
Evidence for complete translational and reflectional invariance in visual object priming

Irving Biederman, Eric E Cooper

Department of Psychology, Elliott Hall, University of Minnesota, Minneapolis, MN 55455, USA

Received 22 March 1990, in revised form 27 August 1990

Abstract. The magnitude of priming on naming reaction times and on the error rates, resulting from the perception of a briefly presented picture of an object approximately 7 min before the primed object, was found to be independent of whether the primed object was originally viewed in the same hemifield, left–right or upper–lower, or in the same left–right orientation. Performance for same-name, different-exemplar images was worse than for identical images, indicating that not only was there priming from block one to block two, but that some of the priming was visual, rather than purely verbal or conceptual. These results provide evidence for complete translational and reflectional invariance in the representation of objects for purposes of visual recognition. Explicit recognition memory for position and orientation was above chance, suggesting that the representation of objects for recognition is independent of the representations of the location and left–right orientation of objects in space.

1 Introduction

An object viewed on one occasion is more quickly and accurately perceived when presented on a second occasion (Bartram 1974; Schacter et al 1990). We examined whether this priming effect is dependent on the object appearing at the same location on the retina and in the same left–right orientation.

People often exhibit good explicit memory for where in their visual field an object has been projected and for its left–right orientation. Indeed, a fundamental principle of cinematography is never to reverse the camera axis. Reversing the axis would, for example, lead film viewers to think a scene showing a man on the left speaking to a woman on the right changed so that the man was on the right facing left. The principle assumes that the viewer has memory for the location and orientation of the visual entities in the scene.

One might conclude from such observations that a single representation specifying object shape, position, and orientation is activated so that if the object appeared at a different location or in a different orientation its recognition speed would not be as great as when the object was at its original position and in the original orientation (Jacoby et al 1989). That is, a change in the viewing conditions would result in less priming than if the original viewing conditions were reinstated. Some loss in recognition speed would then be expected if the object was translated or reflected from its original position. From this account, our ability to appreciate the equivalence of two images of one object when they have been seen at different times and positions and in different orientations in the visual field is achieved through the establishment of multiple representations at different viewing opportunities or through subsequent cognitive inferential processes.

In contrast to a representation specifying an integrated perceptual event is the possibility that shape, for purposes of recognition, is coded independently of position and of left–right orientation (Hummel and Biederman 1992). Translational and reflectional invariance of shape priming would then be expected. There are computational advantages in such coding since representations of the shape of an object would not

have to be duplicated for every retinal position and left-right orientation in which the object was seen. On *a priori* grounds, it would seem that both variables (retinal position and left-right orientation) are arbitrary in that a subsequent encounter with an object would rarely be expected to duplicate the original conditions.

Indirect physiological support for this dissociation between the representation of shape for recognition and memory for position derives from primate cortical ablation experiments establishing different cortical loci for memory of position and for memory of object attributes (Mishkin and Appenzeller 1987; Ungerleider and Mishkin 1982). Specifically, ablation of the inferior temporal (IT) cortex results in gross impairment of the animal's ability to respond on the basis of object identity, irrespective of position, but does not result in significant impairment in the use of a spatial cue. Ablation of the posterior parietal region results in the opposite effects. That the system involved in object recognition might also exhibit translational invariance is suggested by the extremely large receptive fields—up to 26 deg in diameter—found in the IT cortex (Gross 1973). Moreover, Gross and Mishkin (1977) showed that the IT cortex was crucial for interhemispheric transfer of pattern discrimination habits.

There is no question that people can recognize an object when it is in a different location in the visual field. But does their recognition speed suffer from the translation? We employed a priming paradigm in which subjects viewed briefly presented pictures of objects that appeared either to the left or to the right of fixation (experiments I, II, and III) or above or below fixation (experiment IV). We also assessed whether priming invariance exists for left-right reflections (experiment I). Experiments II and IV also compared the priming of identical images with same-name, different-exemplar versions to assess whether the priming was visual. Experiment III measured whether the subjects had explicit memory for position and for orientation under the experimental conditions.

Surprisingly, the efficiency of the presumed translational invariance of shape representations on object naming has received little assessment. One experiment that did examine the effects of modest shifts of object images on same-name, different-exemplar judgments was reported by Ellis et al (1989). On each trial, their subjects viewed two black-and-white photographic slides in quick succession [interstimulus interval (ISI), 60 ms]. On half the trials, those in which a 'same' response was appropriate, the second image was: (a) identical to the first, or (b) projected from another source so that it was displaced horizontally from the first, or (c) rotated in depth by 45°, or (d) a same-name, different-exemplar object. Reaction times (RTs) to translated images were only slightly slower than RTs to the identical images, indicating that the representations for these judgments were translationally invariant.

In general, if translational invariance were manifested in a same-name, different-exemplar task at short ISIs, it would be expected to occur in a longer term priming paradigm with many intervening events. However, two features about the experiment of Ellis et al leave open the possibility that shape representations might not be translationally invariant in a priming paradigm. The first is that their photographic stimuli contained surface features, such as light and dark regions, texture, etc. Such cues could have mediated same-name, different-exemplar judgments. The present experiment used line drawings to reduce the availability of surface cues. The second characteristic about the experiment of Ellis et al is that the magnitude of the translation, 2.6 deg, was small (approximately 22%) relative to the images of 11.8 deg. It was so modest, relatively, that none of their subjects noticed it. (Probably the presence of depth-rotated images, in which occupied regions of the screen were changed with otherwise identical pairs, contributed to the lack of awareness of the translation.) In the present investigation, the translation was larger than the image.

2 Experiment I

Experiment I measured whether the magnitude of priming was reduced when the object appeared in the opposite hemifield and in the opposite left-right orientation to those in which it was originally presented. On the first presentation, the images were presented for only 150 ms, a time sufficiently short to preclude the programming of a saccade to the object.

2.1 Method

2.1.1 Subjects. The subjects were thirty-two native English speakers with normal or corrected-to-normal vision. They participated for payment (\$5 per session) or for research experience points for the Introductory Psychology course at the University of Minnesota.

2.1.2 Stimuli. Each subject named forty-eight briefly presented pictures of objects on each of two blocks of trials, the first being a priming block and the second being a primed block. Each picture was a simple line drawing of a common object with a readily available basic-level name (see Biederman and Cooper 1991a, for sample pictures). The pictures were created through the program Cricket Draw and were shown on a high resolution (1024×768) monitor (Mitsubishi model HL6605) controlled by a Macintosh II computer. The images were centered 2.4 deg to the left or to the right of fixation. The maximum extent of each image could be contained in a circle whose diameter subtended a visual angle of 4 deg, so the closest possible point of any picture to fixation was 0.4 deg.

2.1.3 Procedure. The subject pressed a mouse button to start each trial. A fixation dot was then presented for 500 ms, followed by the presentation of the object picture for 150 ms (a duration too brief to make a second eye fixation). The picture was, in turn, followed by a 500 ms mask, an apparently random arrangement of straight and curved lines. Four such arrangements were used so that the mask varied at random from trial to trial.

On both blocks, the subject named each picture with the basic-level term, eg "piano", as it was shown. To decrease the likelihood of subjects using other names for the stimuli, prior to the presentation of the experimental stimuli, subjects read the names of the objects from their terminal.⁽¹⁾ They were told that these were the names of the objects that they were to see in the experiment. Prior to the experimental trials, subjects were given twelve practice trials with images that were not presented on the experimental trials. Two 'buffer' trials, with images from the practice set, were presented at the beginning and at the end of each block. The subjects had been instructed to respond as quickly and as accurately as possible.

The naming RTs were recorded through a Lafayette voice key. Reaction time and accuracy feedback were displayed after each trial. Storage and display of the image files and experimental sequences, timing, feedback, and response recording were accomplished through a single software package, Picture Perception Lab (Kohlmeyer, in press). A response was recorded as an error if it was the name for another class (but naming variants such as "car" instead of "auto" or "phone" for "telephone" were considered correct), or a false start, or if it was not made within 3 s.

2.1.4 Design. The positions (left or right) and orientations were presented in an apparently random sequential fashion, so that the subject could not accurately anticipate the position of the image. The subjects were instructed to maintain fixation

⁽¹⁾ In a number of experiments in our laboratory (Biederman 1987) we have never found this manipulation to interact with any perceptual variable nor even to have a significant effect on reaction times or on error rates.

which was undoubtedly facilitated by the natural tendency of visual capture and by the presentation of the fixation point 500 ms prior to the presentation of the picture.

On the second block a given picture could appear in either the position identical to that on the priming block or on the other side of fixation. For each position type (same or different), half the images were in their original orientation and the other half were in mirror-image reversed orientation. One quarter of the objects viewed by each subject were in each one of the four position and orientation conditions. The sequences of images were balanced across subjects so that the mean serial position of every object in every condition of position and orientation was equal, with all objects appearing equally often in the four conditions of position and orientation variation. For each object, left and right positions and orientations were also balanced across subjects. Approximately 7 min intervened between the first and second presentations of an object.

2.2 Results and discussion

The results are shown in figure 1. Mean correct-naming RTs and error rates were considerably and significantly lower on the second block than they were on the first block ($t_{31} = 8.68$ and 6.16 both for RTs and for error rates, respectively, and both values of $p < 0.001$), indicating that significant effects of priming did occur. There were no reliable effects on RTs or on error rates for the second trial block of either position or orientation changes. The error terms for the F ratios were constructed by regarding pairs of subjects who were shown the same objects in the same conditions as a single group. This error term reflects the interaction of objects and conditions and thus serves to account for variance due to object differences. For the RTs, the values of $F_{1,15}$ were 1.52 and 0.06 , and the values of p were 0.23 and 0.80 , for position and orientation, respectively. The values of $F_{1,15}$ for the error rates (with an arcsin transformation) were 0.76 and 2.88 , and the values of p were 0.39 and 0.11 , for position and orientation, respectively. The experiment was sufficiently sensitive to detect as significant a 16 ms effect on RTs and a 1.64% effect on error rates due to translation.

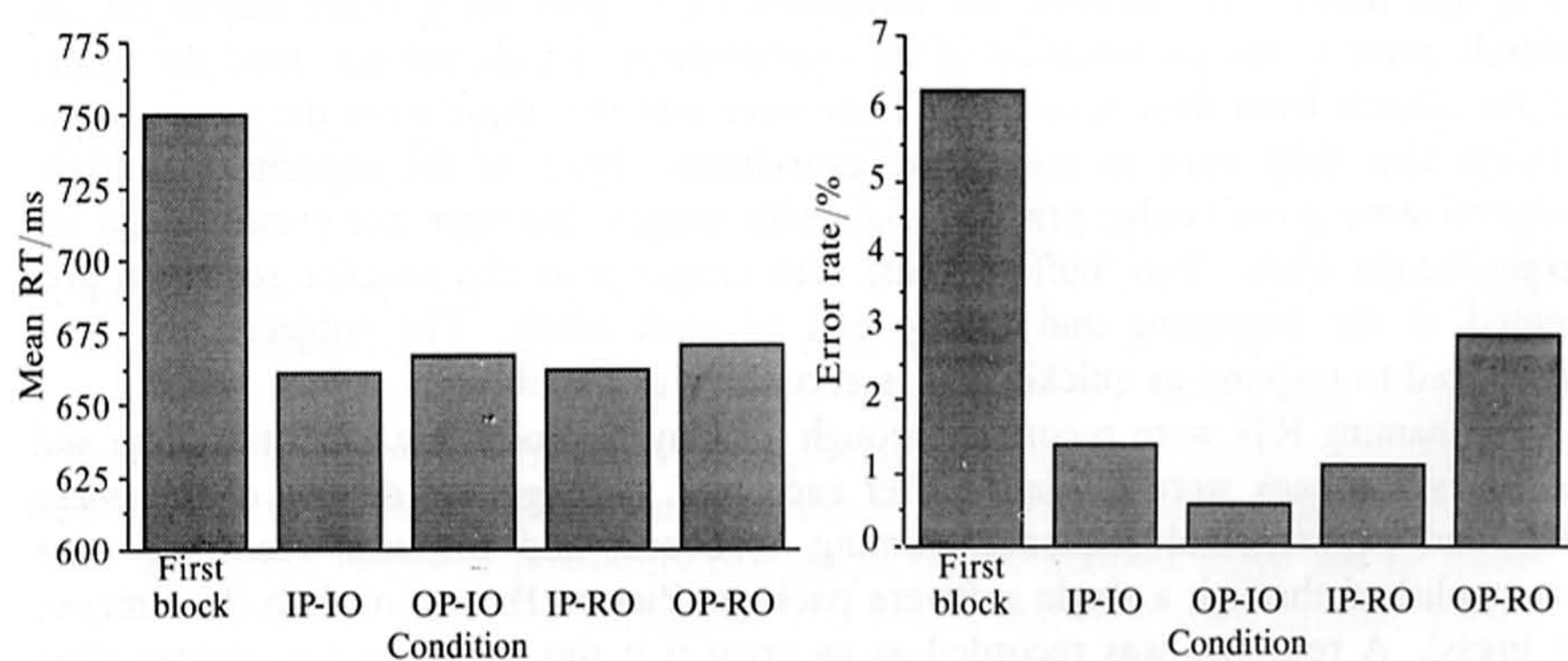


Figure 1. Mean correct-naming reaction-times (RTs) and error rates for the effects of changes in left-right position and mirror-image orientation in experiment I. The conditions for the second block trials were as follows: identical position and orientation (IP-IO); opposite position and identical orientation (OP-IO); identical position and reversed orientation (IP-RO); and opposite position and reversed orientation (OP-RO). The second block data are only for those trials where the object was correctly named on the first block. (Inclusion of those trials where the first block was in error did not alter the pattern of the results, although it did increase the variability.)

Priming with these stimuli revealed complete translation and left-right reflection invariance.⁽²⁾

3 Experiment II

It is possible that all the priming effect manifested in experiment 1 was verbal or conceptual, rather than visual. If a perceptual component of the priming were positionally dependent, but verbal and conceptual components were not, then the results of experiment I could be a consequence of the absence of a visual component to the priming. Experiment II was run to assess whether a component of the priming was indeed visual. This was done by investigating the effect of priming with identical images compared to same-name, different-exemplar images. For example, an upright piano might be displayed on the first block of trials and either that same image or a grand piano would be displayed on the second exposure. An advantage for the identical images on the second block of trials would indicate that the priming was visual.

3.1 Method

This experiment was identical in design and procedure to the first experiment, except that half the objects on the second block were different-shaped exemplars, but with the same name as the objects from the first block. This required that sixty-four objects be used, ie thirty-two pairs with the same names. For example, if an upright piano were shown on the first block, for half the subjects a grand piano would be shown on the second block, and the remaining half of the subjects would be shown the upright piano again. Included in this set were four animals (elephant, bird, dog, and rabbit), for which the different exemplars were generated by picturing different poses of the animal. Orientation variations were not run in experiment II. The exemplar variable (same versus different) occupied the same status in the design as the orientation variable in experiment I.

3.2 Results and discussion

The results are shown in figure 2. Mean correct-naming RTs and error rates were considerably lower (by 85 ms and 4.7%) on the second trial block than on the first block ($t_{31} = 8.68$ and 6.15, for RTs and error rates, respectively, $p < 0.001$ both for RTs and for error rates). Again there was no effect of translation ($F_{1,15} = 0.62$ and 0.82, for RTs and error rates, ns). The experiment was sufficiently sensitive to detect as significant a 39 ms effect on RTs and a 4.52% effect on error rates due to translation.

The different-exemplar trials had RTs that were significantly higher (by 35 ms) than those for the identical trials ($F_{1,15} = 12.13$, $p < 0.004$), establishing that a portion of the priming was indeed visual. (The F ratio for the error rates was less than 1.00.) That a 35 ms effect was highly reliable establishes that the sensitivity of these experimental designs allows detection of effects of a relatively modest magnitude. The interaction between translation and exemplar changes was not significant, both values of $F < 1.00$.

We interpret the advantage of same exemplars over different exemplars as evidence that a component of the priming, amounting to the difference between the two conditions, was indeed visual. An alternative explanation might hold that all the advantage of block two over block one for the same-exemplar condition was a consequence of nonspecific transfer (ie general practice on the task), rather than item-specific priming,

⁽²⁾ The left-right positional variation allowed an assessment of left versus right visual field differences. Performance with objects presented in the left visual field was virtually identical to performance with objects in the right visual field. This was also true in experiment II. A full description of these analyses is presented in Biederman and Cooper (1991b).

combined with an interference effect resulting from the mismatch of exemplars in the different-exemplar condition. Aside from the lack of parsimony in positing two factors, neither the posited nonspecific transfer nor the presumed interference from different-exemplar mismatch receives support from independent observations. Bartram (1974) compared naming performance for completely new objects (requiring different names) to the performance for same-name, different-exemplar objects over a series of trials. Performance with the new objects was approximately equivalent to performance on the first block, indicating an absence of nonspecific transfer in these picture-naming tasks. The putative interference effect might result in the same-name, different-exemplar objects being identified with longer latencies and higher error rates on the second block than new objects with names not used on the first block. But both Bartram's (1974) experiment and one by Biederman and Cooper (1991a) revealed the opposite effect: an advantage for same-name, different-exemplar items over new items in object-naming paradigms.

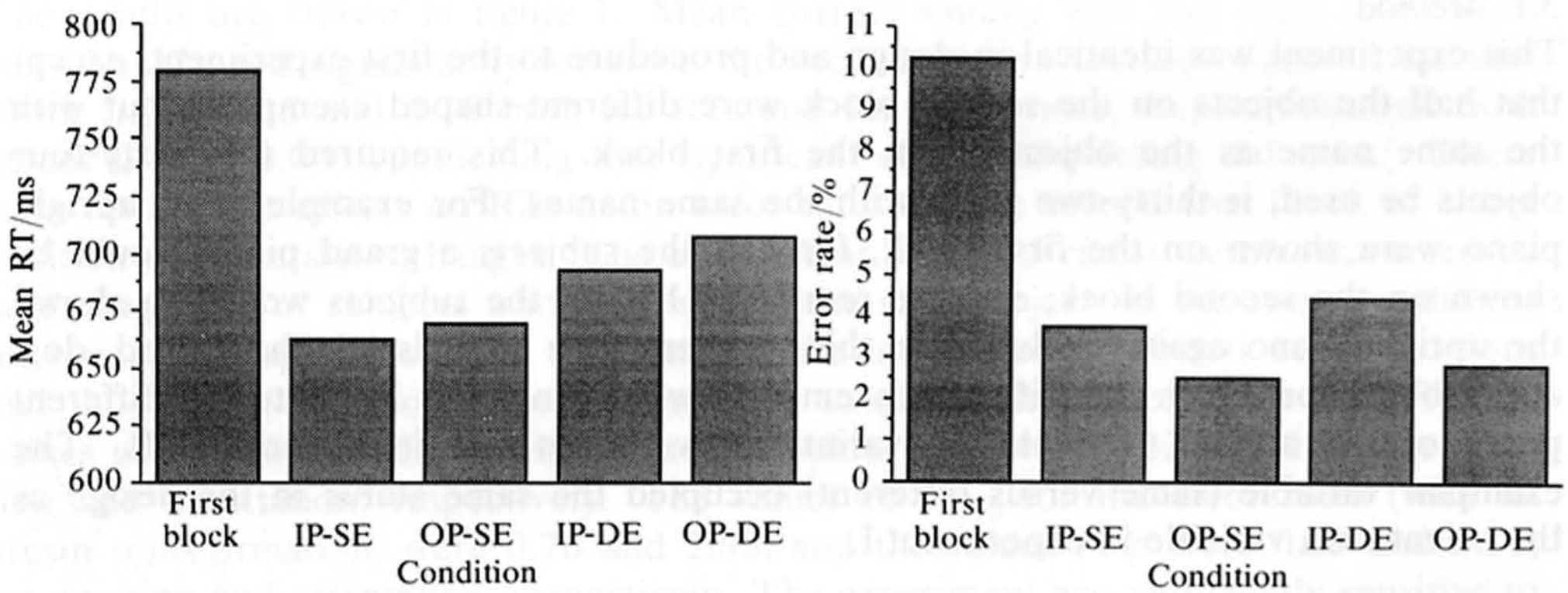


Figure 2. Mean correct-naming reaction-times (RTs) and error rates for the effects of changes in left-right position and exemplar shape in experiment II. The conditions for the second block trials were as follows: identical position, same exemplar (IP-SE); opposite position, same exemplar (OP-SE); identical position, different exemplar (IP-DE); opposite position, different exemplar (OP-DE). The second block data are only for those trials where the object was correctly named on the first block. (Inclusion of those trials where the first block was in error did not alter the pattern of the results, although it did increase the variability.)

4 Experiment III

Experiment III was designed to assess whether there was significant explicit memory for position and for orientation in experiments I and II. It is possible that the absence of an effect of changes in the position and in the orientation of objects was due to the absence of any memory of the original position and orientation of these objects.

4.1 Method

To assess explicit memory for orientation and for position, eight subjects received a first trial block identical to that in experiment I, in which they were instructed to name the pictures as rapidly as possible. These subjects had no reason to expect an explicit memory test for position and orientation on the second block. On the second block of trials, rather than naming each object, they had to judge whether the (a) position and (b) left-right orientation of the object were the same as or different from those in the first block.

4.2 Results

The accuracy of these judgments was above the 50% accuracy expected by chance. The mean percentages for correct naming were 70.0% for position and 59.6% for

orientation ($t_7 = 6.27$ and 2.28 , $p < 0.001$ and < 0.06 , respectively). The above-chance accuracy of these subjects, who had the same instructions as the priming subjects in experiments I, II, and IV, indicates that there was some explicit memory for position and orientation. Unlike the translation variation in the experiment of Ellis et al (1989), these results document that the variations in position (and in orientation) in the present experiments were sufficiently large not only to be readily noticeable, but to yield good explicit memory without subjects having been given specific instructions.

5 Experiment IV

In experiments I and II it was established that left-right translational invariance exists for naming object pictures. In the real world, variations in verticality, upper-lower, are likely to be more significant than left-right variations. Moreover, there are significant functional and anatomical differences between the lower and upper visual fields (Previc 1990). Experiment IV assessed whether there was literally top-down or bottom-up priming (or top-top or bottom-bottom priming) with a design identical to that employed in experiment II.

5.1 Method

The design and method employed was identical to that in experiment II except that the monitor was turned on its side so that its vertical extent could accommodate the top-bottom variation in the images. (Of course the images were rotated 90° from their orientation in the earlier experiments so that they were upright.)

5.2 Results

Figure 3 shows the results for experiment IV. As in the previous experiments, RTs and error rates were markedly lower (by 85 ms and 8.0%) on the second block than the first ($t_{31} = 9.58$ and 6.92 , respectively, $p < 0.001$ for both variables). There was no effect of whether the object occupied the same top-bottom position or not. Mean correct-naming RTs (and error rates) for same and for different positions were 708 ms (4.1%) and 715 ms (4.5%), respectively ($F_{1,15} < 1.00$ for both variables). The experiment was sufficiently sensitive to detect as significant a 44 ms effect on RTs and a 3.9% effect on error rates due to translation.

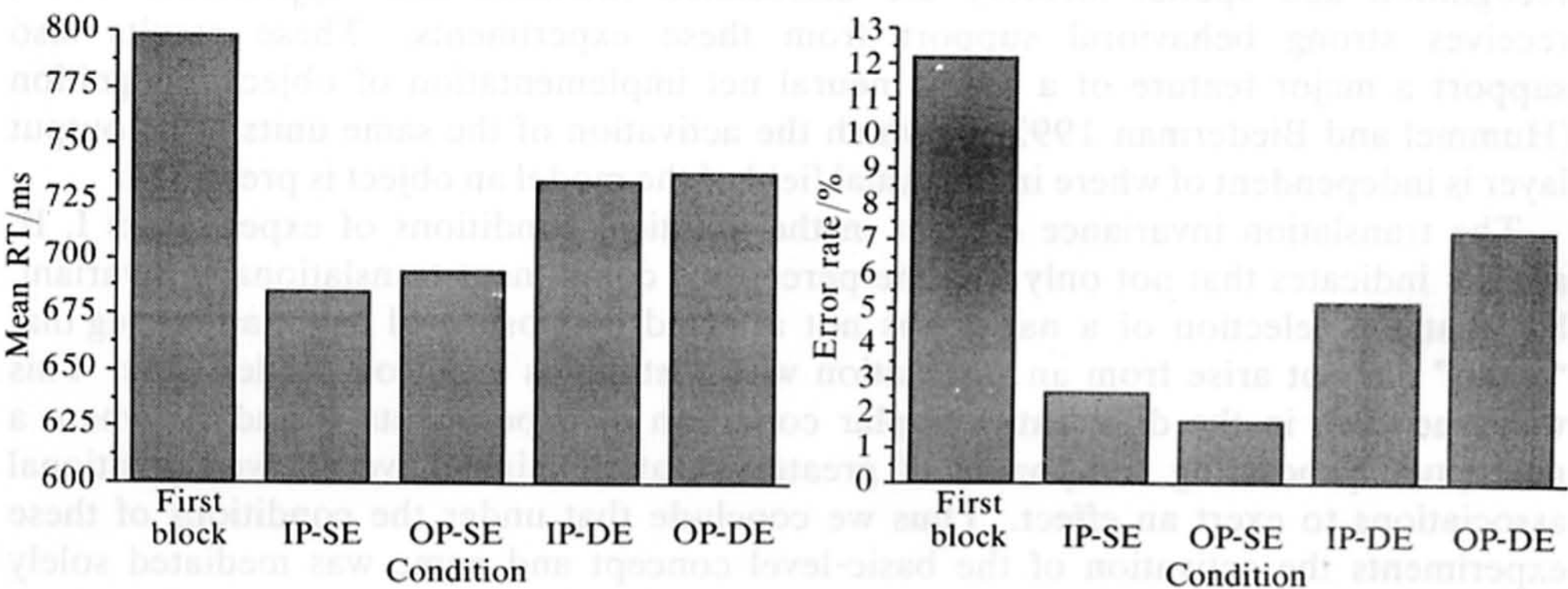


Figure 3. Mean correct-naming reaction-times (RTs) and error rates for the effects of changes in top-bottom position and exemplar shape in experiment IV. The conditions for the second block trials were as follows: identical position, same exemplar (IP-SE); opposite position, same exemplar (OP-SE); identical position, different exemplar (IP-DE); opposite position, different exemplar (OP-DE). The second block data are only for those trials where the object was correctly named on the first block. (Inclusion of those trials where the first block was in error did not alter the pattern of the results, although it did increase the variability.)

Changing the exemplars led to significantly higher RTs and error rates. Mean correct-naming RTs (and error rates) were 689 ms (2.3%) and 733 ms (6.2%) for same and different exemplars, respectively ($F_{1,15} = 13.42$ and 9.51 , $p < 0.01$ for both variables). The values of $F_{1,15}$ for the interaction between position and exemplar were 0.02 and 1.06, ns, for RTs and arcsine error rates, respectively.

6 General discussion

The results of experiments I, II, and IV document complete translation invariance in the naming of briefly presented pictures of objects. That is, viewing the object at a new position resulted in as much priming as when the object was viewed at a different position and in a different orientation. Combined over the three experiments, mean RTs for naming objects at different positions were 6 ms longer and had 1.0% lower error rates than when the objects were in the same position. Matched t -tests ($df = 95$) for the same-position minus different-position scores for all subjects in the three experiments were -0.73 and 1.17 , for RTs and arcsine error rates, respectively, ns.

The advantage of the identical condition over the same-name, different-exemplar condition in experiments II and IV indicated that a component of the priming—the large advantage of the identical condition—was visual. Experiment I also established that the priming was left-right reflectionally invariant. The above-chance accuracy of explicit recognition memory in experiment III for changes in position and in reflection indicated that the lack of an effect on naming RTs and on error rates could not be attributable to the imperceptibility of these changes or to a lack of any memory for them. This dissociation between factors that may affect explicit recognition memory and perceptual processing is a well documented characteristic of repetition priming (eg Jacoby and Dallas 1981; Schacter 1987). It should be noted that the direction of this dissociation in the present investigation differs from the one generally reported in the literature on perceptual priming in which an effect on perceptual processing is not manifested in explicit memory. We found an explicit memory effect that was not manifested in perceptual processing.

The belief, supported by neuropsychological research, that high-level object recognition and spatial memory are dissociated (Mishkin and Appenzeller 1987) receives strong behavioral support from these experiments. These results also support a major feature of a recent neural net implementation of object recognition (Hummel and Biederman 1992) in which the activation of the same units at an output layer is independent of where in the visual field of the model an object is presented.

The translation invariance evident in the identical conditions of experiments I, II, and IV indicates that not only was the perceptual component translationally invariant, but that the selection of a name was not affected by positional associations, eg that “piano” did not arise from an association with a stimulus event on the left side. This was true even in the different-exemplar condition of experiments II and IV where a perceptual processing component of greater duration might have allowed positional associations to exert an effect. Thus we conclude that under the conditions of these experiments the activation of the basic-level concept and name was mediated solely through a positionally-invariant representation of the object.

Acknowledgement. This research was supported by an AFOSR Research Grant 88-0231 to IB and an NSF Graduate Fellowship to EEC. We thank S W Kohlmeier for his assistance with programming.

References

- Bartram D, 1974 "The role of visual and semantic codes in object naming" *Cognitive Psychology* **6** 325–356
- Biederman I, 1987 "Recognition-by-components: A theory of human image understanding" *Psychological Review* **94** 115–147
- Biederman I, Cooper E E, 1991a "Priming contour-deleted images: Evidence for intermediate representations in visual object recognition" *Cognitive Psychology* **23** 393–419
- Biederman I, Cooper E E, 1991b "Object recognition and laterality: Null results" *Neuropsychologia* (in press)
- Ellis R, Allport D A, Humphreys G W, Collis J, 1989 "Varieties of object constancy" *Quarterly Journal of Experimental Psychology A* **41** 775–796
- Gross C G, 1973 "Visual functions of inferotemporal cortex" in *Handbook of Sensory Physiology* volume VII/3 Ed. R Jung (Berlin: Springer) pp 451–482
- Gross C G, Mishkin M, 1977 "The neural basis of stimulus equivalence across retinal translation" in *Lateralization in the Nervous System* Eds S Harnad, R Doty, J Jaynes, L Goldstein, G Krauthamer (New York: Academic Press) pp 109–122
- Hummel J E, Biederman I, 1992 "Dynamic binding in a neural network for shape recognition" *Psychological Review* in press
- Jacoby L L, Baker J G, Brooks L R, 1989 "Episodic effects of picture identification: Implications for theories of concept learning and theories of memory" *Journal of Experimental Psychology: Learning, Memory, and Cognition* **15** 275–281
- Jacoby L L, Dallas M, 1981 "On the relationship between autobiographical memory and perceptual learning" *Journal of Experimental Psychology: General* **110** 306–340
- Kohlmeyer S W, in press, "Picture Perception Lab: A software package for picture perception experiments on the Macintosh II" *Behavioral Research Methods, Instruments, and Computers*
- Mishkin M, Appenzeller T, 1987 "The anatomy of memory" *Scientific American* **256** 80–89
- Previc F H, 1990 "Functional specialization in the lower and upper visual fields in humans: Its ecological origins and neurophysiological implications" *Behavioral and Brain Sciences* **13** 519–575
- Schacter D L, 1987 "Implicit memory: History and current status" *Journal of Experimental Psychology: Learning, Memory, and Cognition* **13** 501–518
- Schacter D L, Dulaney S M, Merikle E P, 1990 "Priming of nonverbal information and the nature of implicit memory" in *The Psychology of Learning and Motivation* volume 26 Ed. G H Bower (New York: Academic Press) pp 83–123
- Ungerleider L G, Mishkin M, 1982 "Two cortical visual systems" in *Analysis of Visual Behavior* Eds D J Ingle, M A Goodale, R J W Mansfield (Cambridge, MA: MIT) pp 549–586