Attention, Perception, & Psychophysics
2009, 71 (3), 490-502
doi:10.3758/APP.71.3.490

Integration of multiple views of scenes

Monica S. Castelhano
Queen's University, Kingston, Ontario, Canada

Alexander Pollatsek
University of Massachusetts, Amherst, Massachusetts

and

Keith Rayner
University of California, San Diego, La Jolla, California

In two experiments, memory was tested for changes in viewpoints in naturalistic scenes. In the key study condition, participants viewed two images of the same scene from viewpoints 40º apart. There were two other study conditions: The two study images were identical or were of different scenes. A test image followed immediately, and participants judged whether it was identical to either of the study images. The scene in the test image was always the same as in a study image and was at least 20º from any study image on *different* trials. Two models were tested: (1) views stored and retrieved independently and (2) views combined at retrieval. The crucial test of these hypotheses involved a comparison (in the key study condition) of the *interpolation* condition (the test image was presented between the two study images and 20º from both) and the *extrapolation* condition (it was 20º from one study image and 60º from the other). Performance in the interpolation condition was far worse than what was predicted by the first model, whereas the second model fit the data quite well. The latter model is parsimonious in that it integrates previous experiences without requiring the integration of the views in memory. We review some of this model's broader implications.

One of the most amazing properties of the visual system is its ability to identify our surroundings, despite the variety of visual conditions that change the image reflected onto the retina (e.g., lighting, color). This is especially apparent in our ability to recognize the same scenes from a number of different views. Although there have been many studies on the recognition and perception of scenes over the past few decades (for a review, see Henderson, 2007), only a few have investigated how scenes are represented across different viewpoints and how the system uses this information in scene recognition (Christou & Bülthoff, 1999; Garsoffky, Schwan, & Hesse, 2002; Hock & Schmelzkopf, 1980; Nakatani, Pollatsek, & Johnson, 2002; Shelton & McNamara, 1997). Christou and Bülthoff focused on this problem by using a navigation task in a virtual-reality setting. Participants were allowed to explore an attic (consisting of multiple rooms) from certain viewpoints. When participants were asked to recognize still images taken from this environment, it was shown that scene recognition was highly viewpoint dependent. However, the scenes were relatively impoverished, because they consisted mainly of rooms defined by planar walls, floors, and ceilings, with only the angles and orientations of these planes able to provide information for distinguishing these views. Shelton and McNamara were also interested in the representation of large, navigable spaces and had participants memorize the relative locations of items from specific viewpoints within an array of objects. Participants were then asked to imagine themselves aligned to different views and to report the relative positions of the objects. Imagined headings aligned with the study views were easier to retrieve than novel headings, and so performance was consistent with a viewpoint-dependent representation of the space. The important thing to note about these two studies is that the space, as represented by the walls or the array of objects, was the primary focus of the investigation.

In other studies (Garsoffky, Huff, & Schwan, 2007; Garsoffky et al., 2002; Huff, Schwan, & Garsoffky, 2007), participants studied dynamic scenes depicting moving players in a soccer game (Garsoffky et al., 2002) or a basketball game (Garsoffky et al., 2007). In these studies as well, it was found that performance was viewpoint dependent. However, as discussed in Huff et al. (2007), the nature of representing dynamic scenes may be likened more to a descriptive "event" representation (Zacks, Tversky, & Iyer, 2001) and so may involve consideration of information not present in the static scenes that are typical of scene-recognition studies. Therefore, it is difficult to conclude solely on the basis of these studies how scenes are repre-

M. S. Castelhano, monica.castelhano@queensu.ca

© *2009 The Psychonomic Society, Inc.*

sented across different viewpoints and how the system uses this information, because the definition of the scene varied so greatly. In the present study, scenes are defined as semantically coherent views of the world, as in the definition put forth by Henderson and Hollingworth (1999).

In contrast to scene recognition, the nature of object recognition has been the subject of numerous investigations, and the theoretical approaches may shed some light on how the visual system processes more complex scenes (for a complete review, see Peissig & Tarr, 2007). Theoretical approaches to the recognition of objects across viewpoints are of two main types: *viewpoint dependent* and *viewpoint invariant.* Although these labels describe the type of recognition performance expected when objects are recognized from new viewpoints, the main tenet of each approach is the nature of object representations. Researchers who support a viewpoint-dependent approach posit that object representations are image based (I. Bülthoff & H. H. Bülthoff, 2003; Tarr, 1995) or prototype based (Edelman, 1999; Ullman, 1989). When one is presented with a new view of a known object, an alignment process takes place to match the present view with the stored view (Tarr, 1995; Tarr & Pinker, 1989; Ullman, 1989); otherwise, recognition is based on viewpoint-independent distinctive features (Tarr, 1995). So, in this approach, object representations consist of multiple snapshots gathered from experience at specific viewpoints. As a result, input from the new view is aligned and relatively monotonic increases in error and response times (RTs) are expected as a function of the distance of this new view from the most similar stored view.

Alternatively, researchers supporting a viewpoint-invariant approach have proposed that the visual system creates a viewpoint-invariant representation of objects (Biederman, 1987; Biederman & Gerhardstein, 1993, 1995; Marr, 1982; Marr & Nishihara, 1978). These researchers have posited that objects are represented as structural descriptions of the spatial relations among simple, volumetric 3-D parts. Thus, the structural description representations can be used to recognize objects from a number of viewpoints, limited only by the extent to which the constituent parts do not change across views (Biederman, 1987; Biederman & Gerhardstein, 1993). Given that the structural descriptions match across views, performance should be relatively consistent across these views of the object, even if they have not been viewed previously. Therefore, it follows that viewpoint-dependent performance is expected when the structural description changes by either the addition of a part previously occluded or the deletion of a part that has become occluded (Biederman & Gerhardstein, 1993). In this case, a new structural description is required for the same objects from a different view to account for the changes in visible parts.

It is not clear how the theoretical approaches based on the use of isolated objects as stimuli generalize to scenes, which consist of multiple objects arranged in specific spatial layouts, all of which can change with shifts in viewpoint. Despite these differences, the constraints placed by each of the approaches on visual recognition can serve as a guide to investigate viewpoint information in scene representations. What is clear from both of these approaches is that, when multiple views of the same object or scene are shown and each view differs drastically (either because of a large change in distance from the new view to the stored view or because numerous part changes have occurred in the structural description), a new representation is created. We surmise that, when a scene is shown at different viewpoints, each view is represented anew due to a number of changes that occur in a scene in terms of both its spatial layout and its object contents.[1]

The most common test of the two prominent approaches to object recognition (viewpoint dependent, viewpoint invariant) has been to examine the extent to which new views of an object can be recognized by generalizing from known views (Biederman & Gerhardstein, 1993, 1995; Peissig & Tarr, 2007; Tarr & Pinker, 1989). If multiple studied views of the object are considered, the visual system must contend with multiple experiences with that object when it judges a new view. If we assume that the previously studied views of the object are stored as multiple representations (i.e., assuming an instance-based approach), there are generally two types of judgment processes that can be taking place to allow the system to decide whether a new view of an object matches the stored views of an object in memory. On the one hand, the new view can be compared (e.g., via alignment) with each of the views stored in memory (Tarr & Pinker, 1989, 1990; Ullman, 1989). Alternatively, all stored views can be somehow combined and compared with the new view (H. H. Bülthoff & Edelman, 1992; Edelman, 1999). H. H. Bülthoff and Edelman showed that the object is more readily recognized when a test image is from an *interpolated* viewpoint (i.e., presented between the two studied views) than when the viewpoint of the test image is *extrapolated* (i.e., outside of the boundaries set by these studied views); they concluded that the latter of the two judgment processes was more likely occurring.

If we consider a case in which multiple views of a scene are observed and stored, there is the question as to how the system uses the stored information from multiple instances of different views to assess a new view of the same scene. On one hand, it could be that, due to the vast number of changes occurring with viewpoint changes, the system is less able to combine old views for comparison with new ones. On the other, in order to be able to extrapolate to new views, it is possible that the system treats individual objects and scenes in the same way and combines the known views of a scene. In the present study, we focused our investigation on examining how a judgment is made. We point out that an investigation into the processes involved in making this judgment does not favor one type of representation (i.e., these judgments can be instantiated with either an image-based or a structural description, or with some other type of scene representation). We think, however, that the way in which information is used in the decision process can shed some light on the representation; we return to these ideas in the General Discussion section.

In the present study, participants were asked to judge whether a test image of a scene was presented either from the same viewpoint as were previously seen images of the scene or from a new one. We chose a viewpoint discrimination task in order to avoid possible confounding effects of distinctive features that can accompany a typical identity-recognition task (especially in the case of very disparate stimuli, such as scenes) (Tarr, Bülthoff, Zabinski, & Blanz, 1997). Our ensuing focus is on how one makes a judgment of "new" in this task (i.e., the scene that has been previously viewed, but never from the viewpoint of the test image). In particular, we were interested in determining whether performance in such a task could be explained by assuming that people compared each of the stored viewpoint-dependent representations of the scene with the test image separately, or whether one needs to posit that some sort of integration of these representations is involved in the decision. In the following, we assume (1) that the views of a scene are all taken from the same vertical height, so that the views differ only in the horizontal viewing angle, and (2) that the difference in horizontal viewing angle between two scenes is ordinarily a good index of how similar a study image is to a test image.

Again, assuming that the views of a scene are stored as separate representations, we would expect memory performance to have two properties. The first is the obvious property that, all else being equal, a *new* judgment is less difficult the greater the difference in the viewing angle between the test image and a particular study image. The second property is that having two different views of the scene in memory (neither of which is the same as the test image) makes it more difficult to make a *new* judgment than if only one of those images is in memory. That is, it would seem that each additional study image has some degree of match to the test image and thus should make a *new* (i.e., mismatch) judgment more difficult. We describe two models (with those two properties) that we will evaluate on the basis of the memory performance observed in the experiments. The first is described in detail below; the second is only sketched, but will be presented in full and tested later. However, we emphasize that a key property of both, and one that we are exploring, is that each stored scene view is a separate memory representation (for reasons stated above). The difference lies in how the results of these judgments are accomplished.

First, consider the independent decision model. This is perhaps the simplest way in which the information from the comparisons between various study images and the test images can be assessed. In this model, each view is stored in memory as a separate entity, and then, when a test image is presented, (1) each study image is compared with the test image independently, and a *yes*/*no* decision is made about whether there is a match between the two; then (2) a "no" (*new*) decision is made on the test image only if each of these separate decisions is "no."

Although this model may seem too simple, it is not unreasonable. First, it has the two properties that we outlined above. Obviously, it predicts that the probability of making an overall decision of *new* would be higher the more dissimilar any given study image is from the test image (all else being equal). It also predicts that introducing an additional study image would make a *new* decision less probable, because a *new* decision is made only if all the individual decisions are "no"; thus, unless the added study image can be rejected as being different from the test image with a probability of 1, the probability of a *new* judgment is lowered by this added study image. Moreover, using the independent decision model with an inverse decision rule (i.e., "Have you seen this scene before—ignoring orientation?") could also give a reasonable account of the typical identity-recognition task discussed earlier (H. H. Bülthoff & Edelman, 1992; Christou & Bülthoff, 1999; Edelman & Bülthoff, 1992; Humphrey & Khan, 1992; Ullman, 1989; Ullman & Basri, 1991). That is, such a model would assume that each study image is compared with the test image and a *yes*/*no* (match/mismatch) decision is made, and that an overall *new* decision is made only if *all* the individual decisions are "mismatch." Thus, such a model would predict that *old* judgments would be more probable as the number of different study images increases, that *old* judgments would be more likely the more similar the test image is to any particular study image, and that interpolated test images are more likely to be judged "the same" as the study images, because they are, on average, closer to them.

The independent decision model, however, makes one fairly simple, testable prediction that does not require any specification of parameters. It involves a comparison between the two key test conditions in Experiment 1: the interpolation condition and the two different extrapolation conditions. In the key study condition in Experiment 1, we showed participants two different images of the same scene from two viewpoints. To be consistent with our terminology below, we refer to the viewing angles of the two study images as 20º and 60º. Participants were told to study these images for an immediate memory test to follow. As stated above, participants were asked to indicate whether the test image was taken from exactly the same viewpoint as was one of the study images (to which they would respond "old") or from a new viewpoint (to which they would respond "new"). There were three conditions for *new* test images. In the interpolation condition, the test image was at the 40º viewpoint (i.e., intermediate between the two study viewpoints), and in the extrapolation test conditions, the test image was at either 0º or 80º (i.e., outside the two study viewpoints). Thus, in the interpolation condition, both memory study images were 20º from the test image, whereas in each extrapolation condition, one memory study image was 20º from the test image and the other was 60º from the test image. Given that the study images were presented sequentially, the order of presentation must be taken into account; it was reasonable to assume that the decision of "mismatch" would become more difficult with a greater lag between the presentations of a study image and the test image. By making one simplifying assumption—that all "near" study images (i.e., 20º from the test image) were equally confusable with the test image and all "far" study images (i.e., 60º from the test image) were also equally confusable with the test image—we were able to model the probabilities of making *new* judgments in the interpolation and extrapolation conditions with the following three equations (where "first" and "second" mean presented first and second, respectively, and "no" means deciding that there is a mismatch between an individual study image and the test image).

Extrapolation first near:

$p(new) = p(\text{“no” to near stimulus when first}) \cdot p(\text{“no” to far stimulus when second})$

Extrapolation second near:

$p(new) = p(\text{“no” to far stimulus when first}) \cdot p(\text{“no” to near stimulus when second})$

Interpolation:

$p(new) = p(\text{“no” to near stimulus when first}) \cdot p(\text{“no” to near stimulus when second})$

These equations made the obvious prediction that performance would be worse in the interpolation condition (i.e., fewer correct "new" responses) than in either extrapolation condition, because one of the component mismatch decisions was more difficult in the interpolation condition than in either of the extrapolation conditions. However, they also imply the following less obvious prediction:

$$p(new, \text{interpolation condition}) \geq p(new, \text{extrapolation first near condition}) \cdot p(new, \text{extrapolation second near condition}).$$

This inequality follows straightforwardly from the three equations above. That is, multiplying the right-hand sides of the two equations for the extrapolation conditions yields the product of four probabilities: the product of two probabilities in the right-hand side for the interpolation condition multiplied by two other probabilities. Hence, because probabilities are less than or equal to 1, this product (the right-hand side of the inequality above) is less than or equal to the left-hand side of the inequality. We should add that this prediction is independent of how one views each mismatch decision as being made: Each decision could be consistent with a simple threshold model or with a signal-detection model. Thus, testing whether Contrast 1 is greater than or equal to zero is our key test of the independent decision model.

Contrast 1:

$$p(new, \text{interpolation condition}) - p(new, \text{extrapolation first near condition}) \cdot p(new, \text{extrapolation second near condition})$$

What are the alternatives to this model? Staying within the general framework of assuming (1) that the memory representations are stored separately and (2) that the decision of *old* or *new* has, as components, independent comparisons of the test image with the various study images, there are clearly many such models. The one we consider here (referred to as the *integrated decision model*) is that the *old* or *new* judgment is produced through a combination of the information from all representations. To allow for this combination, each comparison of the test image with a study image produces a graded or "strength of match" output rather than an *old*/*new* decision. We develop the details of two instances of this class of models below, and, in particular, the latter appears to account for the data quite well. However, because the specification of these models requires significant detail, including assumptions about parameters, we defer discussion of it until later.

Briefly put, Experiment 1 contains both the key test of the independent decision model and the key data for the integrated decision model. Experiment 2 is a control experiment, which contains conditions that check that there are no artifacts influencing the key conditions of Experiment 1 and, therefore, compromising the tests.

EXPERIMENT 1

As indicated above, the focus in Experiment 1 was on how people discriminate a test image from two study images of the same scene shown at different viewpoints. In these *different-viewpoints* study conditions, the two study images were 40º apart. In one key test condition (the interpolation condition), the test image was taken from a viewpoint halfway between them. In the other key test condition (the extrapolation test condition), the test image was 20º apart from one of the study images and 60º from the other. There was also a test condition in which the test image matched one of the study images. As our discussion up to this point, we hope, has made clear, although the independent decision model (or any reasonable model) predicts that the interpolation test image will be harder to judge as *different* than the extrapolation test images will be, the independent decision model makes a simple quantitative prediction about the relationship of these conditions (formulated in Contrast 1).

In addition to the different-viewpoint study condition, in the *identical-viewpoint* study condition, the identical image was presented twice during the study section of the trial. This study condition was added as a baseline control to assess the difficulty of discriminating the test image from the images presented at the various viewpoints seen in the different-viewpoints study conditions.

Method

Participants. Thirty-six University of Massachusetts undergraduates participated for credit toward an introductory psychology course. They all had normal or corrected-to-normal vision.

Stimuli and Apparatus. Stimuli consisted of 48 illustrations of naturalistic indoor scenes rendered in 3-D with Data Becker Home Design 5.0. These scenes (which subtended a visual angle of 40.6º horizontally and 31.2º vertically) were displayed at a resolution of 800 × 600 pixels in 24-bit color on a 19-in. monitor with a 100-Hz refresh rate. Each scene was rendered from five different viewpoints, labeled 0º, 20º, 40º, 60º, and 80º in Figure 1A, which shows an example stimulus at each of these viewpoints.

Design and Procedure. Participants were shown two study images, which were followed immediately by a test image (Figure 1B). The two study images could be either different images of the same scene, one from the 20º viewpoint and one from the 60º viewpoint (different-viewpoints condition), or two identical images of a scene taken from either the 20º or the 60º viewpoint (identical-viewpoint condition). The test image could be identical to a study image (*old* condition) or could be the same scene taken from a different viewpoint than that of either of the study scenes (*new* condition). As indicated above, for the *new* condition, there were two types of *new* viewpoints that could be shown during test.

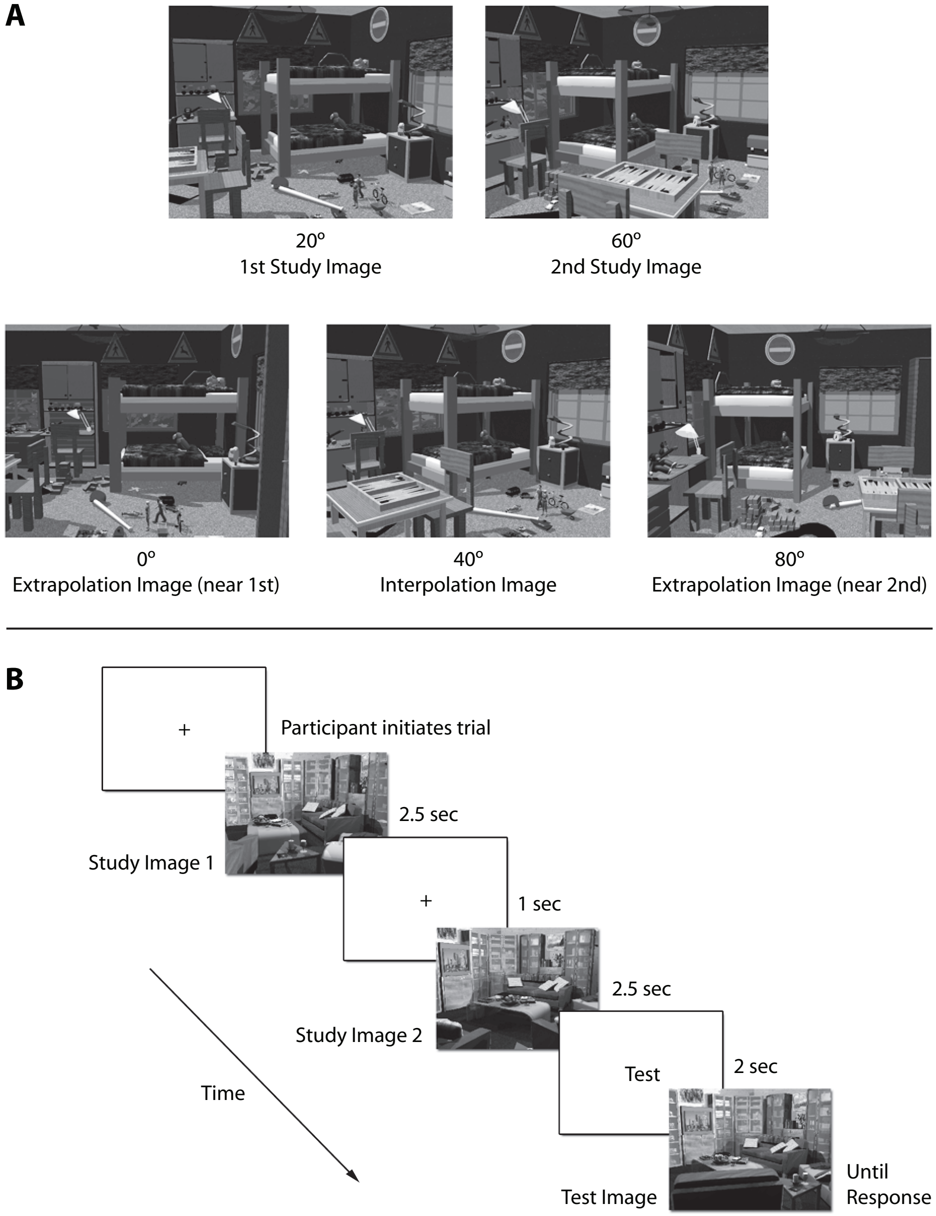

Figure 1. Example of a stimulus at each of its viewpoints used in Experiment 1 (A) and the images seen by a participant on any given trial (B). See text for more details.

In the interpolation condition, the test viewpoint fell between the two study viewpoints (40º). In the extrapolation condition, the test viewpoint fell outside the two study viewpoints shown (0º and 80º). The order and the viewpoints shown in the study and test conditions were counterbalanced across participants, and all 12 combinations of study and test stimuli were used. For the identical-viewpoint study condition, we call tests at the 40º viewpoint *interpolation control* and the average of tests at the 0º and 80º viewpoints *extrapolation control*, because they involved test stimuli at the same viewpoints as those in the interpolation and extrapolation conditions, respectively.

Participants were tested in 48 experimental trials. One third of the trials were *old*, and two thirds were *new*; half of the *new* trials were interpolation trials, and half were extrapolation trials. These were preceded by 15 practice trials. Within each trial, all images were taken from the same scene and for each trial a different scene was used. There was no break between the practice and experimental trials.

At the beginning of each trial, a black fixation cross was shown at the center of a white background. Participants were instructed to look at this fixation cross and to press a key when they were ready to begin a trial. On each trial, the two study images were shown sequentially for 2.5 sec each, with a 1-sec delay between them. Following the presentation of the second image, the word "Test" was displayed for 2 sec at the center of the screen to indicate that the next image would be the test image and required a response. The test image was then displayed until the participant responded. Participants responded as to whether the test image was identical to one of the preceding test images by pressing one of two buttons that corresponded to "yes" (old) or "no" (new). Whenever RT exceeded 2.5 sec, a buzzer sounded, which the participants were warned indi-

Table 1
Accuracy, Percent *Old* Responses, and Mean Response Times (RTs, in Milliseconds) for Experiment 1 by Test Condition

Study and Test Condition	Percent Correct Responses	Percent *Old* Responses	RT
Identical-Viewpoint Study Images			
Correct response *old*	99.3	99.3	944
Correct response *new*			
Extrapolation control	91.3	8.7	948
Interpolation control	88.3	11.7	981
Different-Viewpoints Study Images			
Correct response *old*			
First image same as test image	78.9	78.9	1,185
Second image same as test image	92.3	92.3	1,065
Correct response *new*			
Extrapolation: First image near test image	65.6	34.4	1,227
Extrapolation: Second image near test image	82.9	17.1	1,103
Interpolation	42.0	58.0	1,328

cated that they were responding too slowly. No feedback was given regarding accuracy.

Results and Discussion

The accuracy and RT for each condition are reported in Table 1. The RT pattern mirrors the pattern for accuracy; however, because accuracies were quite low in some conditions, any quantitative tests involving RTs were difficult to interpret; thus the focus is on the accuracy data. Trials that exceeded the 2.5-sec limit placed on RTs were excluded from the rest of the analyses (1.5% of data).

Accuracy. First, consider the identical-viewpoint conditions, which were employed largely to assess whether there was any difference in difficulty of making the discrimination between the two study viewpoints and the three test viewpoints. (Note that all *new* test images in the identical-viewpoint condition were 20º from the study image, and thus the labels *extrapolation* and *interpolation* here merely denote the same test stimuli as those in the corresponding different-viewpoints study conditions.) The overall percent correct in these conditions was quite high (93%), but the error rate was 9.5% higher for the average of the extrapolation and interpolation control conditions than for the *same* test condition [$F(2,70) = 6.74$, $MS_e = 0.017$, $p < .01$]. The 3.0% difference between the extrapolation and interpolation conditions was far from significant ($|t(35)| < 1$).

As indicated above, the results from the different-viewpoints conditions are of central interest, and we turn to them now (see Table 1). As with the identical-viewpoint conditions, on average, the percent correct was higher for the *old* conditions than for the *new* conditions [$F(2,70) = 51.53$, $MS_e = 0.036$, $p < .01$]. Of greater interest, there was a *lag effect*: Memory was poorer when the most relevant study item was the first one shown. This is perhaps seen most clearly in the *old* conditions, for which the error rate was 13.4% higher when the test image was identical to the first study image than when it was identical to the second study image [$t(35) = -3.44$, $p < .05$]. Similarly, in the extrapolation condition, the false alarm rate was 17.3% higher when the "close" study image (the one 20º from the test stimulus) was the first rather than the second study image [$t(35) = -3.94$, $p < .01$]. Of greatest interest, however, is that the percent correct in the interpolation condition was only about half that of the average of the extrapolation conditions [$t(35) = 6.73$, $p < .01$]. Moreover, this is not explainable by the lag effect, because there were 23.6% more false alarms in the interpolation condition than in the extrapolation condition in which the test image was 20º from the first study image (Table 1, extrapolation first image near) [$t(35) = 4.07$, $p < .01$].

RT. RTs were on the order of 1 sec (see Table 1). The RTs reported are for correct trials only. The pattern of data mirrored that of the accuracy data, so that none of the comparisons above were compromised by a speed–accuracy trade-off. However, because of the high error rates in many of the conditions, it is difficult to draw strong inferences about the size of the RT differences. Of greatest interest is that the RTs in the interpolation condition were 158 msec greater than the average of the RTs in the two extrapolation conditions [$t(35) = 4.06$, $p < .001$]; however, the 101-msec difference between the interpolation condition and the extrapolation condition in which the near image was presented first was only marginally significant [$t(33) = 1.81$, $p < .10$]. For the different-viewpoints conditions, the 120-msec lag effect when the correct response was *old* and the 124-msec lag effect for the extrapolation conditions were both significant [$t(35) = 3.08$, $p < .01$; $t(33) = 2.98$, $p < .01$, respectively].

Evaluation of the independent decision model. As indicated above, the key test of the independent decision model involves Contrast 1. That is, the model predicts that the value of the contrast, p(*no*, interpolation condition) − p(*no*, extrapolation first near condition) p(*no*, extrapolation second near condition), should be greater than or equal to zero. However, this prediction is falsified, because the actual value was −15.7%, which is significantly less than zero [$t(35) = -2.74$, $p < .01$].[2] Thus, the independent decision model can be rejected, because performance in the interpolation condition was substantially worse than would be predicted by independent decisions about the match between the test images and each of the two study images. In the framework of this model, the performance cost in the interpolation condition was more than can be explained by the fact that both of the study conditions were close to the test image. Although this suggests that the two study images were, in some sense, combined in memory, we first turn to Experiment 2 before trying to model the data.

EXPERIMENT 2

Experiment 2 was designed primarily to test whether the results observed in Experiment 1 were due simply to decay of the first image or to the second viewpoint's actually interfering with either the storage or the retrieval of the first image. A mixed image condition was added in which the two study images were of different scenes, the test stimulus was an image of one of the two scenes, and the timing of the two presentations and test was the same as in Experiment 1. Thus, any difference in performance between when the tested scene was presented first or second should be due to decay rather than to the specific memory of the viewpoint of the second scene making the viewpoint of the first scene less available. In addition, a difference in performance served as an additional check that the high error rate in the interpolation condition was not an artifact of the angle from which the test stimuli in that condition (40º) were seen in Experiment 1.

Method

Participants. Twenty-four University of Massachusetts undergraduates participated for credit toward an introductory psychology course. They all had normal or corrected-to-normal vision.

Stimuli and Apparatus. The stimuli used in Experiment 2 consisted of the ones used in Experiment 1 plus 48 new scenes created using the same methods described in Experiment 1.

Design and Procedure. The procedure and design were similar to those of Experiment 1, with the exceptions listed below. There was an identical-viewpoint study condition that was the same as in Experiment 1 (with both *old* and *new* test stimuli). However, the different-viewpoints study condition was replaced by the mixed image study condition in which images from two different scenes were presented. One of the scenes was identical to the scene in Experiment 1, and the other was replaced by a perspective view of a different scene. (Half the time, the first study item was replaced, and half the time, the second study item was replaced.) However, the test image was always the scene in Experiment 1 that was 20º from the study item actually presented. Thus, logically, although there was a distinction in the *different* trials between whether the scene tested was presented first or second, there was no difference between the *different* trials analogous to that between the interpolation and extrapolation test conditions. However, in order to directly compare the results from Experiment 2 with those from Experiment 1, the *different* test condition was divided into "extrapolation" and "interpolation," corresponding to which test view was used in Experiment 1. The instructions and the procedure for each trial were identical to those in Experiment 1, but only three practice trials preceded the experimental trials.

Results and Discussion

The accuracy and RTs for Experiment 2 are presented in Table 2. Again, we concentrate on the accuracy results. Also, trials that exceeded the 2.5-sec limit placed on RTs were excluded (1.4% of data).

Accuracy. The mean percent correct in the identical-viewpoint condition was very close to that of Experiment 1 (92.5% vs. 93%, respectively). As in Experiment 1, the error rate in the identical-viewpoint study condition was higher on *old* trials than on *new* trials (97.9% vs. 89.8%), but the difference was not statistically significant [$t(23) = 1.67, p > .10$]. In addition, the difference between the extrapolation and interpolation control conditions was small (1.5%) and in the opposite direction from that in Experiment 1 [$|t(23)| < 1$]. A between-experiments analysis showed no interaction of experiment and test conditions for the identical viewpoint conditions ($F < 1$).

In the mixed image conditions, there was only a 2.4% lag effect on *old* trials [$|t(23)| < 1$], but a 22.3% lag effect on *new* trials (averaged over the extrapolation and interpolation conditions) [$t(23) = -3.92, p < .01$]. The former lag effect was substantially smaller than that for *old*

Table 2
Accuracy, Percent *Old* Responses, and Mean Response Times (RTs, in Milliseconds) for Experiment 2 by Test Condition

	Response Measure		
Study and Test Conditions	Percent Correct Responses	Percent *Old* Responses	RT
Identical Study Images			
Correct response *old*	97.9	97.9	951
Correct response *new*			
Extrapolation control	89.1	10.9	996
Interpolation control	90.5	9.5	973
Mixed Study Images			
Correct response *old*			
First image same as test image	85.8	85.8	1,283
Second image same as test image	88.2	88.2	1,092
Correct response *new*			
Extrapolation: First scene close to test image	59.7	40.3	1,301
Extrapolation: Second scene close to test image	85.4	14.6	1,087
Interpolation: First scene close to test image	63.5	36.5	1,241
Interpolation: Second scene close to test image	82.3	17.7	1,077

different-viewpoints trials in Experiment 1 (where there was a 13.4% lag effect on *old* trials); however, the difference across experiments was only marginally significant [$F(1,58) = 3.48$, $p < .07$]. For the extrapolation *new* trials, the 22.3% lag effect reported above was not significantly different from the 17.3% lag effect in Experiment 1 ($F < 1$). Perhaps most importantly, as shown in Table 2, the average probabilities of being correct in the interpolation and extrapolation conditions in the mixed image study conditions were virtually identical. This indicates that the high false alarm rate in the interpolation condition in Experiment 1 was due to the test image's being "between" the two study images rather than to its being more confusable with one of the test images because of some artifact in the stimuli. Another striking aspect of the data was that performance in the *new* test trials in the mixed study condition was about the same as the performance in the extrapolation test conditions in Experiment 1 ($F < 1$), despite the fact that there was only one study image to distinguish from the test image.

RTs. Again, the RTs were about 1 sec. The RTs reported are for correct trials only. As can be seen in Table 2, the RTs for the *identical* condition did not differ significantly: neither between the *old* and *new* responses [$|t(23)| < 1$] nor between the extrapolation and interpolation conditions [$|t(23)| < 1$]. More interestingly, in the mixed study condition, RTs in the extrapolation conditions were actually 35 msec greater than in the interpolation conditions [however, $t(23) = 1.05$, $p > .20$]. Thus, again, there is no evidence that the test stimuli in the interpolation condition in Experiment 1 were more difficult to judge than those in the extrapolation conditions. As in Experiment 1, there were significant lag effects of 191 msec for correct *old* responses, 214 msec for the extrapolation conditions, and 164 msec for the interpolation conditions [$t(23) = 2.72, 5.47, 2.69$; ps $< .05, .001$, and $.05$, respectively]. Although these lag effects were somewhat larger than those in Experiment 1, the interaction of lag effect with experiment was not close to significant, either for the correct old responses or for the extrapolation test items [$F(1,58) = 1.10, 2.13$; ps $> .20, .10$, respectively].

THE INTEGRATED DECISION MODEL

There is no simple test of the whole class of integrated decision models, because there are many different hypotheses one could have for how a single decision could be made on the basis of the combined amount of agreement between two memory traces and the test image (for examples of such models, see Brainerd & Reyna, 2001; Hintzman, 1986). Our approach, rather than carefully trying to achieve a best fit with numerous models, was to determine whether a fairly simple version of the model, with reasonable assumptions, could explain the data well.

A key construct in our model was the amount of resonance between a study image and the test image. (One can think of this as the strength of "oldness" that the particular memory trace gives to the test image.) For simplicity, we assumed that the strength of resonance each memory trace imparts to the test image is represented as the height of an adjusted normal distribution along an axis representing the viewpoint of the scenes. As mentioned above, we discuss two versions of this class of models. The form of the function in Model 1 was $\exp\{-[(X-\mu)/\sigma]^2/2\pi\}$.[3] The x-axis of this distribution is the viewpoint of the scene, the mean is assumed to be at the viewpoint of the study image, and the y value of this function at any particular value on the x-axis represents the resonance that a test image at that orientation would have with the study image (see Figure 2). To represent the poorer memory for the first study image shown, the variance is set higher. The decision about whether the test image matches either of the study images is made by summing the two matching strengths above and comparing this sum with a threshold value. We assumed that this threshold value had a mean and a standard deviation that was the same across all conditions. If the combined strength from the two study images exceeds the threshold, the response is "yes" (*old*), and if not, the response is "no" (*new*). (The random component can also be viewed as a random variation of the response threshold from trial to trial.)

We started out by trying to fit the data of Experiment 1, and as can be seen in Table 3, the model fit the test conditions in the different-viewpoints study condition quite well. Algebraically, the model we used can be summarized as follows: First, as indicated above, the total "strength," or resonance coming from the two comparisons, is

$$\text{Total resonance} = \exp\{-[(X-\mu_1)/\sigma_1]^2/2\pi\} * \exp\{-[(X-\mu_2)/\sigma_2]^2/2\pi\}.$$

The parameter values we used were as follows: For the first memory stimulus, the mean value (μ_1) was 20º and its standard deviation (σ_1) was 18º; for the second memory stimulus, the mean value (μ_2) was 60º and its standard deviation (σ_2) was 12º. (Thus, the mean values were not free parameters.) Then, the total resonance (TR) at Viewpoint X was compared with a response threshold value (TH), so that the probability of a "yes" response was given by the probability that TR was greater than TH. The value of TH was given by a normal distribution, with a mean of 0.24 and a standard deviation of 0.12. Note that the predicted false alarm rate for the interpolation condition (61.4%) is well within the observed 95% confidence interval for this value (52.2%–66.3%). The one aspect of the data that the integrated decision model did not explain particularly well, however, was the significant lag effect for *old* responses in the different-viewpoints study condition. The model predicted only a 5.6% lag effect, whereas the observed lag effect was 13.4%.

Before we return to the problem in predicting the lag effect, we first turn to the question of whether the same model can fit the identical-viewpoint study condition. The answer is that it can with only one parameter change for these trials. It seems reasonable to assume that the participant is quite aware of the fact that the viewpoints of the study images on a trial were identical, and thus they set up a substantially higher strength threshold criterion for saying "old" on those trials. Holding the other parameters constant and changing the mean response threshold to 0.47 for identical trials, the model predicted a hit rate of 99.7% in the identical-

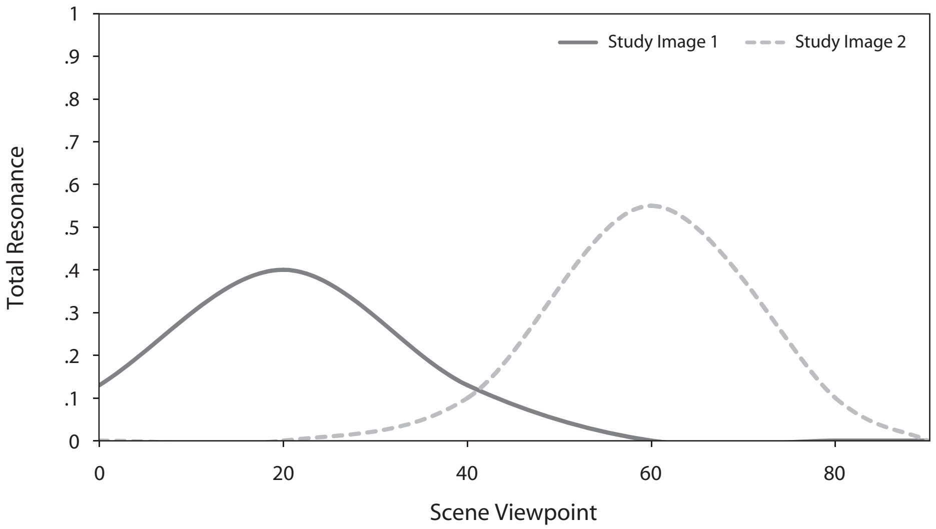

Figure 2. The resonance component of the integrated decision model. The *x*-axis represents the viewpoint of a scene (measured in degrees of angle from the viewpoint arbitrarily chosen as 0º). The mode of the distribution depicted by the solid line represents the viewpoint of Study Image 1, and the height of the curve for a given *x* value represents the resonance that would be produced by this study image and a test image at viewpoint *x*. The distribution depicted by the dashed line represents the analogous resonance function for Study Image 2. The poorer memory for Study Image 1 is represented by a larger variance in the curve. To reach a decision about the degree of match between a test image and the two study images, the two matching strengths are summed and compared with the threshold value. Thus, if the test image is at 40º, as assumed in the text, the combined strengths would be the sum of the *y* values of the two curves at 40º. See text for more details.

viewpoint *old* condition and a false alarm rate of 9.8% in the identical-viewpoint *new* condition, which are extremely close to the observed values of 99.3% and 10.2%, respectively. Given that the data in the *identical* conditions in Experiment 2 were virtually the same as those in Experiment 1 and that Model 2 (below) would make virtually the same predictions in these conditions as would Model 1, we concentrate our modeling efforts below on the different-viewpoints presentation conditions.

We then went on to determine whether Integrated Decision Model 1 could explain the data of Experiment 2, and, if so, whether parameters need to be changed in order to do so. The answer is that an Integrated Decision Model 1 can fit the data quite well, but that a few parameters needed to be changed slightly. In Experiment 1, $\sigma_1 = 18^{\circ}$ for the first memory stimulus, $\sigma_2 = 12^{\circ}$ for the second memory stimulus, and the mean and standard deviation for the threshold value were 0.24 and 0.12, respectively. In Experiment 2, values that gave a fairly good fit were $\sigma_1 = 14^{\circ}$, $\sigma_2 = 18^{\circ}$, and the mean and standard deviation for the threshold value were 0.28 and 0.12, respectively. With these values, the model predicted 87.7% *old* responses when the test image was *old* in the mixed image conditions (observed = 87.5%) and 35.4% and 16.6% *old* responses when the test image was *new* trials when the near stimulus was the first and second presented, respectively (observed = 38.4% and 16.1%). It is also worth noting that, unlike for Experiment 1, the model predicted no lag effect for *old* trials. Thus, although the model correctly predicted that the lag effect was smaller in Experiment 2, it predicted too small a lag effect in both experiments.

Model 1's underprediction of the lag effect is due to the oversimplified "decay function" we chose above. It accounted for increasing decay by increasing the parameter σ in the function $\exp\{-[(X - \mu)/\sigma]^2/2\pi\}$, which can account for the effects of decay for all values of similarity except those for the identical study image and test image: There, the value of the exponent is merely equal to zero, regardless of the σ parameter, because $X - \mu = 0$. In fact, the smaller lag effect predicted above was not due to decay of the study image that matched the test image; instead, it was "right for the wrong reason." That is, the smaller lag effect was due to the fact that the second study image sends a clearer signal of "mismatch," which lowers the probability of responding "old."

This problem for the different-viewpoints conditions with Model 1 was fixable with one minor change. Because we did not have a principled way of adjusting the height of the curve at the mode, we tried what seemed like the simplest alternative: The function for the second memory stimulus was as before, but we used an adjusted value for the first memory stimulus for Model 2 ($K * \exp\{-[(X - \mu)/\sigma]^2/2\pi\}$, where K would be a number slightly less than 1). Indeed, this worked quite well for $K = .93$ (see Table 3). In Experiment 1, for the *same* test conditions, Model 2 predicted 87.2% and 77.3% *old* responses for the short and long lag conditions (observed = 92.3% and 78.9%), and in Experiment 2, the analogous predictions were 90.3% and 86.1% (observed = 89.2% and 85.8%). This improved fit was at no expense to the fit for the *different* conditions in Experiment 1. The adjusted model predicted 60.4% *old* responses in the interpolation condition (observed = 58.0%) and pre-

Table 3
Observed and Predicted Values for Percent *Old* Responses for the Different Study Images Conditions (Experiments 1 and 2)

Study and Test Conditions	Percent *Old* Responses Observed	Percent *Old* Responses Predicted by Model 1	Percent *Old* Responses Predicted by Model 2
Experiment 1			
Correct response *old*			
First image same as test image	78.9	84.3	77.3
Second image same as test image	92.3	89.9	87.2
Correct response *new*			
Extrapolation: First image near test image	34.4	29.5	34.5
Extrapolation: Second image near test image	17.1	6.8	14.4
Interpolation	58.0	61.4	60.4
Experiment 2			
Correct response *old*			
First image same as test image	85.8	87.7	86.1
Second image same as test image	89.2	87.7	90.3
Correct response *new* (average of extrapolation and interpolation)			
First scene close to test image	38.4	35.4	40.9
Second scene close to test image	16.1	16.6	15.8

Note—See text for descriptions of Integrated Decision Models 1 and 2.

dicted false alarm rates of 14.4% and 34.5% in the short- and long-lag extrapolation conditions, respectively (observed = 17.1% and 34.4%); in Experiment 2, it predicted 15.8% and 40.9% *old* responses in the short and long lag conditions, respectively (observed = 16.1% and 38.4%). In this modeling exercise, the standard deviations of the two memory distributions were kept as before (i.e., 12º and 18º), and the values for the mean and standard deviation of the threshold criterion were adjusted only slightly from before (0.26 and 0.15 in Experiment 1, 0.23 and 0.13 in Experiment 2).

We should make clear that we view our modeling work here as exploratory; we made no serious attempts to achieve "best fits," and it is possible that there is a tweak of the parameters that could give an even better fit than the model tested. However, it is not clear whether there is a reason to explore whether this is so. Among other things, there are many free parameters, so that it may seem like an empty exercise to get a perfect fit of the data. However, we adopted the original form of our "degree of match" function to keep things as simple as possible. Our modified version added one more free parameter and thus is a bit less parsimonious, but the qualitative way in which we allowed the parameters to vary made intuitive sense. What seems more to the point, however, is that a careful fitting exercise would be of great value only when one had a more principled way to define parameters and the form of the function describing the similarity of two viewpoints. As it is, we think the main point that is unarguable is that such an integrated decision model provides a quite adequate explanation of our data.

GENERAL DISCUSSION

In the present study, we investigated how different views of a scene are represented and retrieved by the visual system. In Experiment 1, we found that, when asked to discriminate between old and new views of a scene, participants were much more likely to mistake a test image that was an interpolated view between the two study images than they were a test image that was an extrapolated view (i.e., as close to one of the study images as the interpolated view, but not from a view that was "between" those of the two study images). Experiment 2 confirmed that this effect was not due to uncontrolled aspects of the test stimuli and indicated that it was also not due to the memory representation of the second image interfering with access to the first. Contrast 1 indicated that these data cannot be accounted for in the independent decision model that posits (1) that independent study images of each study image are stored and (2) that the decision about whether the test image is *old* or *new* is made by independent decisions about the match between the test image and each study image. In contrast, our modeling efforts indicated that an integrated decision model fit the data reasonably well (with the exception of the lag effects). We emphasize that the integrated decision model posits that the decision about whether the test stimulus is *old* or *new* depends on a summation of two "strengths" representing the degree of match between the test stimulus and each of the two memory representations.

The broader question is whether the major premise of the integrated decision model seems reasonable: that the memory representations of views of scenes invariably are independent and their contributions are combined only at the decision stage. At one extreme, the answer is likely to be "no." That is, if the successive views were shown in close enough sequence so that there was apparent motion, it seems quite unlikely that the views would be stored as separate memory items. Studies with faces (Wallis & Bülthoff, 2001) and objects (Kourtzi, Erb, Grodd, & Bülthoff, 2003) have shown that when static images are shown in succession (as to imply motion through different viewpoints), the different viewpoints are arguably integrated into a single representation. Wallis and Bülthoff investigated whether temporal contiguity allows the visual system to link together separate views of the same object. Participants were presented with five views of a face (from left profile through to right profile) sequentially and were asked to perform the difficult task of matching profile and frontal views of faces during a testing session. Using morphed images as the "in-between" faces, Wallis and Bülthoff presented two different faces in each sequence (one at each profile view and one at a frontal view). They found that participants were more likely to associate faces that were presented within the same sequence than when they were presented between sequences. They concluded that the temporal proximity of the faces as they changed over the views was used by the visual system to associate the different views into a single object identity. In another study, Kourtzi et al. used an fMRI adaptation method to examine whether the implied 3-D structure from a 2-D image of an object produced the reduced activation levels typical of similar stimuli in adaptation studies. They observed adaptation when the two stimuli in a trial were images of the same object rotated 30º in depth. These results imply that, with quick succession of presentation, the visual system assumes that the same object is present, despite appreciable changes in viewpoint.

At the other extreme, if there were days between the two views, it would seem quite likely that they would be stored separately. However, it is clear from this experiment (and from other experiments that use a variety of other stimuli, such as novel objects and faces) that the information presented from multiple views is integrated during some phase of processing and that the integration (regardless of whether it happens at encoding or retrieval) has important implications for recognizing new views of a familiar stimulus and possibly for recognizing new tokens within the same category. In a recent study, Friedman and Waller (2008) found that the relative position of playground equipment could be interpolated after extensive training on two studied views. Thus, it seems that, even with long-term memory, the integration of views can occur and can aid in recognition. We now turn to the consideration of how integration of various experiences of stimuli may link processing across recognition, generalizability, and categorization.

As mentioned in the introduction, there has been considerable debate in the object recognition literature about how objects are represented over their various views. One conjecture to emerge from this debate has been that the system seems to be based on representations of novel stimuli that are viewpoint specific and that, with increased experience, viewpoint-invariant performance is observed. The simplest version of this conjecture was an image-based view that relied on comparing pictorial information of the percept with the internal representation (e.g., Tarr & Pinker, 1989; Ullman, 1989). More recent image-based theories have moved away from this simple instantiation but have maintained the framework of objects' views being represented in separate representations (Peissig & Tarr, 2007). As mentioned above, this notion is also found in the updated recognition-by-components account of Biederman and Gerhardstein (1993, 1995): Disparate views that result in increased error and RTs are accounted for by different structural descriptions of the same object that reflect when parts disappear or appear with changes in viewpoint.

It is clear from the data in the present experiments that, in the implementation of this framework, independence of representations cannot necessarily be extended to independence of processing at retrieval, at least for judgments of whether there is an exact match between the test stimulus and the images in memory. This raises the question of whether this conclusion (nonindependence of decisions) is also likely to apply to the typical object recognition test (i.e., whether a test image matches any study image, ignoring differences in viewpoint) used in the object recognition literature. Clearly, if one used the same interpolation and extrapolation stimuli that we employed in a typical recognition test, one would expect the analogous result: that the probability of *old* responses would be substantially higher in the interpolation condition than in the extrapolation conditions. (Note, however, that in this task, these would be correct rather than incorrect responses.) Moreover, even the independent decision model assumed here—responding "new" only if none of the comparisons of the test image and study image representations yield a "match" outcome—would predict such an outcome.

However, it seems that the visual memory system would naturally try to take advantage of the various experiences with a stimulus by integrating this information in order to make a judgment. It is interesting that participants did not adopt an independent decision rule in the viewpoint recognition task of the present study (for which it may be optimal). This suggests that an integrated decision process is generally what is used in recognizing scenes, and it spills over into this more specialized task as well, thereby, among other things, making the "new interpolation" views seem as if they are *old*.

Taken in conjunction with earlier studies, our data suggest that the mechanisms involved in the recognition of viewpoints and identity for isolated objects also apply to scenes. If one thinks of scenes merely as larger objects, this is not surprising. However, given the much greater complexity of scenes and the much looser spatial relationships among the component objects in scenes than those among components of objects, this similarity might seem surprising. Regardless of how scene representations are conceptualized, these data suggest that, even with an assumed instance-based representational system, how the system integrates this information must be considered in addition to how the system represents this information.

We close by discussing two related issues: the form of the representation of these images in memory and whether the integration of information from the study images occurs only at retrieval or occurs at the level of the scene representation as well. Our description of the images in terms of differing in the viewpoint angle tacitly assumes that the representation of the scene is "visual" and also that the gradient of similarity is related ordinally to the angular disparity between the two views. There are clearly conceptualizations consistent with this. Perhaps the most concrete way is to assume that each scene view is a collection of "visual features," such as visible surfaces, visible objects, and occlusion relationships among objects. As the viewpoint changes, these features change, but the closer that two viewpoints are to one another, the more the features that the images have in common. Moreover, it is not unreasonable to think that two images that are as much as 40º apart share some common features, are thus somewhat confusable with each other, and add "resonance" to the sum that is the central construct assumed in the integrated decision model. Hock and Schmelzkopf (1980) looked at the effect of translation view changes (the camera was centered, and view was changed by panning the camera around) and found consistent results: The overlap in visual features predicted how well a scene was recognized as being the same one across these views. Thus, this study fits the notion that, as long as there is some overlap between the visual features (or layout) of the test and the memory that can be added to the resonance sum, a scene can be recognized reliably from a new viewpoint.

Of course, this may not be the only way that viewpoint differences may be represented. First, there can be large differences between a scene viewed from similar viewpoints, such as when a very significant object is occluded in one viewpoint but is visible enough from another viewpoint to be easily recognizable (akin to a unique feature in object recognition). (We tried to avoid such instances in our stimuli.) Second, it is almost certain that participants engage in some verbal recoding of these visual images, although it is less certain that this verbal recoding plays a significant role when the images are not visible for extended periods of time. Third, it is possible that a representation in a different format (perhaps containing more geometric information than is generally implied by a featural account) is underlying the judgments.

This brings us to the related issue of whether the information from the two study images is integrated only at the response stage or in memory as well (at the level of the representation). We should make clear that we are not claiming that our results rule out such integration in memory; however, our modeling indicates that nothing in our data requires one to make an assumption of integration in the memory store. Indeed, as stated in the introduction, even if we were to apply a structural description as the scene representations (the more integrative of the two object representation theories discussed above), separate representation for each of the views (incrementing by at least 20º) would most likely be necessary. Admittedly, we have little idea of how to test this assumption. However, until data appear that make such an assumption necessary, we think that the integrated decision model is the most parsimonious in two senses. First, it requires one only to posit a single and simple integration mechanism. Second, it does not require a detailed theory of what the memory representations of the images are. In contrast, a theory that posits integration of the study images would need to explicate in detail what the representations are and how these representations are integrated and/or interact.

In sum, we think that our data are clear in rejecting a model of image recognition memory in which study images are stored and compared with the test image independently. In contrast, our data indicate that a model in which (1) images are stored independently, (2) their representations of viewpoint get "fuzzier" over even short memory intervals, and (3) an index of their similarity to the test image is summed in the decision process handles the observed data quite well. It remains to be seen, however, whether this model can also explain situations with more study images or greater delay intervals. It also would require a "deeper" theory to describe accurately the similarity space of the scene images.

AUTHOR NOTE

This work was supported by Grant HD26765 from the National Institutes of Health, by a grant from Microsoft to K.R., and by a grant from the Natural Sciences and Engineering Research Council of Canada to M.S.C. We thank Allison Rex for help with data collection and Neil Macmillan for his helpful comments during our attempts to model the data. We also thank three anonymous reviewers and Daryl Wilson for helpful comments. Address correspondence to M. S. Castelhano, Department of Psychology, 353 Humphrey Hall, Queen's University, Kingston, ON, K7L 3N6 Canada (e-mail: monica.castelhano@queensu.ca).

Note—Accepted by the previous editorial team, when Thomas H. Carr was Editor.

REFERENCES

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

Biederman, I., & Gerhardstein, P. C. (1993). Recognizing depth-rotated objects: Evidence and conditions for three-dimensional viewpoint invariance. *Journal of Experimental Psychology: Human Perception & Performance*, **23**, 1162-1182.

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

Brainerd, C. J., & Reyna, V. F. (2001). Fuzzy-trace theory: Dual processes in memory, reasoning, and cognitive neuroscience. *Advances in Child Development & Behavior*, **28**, 41-100.

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

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

Christou, C., & Bülthoff, H. H. (1999). View dependence in scene recognition after active learning. *Memory & Cognition*, **27**, 996-1007.

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

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

Friedman, A., & Waller, D. (2008). View combination in scene recognition. *Memory & Cognition*, **36**, 467-478.

Garsoffky, B., Huff, M., & Schwan, S. (2007). Changing viewpoints during dynamic events. *Perception*, **36**, 366-374.

Garsoffky, B., Schwan, S., & Hesse, F. W. (2002). Viewpoint dependency in the recognition of dynamic scenes. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **28**, 1035-1050.

Henderson, J. M. (2007). Regarding scenes. *Current Directions in Psychological Science*, **16**, 219-222.

Henderson, J. M., & Hollingworth, A. (1999). High-level scene perception. *Annual Review of Psychology*, **50**, 243-271.

Hintzman, D. (1986). "Schema abstraction" in a multiple-trace memory model. *Psychological Review*, **93**, 411-428.

Hock, H. S., & Schmelzkopf, K. F. (1980). The abstraction of sche-

matic representations from photographs of real-world scenes. *Memory & Cognition*, **8**, 543-554.
HUFF, M., SCHWAN, S., & GARSOFFKY, B. (2007). The spatial representation of dynamic scenes—An integrative approach. In T. Barkowsky, M. Knauff, G. Ligozat, & D. R. Montello (Eds.), *Spatial cognition: Reasoning, action, interaction* (Lecture Notes in Artificial Intelligence, No. 4387, pp. 140-155). Berlin: Springer.
HUMPHREY, G. K., & KHAN, S. C. (1992). Recognizing novel views of three-dimensional objects. *Canadian Journal of Psychology*, **46**, 170-190.
KOURTZI, Z., ERB, M., GRODD, W., & BÜLTHOFF, H. H. (2003). Representation of the perceived 3-D object shape in the human lateral occipital complex. *Cerebral Cortex*, **13**, 911-920.
MARR, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information.* San Francisco: Freeman.
MARR, D., & NISHIHARA, H. K. (1978). Visual information-processing: Artificial intelligence and sensorium of sight. *Technology Review*, **81**, 28-49.
NAKATANI, C., POLLATSEK, A., & JOHNSON, S. H. (2002). Viewpoint-dependent recognition of scenes. *Quarterly Journal of Experimental Psychology*, **55A**, 115-119.
PEISSIG, J. J., & TARR, M. J. (2007). Visual object recognition: Do we know more now than we did 20 years ago? *Annual Review of Psychology*, **58**, 75-96.
SHELTON, A. L., & MCNAMARA, T. P. (1997). Multiple views of spatial memory. *Psychonomic Bulletin & Review*, **4**, 102-106.
TARR, M. J. (1995). Rotating objects to recognize them: A case study on the role of viewpoint dependency in the recognition of three-dimensional objects. *Psychonomic Bulletin & Review*, **2**, 55-82.
TARR, M. J., BÜLTHOFF, H. H., ZABINSKI, M., & BLANZ, V. (1997). To what extent do unique parts influence recognition across changes in viewpoint? *Psychological Science*, **8**, 282-289.
TARR, M. J., & PINKER, S. (1989). Mental rotation and orientation dependence in shape recognition. *Cognitive Psychology*, **21**, 233-282.
TARR, M. J., & PINKER, S. (1990). When does human object recognition use a viewer-centered reference frame? *Psychological Science*, **1**, 253-256.
ULLMAN, S. (1989). Aligning pictorial descriptions: An approach to object recognition. *Cognition*, **32**, 193-254.
ULLMAN, S., & BASRI, R. (1991). Recognition by linear combinations of models. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, **13**, 992-1006.
WALLIS, G., & BÜLTHOFF, H. H. (2001). Effects of temporal association on recognition memory. *Proceedings of the National Academy of Sciences*, **98**, 4800-4804.
ZACKS, J. M., TVERSKY, B., & IYER, G. (2001). Perceiving, remembering, and communicating structure in events. *Journal of Experimental Psychology: General*, **130**, 29-58.

NOTES

1. Here we refer to viewpoint changes that are at least 20º apart. For an example of this type of change, see Figure 1. We are aware that this may not be the case with smaller incremental changes; however, it is unclear at this point where the limit would be, and this question is outside the scope of the present study.

2. The mean value of this contrast (.157) is slightly different from the value computed from the means in Table 1 ($.124 = .420 - .656 \times .829$) because the computation of the means is nonlinear and the average of this computation participant by participant is not necessarily the computation on the averages.

3. This function is an adjusted version of the normal curve. The "adjustment" is in the denominator of the exponent, which is 2π instead of $2\pi\sigma$. This ensures that the modal value of the function does not change drastically as the standard deviation changes. However, it has the unfortunate property that the modal value does not change at all. (This is rectified in Model 2.)

(Manuscript received March 12, 2007;
revision accepted for publication November 11, 2008.)