Constraints on View Combination: Effects of Self-Occlusion and Differences Among Familiar and Novel Views

Alan C.-N. Wong
Vanderbilt University

William G. Hayward
The Chinese University of Hong Kong

The use of multiple familiar views of objects to facilitate recognition of novel views has been addressed in a number of behavioral studies, but the results have not been conclusive. The present study was a comprehensive examination of view combination for different types of novel views (internal or external to the studied views) and different objects (amoeboid objects and objects composed of geons; objects with and without self-occlusion across rotation). The authors found that the advantage gained from the study of 2 views was more than the generalization from each of the studied views presented alone. This facilitation occurred only for internal views but not external views. In addition, the benefits from the study of 2 views diminished when (a) the studied views did not share the same visible features and when (b) the studied views were separated by a small angular difference.

One main problem in understanding human object recognition is how the human visual system can recognize an object despite constant changes in its appearance. One aspect of this problem is view generalization: How can one recognize an object across different viewpoints? Although Marr (1982) asserted that all views of an object give rise to a single object model, current theories propose multiple representations for each object (e.g., Biederman & Gerhardtstein, 1993, 1995; Hummel & Stankiewicz, 1998; Perrett, Oram, & Ashbridge, 1998; Poggio & Edelman, 1990; Tarr & Pinker, 1989). Recent behavioral findings suggest that one view may be insufficient for efficient recognition of the same object across a range of views. It has been found, across a wide range of objects and tasks, that human recognition performance deteriorates systematically as the view of an object to be recognized deviates from its familiar view (Hayward & Tarr, 1997; Lawson, 1999; Lawson, Humphreys, & Watson, 1994; Liu, 1996; Newell & Findlay, 1997; Tarr & Bulthoff, 1995; Tarr & Pinker, 1989; Tarr, Williams, Hayward, & Gauthier, 1998; but see Biederman & Gerhardtstein, 1995). Neurophysiological studies further suggest that humans may encode multiple view-specific representations for an object. Within the primate inferotemporal cortex, cells or neuronal populations have been identified to be most responsive toward particular views of an object or a complex pattern (Logothetis & Pauls, 1995; Tanaka, 1996; Wang, Tanaka, & Tanifuji, 1996).

Although humans need to experience multiple views of an object in order to recognize it efficiently at novel views, most investigations of depth-rotated object recognition have had participants study only one object view. These studies do not address how different views of an object are integrated together and used to recognize an object. However, a small number of studies have been conducted to investigate how multiple views are integrated into the process of object recognition. In a classic study, Bulthoff and Edelman (1992; Edelman & Bulthoff, 1992) showed that studying two views of a novel object leads to better subsequent recognition from some novel views than from others. In each trial, they had participants study two views (0°, 75°) of an amoeba-like or wire-like object. Then the participants saw a series of test images and had to judge whether the images depicted the same object as the one studied. When the test image contained the studied object, it was sometimes rotated around the vertical axis to varying degrees. The most interesting result for the present context is that recognition of the studied object was more accurate at internal views (i.e., views within the inner 75° range spanned by the two studied views) than external views (i.e., views in the outside 285° range). This advantage of internal views over external views was replicated by Kourtzi and Shiffrar (1999), who showed that studying two views of a novel object produced more robust priming to internal views than to external views. They further found that priming reduced when the two studied views shared different visible parts or when the angular difference between them was large. Srinivas and Schwoebel (1998; Schwoebel & Srinivas, 2000) compared the difference in novel view recognition between the study of one view and two views. They found better generalization to external views after the study of two views than after one, but this difference only occurred for bilaterally symmetric objects, not for objects that were asymmetric. The above studies...
provide support for the idea of view combination for recognizing novel views. However, as discussed later, these findings are not sufficient to provide a comprehensive and conclusive account for a number of issues, such as the internal views’ advantage over external views, the effect of occlusion across rotation, and the effect of similarity among studied views.

A small number of theoretical models have postulated that view-invariant recognition is the result of cooperation among a number of view-specific representations. Poggio and Edelman (1990) and Bulthoff and Edelman (1992) described a view interpolation model that uses information specific to training views of an object to generalize recognition to novel views. In a special case of such a model, hidden units represent the training views presented to the network. A novel view introduced to the network activates different view-tuned hidden units to different extents, depending on the similarity between the novel view and an individual unit’s preferred view. The weighted sum of the activities of different units then determines how well the novel view is recognized (in terms of speed and accuracy). This model predicts that the response of an object recognition system to a view is determined by the similarity between the current view and the views experienced before. It also predicts that, given two views represented, a novel view falling between the two studied views is likely to be recognized better than a novel view falling outside, because an internal view activates the two view-tuned units to a greater extent, leading to higher summed activity and thus better performance. Perrett et al. (1998) also proposed a neurophysiological theory describing similar mechanisms of achieving view invariance.

These theories and the behavioral studies noted above suggest that the use of multiple views should facilitate generalization from familiar to novel views, at least in some circumstances. The present study was aimed at systematically studying three questions related to the conditions in which multiple views are useful: (a) Does the study of two views benefit subsequent view generalization to internal views only or to both internal and external views? (b) Does self-occlusion affect the ability of the visual system to use multiple views? and (c) Is the advantage of internal views over external views simply a result of the proximity of studied views to the test view?

Is Extrapolation Possible?

View combination theories (Perrett et al., 1998; Poggio & Edelman, 1990) suggest that multiple view-sensitive representations result in better generalization to internal than external views, because internal views are closer to both studied views and thus cause higher activation of the familiar view representations. Although external views are mostly close only to one of the studied views and are far from the other, the theories do not preclude the possibility that external views can activate both the closer studied view and, to a lesser extent, the further studied view. If this is true, then external views can also benefit from the study of two views, albeit not as much as internal views.

From past studies, it is unclear whether only internal views, or both internal and external views, benefit from the study of two views. Both Bulthoff and Edelman (1992) and Kourtzi and Shiffrar (1999) found more generalization to internal than external views. However, in their studies, participants always studied two views of an object. As they did not have a comparison condition in which only one view was studied, it is unknown whether the amount of generalization obtained with the study of two views was larger than the generalization from each of the studied views presented in isolation. This leads to two possible interpretations of their results. One is that only internal views benefit from the study of two views. The other is that both interpolation and extrapolation occur, but the facilitation is larger for the internal views. Srinivas and Schwoebel (1998) actually found extrapolation indicated by better generalization with two studied views rather one, but this occurred only for bilaterally symmetric objects, which have external views resembling the mirror images of the internal views. Therefore, it is still unknown whether the study of multiple views improves generalization to external views.

The straightforward way to test this possibility of extrapolation is to test whether, for objects without symmetry, there is any advantage of generalization from two views over generalization from each one of the two views presented in isolation. Therefore, in both of our experiments, we introduced a baseline condition in which we measured the amount of generalization from the studied view to the novel views when only one view had been studied. We then examined whether the study of two views improved the generalization. The use of objects without symmetry helped to ensure that the external views did not form any mirror-image relationship with any of the internal views.

Effect of Self-Occlusion

When rotation of an object involves self-occlusion, different views tend to share different visible parts. Theoretically, this could help or hinder generalization processes. On the one hand, studying more than one view could lead to better generalization to novel views because each studied view might only share some of the visible features with the novel views. An internal view, which is more likely to have its features shared with the studied views, tends to benefit more than an external view, which might share many visible features with only one of the studied views and not with the other. This could account for Bulthoff and Edelman’s (1992) finding of an advantage for internal over external views, as occlusions of vertices and protrusions were likely to occur in the wire-like and amoeboïd objects they used (see, e.g., in their Figure 1).

Alternatively, self-occlusion may instead impede the process of combining multiple views. In behavioral studies, a greater recognition cost is generally found when a novel object view has different visible features from the familiar view of the same object (Biederman & Gerhardstein, 1993; Hayward & Tarr, 1997). With self-occlusion, it is expected to be harder to combine views with qualitative differences in visible features. Kourtzi and Shiffrar’s (1999) findings suggested that this may be the case. They found more priming at the novel views when the two studied views shared the same visible parts than when a part of one studied view was occluded at the other studied view. As noted before, however, it is not clear whether this result represents true facilitation above and beyond that expected from the study of any single view. It could be that self-occlusion reduces priming from the nearest view rather than actually hinders the combination of the studied views.
In our experiments, we studied the effect of self-occlusion by generating two kinds of objects (see examples in Figure 1). In the nonoccluded sets, all of the presented views share the same visible parts. In the occluded sets, only at one view (View 3) are all six protrusions of an object visible. At the other four views, only five protrusions are visible. The consequence is that one needs to study two object views (e.g., Views 2 and 4) in order to see all of the protrusions shown in View 3. If self-occlusion is necessary for view combination to occur, then we should find an advantage of studying two views over one only for the occluded set. If, however, self-occlusion impedes view combination, then we should observe a larger advantage for the nonoccluded set than the occluded set.

Similarity Between Studied Views

According to view combination theories, when two views are studied, an external novel view has a great disadvantage compared with an internal view, because in most cases an internal view is close to both studied views, whereas an external view is close to only one of the studied views. Suppose we have two situations: (a) the two studied views are moderately far apart, and the novel view falls in between the studied views; and (b) the two studied views are closer to each other, and the novel view falls outside. The novel view in situation (a) is an internal view, whereas the novel view in situation (b) is external, but they are equalized in terms of their angular difference from the two studied views in their respective situation. Would the same generalization occur in both cases? One objective of our Experiment 2 was to compare generalization with internal and external views when their distance from the studied views was controlled. If the similarity between the novel and the studied views is the only explanation for internal views’ advantage over external views, then we should see both interpolation and extrapolation.

A sequential matching task was used in the two experiments (see Figure 2). Each trial consisted of a short study phase and a test stimulus. For the one-view study condition, the same study view was presented two times in the study phase. For the two-views study condition, different views of the same object were presented in the study phase. Participants then had to judge whether the test...
image depicted the same object as the one shown initially. In the one-view study condition, we expected to see better performance at the studied view than at the novel views. A reduction of this viewpoint cost in the two-views study condition would indicate facilitation from studying two views over the study of each view alone.

Consistent with the previous studies, the sets of possible objects were relatively homogeneous. Two types of objects were used. One type was based on the amoeboid objects of Bülthoff and Edelman (1992) and Logothetis and Pauls (1995). Each object was composed of a central sphere with six protrusions pointing in different directions. The objects have the same central sphere and protrusions similar in shape and different only in terms of length and thickness, and thus the main difference among the objects was in their metric properties and spatial arrangement of protrusions. Another type had protrusions with distinct shapes (called geons; see Biederman, 1987) similar to those objects used in the view combination studies of Kourtzi and Shiffrar (1999) and Schwobbel and Srinivas (2000). The adoption of two kinds of objects allows greater generalizability of results. To avoid bilateral symmetry (which complicates the comparison between interpolation and extrapolation), we used only asymmetric objects. Some objects had features occluded across different views, whereas others had all protrusions visible at all presented views.

Experiment 1

Experiment 1 was designed to test two questions. First, we tested whether the view generalization advantage obtained from the study of two views was greater than the generalization from each studied view seen alone. Second, the effect of self-occlusion on view combination was studied by comparing two sets of objects, one with the same protrusions preserved across rotation, and the other with different visible protrusions across rotation.

Method

Participants. Fifty-three students at The Chinese University of Hong Kong participated in the experiment either for fulfillment of a course requirement or for monetary reward. All reported normal or corrected-to-normal eyesight. Thirty participants saw only the amoeboid objects, and the other 23 saw only the geon objects.

Apparatus and stimuli. A G3 (17-in. cathode-ray tube [CRT]) and two iMac computers (15-in. CRT), each with monitors set to an 800 × 600-pixel resolution, were used. Presentation of stimuli was controlled by RSVP software (Williams & Tarr, 1999). The objects used in the experiment were created and rendered by Carrara software (Version 1.0; MetaCreations).

Figure 1A shows examples of the objects used. Two types of objects were created. Each amoeboid object consisted of a central sphere with six irregular protrusions of different lengths and pointing directions. The central sphere was the same in all objects, and different protrusions were similar in general, so the main difference among objects was in the configuration of protrusions. Each geon object had a central volume with six protrusions of different regular, volumetric shapes and pointing directions. Unlike the amoeboid objects, the geon objects could be discriminated from each other by the shapes and locations of different protrusions. The central part was the same in all objects and was informative of the object view. For each type of object, occluded and nonoccluded sets were created (Figures 1B and 1C). Each object in the nonoccluded set had all protrusions visible across the views shown. Each object in the occluded set had some protrusions visible at some presented views only. As shown in the lower panels of both Figures 1B and 1C, the protrusions marked by squares and circles were clearly visible at some of the views but heavily or totally occluded at other views. Sequential matching involving these homogeneous objects amounted to a within-category, or subordinate-level, recognition.

Four sets of 20 objects formed a total of 80 objects used in the experiment (occluded and nonoccluded sets for both geon and amoeboid objects). The objects were rendered with realistic lighting and shading. The amoeboid objects were given a yellow plastic texture on a black background, whereas the geon objects were given a red plastic texture on a white background. They were about 9 cm (amoeboid objects) and 4 cm (geon objects) large, spanning a visual angle of about 6.9° and 3.1°, respectively, for a participant sitting 75 cm away from the monitor (however, viewing position was not fixed). Static images were created for five views of each object. The five views were created by rotating the rendering camera around the vertical axis passing through the center of the central volume of each object. An angular difference of 20° (amoeboid objects) and 28° (geon objects) existed between adjacent views.1

Slightly different study stimuli were created for the different object sets. For the geon objects, single static views were rendered and presented as study stimuli. For the amoeboid objects, however, objects were presented as an animation rotating in 5° oscillations around a central view. Additional kinetic-depth information was given about the amoeboid objects because of concerns that their lack of distinctive features would produce very poor recognition performance. For simplicity, throughout the rest of the discussion of Experiment 1 we shall refer to study views; the reader should note that for the amoeboid objects these were actually study animations.

Design. There were four sets of objects (amoeboid nonoccluded, amoeboid occluded, geon nonoccluded, geon occluded). Within each set there were two within-participant variables (Study Condition × Test View). The study condition variable consisted of the one-view and two-views conditions. Within the two-views condition, each studied object was presented twice, once at the novel views, and once at the study views. Because this was the condition in which the viewpoint cost was required for looking at the effect of view combination, we increased the angular difference when we ran the experiment on the geon objects.

1 We conducted the experiment first with the amoeboid objects and found that the viewpoint cost was not big enough to start with in one condition with one studied view. Because a sufficient viewpoint cost was required for looking at the effect of view combination, we increased the angular differences when we ran the experiment on the geon objects.
views conditions. The test view variable consisted of three levels: studied, internal, and external.

Within an object set, the trials for different conditions are arranged as follows. For each participant, 10 objects were presented in the one-view study condition and the other 10 in the two-views condition.² In the two-views study condition, both Views 2 and 4 were studied. A test view of either View 2 or View 4 was regarded as a studied test view (Views 2 and 4 of the 10 objects formed 20 trials). A test view of View 3 was regarded as an internal view (View 3 of the 10 objects formed 10 trials), whereas a test view of either View 1 or View 5 was regarded as an external view (Views 1 and 5 of the 10 objects formed 20 trials). In the one-view study condition, participants studied View 2 of five objects and View 4 of the remaining five. When participants studied View 2, and the subsequent test view was also View 2, this view was regarded as a studied view. When the test view was View 3, it was regarded as an internal view. When the test view was View 1, it was regarded as an external view. Note that the naming of internal and external views was arbitrary and served to compare with the two-views condition. When only View 4 was studied, the subsequent Test View 4 was defined as the studied view, whereas the Test View 3 was the internal view and the Test View 5 was the external view. So there were 10 trials with a studied test view, 10 trials with an internal view, and 10 trials with an external view.

Procedure. Each participant saw either 40 amoeboid objects or 40 geon objects. The five views for each of the 40 objects (20 in the occluded set and 20 in the nonoccluded set) formed a total of 200 same trials. An equal number of different trials were created by using each same trial and replacing the object in the test phase with the corresponding view of another randomly chosen object in the same set. Each trial consisted of a study and a test phase. For the amoeboid objects, a fixation appeared for 1,000 ms, followed by presentation of two same or different views of an object (the two views lasted for 2,000 ms and were separated by a 500-ms blank). A mask was then presented for 500 ms, followed by the test phase with a static view of either the same or different object. For the geon objects, a fixation first appeared for 1,000 ms. The study view was then shown for 300 ms, followed by a 750-ms blank and the second view for another 300 ms.³ A mask was then shown for 510 ms, followed by the test phase with a static view of either the same or different object. Participants had to press the appropriate key (1 for same and 2 for different) as accurately and as fast as possible. Once the participant responded, the test view disappeared and the next test trial began. The trials were randomly presented and separated in five blocks of 80 trials each. Four to six practice trials were provided before the experimental trials.

Results

The data from 3 participants were discarded because 2 of them had below-chance-level performance (percentage of error trials = 45% and 41%), whereas the other 1 was below chance level in one of the blocks (percentage of error trials = 44%). The remaining participants were on average correct in 83.94% of trials (SD = 5.37%), and their mean response time was 1,139 ms (SD = 369 ms). Percentages of error trials were calculated only on the same trials in which participants responded within the time range of 250 ms and 4,000 ms. Out of these trials, only those with correct responses were included in response time analyses. This resulted in 14.40% of the same trials being discarded.

The response times and error rates are shown in Figure 3 and Appendix A, respectively. The two measures showed consistent results in general, although response time was more sensitive to differences among conditions.

For each object set (amoeboid nonoccluded, amoeboid occluded, geon nonoccluded, geon occluded), we performed two analyses of variance (ANOVAs) to test the presence of interpola-

2 The objects chosen for each condition were counterbalanced across participants.

3 The presentation time for amoeboid objects was longer because they were thought to be more difficult to encode.

4 We expected the same viewpoint costs for internal and external views when there was only one study view. At least for the response time data, however, the viewpoint cost for the external view was greater than that for the internal view. Such discrepancy may be a result of the specific geometry of the nonoccluded amoeboid objects used. Nevertheless, this initial advantage of the internal views should not be the reason for the occurrence of interpolation but not extrapolation, because the same results were obtained (a) for other object sets when there was no advantage of the internal view over external view to begin with and (b) in Experiment 2 when the same view was examined for the internal and extrapolation conditions.
Figure 3. Average response times in Experiment 1. An asterisk (*) represents a significant Study Condition × Test View interaction either in response time or in percentage of error trials. Two asterisks (**) represent interaction in both measures. An interaction indicates a facilitation of studying two views over one. Error bars represent 95% confidence intervals based on within-participant error.
Geon nonoccluded set. Interpolation was shown by the significant Study Condition × Test View interaction in response time, \( F(1, 19) = 11.31, p < .01 \) (see Figure 3C). Responses at the studied views were faster than at the internal views only when one view had been studied but not with the prior study of two views. The interaction showed the same pattern but did not reach significance in error, \( F(1, 19) = 2.99, p = .10 \).

For extrapolation, no interaction was found. There was a main effect of test view in both response time, \( F(1, 19) = 35.27, p < .01 \), and error, \( F(1, 19) = 23.90, p < .01 \). Responses were faster and more accurate when the test views were studied. The main effect of study condition was significant only in response time, \( F(1, 19) = 5.48, p < .04 \), but not in error (\( F < 1 \)).

Geon occluded set. Facilitation was found for internal but not external views (see Figure 3D). For interpolation, there was a significant Study Condition × Test View interaction in response time, \( F(1, 19) = 6.19, p < .03 \). The viewpoint cost between the studied view(s) and View 3 was smaller when two views had been studied than when only one was studied. This interaction was marginally significant in error, \( F(1, 19) = 4.22, p = .05 \).

For extrapolation, there was a main effect of test view in both response time, \( F(1, 19) = 15.54, p < .01 \), and error, \( F(1, 19) = 33.79, p < .01 \). Responses were faster and more accurate for studied views than external views. The effect of study condition was significant in response time, \( F(1, 19) = 5.45, p < .04 \), but not in error (\( F < 1 \)). No interaction was found.

Discussion

There were two main findings in Experiment 1. First, the study of two views facilitated view generalization to internal views. The viewpoint cost for recognizing novel views after study of one view disappeared or was diminished when (a) two views were studied and (b) the novel view lay between the two studied views. This improvement in generalization did not occur for novel views outside the range spanned by the studied views. In addition, use of external views was statistically indistinguishable from recognition when only one view was studied.

Second, this interpolation occurred for both occluded and non-occluded object sets. The occluded object set in our experiment was similar to the objects used in Bülthoff and Edelman (1992). One may suggest that the internal view advantage over external views in these occluded objects was the result of internal views sharing more visible parts with studied views. However, we also found such an internal view advantage in objects with no occlusion. For these objects, both internal and external views shared the same visible parts with the studied views. Therefore, we can conclude that self-occlusion is not a necessary condition for view combination and interpolation to occur.

Experiment 2

In Experiment 2 and other previous studies, generalization was usually found to favor views between the studied views rather than views outside. One possible explanation for this generalization advantage is that an internal view was always similar to both studied views, whereas an external view was only similar to one of them. Experiment 2 was designed to test whether this alone can explain the absence of extrapolation.

In Experiment 2, the angular difference between test and studied views was controlled. There were three study conditions. In the View 2–View 2 study condition, only one object view (e.g., View 2) was studied. In the View 1–View 2 study condition, the studied views were just one frame from each other (e.g., Views 1 and 2), whereas in the View 2–View 5 study condition, the two studied views were farther apart from each other (e.g., Views 2 and 5). Of particular interest was the recognition performance for the novel View 3 in all of the three study conditions. In the View 1–View 2 condition, View 3 was an external view one frame away from the near studied view (e.g., View 2) and two frames away from the far studied view (e.g., View 1). In the View 2–View 5 condition, View 3 was an internal view one frame away from the near studied view (e.g., View 2) and two frames away from the far studied view (e.g., View 5). In other words, the angular difference between the novel View 3 and the two studied views was equalized in the two conditions. If the similarity between the novel view and the studied views is the sole factor for the advantage of an internal view over an external view, then generalization to the novel View 3 should be comparable in both study conditions.

Experiment 2 also provided a second test for the effect of self-occlusion on view combination. An open question is whether there is any difference in interpolation found for nonoccluded and occluded object sets with a larger angular difference between the studied views. In Experiment 1, the two studied views were always two steps different from each other in the two-views condition. The similar studied views may have provided a highly favorable condition for interpolation such that self-occlusion might not matter. In Experiment 2, the difference between the studied views was three steps in the View 2–View 5 conditions. We expected this larger difference between the studied views would render any effect of self-occlusion more obvious.

Method

Participants. Seventy-three university students participated in the experiment for either partial fulfillment of a course requirement or monetary reward. All reported normal or corrected-to-normal eyesight. Twenty-four participants were presented with the nonoccluded amoeboid object set only, whereas 22 were presented with the occluded amoeboid set only. One participant was presented with both nonoccluded and occluded sets of amoeboid objects. The remaining 26 participants saw both the nonoccluded and occluded sets of geon objects.

Apparatus and stimuli. Two iMac computers (15-in. CRT), each with monitors set to an 800 × 600-pixel resolution, were used. The stimuli were the same as in Experiment 1, except that there were 30 objects instead of 20 in each of the four object sets (amoeboid nonoccluded, amoeboid occluded, geon nonoccluded, geon occluded). Static images were used for both amoeboid and geon objects in both study and test phases.

Design. There were four object sets (amoeboid nonoccluded, amoeboid occluded, geon nonoccluded, geon occluded), and within each set there were two within-participant variables (Study Condition × Test View). The study condition consisted of three levels: View 2–View 2, View 2–View 5, and View 1–View 2. The test view consisted of two levels: View 2 (studied) and View 3 (unstudied).

Within each object set 10 objects were in the View 2–View 2 condition, with another 10 in the View 2–View 5 condition and the remaining 10 in the View 1–View 2 condition. In the View 2–View 2 condition, View 2 was presented twice in the study phase. In the View 2–View 5 condition, Views 2 and 5 were studied. In the View 1–View 2 condition, Views 1 and 2 were studied. In the test phase, the Test View 2 was regarded as a studied view in all three conditions. The Test View 3 was a novel view in the View
2–View 2 condition, an internal view in the View 2–View 5 condition, and an external view in the View 1–View 2 condition. Overall there were 10 trials for each of the six trial combinations (Study Condition × Test View).

Procedure. In each trial, during the study phase, a fixation cross appeared for 1,000 ms. The first study view was then shown for 300 ms, followed by a 750-ms blank and the second study view for another 300 ms. A blank was presented for 500 ms, followed by a 510-ms mask and then the test view of either the same object as that in the study phase or a different object. Participants were asked to press the appropriate key (for same and 2 for different) accurately and as fast as possible. Once the participant responded, the test view disappeared and the second test trial began. The five test views for each of the 60 objects formed a total of 300 same trials.

An equal number of different trials were created by using each same trial and replacing the object in the test phase with the corresponding view of another randomly chosen object in the same set. Only the same trials were analyzed. In total there were 600 trials, which were randomly presented and separated in six blocks of 100 trials each. Six practice trials were provided before the experimental trials.

**Results**

The data from 2 participants were discarded because of near-chance-level performance (percentage of error trials = 50% and 40%). The remaining participants were on average correct in 80.02% of trials (SD = 6.70%), and the average response time was 1,187 ms (SD = 308 ms). Error rates were calculated only on the same trials in which participants responded within the time range of 250 ms to 4,000 ms. Out of these trials only those with correct responses were included in response time analyses. This resulted in 17.63% of the same trials being discarded from response time analyses.

The response times and error rates are shown in Figure 4 and Appendix B, respectively. As in Experiment 1, the two measures showed consistent results in general, although again response time was more sensitive to differences among conditions.

For each object set (amoeboid nonoccluded, amoeboid occluded, geon nonoccluded, geon occluded), we performed two ANOVAs to test the presence of interpolation and extrapolation. Study condition (View 2–View 2 vs. View 2–View 5) and test view (View 2 vs. View 3) were the independent variables in the ANOVA for interpolation; the ANOVA for extrapolation had the same factors but different study conditions were tested (View 2–View 2 vs. View 1–View 2). Facilitation in view generalization from studying two views would be indicated by a significant interaction in the right direction (i.e., the effect of test view in the one-view condition reduced in the two-views condition).

In general, only interpolation for nonoccluded object sets was found (see left panels of Figures 4A and 4C). For these object sets, the longer response time for the novel View 3 than the studied View 2 when one view was studied (View 2–View 2) reduced to a nonsignificant value when Views 2 and 5 were studied.

**Amoeboid nonoccluded set.** Interpolation was revealed in response time by the significant interaction between study condition and test view, \( F(1, 24) = 15.80, p < .01 \) (see Figure 4A). Responses were faster at the studied View 2 than at the novel View 3 when only one view (View 2) had been studied but not when both Views 2 and 5 had been studied. In error rate, there was a significant effect of study condition, \( F(1, 24) = 6.01, p < .03 \), and test view, \( F(1, 24) = 4.32, p < .05 \), but no interaction.

No extrapolation was shown. Only the effect of test view was significant in response time, \( F(1, 24) = 15.89, p < .01 \), and error, \( F(1, 24) = 19.03, p < .01 \). Neither the effect of study condition nor the Study Condition × Test View interaction was significant (\( ps > .20 \)). The viewpoint cost between the studied View 2 and the novel View 3 was about the same with the study of either View 2 alone or both Views 1 and 2.

**Amoeboid occluded set.** No evidence for any interpolation or extrapolation was found (see Figure 4B). For interpolation, only the main effect of test view was significant in response time, \( F(1, 22) = 10.24, p < .01 \). Responses at View 2 were faster than responses at View 3 across both study conditions. In error rate, both the effects of test view, \( F(1, 24) = 14.64, p < .01 \), and study condition, \( F(1, 24) = 4.95, p < .04 \), were significant. Similarly, for extrapolation, only test view had a significant effect in response time, \( F(1, 22) = 15.08, p < .01 \), and error, \( F(1, 24) = 24.36, p < .01 \). There was a viewpoint cost between studied View 2 and novel View 3 in both the View 2–View 2 and View 1–View 2 study conditions, but no evidence of an interaction.

**Geon nonoccluded set.** Signs of interpolation were found (see Figure 4C). The Study Condition × Test View interaction was marginally significant in response time, \( F(1, 23) = 4.09, p = .05 \), and error, \( F(1, 26) = 3.61, p = .07 \). Performance at View 2 was better than at View 3 with the study of only one view (View 2–View 2), but such a difference diminished with the study of two views (View 2–View 5).

No extrapolation was found. There was only the main effect of test view in response time, \( F(1, 23) = 13.44, p < .01 \), and error, \( F(1, 23) = 21.44, p < .01 \). Performance was better at View 2 than at View 3 whether only View 2 or both Views 1 and 2 had been studied.

**Geon occluded set.** No interpolation or extrapolation was found (see Figure 4D). For interpolation, the main effects of test view in response time, \( F(1, 23) = 19.66, p < .01 \), and error, \( F(1, 23) = 22.31, p < .01 \), and study condition in response time, \( F(1, 23) = 4.26, p = .05 \), and error, \( F(1, 23) = 10.10, p < .01 \), were significant. The viewpoint cost between the studied View 2 and the novel View 3 when only View 2 was studied remained when both Views 2 and 5 were studied.

For extrapolation, there was a significant effect of test view in response time, \( F(1, 23) = 36.78, p < .01 \), and error, \( F(1, 23) = 35.38, p < .01 \). Responses were faster and more accurate at View 2 than at View 3 in both study conditions. There was also a significant effect of study condition in error rate, \( F(1, 23) = 5.51, p < .03 \). No interaction was found.

**Discussion**

In this experiment, we showed that interpolation but not extrapolation occurred even though the angular difference between the novel view and the two studied views was controlled. Combining 5 Actually, in the View 2–View 2 condition, half of the trials had View 2 presented twice and half had View 4 presented twice. The same types of trials gave identical results and were pooled together in discussion for simplicity. The same applies to other conditions: In the View 2–View 5 condition, half of the trials had Views 2 and 5 presented and half had Views 1 and 4 presented; in the View 1–View 2 condition, half of the trials had Views 1 and 2 presented and half had Views 4 and 5.

6 This is true for participants introduced with two-object sets.
Figure 4. Average response times in Experiment 2. An asterisk (*) represents a significant Study Condition x Test View interaction in response time. Error bars represent 95% confidence intervals based on within-participant error. v2 = View 2; v3 = View 3; v5 = View 5; v1 = View 1.
results from Experiments 1 and 2, we found two reasons why the study of multiple views benefits generalization to internal views only but not external views. When the studied views were fixed (as in Experiment 1), the internal views were always close to both studied views, whereas the external views were close to only one of the studied views. When we made the external views close to both studied views (as in Experiment 2), we had to bring the studied views closer to each other, too. In this case, they were so similar that not much extra information was provided by two views instead of one. Extrapolation still did not occur. The second finding of this experiment was that, unlike Experiment 1, interpolation did not occur in all conditions. Rather, it was shown only when there was no occlusion between studied views. This difference apparently stands in contrast to the results of Experiment 1, in which interpolation occurred for both occluded and nonoccluded object sets. One speculation for the effect of self-occlusion here but not in Experiment 1 is that the angular difference between the studied views was larger in this experiment. In Experiment 1, even when the two studied views had different visible parts, they were close to each other (Views 2 and 4 were 40° and 56° apart for amoeboid and geon objects, respectively), and combining information from the two views was still possible. In Experiment 2, however, the different visible parts plus a large angular difference (Views 2 and 5 were 60° and 84° apart for amoeboid and geon objects, respectively) made the two studied views very different from each other and thus more difficult to combine.

One point to note about the results is that the studied view in the View 2–View 5 condition seems harder to recognize (based on performance) than the studied view in the other study conditions. One may argue that for the nonoccluded object sets, perhaps the lack of difference between the studied view and the novel view in the View 2–View 5 condition did not reflect genuine interpolation. Instead, this lack of difference might be caused by the drop in performance at the studied views. We should also notice, however, that for occluded object sets, a response time increase at the studied views occurred in some fashion in the View 2–View 5 condition. In this case, however, the response at the novel view was still higher than that at the studied view. Presumably, as the baseline (studied view) rises, all other responses would be expected to rise accordingly. Thus, a rise in difficulty at the studied view does not necessarily imply a smaller likelihood of finding a difference between the studied and novel views.

General Discussion

There were three main findings in the present study. First, the study of two views caused a greater generalization advantage than the generalization from each of the studied views presented in isolation, and this advantage occurred for internal views but not for external views. Second, the exclusive benefit for the internal view remained the same even when the internal and external views were equally far away on average from the two studied views. Third, studying two views sharing the same visible parts led to better generalization than studying two views with different visible parts. Below we discuss each finding in terms of providing both support for and constraints on models of view combination.

The Lack of Extrapolation and Proximity of Studied and Novel Views

Experiment 1 showed that the study of two object views benefits generalization to internal views but not external views. In past studies involving the study of two object views, better performance was found for internal than external views. To our knowledge, however, this is the first study that showed that, for asymmetric objects with or without occlusion, generalization to the external view is no better than when only one object view is studied. The exclusive internal view advantage is consistent with theories of view combination (Perrett et al., 1998; Poggio & Edelman, 1990). According to these theories, an internal view is close to both studied views and thus activates both view-specific representations, leading to a higher level of overall activity. An external view is close to only one of the studied views, thus the summed activity will not be much different from the case when only the closer studied view has been seen.

Srinivas and Schwoebel (1998) also found no extrapolation using asymmetric objects. It should be noted that Experiment 1 of our study provides more conclusive and informative results for three reasons. First, whereas Srinivas and Schwoebel only studied external views and had no control over self-occlusion, we demonstrated in objects with and without self-occlusion that any advantage of view combination applied to internal but not external views. Second, although the external view in their experiments was 80° from the closest studied view, in our Experiment 1 the view difference was much smaller (20° and 28° for amoeboid and geon objects). Even so, however, the external views did not benefit from the study of two views compared with the study of one. Third, even when the external and internal views were equally far away from the studied views (Experiment 2), still no extrapolation was found. As discussed in the next section, this latter result may not be readily predicted by view-based theories of object recognition.

Proximity of Studied Views

The only factor that the above-mentioned theories of view combination (Perrett et al., 1998; Poggio & Edelman, 1990) emphasized concerning the recognition of a novel view is its similarity with previously studied views. If this is indeed the only factor, then we should expect no difference between internal and external views when their averaged distances from the study views are kept the same, as in Experiment 2. The persistent absence of extrapolation in just such a condition suggests another way in which an internal novel view is superior to an external novel view. For an external view to be close to the studied views, the studied views have also to be very close to each other. When the two studied views are too similar to each other, they do not provide much extra information about the object at novel views compared with each of the views alone. Indeed, one possibility is that only one representation is formed because this might be sufficient for representing two very similar views. If the second studied view is too similar to the first studied view, the representation for the first view may suffice for representation of the second view. Therefore, no new representation would be formed for the second view, and generalization to external views would be essentially identical to recognition when only one view has been studied. Poggio
Edelman (1990) considered the possibility of using fewer view-specific representations than the number of views introduced to their network. However, they do not discuss the situations in which this reduction of representations occurs. The results of Experiment 2 here suggest that similarity of studied views could be one reason for formation of fewer representations. Support for this alternative prediction comes from Schwoebel and Srinivas (2000), who showed with bilaterally symmetric objects that studying two dissimilar views led to greater generalization than studying two similar views. They claimed that the dissimilar studied views provide a better specification of object features and thus better view combination.

**Self-Occlusion**

Self-occlusion reduces the improvement in generalization by view combination. Experiment 1 showed better generalization with the study of two views not only for the nonoccluded object sets but also for the occluded sets. On the contrary, self-occlusion reduces the advantage from the study of multiple views, as shown in Experiment 2, when the studied views were more different from each other. We suggest two ways in which self-occlusion can exert an effect, one related to encoding of the studied views and one related to retrieval when seeing a novel view.

First, it may be more difficult to associate two object views that do not share the same visible features. Poggio and Edelman (1990) assumed that no self-occlusion occurs when they tested their network. They noted that their model could, in principle, be applied to cases with self-occlusion by “partitioning the viewpoint space for each object into a set of ‘aspects’” (p. 263). One speculation is that combining views from different aspects (views showing different visible features) is harder than combining views from the same aspect (Tarr & Kriegman, 2001).

Another possible cause for this attenuation of interpolation is that the internal view activates the representations for the two studied views to a lesser degree when self-occlusion occurs. Perrett et al. (1998) speculated, on the basis of neurophysiological findings of faster accumulation of activity in cells of the temporal cortex to whole versus part objects, that occlusion slows down the accumulation of neural evidence and leads to worse behavioral performance. Following this argument, a novel view that shares only some of the features with the stored views would be less well recognized. In the occluded object sets of our study, though, the internal view always contains all of the parts of both studied views. To understand whether Perrett et al.’s (1998) theory can account for the effect of self-occlusion we obtained, we examined the generalization gradient in the one-view conditions in Experiments 1 and 2, separately for nonoccluded and occluded object sets. When only one view was studied, recognition at test views three frames away from the studied view was still quite accurate for the nonoccluded object sets (76%). Recognition for the occluded object sets, however, was much less accurate with a test view separated from the studied view by two frames (66%) and approached chance level with a test view rotated by three frames (57%). Therefore, considerable sensitivity toward internal views was more likely to be retained for the nonoccluded object set, even when the angular disparity was large. It is plausible that an internal view more likely activates both view-specific representations for the nonoccluded object set, resulting in interpolation.

Generalizing View Combination

In the present study, participants were required to discriminate among very similar individual objects, a task regarded typically as subordinate-level recognition. It is likely that for the more general cases of object recognition, in which the objects to be discriminated are less similar to each other, some sort of view generalization mechanisms are still necessary. Hayward and Williams (2000) systematically varied the similarity among novel objects in three different sets and found that the viewpoint-cost function remained roughly the same. In other words, rotating the objects increased the recognition time to the same extent whether the objects were easy or hard to discriminate from each other. Though other researchers found viewpoint invariance with highly dissimilar objects (e.g., Newell, 1998), there remains the possibility that in more general cases of object recognition, there is still space for view generalization to benefit from mechanisms that combine information from multiple experienced views. Tarr and Gauthier (1998) found that in a class-level discrimination task, the experience with one view of an object was generalized to the corresponding view of another object in the same class. It will be interesting to study whether interpolation caused by studying two object views can also be generalized to other objects in the same class.

To conclude, the current work provides a systematic and comprehensive account of the constraints on view combination processes in view generalization. These constraints involve the encoding and retrieval processes during object recognition and should be incorporated into future versions of view combination theories.

**References**


---

**Appendix A**

**Average Percentages of Error Trials in Experiment 1**

<table>
<thead>
<tr>
<th>Object set and study condition</th>
<th>Test view</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Internal</td>
<td>External</td>
</tr>
<tr>
<td><strong>Amoeboid nonoccluded</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>One view</td>
<td>4.11 (1.27)</td>
<td>8.37 (1.92)</td>
</tr>
<tr>
<td>Two views</td>
<td>4.76 (1.20)</td>
<td>3.33 (0.99)</td>
</tr>
<tr>
<td><strong>Amoeboid occluded</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>One view</td>
<td>4.41 (1.59)</td>
<td>10.41 (1.69)</td>
</tr>
<tr>
<td>Two views</td>
<td>10.84 (1.45)</td>
<td>5.81 (1.48)</td>
</tr>
<tr>
<td><strong>Geon nonoccluded</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>One view</td>
<td>4.53 (1.14)</td>
<td>12.19 (2.19)</td>
</tr>
<tr>
<td>Two views</td>
<td>3.50 (1.03)</td>
<td>6.49 (2.49)</td>
</tr>
<tr>
<td><strong>Geon occluded</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>One view</td>
<td>2.50 (0.76)</td>
<td>15.17 (3.36)</td>
</tr>
<tr>
<td>Two views</td>
<td>6.58 (1.57)</td>
<td>12.06 (2.67)</td>
</tr>
</tbody>
</table>

*Note.* Standard errors are in parentheses.

---

**Appendix B**

**Average Percentages of Error Trials in Experiment 2**

<table>
<thead>
<tr>
<th>Object set and study condition</th>
<th>Test view</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>View 2 (studied)</td>
<td>View 3 (novel)</td>
</tr>
<tr>
<td><strong>Amoeboid nonoccluded</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>View 2–View 2</td>
<td>6.84 (1.48)</td>
<td>12.06 (2.67)</td>
</tr>
<tr>
<td>View 2–View 5</td>
<td>6.84 (1.48)</td>
<td>12.06 (2.67)</td>
</tr>
<tr>
<td>View 1–View 2</td>
<td>6.84 (1.48)</td>
<td>12.06 (2.67)</td>
</tr>
<tr>
<td><strong>Amoeboid occluded</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>View 2–View 2</td>
<td>5.00 (1.59)</td>
<td>16.15 (3.07)</td>
</tr>
<tr>
<td>View 2–View 5</td>
<td>5.00 (1.59)</td>
<td>16.15 (3.07)</td>
</tr>
<tr>
<td>View 1–View 2</td>
<td>5.00 (1.59)</td>
<td>16.15 (3.07)</td>
</tr>
<tr>
<td><strong>Geon nonoccluded</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>View 2–View 2</td>
<td>5.00 (1.59)</td>
<td>16.15 (3.07)</td>
</tr>
<tr>
<td>View 2–View 5</td>
<td>5.00 (1.59)</td>
<td>16.15 (3.07)</td>
</tr>
<tr>
<td>View 1–View 2</td>
<td>5.00 (1.59)</td>
<td>16.15 (3.07)</td>
</tr>
<tr>
<td><strong>Geon occluded</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>View 2–View 2</td>
<td>5.00 (1.59)</td>
<td>16.15 (3.07)</td>
</tr>
<tr>
<td>View 2–View 5</td>
<td>5.00 (1.59)</td>
<td>16.15 (3.07)</td>
</tr>
<tr>
<td>View 1–View 2</td>
<td>5.00 (1.59)</td>
<td>16.15 (3.07)</td>
</tr>
</tbody>
</table>

*Note.* Standard errors are in parentheses.