but 6 of the correlations were reliable at \( p < .05 \).

The mean similarity ratings for each of the 66 different pairs of views were entered into separate MDS analyses for each of the 21 objects. Two-dimensional MDS solutions resulted in a close fit with the raw similarity data for each of the 21 objects. The amount of stress for each object’s solution was less than or equal to 0.22 \((M = 0.15)\), indicating a close fit between the distance between the different views in MDS solutions and the raw similarity data (stress of 0 represents a perfect fit). Similarly, the proportion of variance \((R^2)\) in the similarity ratings accounted for by the differences in the scaling solutions were all above .65 \((M = .86)\), also suggesting a close fit between the raw similarity data and the MDS solutions.

After examining the 21 MDS solutions, we observed that the perceived similarity between object views depended on at least two dimensions (see Figure 2 for an illustration of a prototypical MDS solution). One dimension seemed to reflect the degree to

Figure 2. This plot represents a typical multidimensional scaling solution obtained in Experiment 1. Views that are closer together in the plot were rated as being more similar to one another than were more distant views in the plot.
which the object's main axis of elongation was foreshortened (the main axis refers to the axis of symmetry of the object). A second dimension seemed to reflect the degree to which the object faced toward or away from the viewer. Interestingly, the MDS solutions for 19 of the 21 objects suggested that similarity was not strongly dependent on the angular disparity between views. Indeed, a negative correlation (i.e., $-0.45$) was observed between angular disparity and the similarity ratings for the 21 objects, suggesting that whereas views separated by smaller angular disparities tended to be perceived as more similar than those separated by larger angles, there was considerable variability in perceived similarity that was not accounted for by angular disparity. Whereas similarity ratings were highly correlated with angular disparity for 2 objects, it is not clear at this point what may have distinguished these 2 from the other 19 objects. To ensure consistency in our item set, we replaced these 2 objects with 2 other objects that yielded a pattern of view similarity that was consistent with the original set of 19 objects.

Experiment 2

Experiment 2 was designed to examine whether the measure of similarity derived from the similarity ratings in Experiment 1 plays a role in the mapping of different views of objects. To address this issue, we examined whether the recognition of an object in a novel view is dependent on its similarity to representations of existing views of the object in memory. Specifically, we predicted that if a participant had studied a particular view (90°, 60°, or 30° view) of an object, the degree of generalization to a novel view (270° view) would be governed by the similarity between the novel view and the studied view. We chose the 90°, 60°, and 30° views to reflect differing degrees of similarity to the novel 270° view. Furthermore, the 90°, 60°, and 30° study views were also chosen to unconfound the effects of angular disparity and similarity, such that the 90° view was most similar to the 270° test view but also reflected the greatest degree of rotation from the 270° view. Similarly, the 30° view was least similar to the 270° test view but also reflected the smallest degree of rotation from the 270° view. This was done to ensure that any effects of similarity on recognition could not be attributed to the angular disparity between views.

Method

Participants. Twenty-seven Boston College undergraduates participated to partially fulfill course requirements.

Materials. The 21 objects from Experiment 1, along with 21 additional objects constructed using the same methods as described above, were used in Experiment 2. The 21 additional objects served as nonstudied objects, whereas the 21 objects from Experiment 1 served as studied objects. The similarity ratings obtained in Experiment 1 were used to select three study views of each object that were either similar (90°), less similar (60°), or dissimilar (30°) to a test view (270°), with the constraint that these views differed in angular disparity in the manner described above.

Analyses of the mean similarity ratings obtained in Experiment 1 suggested that the perceived similarity between the 90° and 270° views ($M = 5.51$) was greater than the similarity between the 60° and 270° views ($M = 3.83$). $F(1, 20) = 84.42$, $MSE = 0.35$. A numerically greater similarity rating was observed for the 90° and 270° views for all but one of the 21 objects. Similarly, the perceived similarity between the 60° and 270° views ($M = 3.83$) was greater than the similarity between the 30° and 270° views ($M = 2.46$), $F(1, 20) = 60.20$, $MSE = 0.33$. Again, a numerically greater similarity rating was observed for the 60° and 270° views compared with the 30° and 270° views for all but one of the 21 objects. Note, however, that the angular disparity between the 90° and 270° views is 180°, whereas the disparity between the 60° and 270° views is 150°, and the disparity between the 30° and 270° views is 120°.

Design. A single-factor (study view: similar, less similar, dissimilar) within-participants design was used. Either the 90°, 60°, or 30° view of each of the 21 objects was presented to each participant during the study phase. The 270° view of each of the 21 studied objects was presented during the test phase along with the 270° view of 21 additional nonstudied objects. Thus, of the 42 objects presented at test, 7 objects were presented in similar views at study and test (i.e., the 90° and 270° views), 7 objects were presented in less similar views at study and test (i.e., the 60° and 270° views), 7 objects were presented in dissimilar views at study and test (i.e., the 30° and 270° views), and 21 objects were nonstudied. Note that because the same view was presented at test for the similar, less similar, and dissimilar conditions, any differences in performance observed for responses at test are necessarily due to differences in the previous experience with the objects at study. Three counterbalanced study lists were constructed to ensure that each of the studied objects was in the similar, less similar, and dissimilar conditions an equal number of times across participants.

Procedure. Participants were tested individually on a Macintosh Quadra computer. At study, participants were told that they would see novel 3-D objects one at a time on the computer screen for 5 s each. Participants were asked to pay careful attention to each object while it was on the screen and to then invent a possible function for each object and write a brief description of this function on a sheet of paper before pressing a key to view the next object. This “deep-level” encoding task was chosen to encourage participants to pay attention to the structure of each object and to promote reliable recognition memory performance (Schacter, Cooper, & Delaney, 1990; Schacter, Cooper, Delaney, Peterson, & Tharan, 1991). Following the study phase, participants engaged in a 5-min distractor task that involved solving anagrams.

At test, objects appeared for a maximum of 5 s, and participants were instructed to press a green key if an object had been seen during the study phase (studied) or a red key if an object had not been seen during the study phase (nonstudied) as quickly and as accurately as possible. Participants were told to ignore any changes in view between the first and second appearances of an object and to respond studied if they had seen an object previously, regardless of whether or not it was seen from a different point of view. A separate random order of presentation at both study and test was determined for each participant. The random presentation of the stimuli and the recording of response accuracy and latency measures were controlled by PsychLab (Quebec, Canada) software.

Results and Discussion

For this and each of the following experiments reported in this article, both accuracy and latency data are reported. Recognition accuracy was measured as the proportion of hits minus false alarms. Only latency data for correct responses were analyzed. Analyses were performed by both participant and item variability, but because of the consistency between these analyses, only the analyses by participant variability are reported, unless stated otherwise. Similarly, whereas analyses comparing the accuracy in each condition with chance performance revealed that performance was significantly above chance in every condition of each experiment, the analyses are not reported to reduce the redundancy and improve the readability of the text. The criterion for reporting results as significant was .05. The error bars in all of the figures represent 95% confidence intervals. For within-participant analy-
ses, the 95% confidence intervals were calculated using the interaction mean square error for the main effect of study view (see Loftus & Masson, 1994). For between-participants analyses, the 95% confidence intervals were calculated using the within-condition mean square error from the omnibus analysis of the study-view conditions. For the accuracy data, outliers were defined as participants with mean accuracies more than two standard deviations below the grand mean for the studied conditions. Similarly, for the latency data, outliers were defined as participants with mean latencies more than two standard deviations above the grand mean for the studied conditions. However, no participants were identified as outliers by these criteria. Individual reaction times greater than 3,000 ms were excluded from latency analyses.

As indicated by Figure 3, there was a significant effect of study view for the corrected recognition data, $F(2, 52) = 25.25$, $MSE = 0.02$. Planned comparisons indicated a significant advantage for the similar condition (i.e., 90° and 270° views) over the less similar condition (i.e., 60° and 270° views), $F(1, 26) = 22.37$, $MSE = 0.01$, and a significant advantage for the less similar condition (i.e., 60° and 270° views) over the dissimilar condition (i.e., 30° and 270° views), $F(1, 26) = 8.83$, $MSE = 0.02$, suggesting improved recognition memory performance when study and test views of objects were similar to one another. Analyses of latency data indicated no significant differences between the three study-view conditions, $F_s \leq 1$, suggesting an absence of any speed-accuracy trade-offs. The mean latencies (in ms) for the similar, less similar, and dissimilar conditions were 1,202, 1,269, and 1,270, respectively.

Importantly, the advantages for similar over dissimilar study and test views occurred despite the fact that greater angles of disparity separated the similar views than the dissimilar views. This suggests that the mapping of different views to the same object occurs by a computation of perceptual similarity between the views. Note that whereas the views in Experiment 2 were chosen to contrast the effects of similarity and angular disparity, similarity may often be confounded with angular disparity, as suggested by the correlation observed between these two variables in Experiment 1. Thus, it may be that the typical "disparity" effects reported in the literature may be due instead to the perceptual similarity between views (e.g., Tarr, 1995).

Given that temporally separated views of objects are mapped to the same object most easily when the views are similar to each other, how does similarity affect the view combination process? As we pointed out earlier, similarity may have two different effects on the view combination process. First, for view combination to occur, two views must first be mapped to the same object before they can be combined. If it is impossible to map dissimilar views to the same object when they are temporally separated, similar views should, when combined, produce better generalization than would dissimilar views if the similar views provide more structural information about the object than is provided by either view presented in isolation. However, if dissimilar views can be combined, because of temporal contiguity between the views, dissimilar views should produce better generalization than would similar views. This is because dissimilar views presumably have different sets of visible features of objects that, when combined, would allow for greater generalization to novel views. We examined these two hypotheses in Experiments 3–5. Experiments 3A–3B were designed to investigate whether exposure to dissimilar views of objects improves generalization to novel views when view combination is facilitated by temporal contiguity between views. Specifically, we addressed whether exposure to two dissimilar views (30° and 90° views) would produce better generalization to a third view (210° view) compared with exposure to two similar views (0° and 30° views). Further, the issue was whether exposure to any two views (either 0° and 30° or 30° and 90° views) is sufficient to produce the view combination advantage over and above the generalization obtained from exposure to any one of the views (0°, 30°, or 90° views). Experiment 3A was therefore designed to observe baseline transfer from a new set of study conditions.

![Figure 3.](image-url) The mean corrected recognition scores for the similar, less similar, and dissimilar conditions of Experiment 2. The false-alarm rate was .31. The error bars are within-participant, 95% confidence intervals calculated from the main effect of study view (see Loftus & Masson, 1994).
views (0°, 30°, or 90° views) to a novel view (210° view) at test. Experiment 3B then examined whether the combinations of 0° and 30° views or 30° and 90° views produced generalization over and above the generalization observed from these views in isolation.

**Experiments 3A and 3B**

**Method**

**Participants.** Thirty Boston College undergraduates participated in Experiment 3A, and an additional 30 students participated in Experiment 3B to partially fulfill course requirements.

**Materials.** For counterbalancing purposes, 18 of the 21 studied objects used in Experiments 1 and 2 were used in Experiments 3A and 3B. Additionally, 18 of the nonstudied objects used in Experiment 2 were used in Experiments 3A and 3B. The 0° and 30° views were rated as being more similar to each other in Experiment 1 (M = 6.44) than were the 30° and 90° views (M = 3.13), F(1, 17) = 194.08, MSE = 0.57. The 0° and 30° views were numerically more similar than were the 30° and 90° views for all 18 objects. The 210° test view was an extrapolated view that was 150° away from the nearest of the similar study views and 120° away from the nearest of the dissimilar study views. The 210° test view was also dissimilar to all of the study views. The mean ratings of similarity between the 0°, 30°, and 90° views on the one hand, and the 210° view on the other hand, were 2.94, 3.18, and 2.09, respectively.

**Design.** For Experiment 3A, a single-factor (study view: 0°, 30°, or 90° views) within-participants design was used. At study, participants were presented with either the 0°, 30°, or 90° views of each of the 18 objects. At test, all participants were exposed to the 210° view of the 18 studied objects along with the 210° view of 18 nonstudied objects. Thus, of the 36 objects presented at test, 6 were seen from the 0° view at study, 6 were seen from the 30° view at study, 6 were seen from the 90° view at study, and 18 were nonstudied objects. Three counterbalanced study lists were constructed to ensure that each of the studied objects was in each of the three conditions an equal number of times across participants.

For Experiment 3B, a single-factor (study views: similar pairs or dissimilar pairs) within-participants design was used. At study, participants were presented with two consecutive views of each of the 18 objects that were either similar or dissimilar. At test, all participants were presented with the 210° view of the 18 studied objects along with the 210° view of 18 nonstudied objects. Thus, of the 36 objects presented at test, 9 were in the similar condition, 9 were in the dissimilar condition, and 18 were nonstudied objects. Two counterbalanced study lists were constructed to ensure that each of the studied objects was in the dissimilar and similar conditions an equal number of times across participants.

**Procedure.** Pilot data for Experiment 3A collected from 8 participants using the same presentation-duration and encoding-task procedures as in Experiment 2 indicated that recognition accuracy was at ceiling. Presumably this increase in performance was due to the fact that in Experiment 3A, a single view of each object was presented twice consecutively rather than once, as in Experiment 2. Two presentations of each object view (i.e., 0° and 0°, or 0° and 30°, or 0° and 90° views) were necessary to hold the amount of exposure to objects at study constant across Experiments 3A and 3B, because in Experiment 3B, two different views of the same object were presented to examine the effects of view combination. To avoid ceiling effects, we reduced presentation durations and chose a shallower encoding task for Experiments 3A and 3B. Participants were asked to indicate whether each object that appeared on the computer screen looked more like a tool (e.g., screwdriver, computer, or vehicle) or more like a support (e.g., chair, desk, or coatrack) by pressing either a green key or a red key on the keyboard. Presentation times were decreased from 5 s each to 1 s each. So, each object was seen for a maximum of 2 s: 1 s for each view of each object. Experiment 3A was thus conducted to establish the baseline measure of transfer from the 0°, 30°, and 90° views to the 210° view. In Experiment 3B, we examined whether there was an advantage due to view combination over and above any transfer from the 0°, 30°, or 90° views alone. In Experiment 3B, two views (either 0° and 30° or 30° and 90° views) were presented consecutively. In all other respects, the procedures for Experiments 3A and 3B were exactly the same. Participants were asked to make a separate response for each view of each object. The pairs of object views were separated by a 500-ms blank screen and were presented in a random order determined separately for each participant.

**Results and Discussion: Experiment 3A**

Because the 0°, 30°, and 90° views were all relatively dissimilar to the 210° view, we did not expect to find differences in recognition performance for the three study-view conditions. However, the effect of study view approached significance in analyses by participant variability, F(2, 58) = 2.99, MSE = 0.04, p < .06, and by item variability, F(2, 34) = 2.64, MSE = 0.03, p < .09. Individual comparisons indicated that the advantage for the 0° view over the 90° view approached significance by both participant variability, F(1, 29) = 2.98, MSE = 0.05, p < .10, and by item variability, F(1, 17) = 3.05, MSE = 0.03, p < .10, suggesting that participants were better able to recognize the 210° view of test objects after seeing the 0° view at study than after seeing the 90° view at study. Similarly, a significant advantage was observed for the 30° view condition over the 90° view condition in an analysis by participant variability, F(1, 29) = 4.62, MSE = 0.04, and this effect approached significance in an analysis by item variability, F(1, 17) = 3.70, MSE = 0.03, p < .07, suggesting that the 210° test view was recognized better after seeing the 30° view at study than after seeing the 90° view at study. There was no significant difference between the 0° view and the 30° view conditions, F < 1.

The same pattern of findings was also observed in analyses of the latency data. There was a significant effect of study view, F(2, 56) = 7.30, MSE = 49,157.34, and there were significant advantages for the 0° view condition over the 90° view condition, F(1, 28) = 12.38, MSE = 34,594.45, and for the 30° view condition over the 90° view condition, F(1, 28) = 9.70, MSE = 64,887.47, but no significant difference between the 0° view condition and the 30° view condition, F < 1. The mean latencies (in ms) for correct responses at test were 1,131, 1,098, and 1,304 for the 0°, 30°, and 90° view conditions, respectively. Interestingly, analyses of the mean similarity ratings obtained in Experiment 1 indicate that the perceived similarity between the 0° and 210° views (M = 2.94) is significantly greater than the similarity between the 90° and 210° views (M = 2.09), F(1, 19) = 8.24, MSE = 0.88, and the similarity between the 30° and 210° views (M = 3.18) is also greater than the similarity between the 90° and 210° views, F(1, 19) = 12.92, MSE = 0.92. Thus, the recognition advantages for the 0° and 30° view conditions over the 90° view condition are consistent with the similarity data between these three study views and the 210° test view. However, this pattern of findings is qualified by the observation that, whereas there was no significant difference between mean accuracy or latency measures for the 0° and 30° conditions, Fs < 1, an analysis of the similarity ratings indicated that the difference between the ratings for the 0° and 210° views (M = 2.94) and the 30° and 210° views (M = 3.18) approached significance, F(1, 19) = 4.13, MSE = 0.14, p < .06. This inconsistency between the recognition and similarity data may be due to insufficient statistical power in the recognition data.
Taken together, as in Experiment 2, these data suggest that the recognition of an object seen from a novel view is predicted by the perceived similarity between the novel view and previously encountered views. The issue, then, is whether recognition performance would be enhanced by exposure to two of these views at study, over and above generalization due to similarity alone. To address this issue, we presented a new group of participants with the 0° and 30° views consecutively or with the 30° and 90° views consecutively in Experiment 3B. If view combination enhances generalization to new views of objects, the combination of the 0° and 30° views or of the 30° and 90° views should produce better generalization to the 210° view, as shown in Srinivas and Schwoebel (1998). Furthermore, if view combination is facilitated for dissimilar views (i.e., when the mapping is easier because of the temporally contiguous presentation of views), exposure to the 30° and 90° views should produce better generalization to the 210° view than would exposure to the 0° and 30° views.

Results and Discussion: Experiment 3B

There was no significant difference between the recognition performance in the similar (M = .42) and dissimilar (M = .42) conditions, F < 1. There was also no significant difference between the mean latencies for the similar (M = 993) and dissimilar (M = 952) conditions, F < 1. These findings suggest either that both conditions showed equivalent view combination effects or that both conditions failed to result in view combination effects.

The critical question, then, is whether or not the combination of the 0° and 30° or the 30° and 90° views in Experiment 3B produced recognition performance beyond generalization from the perception of either the 0°, 30°, or 90° view alone in Experiment 3A. To examine this issue, we compared the mean collapsed recognition performance for the similar and dissimilar view conditions of Experiment 3B with the recognition observed in the 30° view condition of Experiment 3A. Because there were significant differences between the 0°, 30°, and 90° view conditions of Experiment 3A, we chose the 30° view condition for comparison because the best recognition performance was observed in this condition, and thus, it provided the most stringent test of view combination effects. As indicated in Figure 4, no significant differences were observed, Fs < 1. Thus, it appears that exposure to two different views of each object at study in Experiment 3B did not produce improved recognition performance above that observed after experience with only one view at study. This finding, then, represents a failure to replicate the earlier generalization that we had obtained to extrapolated novel views of objects (Srinivas & Schwoebel, 1998).

This lack of generalization to extrapolated views was unexpected, but we note one difference between Experiments 3A and 3B and our previous experiments reported in Srinivas and Schwoebel (1998) that could account for the discrepancy. Compared with our previous experiments, Experiments 3A and 3B measured extrapolation to views with greater angular disparities from previously experienced views. Specifically, in Experiment 3B, the test view was rotated by 120° from the nearest studied view in the dissimilar condition and by 150° from the nearest studied view in the similar condition. In contrast, in our earlier experiment (Srinivas & Schwoebel, 1998), the extrapolated test view was rotated by 80° from the nearest study view. Thus, it is possible that view combination allows a gradient of generalization to extrapolated views, with greater generalization to extrapolated views that are close to experienced views than to extrapolated views that are distant from experienced views. Thus, the failure to observe view combination effects in Experiment 3B may be because the test view fell outside of the range of views facilitated by view combination. Experiments 4A and 4B were designed to explore this possibility by measuring the effects of view combination on the recognition of novel extrapolated views that were closer to experienced views. Experiments 4A and 4B were therefore another.
The similarities between study and test views were very similar across the test view were smaller in Experiment 4A compared with Experiment 3A, both study and test to examine whether or not view sensitivity would be 4B to partially fulfill course requirements.

Method

Participants. Sixteen Boston College undergraduates participated in Experiment 4A, and an additional 16 students participated in Experiment 4B to partially fulfill course requirements.

Materials. The views and objects were exactly the same as in Experiment 3A and 3B. The similar views were the 0° and 30° views of each object, as in Experiment 3B. The mean similarity rating for these views obtained in Experiment 1 was 6.44. However, the dissimilar views were the 30° and 210° views of each object. The mean similarity rating for the 30° and 210° views was 3.18 and was significantly lower than the mean similarity for the 0° and 30° views, F(1, 17) = 91.28, MSE = 1.17. The 0° and 30° views were numerically more similar than were the 30° and 210° views for all but one object. The 90° view of objects was presented during both study and test to examine whether or not view sensitivity would be observed. The 90° test view of each object was 60° away from the nearest view in both the similar and dissimilar study views. The mean ratings of similarity between the 0° and 90° views, the 30° and 90° views, and the 210° and 90° views were 2.40, 3.13, and 2.09, respectively. Note that, whereas the disparities between the nearest study view and the extrapolated test view were smaller in Experiment 4A compared with Experiment 3A, the similarities between study and test views were very similar across the two experiments. In Experiment 3A, the mean similarities between the 0° and 210° views, the 30° and 210° views, and the 90° and 210° views were 2.94, 3.18, and 2.09, respectively.

Design. For Experiment 4A, a single-factor (study view: 0°, 30°, 90° or 210° views) within-participants design was used. At study, participants were presented with either the 0°, 30°, 90°, or 210° views of each of the 20 objects. Twenty objects were required for counterbalancing purposes. At test, all participants saw the 90° view of both the 20 studied objects along with the 90° view of the 20 nonstudied objects. Thus, of the 40 objects presented at test, 5 were seen from the 0° view at study, 5 were seen from the 30° view at study, 5 were seen from the 90° view at study, 5 were seen from the 210° view at study, and twenty were nonstudied objects. Four counterbalanced study lists were constructed to ensure that each of the studied objects was in the four conditions an equal number of times across participants.

For Experiment 4B, a single-factor (study views: similar pairs or dissimilar pairs) within-participants design was used. At study, participants were presented with two consecutive views of each of the 20 objects that were either similar or dissimilar. At test, all participants were presented with the 210° view of the 20 studied objects along with the 210° view of the 20 nonstudied objects. Thus, of the 40 objects presented at test, 10 were in the similar condition, 10 were in the dissimilar condition, and 20 were nonstudied objects. Two counterbalanced study lists were constructed to ensure that each of the studied objects was in the dissimilar and similar conditions an equal number of times across participants.

Procedure. The procedures were the same as in Experiments 3A and 3B.

Results and Discussion: Experiment 4A

There was a significant effect of study view, F(3, 45) = 12.89, MSE = 0.03. Planned comparisons indicated a significant advantage for the 90° view condition (M = .61) over the mean accuracy for the combined 0°, 30°, and 210° conditions, F(1, 15) = 95.00, MSE = 0.01, and no significant difference between the 0° (M = .25), 30° (M = .35), and 210° (M = .28) view conditions, F < 1. This pattern of results suggests that recognition memory was best when exactly the same view was seen at both study and test, thus replicating previous findings of viewpoint sensitivity (e.g., Srinivas, 1995; Srinivas & Schwoebel, 1998; Tarr, 1995). Analyses of latency data revealed the same pattern of results. There was a significant effect of study view, F(3, 45) = 5.11, MSE = 26,508.44, a significant advantage for the 90° view condition over the mean of the combined 0°, 30°, and 210° view conditions, F(1, 15) = 12.63, MSE = 16,159.26, and no significant differences between the 0°, 30°, and 210° study conditions either by participant variability, F(2, 30) = 1.83, MSE = 27,643.02, p < .18, or by item variability, F < 1. The mean latencies (in ms) for correct responses at test were 909, 984, 763, and 875 for the 0°, 30°, 90°, and 210° view conditions, respectively. This pattern of data, once again, replicates previous findings suggesting sensitivity to changes in viewpoint.

Although no significant differences were detected between the 0°, 30°, and 210° study conditions, it should be noted that the numerical differences between the accuracy scores in these conditions were consistent with the ratings obtained in Experiment 1 for the perceived similarity between these three views and the 90° test view. The mean corrected recognition scores for the 0°, 30°, and 210° conditions were .25, .35, and .28, respectively. The mean similarity ratings for the 0° and 90° views, the 30° and 90° views, and the 210° and 90° views were 2.40, 3.13, and 2.09, respectively. Analyses of the mean similarity ratings indicated significantly greater similarity between the 30° and 90° views than between the 0° and 90° views, F(1, 19) = 17.20, MSE = 0.31, and the 210° and 90° views, F(1, 19) = 17.29, MSE = 0.62. There was no significant difference between the 0° and 90° views and the 210° and 90° views, F(1, 19) = 2.46, MSE = 0.40, p < .13. We now turn in Experiment 4B to whether perception of consecutively presented similar views (i.e., 0° and 30° views) or dissimilar views (i.e., 30° and 210° views) promoted generalization to a novel 90° test view beyond the generalization observed after experiences with the 0°, 30°, or 210° view in isolation.

Results and Discussion: Experiment 4B

To compare the recognition performance observed in the similar and dissimilar views conditions of Experiment 4B with the performance with isolated views in Experiment 4A, we performed two between-participants analyses on the corrected recognition data of each experiment. As indicated in Figure 5, a significant advantage was observed for the dissimilar views condition (M = .45) in Experiment 4B over the mean performance of the combined 0°, 30°, and 210° conditions (M = .29) in Experiment 4A, F(1, 30) = 5.58, MSE = 0.04. These results suggest that a recognition advantage due to view combination occurred for the dissimilar views condition of Experiment 4B. That is, experience with the 30° and 210° views at study appeared to improve recognition
memory performance for the 90° test view beyond the performance observed after experience with any one of the study views (i.e., 0°, 30°, or 210° views) in isolation. In contrast, no significant difference was observed between the performance in the similar views condition (M = .35) in Experiment 4B and the mean recognition performance for the combined 0°, 30°, and 210° view conditions (M = .29) of Experiment 4A, F < 1, suggesting that experience with the 0° and 30° views did not improve recognition for the 90° test view beyond the performance observed after experience with either the 0°, 30°, or 210° views in isolation. Thus, no view combination effects were observed for the similar view conditions in either Experiment 3B or 4B, despite the fact that the disparity between the nearest study and test views was reduced from 150° in Experiment 3B to 60° in Experiment 4B. The advantage due to view combination, therefore, appears to be dependent on experience with dissimilar views. This conclusion is strengthened by the analysis that indicated a significant advantage for the dissimilar views condition over the similar views condition of Experiment 4B, F(1, 15) = 4.80, MSE = 0.02. Analysis of the latency data indicated no significant difference between the mean latencies in the similar views (M = 1,267) and dissimilar views (M = 1,272) conditions, F < 1, suggesting an absence of any speed-accuracy trade-offs.

Together, the results of Experiments 3A–4B suggest two important aspects of view combination. First, view combination produces generalization to extrapolated views when the views are relatively close to experienced views (Experiments 4A and 4B) but not when they are distant from experienced views (Experiments 3A and 3B). Thus, the generalization produced from view combination reduces as a function of angular disparity of the extrapolated view from the nearest experienced view. So, whereas the initial mapping of views onto the same object appears to depend on view similarity rather than angular disparity, the generalization advantage due to view combination appears to depend on the angular disparity between novel views and previously experienced views. Second, view combination improves generalization to extrapolated views only when the combined views are sufficiently dissimilar. In other words, when dissimilar views are presented contiguously in time, they allow better transfer to novel views, presumably because of greater variability of features that are visible in the two views. In our last set of experiments, we addressed whether view combination effects are observed for dissimilar views when the combination process is more difficult (i.e., when views are separated by time and intervening distractor objects) or whether the advantage for dissimilar over similar views observed in Experiment 4B disappears in Experiment 5B because of an inability to map the dissimilar views onto the same object for view combination. Experiment 5A, like Experiment 4A, was designed to establish baseline measures of generalization to the 90° novel-test view after experience with either the 0°, 30°, 90°, or 210° view in isolation.

Experiments 5A and 5B

Method

Participants. Sixteen Boston College undergraduates participated in Experiment 5A and an additional 16 students participated in Experiment 5B to partially fulfill course requirements.

Materials, design, and procedure. For Experiment 5A, the stimuli, design, and procedure were exactly the same as in Experiment 4A, except that in Experiment 5A, the two study views of each object were separated by time and intervening objects. This was achieved by presenting the first view of each pair during an initial study block and the second view of each pair during a second study block. There was an approximate 1-min delay between the first and second study blocks. Thus, each pair of study views was separated by at least 1 min.

For Experiment 5B, the materials, design, and procedure were exactly the same as in Experiment 4B, except that study views (either 0° and 30°...
or 30° and 210° views) of each object were presented during two study blocks to examine the effects of view combination.

**Results and Discussion: Experiment 5A**

There was a significant effect of study view, $F(3, 45) = 11.12$, $MSE = 0.03$. Planned comparisons indicated a significant advantage for the 90° view condition ($M = 0.64$) over the mean accuracy for the combined 0°, 30°, and 210° conditions, $F(1, 15) = 60.75$, $MSE = 0.01$, and no significant difference between the 0° ($M = .28$), 30° ($M = .38$), and 210° ($M = .37$) view conditions, $F \leq 1$. As in Experiment 4A, this pattern of results suggests that recognition memory was best when exactly the same view was seen at both study and test. Analyses of latency data revealed the same pattern of results. There was a significant effect of study view, $F(3, 45) = 2.87$, $MSE = 25,783.79$, a significant advantage for the 90° view condition over the mean of the combined 0°, 30°, and 210° view conditions, $F(1, 15) = 11.22$, $MSE = 9,586.77$, and no significant differences between the 0°, 30°, and 210° study conditions by participant variability, $F < 1$, or by item variability, $F(2, 32) = 1.72$, $MSE = 68,533.43$, $p < .20$. The mean latencies (in ms) for correct responses at test were 983, 897, 820, and 928 for the 0°, 30°, 90°, and 210° view conditions, respectively. This pattern of data, once again, replicates previous findings suggesting sensitivity to changes in viewpoint.

Although no significant differences were detected between the 0°, 30°, and 210° study conditions, it should be noted that the numerical differences between the accuracy scores in these conditions were generally consistent with the ratings obtained in Experiment 1 for the perceived similarity between these three views and the 90° test view. The mean corrected recognition scores for the 0°, 30°, and 210° conditions were .28, .38, and .37, respectively. The mean similarity ratings for the 0° and 90° views, the 30° and 90° views, and the 210° and 90° views were 2.40, 3.13, and 2.09, respectively. In Experiment 5B, we examined whether perception of temporally separated similar views (i.e., 0° and 30° views) or dissimilar views (i.e., 30° and 210° views) promoted generalization to a novel 90° test view beyond the generalization observed after experiences with the 0°, 30°, or 210° view in isolation.

**Results and Discussion: Experiment 5B**

Unlike Experiment 4B, there was no significant advantage for the dissimilar views condition ($M = .43$) over the similar views condition ($M = .42$), $F < 1$, suggesting that experience with either the similar or dissimilar views produced equivalent recognition memory performance for test objects seen from a novel view. Analysis of the latency data also indicated no significant difference between the mean latencies in the similar ($M = 1.052$) and dissimilar ($M = 1.082$) conditions, $F < 1$.

To examine whether exposure to the similar or dissimilar views at study produced improved recognition memory relative to the memory performance achieved after exposure to any one of the study views in isolation (i.e., either the 0°, 30°, or 210° view), we compared the mean corrected recognition scores from the combined similar and dissimilar conditions of Experiment 5B with the mean of the combined 0°, 30°, and 210° view conditions of Experiment 5A. As indicated in Figure 6, there was no significant advantage for the combined similar and dissimilar conditions ($M = .43$) over the combined 0°, 30°, and 210° view conditions ($M = .34$), by participant variability, $F(1, 30) = 1.56$, $MSE = 0.04$, $p < .22$, or by item variability, $F = 1$. These findings replicate the failure to observe view combination effects after experience with similar views in Experiments 3B and 4B, and they suggest that the visual information provided by similar views may be too redundant to derive recognition advantages due to view combination. Additionally, these findings suggest that view combination may not occur when dissimilar views are separated by...
time and intervening objects. In other words, it may be that dissimilar views cannot be mapped onto the same object when the views are separated by time and intervening objects, and thus, no view combination is possible. Note that the failure to observe view combination effects in Experiment 5B is inconsistent with our previous observation of view combination effects after experience with temporally separated views (Srinivas & Schwoebel, 1998). However, Srinivas and Schwoebel (1998) did not compare the recognition performance observed after experience with temporally separated views with the performance obtained in an appropriate control condition (e.g., Experiment 5A of this article) in which pairs of single-study views were separated by time and intervening objects. Thus, it is not clear if view combination resulted in performance above and beyond that achieved after experience with isolated views. Regardless, whereas the data above suggest that experience with temporally separated dissimilar views does not result in view combination effects, it may be that views of intermediate similarity can be mapped onto the same object to support view combination. For the purposes of this article, however, the critical point is that there do appear to be temporal constraints on the view combination process.

General Discussion

The present set of experiments yielded four main findings. First, the similarity between depth-rotated views of novel objects does not appear to be simply a function of angular disparity. Second, the recognition of objects seen from a novel view is predicted by the similarity between the novel view and previously experienced views. Third, generalization advantages due to the combination of views are affected by the similarity between the combined views when these views are presented contiguously in time. Fourth, generalization advantages due to the combination of previously experienced views are not observed when these views are separated by time and intervening distractor objects. Each of these findings is discussed in turn.

View Similarity

Whereas several researchers have made assumptions about the similarity between different views of objects (e.g., Hayward, 1998; Lawson & Humphreys, 1996; Lawson, Humphreys, & Watson, 1994; Srinivas & Schwoebel, 1998), Experiment 1 of the present set of experiments represents the first attempt to obtain an empirically derived measure of the perceived similarity between different views of depth-rotated objects. The findings of Experiment 1 suggest that view similarity is not simply a function of the angular disparity between different views of an object. Rather, for this object set, the findings suggest that the perceived similarity between views of an object may depend on how the axis of elongation is oriented with respect to the viewer (i.e., the degree of foreshortening) and the degree to which the object faces toward or away from the viewer.

Similarity and Object Recognition

The pattern of results observed in Experiments 2, 3A, 4A, and 5A suggests that the recognition of an object seen from a novel view may be dependent on the similarity between the novel view and a previously encountered view of the object. Thus, the current findings provide support for the prediction that more similar views will be more easily recognized as views of the same object than will less similar views. The current findings support previous assumptions of a relationship between perceived view similarity and object recognition performance (Hayward, 1998; Lawson & Humphreys, 1996; Lawson et al., 1994; Srinivas & Schwoebel, 1998).

Importantly, the observed recognition advantages for test views that were similar to study views were not limited to test views that were mirror images of study views. Previous studies have demonstrated advantages for views that were mirror images of the outline shape of studied views relative to views rotated away from the studied views by the same amount that were not mirror images of studied views (Biederman & Gerhardtstein, 1993; Hayward, 1998). Hayward (1998) also observed advantages on sequential matching and naming tasks for mirror images of studied views that were created by 180° rotations from studied views relative to views that were rotated 60° from studied views but that were not mirror images of studied views. Consistent with these findings, Vetter, Poggio, and Bülthoff (1994) have shown that, for bilaterally symmetrical objects, it may be possible to represent "virtual views" that correspond to mirror images of the outline shape of a studied view. The present results extend these previous findings and suggest that deviations from a linear relationship between recognition performance and angular disparity may not be limited to cases in which study and test views are mirror images. Instead, the present results suggest that recognition performance is better predicted by the perceived similarity of even non-mirror-image views than by the angular disparity between views. However, there is an interesting alternative account of the present results that is consistent with the findings of Vetter et al. and with reported disparity effects (e.g., Tarr, 1995). Specifically, the pattern of results observed in Experiments 2, 3A, 4A, and 5A is also predicted by the angular disparity between the mirror image of a studied view (i.e., a virtual view) and the test view. On the basis of the correlation observed between angular disparity and perceived similarity in Experiment 1, we suggest that this alternative account is complicated by the possibility that the angular disparity between a virtual view and a test view may also be confounded with the perceptual similarity between the two views. Thus, this account requires further examination.

Constraints on View Combination

Previous findings have suggested that experience with more than one view of an object at study results in later recognition memory performance that is better than the memory performance observed after experience with the same study views presented in isolation (Bülthoff & Edelman, 1992; Srinivas & Schwoebel, 1998). These observations suggest that representations of previously seen views, rather than being static and isolated from one another, are combined or integrated with one another in ways that facilitate later recognition of objects when they are seen from novel views. The current findings suggest three constraints on the view combination process.

First, view combination does not appear to result in a 3-D viewpoint invariant representation. The findings of Experiments 3A-4B suggest that view combination may produce a gradient of
generalization to novel extrapolated views that decreases as a function of increasing angular disparity between the novel view and the nearest studied view. Further evidence that view combination does not result in viewpoint invariant representations was obtained in Experiment 4B. Experience with the dissimilar study views in Experiment 4B resulted in better recognition of a novel view at test \( (M = 0.45) \) than did experience with either the 0°, 30°, or 210° views in isolation at study \( (M = 0.29) \). However, the recognition performance in the dissimilar condition was nevertheless significantly lower than that observed when exactly the same views (90° views) of objects were seen at study and test \( (M = 0.61) \), \( F(1, 30) = 5.77, MSE = 0.04 \). Thus, the present findings are consistent with the proposal that features available in the 2-D representations of previously experienced views may, through linear combination, produce greater generalization to a limited range of novel views. We speculate that the generalization gradient may be limited to views within a given range of angular disparity from the studied views because, in general, views rotated further from studied views are more likely to consist of features that are qualitatively different from studied views in part because of self-occlusions, as well as newly visible features. We thus predict that the extent of the generalization gradient depends at least partly on the degree to which features distinguishing objects are visible across rotations in depth (Liter, 1998).

Second, the findings of Experiment 4B suggest that the view combination effect may depend on the similarity between previously experienced views. It appears that study views, when presented contiguously in time, must be sufficiently dissimilar from one another for view combination effects to occur. Thus, it may be that the information gained from experience with a more dissimilar set of views allows for more complete specification of the features of an object and, therefore, improves recognition memory performance for subsequently presented novel views.

Third, the findings of Experiment 5B suggest that view combination does not occur when study views are separated by time and intervening objects. Thus, it appears that there are temporal constraints on the view combination process. Presumably, either the difficulty in mapping temporally separated dissimilar views to the same object or the temporal constraints on the view combination process itself result in the failure to observe view combination effects after experience with dissimilar views separated by time and intervening objects. It is important to note that the observation of a view combination effect for the dissimilar views condition of Experiment 4B, but not for Experiment 5B, suggests that the improved recognition memory performance observed in Experiment 4B was not due to the encoding of isolated dissimilar views but rather to the combination of these views, as proposed by Bulthoff and Edelman (1992).

In conclusion, findings from the present set of experiments suggest that it is easier to recognize objects seen from novel views when these views are similar to previously experienced views. However, although dissimilar views are more difficult to map onto the same object, if this mapping occurs, then experience with dissimilar views presented contiguously in time results in greater generalization to a limited range of novel views through view combination. Finally, the recognition advantages due to view combination appear to depend on the views appearing in close temporal succession.

References


Srinivas, K. (1995). Representation of rotated objects in explicit and


Received April 29, 1998
Revision received January 27, 2000
Accepted February 2, 2000