Is entry-level recognition viewpoint invariant or viewpoint dependent?

JANICE E. MURRAY
University of Otago, Dunedin, New Zealand

Subjects decided whether an object drawing matched the entry-level name that immediately preceded it in a name–object sequence. When objects in the stimulus set were visually similar with respect to global shape and configuration of parts, response time increased linearly from 0° to 120° for both match and mismatch trials. Similar effects of orientation were found on match trials when objects in the stimulus set were visually dissimilar. No effects of orientation were observed when name and drawing did not match in the visually dissimilar condition. The results are consistent with the view that, in a variety of viewing situations, the initial identification of an object at the entry level is accomplished by viewpoint-dependent mechanisms.

Time to name line drawings of familiar objects rotated in the picture plane is a function of the angular disparity between the rotated drawing and upright; the farther away the drawing is rotated from the upright, the longer the response time is (e.g., Jolicoeur, 1985; McMullen & Jolicoeur, 1992; Murray, Jolicoeur, McMullen, & Ingleton, 1993). This effect has widely been interpreted as reflecting a mechanism of object recognition that is viewpoint dependent (Jolicoeur, 1990; Tarr & Pinker, 1989). Representations may be stored in memory at the most typical orientation (e.g., upright) or at a small number of familiar orientations. In order to identify an object at a viewpoint that deviates from the canonical view(s), the object must be brought into alignment with the familiar viewpoint through some normalization process.

Hamm and McMullen (in press) explored the possibility that the need for viewpoint-dependent representations and normalization may depend on the level of identification required. Objects can be identified at a general level (superordinate; e.g., "animal"), a specific level (subordinate; e.g., "Dalmatian"), or an intermediate level (basic; e.g., "dog"). In their experiments, Hamm and McMullen asked subjects to match names and rotated line drawings of objects at the three levels of abstraction. Significant effects of orientation were observed for subordinate-level matching when name and object matched. Orientation effects were also obtained on mismatch trials when subordinate names and mismatched objects shared identity at the basic level. In contrast, no effects of orientation were found for either superordinate- or basic-level matching.

In their account of the results, Hamm and McMullen argued that identification decisions regarding basic- and superordinate-level identification were made on the basis

of basic-level representations that are viewpoint invariant. In keeping with previous results, they argued that identification of objects at the more specific subordinate level would require additional perceptual processing (Jolicoeur, Gluck, & Kosslyn, 1984). In the case of rotated objects, Hamm and McMullen argued that this additional processing would include normalization. Normalization would be necessary to determine the spatial relations among the parts of the object (McMullen & Jolicoeur, 1992) in order to discriminate among subordinate members of a basic category, where exemplars are typically similar in terms of parts and may only be distinguishable on the basis of part relations.

In assigning special status to basic-level representations, Hamm and McMullen are entirely consistent with previous work by Rosch, Mervis, Gray, Johnson, and Boyes-Braem (1976), who demonstrated that the initial level of identification of upright objects is generally at the basic level.[1] The new possibility suggested by Hamm and McMullen's results is that the initial identification at the basic or entry level is viewpoint invariant, with viewpoint-dependent recognition occurring only when a more specific level of identification is required. In apparent contradiction with the notion that entry-level recognition is viewpoint invariant is the abundant evidence showing orientation effects in naming tasks (e.g., Jolicoeur, 1985, 1988; Jolicoeur & Milliken, 1989; McMullen & Jolicoeur, 1992; Murray et al., 1993), where it has been largely assumed that identification is at the entry level (Tarr, 1995). However, there is no specific control over the level of identification in naming tasks, and Hamm and McMullen contend that consideration of the acceptable names for the objects suggests many instances consistent with subordinate-level classification.

The generality of the finding that entry-level identification is accomplished by means of viewpoint-invariant mechanisms needs to be considered. In both the subordinate- and the entry-level verification conditions in Hamm and McMullen's experiments, the stimuli consisted of 12

exemplars for each of the six categories used (car, boat, aircraft, bird, dog, and bug). A comparison of the conditions suggests that important differences existed between the two verification tasks. The discriminations required to perform the task in the subordinate-level verification condition were necessarily made among objects that were highly similar in both global shape and spatial relations. For each exemplar-specific verification, there were 12 pictures with different subordinate identities whose part composition and overall visual structure would be highly consistent with the target name. This difficult within-category discrimination would rule out any simple part(s) or global shape being diagnostic in performance of the subordinate-level verification task.

What is notable about the between-category discriminations required in the entry-level condition of Hamm and McMullen's experiment is the restricted nature of the stimulus set. In the entry-level verification task, there is no need to discriminate between the visually similar members of each of the basic categories, since the discrimination is at the category level. Accordingly, the task essentially required only six discriminations. In addition, the six basic categories shared very little with each other by way of global shape and spatial relations among parts. The combined effect of between-category visual dissimilarity and the small number of distinctive visual shapes to be discriminated would effectively make the discrimination much easier in the entry-level condition and allow for the use of a feature- or part-based processing strategy for recognition. For example, detection of a shape with four parallel elongated structures would be sufficient to accurately identify a dog in the stimulus set, since no other category would share this collection of parts or structural description. There is ample evidence to indicate that when members of the stimulus set can be discriminated from all others on the basis of unique attributes, viewpoint invariance in recognition may be obtained (e.g., Jolicoeur, 1985; Murray et al., 1993). In the entry-level condition of Hamm and McMullen, knowledge of the viewpoint-invariant information could be acquired quickly given the repeated experience (12 exemplars per category) of the small number of shape types to be discriminated.

It can be argued that the conditions under which viewpoint-invariance was exhibited in the entry-level condition do not fully capture the experience of everyday object recognition. First, there any many instances when entry-level discriminations must be made between objects that are visually similar but are from different categories. For example, is that object gliding in the sky an airplane or a Royal Albatross, that object in the paddock a pony or a donkey? Additionally, independent of visual similarity, everyday entry-level recognition is often accomplished in situations where the number of candidate objects is potentially high. In these various contexts, entry-level recognition by viewpoint-invariant mechanisms prior to normalization may not be possible. This possibility was investigated in the present experiment using the name–picture matching task. In one condition, entry-

level recognition was assessed under conditions in which objects within the stimulus set were from different basic categories but overlapped significantly in global shape and configuration of parts. In a second condition, objects were also from different basic categories but were highly variable in global shape and part relations. In both conditions, a relatively large number of unique name–object pairs was tested.

METHOD

Subjects

The subjects were 100 undergraduate students of the University of Otago who participated in the experiment as part of their course. English was the first language of all participants, and no subject had previous experience with the stimuli. The single testing session for each subject was approximately 15 min in duration.

Stimuli and Apparatus

The stimuli were line drawings selected from Snodgrass and Vanderwart (1980). The line drawings depicted objects that had high name agreement (91%) at the entry level of classification.[2] The names of two objects, *sweater* and *baby carriage*, were changed to *jumper* and *pram*, respectively, to reflect the entry-level names for these items in New Zealand culture. Each drawing was produced at orientations of 0°, 60°, 120°, 240°, and 300° of clockwise rotation.

Two sets of 36 drawings were created as shown in Table 1. Within each set, half the stimuli were used on trials where a name that preceded the drawing matched the drawing (match), and half were used on trials where the preceding name and drawing did not represent the same object (mismatch). One set of stimuli comprised objects that were visually similar. Across the entire set, there was a large number of objects that had a high degree of overlap in global shape and shared a number of similar parts in similar arrangements. This was achieved by careful selection of objects from a small number of semantic categories (e.g., animals, birds). Members of semantic categories were assigned in equal numbers to the match and mismatch groups. For each mismatch name–drawing pair, the similarity in visual structure was specifically

Table 1
Names of the Objects Used in the Match and Mismatch Trials for the Two Conditions of Visual Similarity

Visually Similar		Visually Dissimilar	
Match	Mismatch	Match	Mismatch
donkey	cat/monkey	airplane	giraffe/lamp
swan	owl/eagle	frog	iron/pram
shirt	dress/coat	telephone	window/kangaroo
seal	sheep/goat	car	penguin/hat
mouse	rhinoceros/elephant	bear	grasshopper/bicycle
chair	desk/dresser	foot	flower/kettle
squirrel	dog/fox	boat	trafficlight/owl
penguin	duck/chicken	duck	crab/cannon
lamp	bed/couch	tree	drum/helicopter
bear	deer/cow	wagon	lightbulb/chair
fly	ant/spider	fly	tie/spider
church	house/barn	jumper	bell/alligator
camel	horse/lion	mushroom	pie/truck
jumper	boot/shoe	gun	television/shoe
airplane	bicycle/motorcycle	basket	rollerskate/bed
gorilla	frog/snail	doll	fish/church
rabbit	tiger/zebra	apple	handbag/cake
helicopter	car/bus	piano	clock/glass

Note—Each drawing listed for match trials was presented as both the name and the drawing in the name–picture sequence. The name and subsequent drawing for mismatch trials are as given.

maintained; objects represented by the name and drawing were from the same semantic category and visually similar (e.g., bicycle–motorcycle).

The second set of stimuli comprised objects that were visually dissimilar. These stimuli were selected from a variety of semantic categories and had relatively low overlap in global shape, parts, and part structure. This dissimilarity in visual structure was emphasized in the specific name–drawing pairs of the mismatch condition (e.g., fish–church). Fourteen drawings were common to the similar and dissimilar stimulus sets.

On the basis of the norms provided by Snodgrass and Vanderwart (1980), mean ratings of visual complexity and familiarity were calculated for each match and mismatch group in the two sets of stimuli. A comparison among the four means for each measure revealed no significant differences.

The drawings were presented on a VGA color monitor as black line drawings against a light gray background, and they subtended an average of 3° of visual angle. Stimulus presentation and collection of responses were controlled by an IBM-compatible computer using Micro Experimental Laboratory software (Schneider, 1988).

Procedure

The subjects were randomly assigned to either the visually similar or the visually dissimilar condition. All other aspects of their treatment were the same.

Each trial began with the presentation of an entry-level name of an object in the center of the screen for 1,500 msec. The screen was then blanked for 500 msec. Following this interval, the screen was turned on at the top of the refresh cycle to reveal a drawing of an object. The drawing remained on until the subject responded. The subjects were instructed to decide whether the drawing and the preceding name represented the same object. Match responses were made by pressing the B key on the keyboard with the index finger of the left hand; mismatch responses were made by pressing the N key with the index finger of the right hand. Latency was measured from the onset of the drawing to the subject's keypress. Eighteen match pairs and 18 mismatch pairs were each presented once in random order in a single block of trials. Within each match and mismatch group, 6 pairs were presented at 0°, and 3 pairs were presented at each of 60°, 120°, 240°, and 300°. To control for any item-specific effects on orientation, each drawing occurred at each orientation equally often across the entire group of subjects. The 36 experimental trials were preceded by 8 practice trials. Both accuracy and speed of response were emphasized, and the subjects were provided with accuracy and response time (RT) feedback for 1 sec at the end of each trial.

RESULTS

An outlier criterion excluding RTs less than 200 msec and greater than 1,650 msec resulted in a loss of less than 1% of the data. RTs for clockwise and counterclockwise rotations of the same magnitude were averaged. Thus, 60° and 300° RTs were averaged to represent 60° departures from upright, and 120° and 240° RTs were averaged to represent 120° departures from upright. Mean RTs for correct responses were calculated for each subject in each visual similarity and match/mismatch condition at each orientation. The mean RTs for each of the conditions are displayed in Figure 1.

The results for the two visual similarity groups were analyzed separately. In each analysis of variance, trial type (match, mismatch) and orientation (0°, 60°, and 120°) served as within-subjects factors.

Response Times

Visually similar. Response latencies were slower on trials where the name and drawing did not match (772 msec)
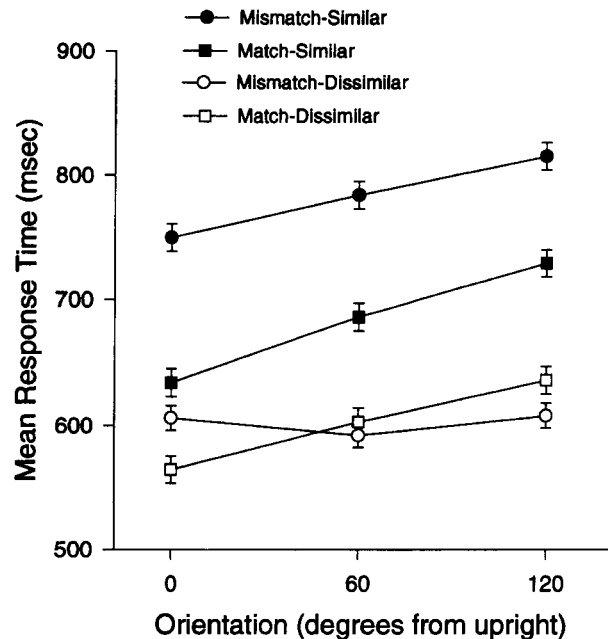

Figure 1. Mean naming times at each orientation for match and mismatch drawings in the visually similar and visually dissimilar conditions. Standard error bars are within-subjects errors (Loftus & Masson, 1994) calculated separately for each function.

than on trials where the two depicted the same object (683 msec) [$F(1,49) = 79.29$, $MS_e = 9,494$, $p < .001$]. As is evident in Figure 1, RT increased with increasing departure from the upright [$F(2,98) = 24.35$, $MS_e = 6,565$, $p < .001$]. The component of the orientation effect reflecting the linear effect of orientation was assessed by applying the contrasts -1, 0, and 1 to the results for the three orientations. This revealed a strong linear component [$F(1,49) = 54.99$, $MS_e = 5,801$, $p < .001$]. The residual variance in response latencies after removal of the variance for the linear trend was not significant ($F < 1$). There was no significant difference in the orientation effect across match and mismatch trials ($F < 1$). The linear component of orientation effect also did not differ significantly across trial type as assessed by the orientation$_{linear}$ $\times$ trial type interaction [$F(1,49) = 1.89$, $MS_e = 6,234$, $p > .15$]. The least squares estimate of rate of normalization was 1,389°/sec for match trials, and 1,846°/ sec for mismatch trials.

Visually dissimilar. The mean response latency for match (601 msec) and mismatch trials (602 msec) did not differ ($F < 1$). As was the case for the visually similar condition, RT increased as a function of orientation [$F(2,98) = 6.81$, $MS_e = 5,291$, $p < .01$]. The linear component of the orientation effect was also significant [$F(1,49) = 13.71$, $MS_e = 5,099$, $p < .001$]. As suggested by Figure 1, the effect of orientation depended on trial type [$F(2,98) = 6.04$, $MS_e = 5,538$, $p < .01$], with this difference reflected additionally in the interaction of trial type and linear trend across orientation [$F(1,49) = 9.64$,

$MS_e = 6,380, p < .01$]. Further investigation of this interaction determined that there was no orientation effect when name and drawing mismatched ($F < 1$). In contrast, when name and drawing did match, there was a significant effect of orientation [$F(2,98) = 10.49$, $MS_e = 6,275, p < .001$]. The linear component of this effect was significant [$F(1,49) = 19.11$, $MS_e = 6,869$, $p < .001$], with no significant residual ($F < 1$). The normalization rate was $1,667°/sec$.

Error Rates

Similar analyses for error data for the visually similar and dissimilar conditions revealed only two significant effects. In the visually similar condition, the 12% error rate on mismatch trials was higher than the error rate on match trials at 6% [$F(1,49) = 24.25$, $MS_e = 0.0141, p < .001$]. The reverse pattern was exhibited in the visually dissimilar condition. An error rate of 7% was found on match trials, whereas an error rate of 5% was found on mismatch trials [$F(1,49) = 12.42$, $MS_e = 0.0044, p < .001$].

DISCUSSION

The results clearly support the notion that entry-level recognition is not exclusively viewpoint invariant. In the visually similar condition, the context of the entry-level recognition task demanded fine discriminations between objects of similar shape, and recognition was dependent on deviation from upright for both match and mismatch trials. When objects have similar features in similar spatial relationships, viewpoint-invariant information about global shape or constellations of features simply does not serve to distinguish one object from all others in the stimulus set. Similarity in overall visual structure is typified by subordinate-level discriminations, and it was indeed at that level that Hamm and McMullen observed effects of orientation on time to identify objects. The present results suggest that Hamm and McMullen's orientation effect at the subordinate level occurred because of the visual similarity factor, rather than the level of identification per se; visual similarity among objects to be discriminated at either a subordinate or an entry level precipitates use of a mechanism of recognition that is viewpoint dependent.

The present results also indicate that viewpoint-dependent entry-level recognition is not restricted to contexts in which the objects to be discriminated are similar in global shape and part composition. In the visually dissimilar condition, the objects in the experimental set varied widely in their overall shape, yet a strong orientation effect was observed when name and drawing matched. In a context where set membership is not restricted in terms of number or type of objects, the observed global shape and part configuration of any object requiring identification could be consistent with any number of potential candidates. Thus, in order to respond with confidence on a match trial, the exact entry-level identity of the object must be known. The data suggest that this knowledge of object identity is obtained only after normalization of the object to upright. In contrast, on mismatch trials, knowledge of the entry-level identity of the object is not required to provide an accurate response. When gross differences in overall shape and part relations exist between the target object and the presented object, this viewpoint-invariant information would be sufficient to dismiss the object as a match independent of orientation, as was found.

The results are entirely consistent with the proposal that recognition can be considered along a continuum spanning two mechanisms of recognition (Edelman, 1991; Farah, 1992; Jolicoeur, 1990; Tarr, 1995; Tarr & Bulthoff, 1995). Judgments requiring fine visual discriminations are accomplished by viewpoint-dependent mechanisms. This is most certainly the case when subordinate-level judgments are required

(Hamm & McMullen, in press), but, as evidenced in the present study, also when entry-level judgments are made in contexts where the objects are visually similar. Importantly, viewpoint-dependent mechanisms are also engaged in contexts where the objects are visually dissimilar and set membership is unrestricted and unknown. At the opposite end of the continuum, viewpoint-invariant mechanisms are engaged when gross category discriminations are all that is necessary to perform the task and when diagnostic viewpoint-invariant features are clearly available (e.g., mismatch, visually dissimilar; see Hamm & McMullen, in press). Considered together, the present results and those of Hamm and McMullen demonstrate that entry-level recognition is neither exclusively viewpoint dependent nor exclusively viewpoint invariant. Whether viewpoint-dependent or viewpoint-invariant mechanisms are engaged will depend on the nature of the task, the context, and familiarity with and similarity among the stimuli. In consideration of the conditions under which real-world recognition occurs, the present findings make it clear that, in a wide variety of viewing circumstances, normalization of an object image for matching with a viewpoint-dependent representation is necessary for successful entry-level recognition to occur.

REFERENCES

EDELMAN, S. (1991). *Features of recognition* (Tech. Rep. CS-TR10). Rehovot, Israel: Weizmann Institute of Science.

FARAH, M. J. (1992). Is an object an object an object? Cognitive and neuropsychological investigations of domain-specificity in visual object recognition. *Current Directions in Psychological Science*, 1, 164-169.

HAMM, J. P., & MCMULLEN, P. A. (in press). Effects of orientation on the identification of rotated objects depend on the level of identity. *Journal of Experimental Psychology: Human Perception & Performance*.

JOLICOEUR, P. (1985). The time to name disoriented natural objects. *Memory & Cognition*, 13, 289-303.

JOLICOEUR, P. (1988). Mental rotation and the identification of disoriented objects. *Canadian Journal of Psychology*, 42, 461-478.

JOLICOEUR, P. (1990). Identification of disoriented objects: A dual-systems theory. *Mind & Language*, 5, 387-410.

JOLICOEUR, P., GLUCK, M., & KOSSLYN, S. M. (1984). Picture and names: Making the connection. *Cognitive Psychology*, 16, 351-364.

JOLICOEUR, P., & MILLIKEN, B. (1989). Identification of disoriented objects: Effects of context of prior presentation. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 15, 200-210.

LOFTUS, G. R., & MASSON, M. E. J. (1994). Using confidence intervals in within-subject designs. *Psychonomic Bulletin & Review*, 1, 476-490.

MCMULLEN, P. A., & JOLICOEUR, P. (1992). The reference frame and effects of orientation on finding the top of rotated objects. *Journal of Experimental Psychology: Human Perception & Performance*, 18, 807-820.

MURRAY, J. E., JOLICOEUR, P., MCMULLEN, P. A., & INGLETON, M. (1993). Orientation-invariant transfer of training in the identification of rotated natural objects. *Memory & Cognition*, 21, 604-610.

ROSCH, E., MERVIS, C. B., GRAY, W. D., JOHNSON, D. M., & BOYES-BRAEM, P. (1976). Basic objects in natural categories. *Cognitive Psychology*, 8, 382-439.

SCHNEIDER, W. (1988). Micro Experimental Laboratory: An integrated system for IBM PC compatibles. *Behavior Research Methods, Instruments, & Computers*, 20, 206-217.

SNODGRASS, J. G., & VANDERWART, M. (1980). A standardized set of 260 pictures: Norms for name agreement, image agreement, familiarity, and visual complexity. *Journal of Experimental Psychology: Human Learning & Memory*, 6, 174-215.

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., & BULTHOFF, H. H. (1995). Is human object recognition better described by geon structural descriptions or by multiple views? Comment on Biederman and Gerhardstein (1993). *Journal of Experimental Psychology: Human Perception & Performance*, 21, 1494-1505.

TARR, M. J., & PINKER, S. (1989). Mental rotation and orientation-dependence in shape recognition. *Cognitive Psychology*, **21**, 233-282.

NOTES

1. Subsequent work by Jolicoeur et al. (1984) determined that, in general, objects were initially classified at the basic level, whereas atypical members of a basic-level category were identified initially at a more specific level (e.g., calling a penguin "penguin," rather than "bird").

This "entry-level" characterization more accurately describes the level of abstraction used to identify objects initially, and it will be used in preference over "basic level."

2. Snodgrass and Vanderwart (1980) used the term *basic level* but included atypical concept classifications, such as "penguin," under that term. Thus, their classification is entirely consistent with entry-level classification.