Perception, 1999, volume 28, pages 601 – 615

DOI:10.1068/p2924

Encoding 'regular' and 'random' sequences of views of novel three-dimensional objects

Karin L Harman, G Keith Humphrey¶
Department of Psychology, University of Western Ontario, London, Ontario N6A 5C2, Canada;
e-mail: keith@julian.uwo.ca
Received 10 July 1998, in revised form 28 January 1999

Abstract. When we look at an object as we move or the object moves, our visual system is presented with a sequence of different views of the object. It has been suggested that such regular temporal sequences of views of objects contain information that can aid in the process of representing and recognising objects. We examined whether seeing a series of perspective views of objects in sequence led to more efficient recognition than seeing the same views of objects but presented in a random order. Participants studied images of 20 novel three-dimensional objects rotating in depth under one of two study conditions. In one study condition, participants viewed an ordered sequence of views of objects that was assumed to mimic important aspects of how we normally encounter objects. In the other study condition, participants were presented the same object views, but in a random order. It was expected that studying a regular sequence of views would lead to more efficient recognition than studying a random presentation of object views. Although subsequent recognition accuracy was equal for the two groups, differences in reaction time between the two study groups resulted. Specifically, the random study group responded reliably faster than the sequence study group. Some possible encoding differences between the two groups are discussed.

1 Introduction

A persistent problem in understanding object recognition is how we are able to recognise objects despite changes in their perspective projection from one encounter to the next. Such shape constancy indicates that our representations of objects are at least somewhat viewpoint-invariant. Some theorists have suggested that viewpoint invariance is established through various geometrical computations that are carried out by the visual system when presented with an object. While there is evidence that the visual system performs such computations (eg Biederman 1995; for review see Jolicoeur and Humphrey 1998; Wallis and Bülthoff 1999), Miyashita (1993) and Stryker (1991) have proposed another way that invariant representations might be formed. This proposal is based on results of recent neurophysiological research by Miyashita and colleagues (eg Sakai and Miyashita 1991; for review see Miyashita 1993). Using single-cell recording techniques, they have found that cells in the inferotemporal cortex of the temporal lobe of monkeys form associations between visual stimuli based on the temporal contiguity of the stimuli. Miyashita (1993) and Stryker (1991) have proposed that a mechanism in the temporal cortex that associates successively presented stimuli may play a fundamental role in constructing view-invariant representations of objects. In our everyday visual ecology, different views of an object are nearly always encountered in succession, transformed by object or observer movement. If the visual system contains an associational mechanism, as suggested by Miyashita's research, then successive views of an object would automatically become associated. Such associations could lead to the formation of representations of objects that are relatively independent of orientation.

Several neural-network simulations have used temporal associations in modelling the development of view-invariant representations of objects (Bartlett and Sejnowski 1998; Becker 1997; Edelman and Weinshall 1991; Foldiak 1998; Parga and Rolls 1998; Wallis 1998; Wallis and Baddeley 1997; Wallis et al 1993). For example, Wallis and colleagues

¶ Author to whom all correspondence and requests for reprints should be addressed.

(Wallis and Baddeley 1997; Wallis et al 1993) constructed a model of cortical visual processing with the purpose of simulating processes that occur in object recognition. Their network incorporated a layer of 'cells' that were sensitive to the weighted sums of previous neuronal activity to associate information that had been provided over time. These researchers trained the network with sequences of views of faces rotated in depth, and found that, when two views of the same face were presented in close temporal succession, the model associated the two views together. In this way, a cell in the network generalised its response to two views of the same face, leading to viewpoint-independent 'recognition'.

Other models, such as that of Becker (1997), have also stressed the importance of temporal associations when constructing a network that recognised faces. Becker implemented a hierarchical network where clustering units at a lower level coded temporal information about the input. At a higher layer, gating units used this temporal information to associate views that occurred in close temporal succession. After learning a series of views of faces rotated in depth, the clustering units became specialised for detecting specific views of faces, whereas the gating units became specialised for detecting particular facial features over a limited set of views. Thus, temporal contiguity may have been coded at a lower visual area that then fed into a processing layer that coded structural similarity information. At some point, these two properties of the input may have been combined, resulting in the ability to integrate images together that 'belonged' to the same object, while maintaining the ability to distinguish between other images that represented different objects. The applicability of such neural networks to the understanding of human object recognition is limited, however, unless the resultant data are confirmed by behavioural investigations.

To mimic the way that we normally encode objects, Lawson et al (1994) required participants to study depth plane rotations of objects that occurred either in a regular sequence, or in a random order. Their stimuli were line drawings of familiar objects that were presented rapidly either in a sequence of views or in a series of random views. Higher naming accuracy was found for the objects that were studied in a sequence of views than for the objects that were studied via a set of random views (Lawson et al 1994). These findings supported the notion that the accuracy of identification of familiar objects benefited from studying a sequence of views that could be temporally *and* structurally integrated.

A related study conducted by Wallis (1996) pitted temporal integration against structural integration in face recognition. If the key to constructing a stable object representation under everyday viewing depends upon integrating views of objects that are temporally contiguous, then different objects presented in close temporal succession should also be associated. Associating views of different faces may result, therefore, in a decrease in recognition accuracy for these faces. In Wallis's study, participants viewed each face at several orientations in depth but the views were of different faces and thus were not highly structurally similar. The test task required that participants decide whether two faces were the same or different in three scenarios. The same face may have been shown in different orientations, two different faces from the same rotation sequence may have been shown, or two different faces from two different sequences could have been shown. Results indicated that participants confused different faces from the same sequence more than they confused different faces shown from different sequences. These findings suggested that the views of faces were associated together on the basis of their close occurrence in time, and this association in turn affected participants' subsequent recognition ability.

The present research was motivated by the recent proposals, mentioned above, that regular temporal sequences of images of objects contain information that aids in the process of representing and recognising objects. In particular, we examined whether seeing a series of perspective views of objects in sequence leads to more efficient recognition than seeing the same views of objects but presented in a random order. Furthermore, to investigate how encoding affected representation *construction* we studied the recognition of novel three-dimensional (3-D) objects. In the previous behavioural research outlined above, familiar objects have been used (Lawson et al 1994) or faces (Wallis 1996) to investigate object recognition under sequential viewing conditions. Studying familiar objects may have encouraged the access of stored contextual information such as labels, semantic associations, etc that may have affected recognition performance. On the basis of recent research and theorising, it was expected that seeing coherent temporal sequences of views of novel objects would result in more efficient recognition than viewing random sequences.

2 Experiment 1

In experiment 1 we examined recognition in three groups of participants who studied a set of object views in different orders. For each object, a set of seven views was studied either in a regular sequence of views, in a random order of views, or in an order where views and objects were randomised. If studying object views in a sequence was similar to the way that we normally encounter objects, then this type of presentation may have resulted in the construction of a robust stored representation. A representation created under such viewing conditions may be accessed more efficiently upon subsequent presentation of the object than representations that were constructed by studying a random sequence of object views. On the basis of such notions, participants who studied the object views in a sequence should demonstrate better recognition performance than would participants who were presented object views in a random order.

2.1 *Method*

2.1.1 *Participants.* Thirty-six students (twenty-one females and fifteen males) from the University of Western Ontario volunteered to participate and were paid for their time. All participants had normal or corrected-to-normal acuity and were naive to the experimental design and to the objects used in the experiment. Age of participants ranged from 18 to 27 years, with a median age of 20 years.

2.1.2 *Materials.* The stimuli used in all the reported experiments were 40 grey-scale computer images of novel three-dimensional clay objects that had been used in previous research (Humphrey and Khan 1992). All objects had a main axis of elongation and 'geon'-like (Biederman 1987) parts that were attached to a central body (see figure 1 for examples). Nine views of each object were obtained with a CCD video camera interfaced to an Apple Macintosh microcomputer. The object images were recorded from a 20° angle of elevation. The CCD camera was equipped with a zoom lens (8 – 48 mm) and was 122 cm from the objects. The focal length of the lens was set at 35 mm. At this setting the images were minified relative to the real objects by 10%. No further scaling or reprocessing of the images was done.

Each object view was separated by a 20° rotation about the vertical axis. The views included a 50°, 70°, 90°, 110°, 130°, 150°, 170°, 190°, and 210° rotation. The axis of elongation of the 90° rotation was parallel to the line of sight of the participant, and was therefore a foreshortened view. The 70° view was a slightly foreshortened three-quarter view. The 50° and 130° rotations approximated three-quarter views and were nearly mirror images of one another. The 150° and 190° rotations were side views of the object (see figure 2 for examples). The 50° and the 210° views were used only in the test session of experiment 2 (see figure 2 for examples).

The images were viewed from a distance of 60 cm. The visual angle of the images varied depending on the particular view. For the views in which the long axis of the object was perpendicular to the line of sight, the mean visual angle was 7.4 deg for the

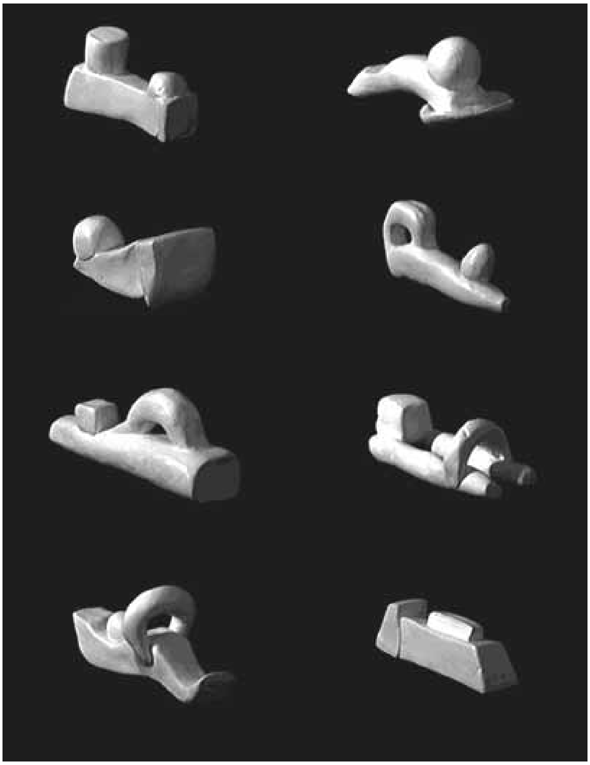

Figure 1. Examples of the novel objects used in the current study.

X dimension and 3.3 deg for the *Y* dimension. For images in which the axis of elongation of the object was parallel to the line of sight, the mean visual angle was 2.8 deg for the *X* dimension and 4.4 deg for the *Y* dimension.

2.1.3 *Design.* Study condition was a between-participants variable and test angle was a within-participant variable. Participants were randomly assigned to one of three study conditions and these groups differed in terms of the order in which the views of the objects were presented. The study groups included: same object/sequence of rotations (SS); same object/random rotations (SR); or random object/random rotations (RR). The SS group viewed the seven orientations of a given object in a sequenced order, that is 70°, 90°, 110°, 130°, 150°, 170°, and 190°. The SR group viewed the seven orientations of a single object in a row, but these orientations were randomised. For example, a participant may see a sequence, such as the 90°, 130°, 50°, 110°, 170°, 70°, 150° views of an object, followed by a different sequence of views for the next object and so on. Within each of these two study conditions, the objects were presented in a random order. The RR group was presented the views and objects in random order, so that object A at 130° may have been seen first, then object K at 90° was seen next, then object D at 110° and so on (see figure 2). Twelve participants were included in each group. The within-participant condition was test angle, and recognition performance was measured for the 70°, 130°, and 190° rotations.

2.1.4 *Procedure*

Study session: Participants were seated in a darkened room and a chin-rest with forehead and lateral head stops was used to restrict head movements. After preliminary instructions and a practice session, the experimenter initiated the study session. Participants were instructed to look at each image as it appeared on the screen and try to remember it, and were told that they would subsequently be tested on their recognition of the objects.

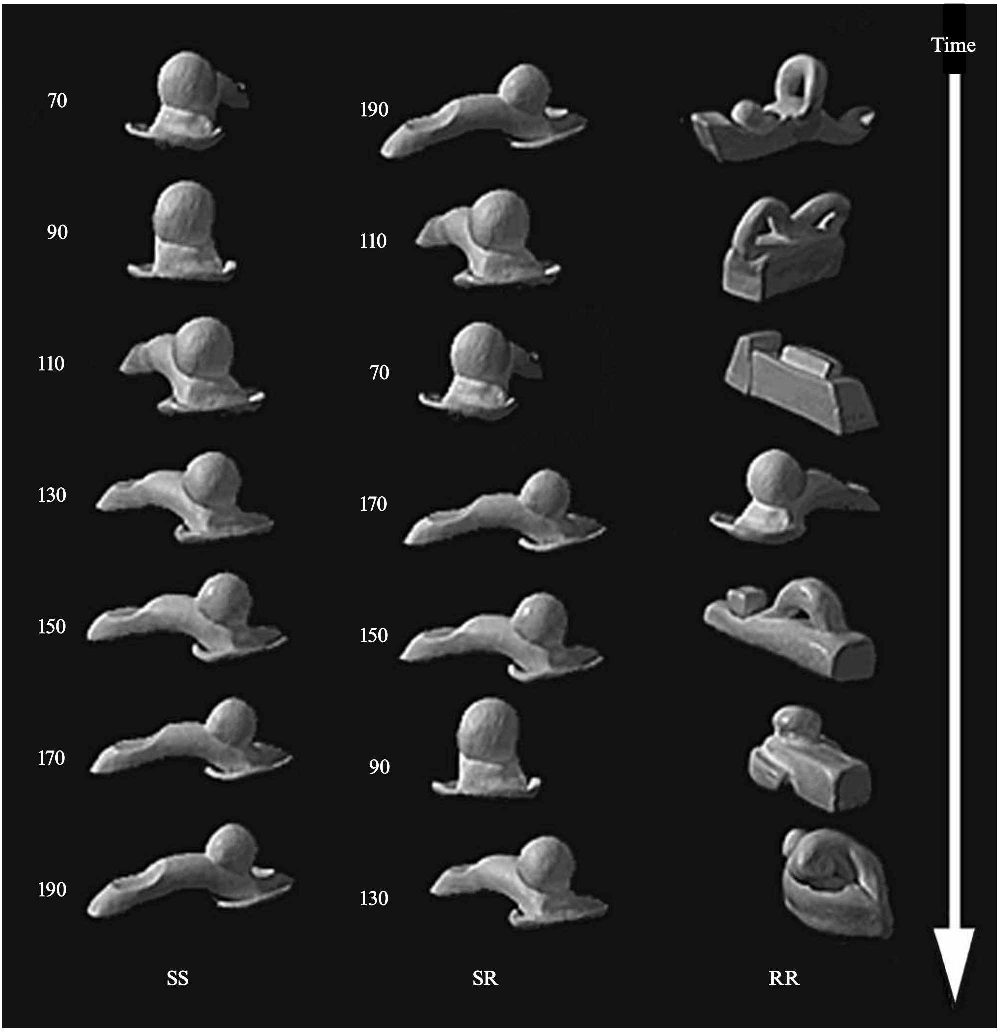

Figure 2. Examples of the three study conditions of experiment 1. Left: regular sequence of views (SS); middle: same-object random-views condition (SR); right: random-objects random-views condition (RR).

The participants then viewed the seven views of all 20 objects. The order in which they saw the views depended on the study condition to which they were assigned. The presentation of the 20 objects was repeated three times so that the participants saw each view of each object three times throughout the study session. Thus, the study session was composed of 420 images (20 objects × 7 views of each object × 3 repetitions). Within each set of presentations, the order of presentation of the 20 objects was randomised. Each study image was presented for 1 s with a 750 ms interstimulus interval (ISI). This ISI did not result in the perception of apparent motion and was used so that the effects of temporal sequencing of object views could be investigated without the possible influence of apparent motion in this experiment. Over the entire study session, each view of each object was presented for a total of 3 s.

Test session: Each test trial was composed of a 1000 ms fixation cross, followed by a 100 ms blank screen and then presentation of a test image. Upon appearance of the test image, participants were required to press keys on a keyboard to indicate whether they had studied the particular object shown, or whether they had not studied the object (an old/new decision). The test image was displayed until the participant made a response. Reaction time and accuracy were recorded by a Macintosh IIci microcomputer. After the participant's response, an interval of 500 ms was followed by the next fixation cross, signaling the next trial. This procedure continued until the participant responded to the 20 old objects and the 20 new objects at the three test orientations (70°, 130°, and 190°) (see figure 3).

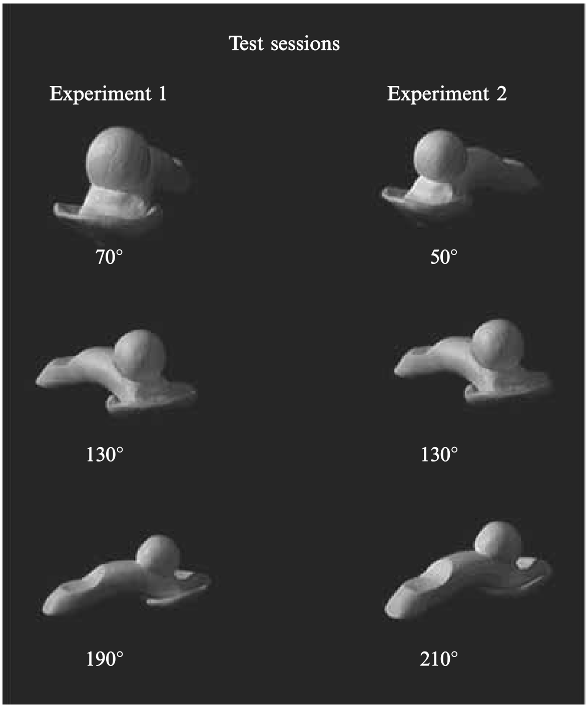

Figure 3. Examples of views used in the test sessions of experiments 1 (left) and 2 (right).

2.2 *Results*

Four separate mixed-model ANOVAs, one for reaction time and one for accuracy of old-object decisions and one for RT and one for accuracy of new object decisions, were run on the resultant data. Study condition (SS, SR, and RR) was a between-participant variable and test angle (70°, 130°, and 190°) was a within-participant variable. In this and all of the following experiments, reaction times for any subject that were greater than 3 standard deviations of their overall mean were removed from the analyses. This procedure led to removal of less than 1% of the trials in all of the experiments.

2.2.1 *Old-object accuracy.* The statistical analysis did not reveal a main effect of study group. Whether a participant studied the object views in a sequence or randomly did not affect recognition accuracy. Testing angle, however, did have a significant effect on accuracy ($F_{2,33} = 9.2, p < 0.0005$). The 130° view was responded to most accurately, followed by the 190°, and then the 70° views (figure 4). No other effects were statistically significant.

2.2.2 *Old-object response times.* The analysis of the response times, unlike that of the accuracy data, revealed a significant main effect of study condition ($F_{2,33} = 4.7$, $p < 0.05$). Interestingly, the group that viewed the object images in a completely random order (RR) responded significantly faster to the test objects than the group that studied the images in a sequenced order (SS) (Neumann – Keuls a posteriori test, $p < 0.05$). The performance of the SR group did not differ significantly from that of the other two groups (see figure 5).

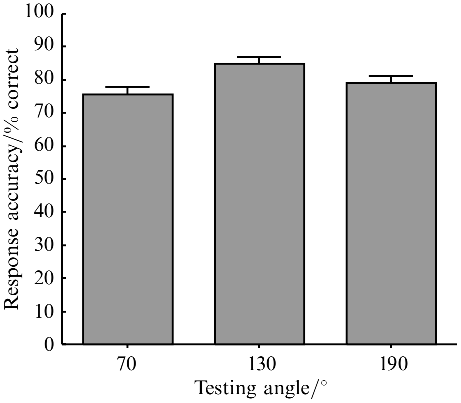

Figure 4. Experiment 1. Main effect of test view on response accuracy to old objects. The error bars indicate, in this and all following figures, +1 standard error of the mean.

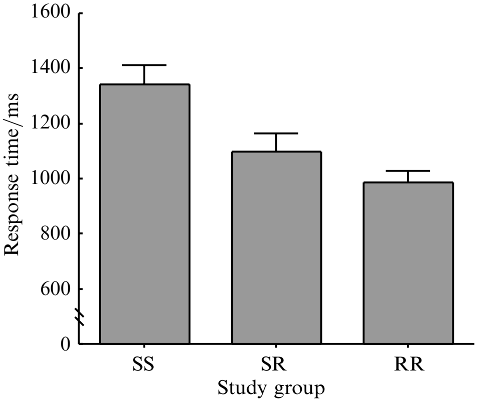

Figure 5. Experiment 1. Main effect of study group on response times to old objects.

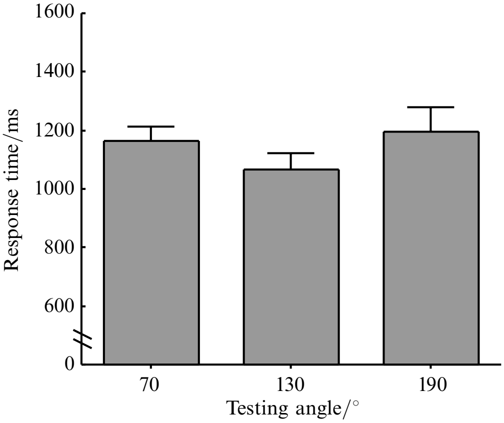

Figure 6. Experiment 1. Main effect of test view on response times to old objects.

In addition, a trend toward a main effect of testing angle was revealed ($F_{2,33} = 2.7$, $p < 0.08$). The 130° view was responded to faster than the other two angles (figure 6). No interactions were statistically significant.

2.2.3 *New-object accuracy.* There were no main effects or interactions as a result of the analyses on accuracy for the new-object decisions. The SS group responded to the new objects with 85% accuracy (SE = 3.6%), the SR group responded with 84% accuracy (SE = 3.4%), and the RR group responded with 82% accuracy (SE = 5.0%).

2.2.4 *New-object response times.* There were no main effects or interactions as a result of the analyses on response times for the new-object decisions ($M_{\mathrm{SS}} = 1665.4$ ms, $\mathrm{SE}_{\mathrm{SS}} = 178.3$ ms; $M_{\mathrm{SR}} = 1469.4$ ms, $\mathrm{SE}_{\mathrm{SR}} = 156.2$ ms; $M_{\mathrm{RR}} = 1445.2$ ms, $\mathrm{SE}_{\mathrm{RR}} = 106.3$ ms).

2.3 *Discussion*

There are two interesting results from this study. First, and most surprisingly, we found that participants who were presented with a regular sequence of views of novel objects during training took *longer* to recognise a subsequent presentation of the objects than did subjects who were trained with a random sequence of object views. The accuracy of participant's responses, however, was not affected by the study condition. The reaction-time difference in these two study conditions was not the expected outcome.

We suggested, on the basis of recent research and theorising, that studying a sequence of views would lead to more efficient recognition than studying a random sequence of views. The results, however, clearly show that, at least under the conditions of the present experiment, viewing coherent temporal sequences during learning did not enhance object recognition.

The second result of interest was the finding that participants responded more efficiently to the 130° view than to the other test orientations. The finding of a recognition benefit for the 130° view in our experiment may have been a result of this view sharing more features with all the other studied views than did the 70° or 190° views. Perhaps the 130° view could be considered to be a 'canonical' (Palmer et al 1981), or 'nonaccidental' view (Biederman 1987; Lowe 1985), given that a minimal number of features were occluded. This result supported other findings of increased accuracy for canonical views of objects (eg Humphrey and Jolicoeur 1993; Palmer et al 1981; for review see Jolicoeur and Humphrey 1998). This result is also consistent with results of other studies of canonical views of novel objects (eg Bülthoff and Edelman 1992; Cutzu and Edelman 1994; Humphrey and Khan 1992).

Although we did not find that viewing a regular sequence of views of objects led to better recognition performance than viewing objects in a random sequence, it is possible that a recognition benefit from viewing a regular sequence would be apparent when generalisation of learning was tested. For example, the representation that was constructed during encoding a regular sequence of views may not have been accessed quickly, but it may have been required for recognising novel views of the study objects. Because previous investigations (eg Lawson et al 1994; Wallis 1996) found that studying a regular sequence of views resulted in higher accuracy than studying a random sequence of views, perhaps an accuracy difference between our two study groups would occur if recognition were tested with novel views of these study objects. This possibility was investigated in experiment 2.

3 Experiment 2

To determine whether studying an ordered sequence of views of an object led to better generalisation of learning than studying a random sequence of views, we tested how well participants could recognise *novel* views of the study objects. Perhaps the representation constructed in the SS study condition was more view-independent than the representation that was constructed in the RR study condition. To this end, we tested participants' recognition of novel views of objects, which were views that fell either in the middle of the study sequence or fell outside of the rotations presented in the study session. Presumably, generalisation of learning would be enhanced if view-independent representations were stored. If the SS group recognised the novel views of the study objects with greater efficiency than the RR group, then this would suggest that studying a sequence of views leads to a more viewpoint-independent representation. That is, it is possible that, when participants study a sequence of views, view-dependent representations would be linked and provide the input to a view-independent representation. However, object views that were studied in a random order may favour the construction of a set of discrete, view-dependent representations owing to the lack of structural and temporal associations between study views.

3.1 *Method*

3.1.1 *Participants.* Participants were forty-two volunteers (twenty-six females and sixteen males) from the University of Western Ontario who were given course credit for their participation. None had participated in experiment 1 and all participants had normal or corrected-to-normal acuity. Ages ranged from 18 to 28 years, with a median age of 19 years.

3.1.2 *Materials.* The stimuli used in the present experiment were the same as in experiment 1. In the present study, however, only six object views were studied: 70°, 90°, 110°, 150°, 170°, and 190°. The test views were a 20° rotation about the vertical axis from the closest studied views. Two of these test views fell outside the angles of the viewing rotation (50° and 210°) and one test view was an orientation that was in the middle of the study views (130°) (see figure 3).

3.1.3 *Design and procedure.* In experiment 2, the number of study rotations was reduced from seven to six. In addition, the SR group was not included in the study in view of the finding in experiment 1 that the results from this group did not differ from the results of either the SS or the RR group. The experimental procedure was the same as in experiment 1.

3.2 *Results*

Four mixed-model ANOVAs were run on the resultant data with study condition (SS or RR) as a between-participants variable and test angle (50°, 130°, and 210°) as a within-participants variable. Old-object and new-object decisions were analysed separately.

3.2.1 *Old-object accuracy.* Accuracy data revealed no main effect of study group. However, there was a main effect of testing angle ($F_{2,40} = 11.40$, $p < 0.0001$). Neuman–Keuls a posteriori tests indicated that recognition accuracy differed significantly ($p < 0.01$) for all three angles. The 130° angle was recognised most accurately, then the 210° view, followed by the 50° view (figure 7). Therefore, even when the 130° view was not studied, it was recognised more accurately than the other two angles.

3.2.2 *Old-object response times.* The ANOVA that was run on the response-time data indicated a main effect of study group ($F_{1,40} = 4.28$, $p < 0.05$) and a trend towards a main effect of angle ($F_{2,40} = 2.61$, $p < 0.08$). The main effect of study group was due to the significantly faster response times of the RR group than those of the SS group (figure 8). There was not a statistically reliable interaction between study group and test angle. The trend towards a main effect of viewing angle was due to the slightly faster response time to the 210° rotation (975 ms) than to the 130° (990 ms) and 50° (1005 ms) views.

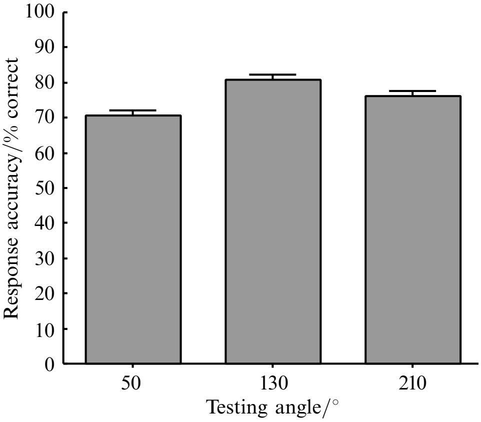

Figure 7. Experiment 2. Main effect of test view on response accuracy to old objects.

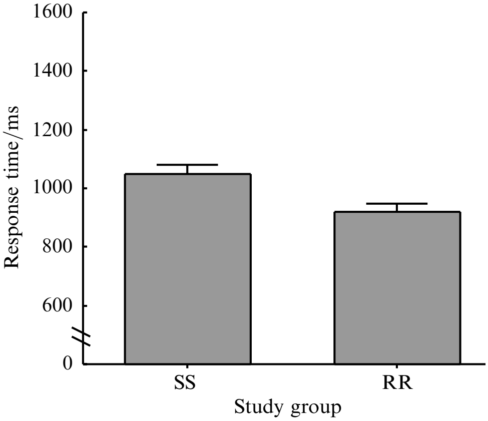

Figure 8. Experiment 2. Main effect of study group on response times to old objects.

3.2.3 *New-object accuracy.* There were no significant differences in the accuracy with which the two study groups responded to the test objects ($M_{\mathrm{SS}} = 79.6\%$, $\mathrm{SE}_{\mathrm{SS}} = 1.9\%$; $M_{\mathrm{RR}} = 81.8\%$, $\mathrm{SE}_{\mathrm{RR}} = 1.9\%$). In addition, no interactions between conditions were statistically reliable.

3.2.4 *New-object response times.* Results indicated that there was a significant main effect of study group on reaction time to new objects ($F_{1,40} = 4.36$, $p < 0.04$). This finding was a result of the RR group responding to the new objects significantly faster ($M_{\mathrm{RR}} = 1032$ ms, $\mathrm{SE}_{\mathrm{RR}} = 38.9$ ms) than the SS group ($M_{\mathrm{SS}} = 1203$ ms, $\mathrm{SE}_{\mathrm{SS}} = 41.2$ ms). No other main effects or interactions were statistically significant.

3.3 *Discussion*

The main effect of study group on the reaction time in experiment 2 replicated the results from experiment 1, and demonstrated once again that studying a regular sequence of views of objects *slowed* object recognition relative to a study condition in which views of objects were presented in a random order. Unlike experiment 1, there was no suggestion that the 130° view was recognised faster than the other two test views. Thus, although this view still benefited from more accurate recognition, it was not recognised any faster than the other test views.

If a view-independent representation were constructed as a result of studying a regular sequence of views, then novel views of objects would have been recognised more efficiently by the SS group. This expectation was not supported by the results. It should be noted, however, that the novel test views used in this experiment were not greatly different from studied views (20° difference). Therefore, conclusive interpretations about the relative view dependence of the representations may not be justified.

4 Experiment 3

The results of experiments 1 and 2 indicate that studying views of objects in a regular sequence slowed the speed of recognition of novel objects relative to their recognition after studying objects and views in a random order. It is possible, however, that studying a sequence of views of an object would only facilitate subsequent recognition relative to random study when movement cues were included during the encoding process. The notion that motion cues may have led to more efficient recognition than encoding images without motion cues was investigated by Hill et al (1997). They tested participants' recognition of a regularly rotating sequence of views of faces that were seen with smooth apparent motion, or a random sequence of views that were perceived without smooth apparent motion. Results demonstrated that recognition accuracy was better for the animated sequences than for the unanimated sequences, implying that structure-from-motion information aided face recognition (Hill et al 1997). On the basis of such results it is possible that for recognition to benefit from studying a sequence of views, information from motion would be required. In our previous study sessions, the interstimulus interval and the stimulus duration were both long enough to ensure that no apparent motion between views was perceived. Thus, the motion cues that we may encounter in a natural setting were not present.

The importance of motion in constructing object representations was investigated more directly in another study. Kourtzi and Schiffrar (1997) found, using a priming task, that apparent motion enhanced the recognition of new-object orientations in the picture plane. Relative to a condition in which two static views of objects were presented without apparent motion, the same two views presented with apparent motion led to enhanced recognition of new-object orientations when those orientations fell within the path of apparent motion.

In experiment 3, we altered the stimulus duration and interstimulus interval between study views during both the sequenced and the random conditions. Participants in the sequence study condition viewed a series of images that rotated about the vertical axis. Participants in the random condition viewed images without smooth apparent motion owing to the structural dissimilarity of successive images. If perception of apparent motion aided in the construction of a representation that could be accessed efficiently, reaction time performance in the SS group might be faster than the reaction-time performance of the RR group.

4.1 *Method*

4.1.1 *Participants.* Participants were thirty volunteers (eighteen females and twelve males) from the University of Western Ontario who were paid for their participation. None had participated in previous experiments and all had normal or corrected-to-normal acuity. Ages ranged from 18 to 32 years, with a median age of 21 years.

4.1.2 *Materials.* Stimuli used were the same as in experiment 1. The testing angles used were the 70°, 130°, and 190° views. All three test views were seen during study.

4.1.3 *Design and procedure.* As in experiment 2, only the SS and RR study groups were included in experiment 3. The study and test instructions and procedures were identical to those in the previous experiments. Stimulus duration was set at 750 ms and the ISI was set at 50 ms. The stimulus duration and ISI were altered to lead to a percept of motion during the sequenced-rotations study session. There were four blocks of study trials instead of three blocks in the previous experiments. The number of blocks was increased to four to ensure that the total viewing time per object image would be the same for all of the experiments (3 s per image).

4.2 *Results*

Four mixed-model ANOVAs were run on the resultant data, one on the accuracy data and one on the RT data for both old and new decisions.

4.2.1 *Old-decision accuracy.* The accuracy data revealed a main effect of testing angle ($F_{2,28} = 8.8$, $p < 0.0005$). As in experiment 1, the 130° angle was recognised more accurately (80.6%) than the other two angles (70°, 70.6%; 190°, 75.1%). No other effects were significant.

4.2.2 *Old-decision reaction time.* There was a main effect of testing angle ($F_{2,28} = 8.10$, $p < 0.005$). The 130° angle was responded to significantly faster (1068.4 ms) than the other two angles (1146.9 ms and 1166.2 ms). Unlike experiments 1 and 2, there was no main effect of study group ($p > 0.05$). There was, however, a significant study group × testing angle interaction ($F_{2,28} = 6.36$, $p < 0.05$). Simple effects analyses indicated that this effect was due to the faster response of the random group at the 190° angle ($p < 0.05$). The two groups responded with equal speed to the other two angles (figure 9).

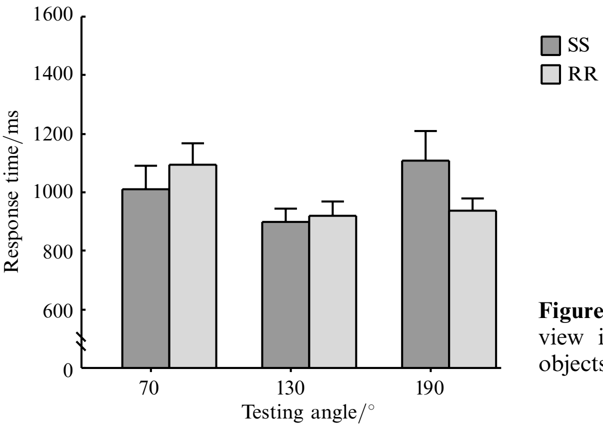

Figure 9. Experiment 3. Study group × test view interaction on response times to old objects.

4.2.3 *New-object accuracy.* There were no main effects of our variables nor interactions between variables on accuracy ($M_{\mathrm{SS}} = 81.3\%$, $\mathrm{SE}_{\mathrm{SS}} = 2.0\%$; $M_{\mathrm{RR}} = 75.3\%$, $\mathrm{SE}_{\mathrm{RR}} = 2.3\%$).

4.2.4 *New-object response times.* There were no main effects of our variables nor interactions between variables on response time ($M_{\mathrm{SS}} = 1126.1$ ms, $\mathrm{SE}_{\mathrm{SS}} = 44.7$ ms; $M_{\mathrm{RR}} = 1230.2$ ms, $\mathrm{SE}_{\mathrm{RR}} = 63.6$ ms).

4.2.5 *Comparison of experiments 1 and 3.* To investigate whether recognition reaction time for old objects was changed for the SS group as a result of apparent-motion cues, we compared the results from experiments 1 and 3. A mixed-model ANOVA [4 groups (SS1, RR1, SS2, RR2), and 3 test angles] was run on the RT data from experiments 1 and 3 and revealed a main effect of study group ($F_{3,51} = 5.807$, $p < 0.002$). An a posteriori (Neuman – Keuls, $p < 0.05$) analysis demonstrated that the SS group from experiment 1 differed significantly from all other groups in their reaction times to test objects. That is, the SS group from experiment 1 responded significantly slower to test objects than the other three test groups. The other three groups did not differ significantly from one another.

Not surprisingly, there was also a significant effect of test angle ($F_{3,51} = 7.107$, $p < 0.001$) as the 130° view was responded to faster than the other two test views.

4.3 *Discussion*

Results from experiment 3 were consistent with the results from experiments 1 and 2 with regard to the accuracy with which the participants responded to the testing angles. That is, the 130° angle was responded to more accurately by both groups, and thus reinforced the notion of a recognition benefit for a canonical view or nonaccidental view over the other test views.

Unlike experiments 1 and 2, the SS and RR groups in the present experiment responded with equal speed to the 70° and 130° angles. The only difference in reaction time between the two groups was at the 190° angle, and here, as in experiments 1 and 2, the SS group responded more slowly than the RR group. Thus, adding apparent motion to the study session of the sequence group allowed these participants to respond as fast as the random group to two of the test angles. This finding was reinforced by our comparison of the data from experiments 1 and 3—apparent motion did decrease the recognition latency in the SS group relative to the latency of the SS group in experiment 1. Thus, as in Kourtzi and Schiffrar's (1997) study, presentation of the regular image sequence with apparent motion produced a recognition benefit relative to the presentation of the same image sequence without motion cues. It could be that this benefit depended on information contained in structure-from-motion, although another possible interpretation concerning the potential reasons that apparent motion facilitated performance for the SS group will be outlined in the general discussion. It is important to note, however, that we still did not observe the expected *advantage* of studying the objects in a sequence of rotations in the SS group relative to the RR group in the present experiment. That is, performance in the SS group was not significantly more efficient than the performance of the RR group.

5 General discussion

The present research was motivated by recent proposals suggesting that regular temporal sequences of views of objects contain information that can aid in the process of representing and recognising objects. In particular, we examined whether seeing a series of perspective views of objects in sequence led to more efficient recognition than seeing the same views of objects presented in a random order. The results of the experiments were not as we expected on the basis of present research and theorising. The results of experiments 1 and 2, and to some extent experiment 3, all show that, rather than enhancing object recognition, viewing a coherent temporal sequence of images of novel objects during training actually led to less efficient recognition, as indicated by decision latency, than did viewing a random sequence of object images.

One possible reason for the slower decision latency in the SS condition is that seeing a coherent sequence of object views leads to less attention or effort during object encoding than does seeing a constantly changing set of object views. The SS condition involved the consecutive presentation of similar images of each object. This may have resulted in less effort or attention being expended in learning about each view of the objects as many of the views of each object were highly similar. Consequently, the images may not have been well encoded and the longer decision latency could have been a reflection of this relatively poor encoding. In contrast, each successive view in the RR condition was highly dissimilar and this could have resulted in more attention being devoted to the encoding of each view.

This interpretation of the latency difference in the SS and RR conditions is similar to the deficient-processing explanations for the so-called 'spacing effect' in word list learning. That is, items in a word list that are repeated in a massed fashion during study are recognised with lower accuracy than repeated items that are distributed with intervening items throughout a list. It has been suggested that the spacing effect occurs because less attention is given to repeated items that are grouped together than when they are spaced apart (cf Greene 1992). Of course, it is clear that any decrease in effort or attention on the part of the SS group was not accompanied by a decrease in accuracy in any of the experiments. This is not the usual finding in word list learning, as naming accuracy is typically affected by massed versus distributed repetitions (Greene 1992). The accuracy measure in the present experiments may not have been highly sensitive, however, given that the test images were displayed until the subjects responded.

The more similar decision latency performance of the SS and RR conditions in experiment 3 than in experiment 1 could be the result of more effortful or attentive encoding in the SS condition in experiment 3. Such an explanation suggests that it is not information contained in apparent motion per se that improved performance in the SS condition in experiment 3. Rather, the apparent motion and the shorter exposure duration of each image in experiment 3 may have been more successful in maintaining attention in the SS condition.

Another related suggestion for the relatively long decision latencies in the SS condition also depends on the similarity of image structure in the successive views of the objects seen in this condition. Tarr and Gauthier (1998) suggested that views of objects rotated in depth may be stored better when there is significant qualitative change in the image structure of the views. In their experiments, they found that views of objects that were similar to trained views were stored relatively poorly compared to views that were not as similar to the trained view. Tarr and Gauthier proposed that the greater the qualitative dissimilarity between one view and another view of an object, the greater the likelihood that both views will be represented in visual memory. This suggestion corresponds in some respects to the distinctiveness hypothesis of encoding, in which it has been proposed that items that are distinctive from one another in a study session will be remembered better than items within a session that are more similar to one another (eg Eysenck 1979). In the present experiments one could generalise such an explanation to the encoding of successive views of objects. It may be that the SS condition promoted the encoding of relatively few views because of the high similarity among many of the successive images. In contrast, the RR condition may have promoted the encoding of many views because of the successive, large changes in image structure from one image to the next.

The suggestions we have offered to account for the longer decision latencies in the SS condition than in the RR condition are certainly speculative and tentative at this point. If further investigations bear such suggestions out, then artificial neural networks that attempt to simulate human performance after training with coherent and random sequences of object views may need to include some factors that modulate encoding. For example, if the longer decision latency in the SS condition results from less attention being devoted to each view of an object during encoding than is devoted in the RR condition, then neural-network models may need to incorporate an attentional-gain parameter to account for the present results and perhaps more generally to model human performance in such tasks. Such a parameter could be implemented as a decreasing learning rate that depends on the spatial redundancies in consecutive images. Further behavioural and neural-network research is needed on the role of different temporal sequences of views of objects in the representation and recognition of objects.

We will end by noting that although the explanations for the poorer performance in the SS condition depend on the participants' recognition of the *similarities* among the successive views of each object, we have not said anything about the nature of the similarity or how it is computed. It is possible, and perhaps likely, that the participants in the SS condition know that they are seeing the same object from different perspectives and it is this knowledge that leads to the decrease in attention and consequently to relatively poor encoding. One can ask then, how do the participants in the SS condition 'know' that they are seeing the same object in different orientations? Is the description of the image based on relatively low-level measurements of the input, or on a higher-level encoding such as a structural description. One possibility is that, with the relatively heterogeneous 'geon'-like objects as used here (see figure 1), they are forming some sort of structural description that is the same for many of the views—that is, it shows some reasonable degree of viewpoint invariance. If indeed this is the case, then our results could be seen to be consistent with structural description approaches to object recognition (eg Biederman 1987; for review see Jolicoeur and Humphrey 1998; Wallis and Bülthoff 1999). In general such approaches do not emphasise the role or importance of temporal correlations in the input because somewhat viewpoint-invariant descriptions can be computed that are based on static views of objects. Of course, similarity may not be computed on the basis of high-level descriptions and there is bound to be overlap in many types of measurements made on successive images in the SS condition. Whatever the nature of the similarity, our suggestion is that this factor may result in a decrease in attention when humans are shown successive images of objects.

Acknowledgements. This research was supported by a postgraduate scholarship from the Natural Sciences and Engineering Research Council of Canada (NSERC) to K L Harman and by a NSERC research grant to G K Humphrey. Portions of the research were presented at the 1998 Annual Meeting of the Association for Research in Vision and Ophthalmology and at the 1998 Annual Meeting of the Canadian Society for Brain, Behaviour, and Cognitive Science. Thanks to two anonymous reviewers whose comments and criticisms of an earlier draft of this paper were most helpful. Thanks also to Tom James and Ken McRae for the time they spent discussing the issues raised by this research.

References

Bartlett M S, Sejnowski T J, 1998 "Learning viewpoint-invariant face representations from visual experience in an attractor network" *Network: Computation in Neural Systems* **9** 399 – 417

Becker S, 1997 "Learning temporally persistent hierarchical representations", in *Advances in Neural Information Processing Systems 9* Eds M Mozer, M Jordan, T Petsche (Cambridge, MA: MIT Press) pp 824 – 830

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

Biederman I, 1995 "Visual object recognition", in *An Invitation to Cognitive Science: Visual Cognition* Eds S M Kosslyn, D N Osherson (Cambridge, MA: MIT Press) pp 121 – 165

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 of the USA* **89** 60 – 64

Cutzu F, Edelman S, 1994 "Canonical views in object representation and recognition" *Vision Research* **34** 3037 – 3056

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

Eysenck M W, 1979 "Depth, distinctiveness, and elaboration", in *Levels of Processing: An Approach to Memory* Eds L Cermak, F I M Craik (Hillsdale, NJ: Lawrence Erlbaum Associates) pp 89 – 118

Foldiak P, 1998 "Learning constancies for object perception", in *Perceptual Constancy: Why Things Look as They Do* Eds V Walsh, J Kulikowski (Cambridge: Cambridge University Press) pp 144 – 172

Greene R L, 1992 *Human Memory: Paradigms and Paradoxes* (Hillsdale, NJ: Lawrence Erlbaum Associates) pp 145 – 151

Hill H, Schyns P G, Akamatsu S, 1997 "Information and viewpoint dependence in face recognition" *Cognition* **62** 201 – 222

Humphrey G K, Jolicoeur P, 1993 "An examination of the effects of axis foreshortening, monocular depth cues and visual field on object identification" *Quarterly Journal of Experimental Psychology A* **46** 137 – 159

Humphrey G K, Khan S C, 1992 "Recognizing novel views of 3D objects" *Canadian Journal of Psychology* **46** 170 – 190

Kourtzi Z, Schiffrar M, 1997 "One-shot invariance in a moving world" *Psychological Science* **8** 461 – 466

Jolicoeur P, Humphrey G K, 1998 "Perception of rotated two-dimensional and three-dimensional objects and visual shapes", in *Perceptual Constancy: Why Things Look as They Do* Eds V Walsh, J Kulikowski (Cambridge: Cambridge University Press) pp 69 – 123

Lawson R, Humphreys G W, Watson D G, 1994 "Object recognition under sequential viewing conditions: evidence for viewpoint-specific recognition procedures" *Perception* **23** 595 – 614

Lowe D G, 1985 *Perceptual Organization and Visual Recognition* (Hingham, MA: Kluwer Academic)

Miyashita Y, 1993 "Inferior temporal cortex: where visual perception meets memory" *Annual Review of Neuroscience* **16** 245 – 263

Palmer S E, Rosch E, Chase P, 1981 "Canonical perspective and the perception of objects", in *Attention & Performance* Eds J Long, A Baddeley (Hillsdale, NJ: Lawrence Erlbaum Associates) pp 135 – 151

Parga N, Rolls E, 1998 "Transform-invariant recognition by association in a recurrent network" *Neural Computation* **10** 1507 – 1525

Sakai K, Miyashita Y, 1991 "Neural organization for the long-term memory of paired associates" *Nature (London)* **354** 152 – 154

Stryker M P, 1991 "Temporal associations" *Nature (London)* **354** 108 – 109

Tarr M J, Gauthier I, 1998 "Do view-dependent mechanisms generalize across members of a class?" *Cognition* **67** 73 – 110

Wallis G, 1996 "Presentation order affects human object recognition learning", in Technical Report No. 36, Max-Planck-Institut für biologische Kybernetik, Tübingen, Germany, pp 1 – 8

Wallis G, 1998 "Spatio-temporal influences at the neural level of object recognition" *Network: Computation in Neural Systems* **9** 265 – 278

Wallis G, Baddeley R, 1997 "Optimal, unsupervised learning in invariant object recognition" *Neural Computation* **9** 883 – 894

Wallis G, Bülthoff H, 1999 "Learning to recognize objects" *Trends in Cognitive Science* **3** 22 – 31

Wallis G, Rolls E, Foldiak P R, 1993 "Learning invariant responses to the natural transformations of objects" *Proceedings of the International Joint Conference on Neural Networks* volume 2 (New York: IEEE) pp 1087 – 1090