The use of word–picture verification to study entry-level object recognition: Further support for view-invariant mechanisms

STEFANO A. DECARO Fooyin University of Technology, Kaohsiung Hsien, Taiwan

and

ADAM REEVES Northeastern University, Boston, Massachusetts

There are substantial logical and empirical reasons for rejecting the popular view that plane-misoriented objects are identified after normalization of global orientation. Our subjects determined the entry-level identity of common objects (line drawings) *before* determining basic orientation (upright vs. rotated)— as they must if they are to begin to know how to restore the image to the canonical upright. We used a description-picture matching procedure in which 75% of the trials involved a mismatch in identity, orientation, or both, so that comparisons could be based on one response (*no*). Times to verify identity were faster than times to verify orientation and did not increase with rotations in the picture plane. That mismatch objects were positively identified at the entry level was shown both in a surprise recognition test for object names and through the transfer of priming from the matching task to a subsequent object-naming task. We conclude that classic mental rotation-like effects on naming times do not reflect early object encoding and recognition processes, which are view invariant, but may stem from double-checking at a postrecognition verification stage.

The process of object recognition through vision inevitably confounds object-specifying information (such as surface, shape, color, and local features) with incidental information specifying the object's orientation (typically, lighting and viewing angle). Thus, a critical question for theory is whether recognition processes can be brought to bear directly on the relevant, object-specifying properties or whether incidental information, such as orientation, must be encoded first. The increase in recognition latency associated with plane-rotated depictions of common objects (e.g., Jolicœur, 1985) has been thought to imply that orientation information is extracted early and then used to transform the object image to match its canonical (upright) view prior to recognition.

However, if recognition can be based on view-invariant properties, successful encoding of object class may precede the extraction of orientation. Psychometric data from a backward-masking experiment which used line drawings of common objects showed that critical stimulus onset asynchronies were 2-19 msec shorter when identity, rather than orientation, was verified and that it was orientation matching, not identity matching, that was most sensitive to rotations in the picture plane (DeCaro & Reeves, 2000). The response times (RTs) and error patterns from an experiment in which subjects matched pictures against orientation-name descriptions (such as upright car) also showed a perceptual advantage for determining object identity over the simplest classification of orientation, rotated versus upright (DeCaro, 1998). These results agree with, and extend to pictures, data from an early experiment by Corballis, Zbrodoff, Shetzer, and Butler (1978), in which subjects were faster to verify the identity of alphanumeric characters than to verify the characters' angular orientations.

Theories that place the process of object identification before the process of orientation identification are naturally compatible with view-invariant accounts of object recognition, because early view-invariant object recognition cannot support the determination of orientation relative to an external frame. Rather, the visual–cognitive system must derive either a viewer-centered or an environmentcentered description of the object, and this is likely to take more time or require additional processes, such as mental rotation. Although view-invariant accounts should predict that orientation is extracted after identity, findings to this effect need to be supplemented by more direct support for the use of view-invariant mechanisms in visual object

This article is based on a dissertation presented by the first author in partial fulfillment of the requirements for the Ph.D. degree at Northeastern University. Portions of this article were presented by the second author at the first annual meeting of the Vision Sciences Society, Ft. Lauderdale, FL. We thank Michael Corballis and Ruth Maki for their careful reviews and encouraging comments on the original draft of this article. Correspondence concerning this article should be addressed to S. A. DeCaro, Department of Applied Languages, Fooyin University of Technology, 151 Chinhsueh Road, Ta-Liao Hsiang, Kaohsiung Hsien 831, Taiwan (e-mail: sdecaro@lynx.dac.neu.edu).

recognition. To this end, we present new experiments based on the description-depiction classification procedure introduced by DeCaro (1998), and we address important issues regarding its use.

Word-Picture Matching

The procedure we employed in this study is a variant of the word-picture matching task often used to probe recognition processes. In the typical word-picture experiment, subjects view a word and then a picture in sequence, and the task is to report whether the picture matches the word on the basis of specified criteria (e.g., Hamm & McMullen, 1998). In our *description-picture* experiment, subjects view an orientation-name description (such as UPRIGHT TREE), followed by a line-drawing depiction of a common object (see Figure 1). The critical trials are those involving a partial mismatch, either in the depicted object's name (an *identity mismatch*) or in the depicted object's orientation (an orientation mismatch). The relative time needed to determine identity and orientation is then inferred from the time taken to correctly report *no* to each of these types of mismatches. Thus, the description-picture procedure is unconventional, because it uses *pairs* of words to probe two types of encoding on each trial and because the match (yes) trials are eventually discarded and only the mismatch (no) cases are used to measure performance.

A methodological advantage of the description-picture procedure is that the critical trials are compared by using the same response, task, and subject. In addition, these comparisons are supplemented by performance in the third, *global mismatch* condition, in which both identity and orientation can signal a mismatch. The global mismatch cases, by themselves, cannot indicate which element is verified first, but they can reveal whether performance in the fastest partial-mismatch condition is influenced by the second mismatch. Such effects are shown by an increase in error rates and RTs to verify partial mismatches, relative to error rates and RTs to verify global mismatches termed a *global mismatch effect*.

The data reported by DeCaro (1998) showed clearly that (1) subjects were faster to detect a mismatch object than a mismatch orientation and (2) identity mismatches rarely elicited a false ves response, whereas a significant number of orientation mismatch cases were incorrectly classified as matches. Both findings point to a primacy for extracting object class: Identity is encoded before orientation, and when orientation is encoded, identity information tends to dominate the response. In this article, we first consider two potential processing models that might explain the original data: a *shift model*, in which the process of specifying orientation is completely time-shifted relative to the process of specifying identity, and a cross *model*, in which the process of specifying orientation begins first but ends sufficiently later to give rise to overall slower mean RTs. These two processing models are distinguished in Experiment 1 with psychometric functions.

We also address a methodological issue concerning specificity in matching tasks. Naming responses have the advantage of specifying a level of classification—typically, the *basic level* (Rosch, Mervis, Gray, Johnson, & Boyes-Braem, 1976), which is also the entry point for most objects (see Jolicœeur, Gluck, & Kosslyn, 1984). Although the descriptions in our task are also at the basic or entry level, it is not clear whether the object depictions are classified



Figure 1. Overview of the description-picture matching procedure. The task is to report *yes* if both the identity and the orientation of the depicted object match the preceding description and to report *no* otherwise. This example illustrates an *identity mismatch* trial.

at this level of specificity, because global shape information alone may be sufficient to signal an identity mismatch. In this study, we employed a recognition test (Experiment 2) and an intertask priming procedure (Experiment 3B) to investigate the specificity of the verification judgments. Finally, in Experiment 3A, we will present a modified version of the description–picture procedure that permitted a direct test of rotation effects on verification times.

GENERAL METHOD

All the subjects in this study were undergraduate students, who participated in partial fulfillment of their psychology course requirements at Northeastern University. To be eligible for participation, students were required to be native speakers of English and to demonstrate at least 20/30 Snellen acuity at the time of the experiments. The various description–picture sessions were always preceded by a block of 32 practice trials, during which simple alphanumeric characters were displayed and matched against preceding descriptions, such as UPRIGHT NUMBER or ROTATED LETTER. In this way, our subjects were sufficiently prepared for the experiments without viewing any object depictions.

The testing items in this study were line drawings of common objects taken from the Snodgrass and Vanderwart (1980) picture set (121 items) and the Microsoft ClipArt gallery (7 items). These depictions were judged beforehand to have a standard upright view. Sixteen students in a pilot study were asked to point to the "top" side of each object once with a computer mouse. An object was considered to have a canonical view in the picture plane if all the students selected the same side of the object, regardless of the object's orientation in the display.

Our methods in this study speak primarily to *entry-level* (rather than basic-level) identification, because some of the word probes were subordinate-level descriptions (e.g., *rocking chair* and *wine glass*). We did not limit our descriptions to the basic level, because we were concerned with early recognition and many objects are first identified (have their entry point) at the subordinate level because they are atypical members of basic-level categories. Instead, our match and mismatch names (such as *car, goat, typewriter, penguin*) reflected the types of responses we observed in our own naming experiments and in Snodgrass and Vanderwart (1980). However, our results should still apply to basic-level object recognition, because the basic and entry levels for common objects tend to overlap and because many of our descriptions used basic-level names.

The subscripts RT and ER are used throughout to denote analyses based on RTs and error rates, respectively. The p values reported for effects involving within-subjects factors are based on the Geisser–Greenhouse lower bound adjustment to degrees of freedom (e.g., Lewis, 1993).

EXPERIMENT 1 Replication and Further Analysis

We will begin with a replication of the original description-picture experiment, in which subjects are tested with upright (0°) and upside-down (180°) novel depictions of common objects. The procedures, items, and equipment were identical to those described in DeCaro (1998), except that here, we include the preliminary grouped data, in which description order (e.g., UPRIGHT CAR vs. CAR UPRIGHT) was tested across two groups of 8 subjects. This

replication serves primarily to facilitate our psychometric analysis and subsequent discussion of the original data.

Method

Subjects. Sixteen undergraduate students participated in Experiment 1.

Stimuli and Equipment. The test displays were 96 black line drawings of common objects, presented against a white background on a 15-in. ViewSonic monitor with 640×480 pixel resolution. Responses were recorded with a two-button serial mouse (Segalowitz & Graves, 1990) attached to an IBM-compatible 80486 microcomputer. The computer controlled the output of the video display and measured RTs through a software millisecond timer with 1-msec precision (Graves & Bradley, 1987, 1991).

Procedure. The subjects held the computer mouse with both hands, using their left and right thumbs to enter each response. Response button assignments were set by each subject according to his or her preference (since there was no need to counterbalance). Each experimental trial began with the presentation of the description, which appeared directly above a black ring centered on an otherwise white display. (The ring subtended 10° of visual angle, which was just large enough to contain the tallest and widest objects). The description specified an orientation, either the word UPRIGHT or the word UPSIDE-DOWN, and the name of a familiar object. For half the subjects, the description first specified orientation (e.g., UPSIDE-DOWN TREE). For the remaining subjects, the description first specified identity (e.g., TREE UPSIDE-DOWN). The subject continued the trial by pressing a mouse button, which erased the entire display and triggered the object to appear 500 msec later. The subjects were instructed to respond as quickly as possible without making errors. Response timing began with the onset of the picture; the duration of the picture was equal to the response time.

A Latin-square design was used to fully counterbalance objects across subjects with the 4 (match conditions) $\times 2$ (depicted orientations) $\times 2$ (description formats) = 16 conditions. Each subject was assigned a unique trial schedule and, therefore, participated in a single block of 96 trials comprising 96 different objects. For each subject, (1) the order of trials within the schedule was randomized without constraint, and (2) mismatch names were randomly selected without replacement from a list of names of nondepicted objects.¹

Results

The mean correct RTs and error rates from each group of subjects were submitted to a two-way, mixed analysis of variance (ANOVA) to measure the within-subjects effect of *mismatch* (orientation, identity, or global), the betweensubjects effect of *description order* (orientation–name or name–orientation), and the interaction. (Excluding the global match [*yes*] data from the ANOVA ensures that response effects cannot confound the results. However, RTs and error rates for all four match conditions are presented in Table 1 for completeness.)

Effects of description order. The results of the ANOVA showed that description order had no effect on performance. When orientation was specified first, the mean RT and error rate were 569 msec and 3.3%, respectively, as compared with 571 msec and 3.1% when name was specified first [$F_{\rm RT}(1,14) < 1$, $MS_{\rm e} = 30,757.67$, $F_{\rm ER}(1,14) < 1$, $MS_{\rm e} = 4.70$, $p_{\rm S} > .75$]. These null effects can be interpreted unambiguously because description order did not interact with type of mismatch [$F_{\rm RT}(2,28) < 1$, $MS_{\rm e} = 1,858.99$, $F_{\rm ER}(2,28) < 1$, $MS_{\rm e} = 6.72$, $p_{\rm S} > .40$]. In light

Table 1 Mean Correct Response Time (in Milliseconds), Standard Deviation, and Percentage Error as a Function of Identity and Orientation Match in Experiments 1–3A

	Orientation							
	N	Mismatch						
Identity	М	SD	%E	М	SD	%E		
	Experime	ent 1 (n	= 16)					
Match	580	133	9.9	675	125	7.8		
Mismatch	517	84	1.0	518	96	0.8		
	Experime	ent 2 (<i>n</i>	= 12)					
Match	536	99	6.9	667	118	10.8		
Mismatch	535	116	3.1	513	84	0.3		
	Experime	nt 3A (<i>n</i>	= 16)					
Match	616	141	7.2	705	149	7.9		
Mismatch	532	91	0.6	516	73	0.6		

of these results, the orientation-name format was used in subsequent experiments, because it conforms to the adjective-noun grammar of the English language and does not appear to bias the outcome.

Effects of mismatch. Type of mismatch had large effects on both RTs and errors $[F_{RT}(2,28) = 70.80, MS_e = 1,858.99, and F_{ER}(2,28) = 37.85, MS_e = 6.72, ps < .001].$ The planned comparisons between identity mismatches (M = 517 msec, 1.0% errors) and orientation mismatches (M = 675 msec, 7.8% errors) showed that the subjects were faster and more likely to correctly verify a mismatch in identity than a mismatch in orientation $[F_{RT}(1,28) = 106.90, MS_e = 1,858.99, \text{ and } F_{ER}(1,30) = 54.60, MS_e = 6.72, ps < .001].$ Moreover, it is evident from Table 1 that responses to identity mismatches were as fast and as accurate as responses to global mismatches (Fs < 1).

Psychometric Analysis and Discussion

The results of Experiment 1 were consistent with those reported in DeCaro (1998) in both the magnitudes and the patterns of effects on RTs and error rates. Table 1 shows that performance on different-identity trials (i.e., identity mismatch and global mismatch) did not vary with orientation match. Thus, the increased RTs on orientation mismatch trials can be attributed to the time needed to extract a viewer-centered description of the object-for example, to locate "top" (Rock, 1973). If so, orientation encoding entirely follows extraction of object class (the shift model). It is, however, still possible that the object's orientation is extracted before the object has been identified but that identity is encoded more rapidly and, so, will eventually give rise to faster object identification. Such early orientation encoding may even work to propel the process of object identification by guiding transformations of rotated objects (e.g., Tarr & Pinker, 1991). The cross model implies that the underlying distribution of latencies to verify orientation is more variable than is the distribution of latencies to verify identity. However, such increased variability in the observed RTs would not by itself prove the

cross model, since the variability must be sufficient to show crossing.

The smooth curves in Figure 2 are logistic functions fit to the mean data for the two partial-mismatch conditions. We used curve fitting to extrapolate beyond the reaction limits of our subjects, where any crossing is likely to be most evident. This particular psychometric function was chosen for its simplicity: Parameter a controls the location, and b controls the slope (or shape).² A higher value of b corresponds to a steeper cumulative function and, therefore, to decreased variability in the generating process. The ordinate in Figure 2 is the cumulative proportion of subjects' total correct judgments. This measure, rather than *proportion correct*, was used so that each subject's two curves would always extend from 0 to 1, permitting precise comparisons of encoding rates and curve locations despite differing error rates.

Consistent with the grouped data plotted in Figure 2, the individual psychometric functions fit to each subject's data showed that the rate of encoding for orientation was slower than the rate of encoding for identity (in 12 out of 16 cases). This was confirmed by a paired *t* test of the 32 estimates of the slope parameters, whose means, b = 5.83 (for orientation) and b = 7.20 (for identity), differed significantly [t(15) = 2.62, p < .05]. However, the times needed to verify orientation did not vary enough to maintain the supposition that orientation is determined before identity. As is shown in Figure 2, any difference in slope



Figure 2. Mean proportion of total mismatches detected as a function of time in Experiment 1. Psychometric curves are best-fitting logistic functions, $y = (x/a)^b / [1 + (x/a)^b]$, where *a* is the response time corresponding to 50% correct and *b* determines the slope. The response times above 1,000 msec were too few and too scattered to include in this plot. Error bars are ± 1 standard error.

was countered by a large difference in the locations of the two functions: The respective mean values of *a* were 616 and 494 msec [t(15) = 8.94, p < .001]. That the two curves in Figure 2 do not cross shows that the increase in RTs for verifying orientation probably reflects a genuine lag, or *shift*, in the time to encode orientation—not merely a slower rate of processing.

EXPERIMENT 2 Recognition for Mismatch Objects

It is assumed that name verification reflects a fairly efficient process of object recognition. However, consider the identity mismatch trial illustrated in Figure 1. Hamm and McMullen (1998) point out that, in such a case, the depiction is identified minimally as "not a . . ." (in this case, "not a rabbit"). It is known that objects that share a basic level tend to be visually similar, whereas objects from different basic categories are often clearly distinguishable in terms of global shape (Rosch et al., 1976). Given the heterogeneity of the objects in our study, it is likely that global shape information would have been sufficient to signal an identity mismatch on many trials, thereby obviating the need for classification at the more specific entry level. Indeed, the subjects may have successfully rejected mismatch objects by comparing the depictions and mismatch names at the superordinate level (e.g., "the depicted object is not an *animal* and, therefore, not a rabbit").

Our first approach to testing the specificity in description-picture matching was to measure the subject's recognition of mismatch objects. To illustrate this approach, suppose that a subject viewed the picture of a tree without the word "tree" in the matching task, as in Figure 1. Further suppose that the subject correctly reported a mismatch, because the picture did not match the preceding "rabbit" description (i.e., not a guess). The question is whether the picture was identified merely as "not a rabbit." If yes, it would be impossible to recall having seen a tree when probed with a word stimulus, "tree" (rather than the original depiction), because our pictures and words can match only at the entry or basic levels. Data inconsistent with guessing therefore provide partial support that mismatch objects are verified at a relatively high level of specificity.

Method

Subjects. Twelve undergraduate students participated in Experiment 2.

Procedure. The equipment, materials, and procedures were identical to those used in Experiment 1, but the subjects were also given a surprise recognition test about 2 min after completing the description-picture matching trials. During the recognition test, the subjects were probed with the names of common objects and were asked to press the "Y" key if a name referred to an object depicted in the description-picture task, and to press the "N" key otherwise. The probes consisted of the 48 names of objects that had been depicted on identity mismatch and global mismatch trials and 48 *foils*—items from the list of mismatch names that had not already served as descriptions on the mismatch trials. Hence, none of the 96 name probes were presented to the subject prior to the recognition test, but half of the names referred to objects that were *depicted* in the description–picture task. After entering each *yes/no* response, the subjects rated how confident they were in their decisions, using the following scale: 0 = don't know (guessing); 1 = barely confident; 2 = reasonably confident; 3 = very confident.

Results and Discussion

The results of the description-picture task are summarized in Table 1. Mean correct RTs and error rates were submitted to a one-way repeated measures ANOVA to measure the effects of mismatch, which were significant $[F_{\rm RT}(2,22) = 43.81, MS_{\rm e} = 1,910.17, \text{ and } F_{\rm ER}(2,22) =$ $33.75, MS_e = 10.35, ps < .001$]. Identity mismatches $(M = 535 \operatorname{msec}, 3.1\% \operatorname{errors})$ were verified faster and more accurately than orientation mismatches [M = 667 msec], 10.8% errors; $F_{\rm RT}(1,22) = 54.87$, $MS_{\rm e} = 1,910.17$, and $F_{\rm ER}(1,22) = 33.84$, $MS_{\rm e} = 10.35$, ps < .001], but somewhat less accurately than global mismatches [M = 0.3%]errors; $F_{\text{ER}}(1,22) = 4.48$, $MS_{\text{e}} = 10.35$, p = .06]. This marginal global mismatch effect on error rates resulted from a slight increase in the number of errors on identity mismatch trials, as compared with the number of errors we obtained in Experiment 1. However, the results again showed no global mismatch effect on RTs $[F_{RT}(1,22) =$ $1.57, MS_e = 1,910.17, p = .24$].

Individual scores on the recognition test ranged from a low of 61.5% correct to a high of 90.6% correct, with a mean of 74.6%, which was well above chance performance [t(11) = 9.14, p < .001]. The lowest score, 61.5%, is itself significantly better than chance (z = 2.35, p < .05, by normal approximation to the binomial). Thus, guessing can be rejected as the sole explanation of performance for all 12 subjects.

The confidence ratings for correct responses (M = 2.46)were significantly higher than the ratings for incorrect responses [M = 2.03; t(11) = 4.84, p < .001], implying that the subjects were somewhat aware of their own errors but were adopting a conservative response criterion. To separate memory performance and response bias, each subject's hit rate (yes to a probe naming a depicted object) and false alarm rate (yes to a foil) were converted to d^{-1} and β . The d[,] provides a measure of stimulus discrimination that is independent of response bias, β (Macmillan & Creelman, 1991). Values of β greater than 1 indicate a bias toward no responses, and indeed, the average value of β in our experiment was 3.02. This agrees with the selfrating data and confirms that the subjects tended to assume that they had not seen any of the objects named in the recognition test unless the name probe elicited a strong sense of familiarity.

Recognition sensitivity (as measured by d·) ranged from a low of 0.87 to a high of 2.92, with a mean over subjects of 1.60, well above the chance level of zero [t(11) =8.58, p < .001]. When hit and false alarm rates were first averaged across subjects instead, the resulting *pooled* d· was 1.44, still well above zero. We should emphasize that the subjects completed the description-picture matching trials not knowing that they would have to remember the depicted objects (all of which were intermixed with potentially interfering object names). These results indicate intact recognition for entry-level identity (at least for most subjects) and rule out the possibility that mismatch objects were identified as "not a . . ." and nothing more specific. Thus, the findings of Experiment 2 support the conclusion of Experiment 1 (DeCaro, 1998) that identity is encoded before orientation.

In word-picture matching tasks, it may be possible to reject a mismatch object on the basis of prominent visual features, such as global shape. Yet the findings of Experiment 2 favor the conclusion that the subjects rejected mismatch objects at a semantic level—specifically, the entry level. It is supposed that the subjects retrieved the entry-level identity of the depicted object and then compared it with the description, also at the entry level. During the surprise recognition test, a positive name probe would activate the same entry-level concept triggered earlier by the picture, signaling that the name referred to one of the depicted objects.

EXPERIMENT 3A Verifying Upright and Multirotated Objects

The verification times in the description-picture task were consistently compatible with earlier encoding for identity, rather than for general orientation, which provides partial support for view-invariant accounts of visual object recognition. Using a backward-masking procedure and picture-word verification, DeCaro and Reeves (2000) also showed that identity was determined before orientation and, in addition, that processing time to extract identity was relatively flat from 60° to 180° of rotation. Hamm and McMullen's (1998) word-picture experiment varied the level of classification of a match (e.g., *collie* vs. *dog* vs. animal) and showed that subordinate-level (e.g., col*lie*) decision times were most affected by rotations from the upright, whereas basic-level and superordinate-level decision times were about equally least affectedalthough still not flat through 180° of rotation.

Whereas the word–picture procedure used by Hamm and McMullen (1998) probed the classification of object class at different levels of specificity and varied orientation in the picture plane, the description–picture procedure used by DeCaro (1998) probed the classification of both object class and orientation, but it could not properly measure the effects of rotations in the plane. First, there are only two spatial views, but more views are required to test for progressive effects of rotation. Second, the RTs to verify orientation mismatches when objects are upright reflect the time to determine that objects are *not rotated*. Conversely, when objects are *not upright*.

An experiment by Maki and Braine (1985) suggested that there may be encoding differences between these two types of classifications. Their subjects first learned to as-

sociate the spatial words up, down, right, and left to objects presented at 0°, 180°, 90°, and 270°, respectively. In a subsequent word-picture verification task, the subjects saw a spatial word followed by a picture of an object. The results showed that matches were reported faster than mismatches when pictures were upright, whereas matches were reported slower than mismatches when pictures were upside-down. Maki and Braine attributed this interaction effect to response compatibility, resulting from an implicit tendency to associate upright with true and upside-down with false. If so, direct comparisons between upright and rotated depictions in the description-picture procedure may be confounded by one or more nuisance variables. Our solution was to add multiple rotated views to the design without changing the task. By comparing the partialmismatch trials within the rotated-view conditions, it is then possible to properly measure the effects of angular rotation on verification times.

Method

Subjects. Sixteen undergraduate students participated in Experiment 3A.

Stimuli and Equipment. The set of pictures was increased to 128 line drawings of common objects, which included the items used in Experiments 1 and 2. Misoriented versions of each object were created by rotating the standard version 45°, 90°, 135°, and 180° in the picture plane. The equipment was as described in Experiment 1.

Procedure. The procedures were generally the same as those in Experiment 1, with a few exceptions. First, because misoriented objects were not always rotated 180°, the description UPSIDE-DOWN was changed to ROTATED. Second, for each subject, 32 unique objects were randomly assigned to each of the four match conditions. Within each match condition, 50% of the items were randomly selected to appear at 0°, and the remaining objects were randomly selected to appear at one of the four rotated views, with the constraint that an equal number of objects was assigned to each view. The subjects were instructed that any nonupright object should be classified as *rotated* when matching orientation—since the *degree* of rotation did not matter. The final ordering of trials was completely random, and—as in the first two experiments—the objects were presented without repetition.

Results and Discussion

The mean correct RTs and error rates were submitted to a two-way repeated measures ANOVA to measure the effects of *mismatch* and *view* and the interaction. The main effects of mismatch were significant $[F_{RT}(2,30) = 45.16,$ $MS_{\rm e} = 13,503.15, F_{\rm ER}(2,30) = 15.11, MS_{\rm e} = 76.37, ps < 100$.001]. Table 1 shows that these effects were consistent with those obtained in Experiment 1, despite the change to multirotated views. The main effects of view were not significant $[F_{RT}(4,60) = 0.87, MS_e = 10,589.36, \text{and } F_{ER}(4,60) =$ 1.36, $MS_{e} = 67.73$, ps > .25]. However, the mismatch × view interaction was marginally significant even with the conservative adjustment to degrees of freedom $[F_{RT}(8, 120) =$ 3.26, $MS_e = 7,239.72$, p = .09; there were too few errors at this level of analysis to test for effects on error rates]. We therefore reexamined the effects of view on just the partial-mismatch trials, using post-hoc contrasts.

Figure 3 shows that the difference in RTs across the orientation mismatch and identity mismatch trials varied



Figure 3. Mean times to correctly verify identity and orientation mismatches as a function of spatial orientation (Experiment 3A). Error bars are 1 standard error.

with angular view. First, on identity mismatch trials, 0° views were verified faster than the four rotated views (45°–180°), whereas on orientation mismatch trials, 0° views were verified slower than the four rotated views $[F_{\rm RT}(1,120) = 21.57, MS_{\rm e} = 7,239.72, p < .001]$. Second, times to verify identity were essentially flat across the four rotated views $[F_{\rm RT}(1,120) = 0.08, MS_{\rm e} = 7,239.72, p = .78]$, whereas times to verify orientation *decreased* with increasing rotation from the upright $[F_{\rm RT}(1,120) = 4.96, MS_{\rm e} = 7,239.72, p < .05]$, as was shown in tests of the linear trend across the 45°, 90°, 135°, and 180° view conditions.

A separate analysis was carried out to determine whether upright/rotated interacted with mismatch/match. Our results replicated Maki and Braine (1985), in that global matches were verified faster than mismatches when objects were upright, whereas global matches were verified slower than mismatches when objects were rotated. However, the response compatibility hypothesis (associating upright with true and upside-down with false) cannot account for the interaction. The left side of Figure 4 shows the match and mismatch cases plotted as a function of whether objects were upright or rotated in the display. Response compatibility predicts that mismatch responses to a rotated object will be faster than mismatch responses to an upright object. The critical result in Figure 4 (left) is that the opposite effect occurred on identity mismatch (Δ) trials. When the data were, instead, classified according to spatial word, as shown on the right side of Figure 4, the cause was obvious: Times to verify *rotated* were longer than times to verify upright, regardless of match and regardless of view. An inspection of the data from Experiment 1 showed the same effect with the words upsidedown and upright.

We therefore reasoned that the difference between upright and nonupright views on identity mismatch trials shown in Figure 3 is an artifact of spatial word, because the description was UPRIGHT for 0° views and ROTATED for the 45°-180° views (owing to the orientation match). If so, the global-mismatch cases should have manifested the opposite effect (owing to the orientation mismatch). When the global mismatch cases were also classified according to the object's physical orientation, such a pattern was found. Consequently, the identity mismatch and global mismatch cases were submitted to a new ANOVA, but the data were explicitly classified according to spatial word (UPRIGHT vs. ROTATED) and depicted view (0° to 180°), ignoring whether there was a match or a mismatch in orientation. The ANOVA confirmed a significant main effect of spatial word $[F_{RT}(1,15) = 23.36, MS_e = 3,755.38, p < 10^{-10}]$.001]. As was predicted, RTs were faster on trials with the word upright (M = 505 msec) than on trials with the word rotated (M = 542 msec). Neither the main effect of view $[F_{\rm RT}(4,60) = 1.68, MS_{\rm e} = 4,572.20]$ nor the word \times view interaction $[F_{RT}(4,60) = 0.89, MS_e = 4,083.51]$ was significant (ps > .20).

The veridical effect of rotation on times to verify object identity is shown by combining the identity mismatch and global mismatch cases, thereby canceling the effect of spatial word. Corrected matching times across the 0°–180° range of views were 511, 534, 534, 553, and 524 msec. This effect of plane rotation on RT was minimal, amounting to 0.1 msec/deg. If orientation encoding or mental rotation had preceded identification, one might have expected a robust orientation effect with our novel (nonrepeated) depictions (Jolicœur, 1990; Jolicœur & Milliken, 1989).

Experiment 3A produced two important findings. The first was that rotations in the picture plane did not increase



Figure 4. The effects of spatial orientation (left panel) and spatial word (right panel) on times to correctly verify matches and mismatches in Experiment 3A.

the time required to determine object identity. On the contrary, the time needed to determine orientation decreased with rotation up to 180°, presumably because it is easier to classify a misoriented object as *rotated* the further it is transformed from the canonical view. However, more time was still required to determine that an upside-down (180°) object was rotated than to determine the object's identity. Thus, the conclusion from Experiment 1 that identity precedes orientation is true regardless of whether subjects must detect small (45°), medium (90°), or large (135° and 180°) rotations in the picture plane.

This experiment also demonstrated the influence of spatial words in word–picture matching tasks. The results showed that the interaction between spatial dimension and match was due to a main effect of spatial word, and not to response compatibility. The details in Maki and Braine (1985) showed that their interaction effect was also consistent with a simpler main effect of word. We therefore choose the more parsimonious spatial word effect to explain both results. The fact that *upright* can be verified faster than *upside-down* and *rotated* points to differences in the encoding and interpretation of certain spatial labels. These differences should be considered when interpreting performance in word–picture tasks that involve spatial judgments, such as location, position, and orientation.

EXPERIMENT 3B Object Naming

The addition of multirotated objects in Experiment 3A provided an opportunity to test the specificity of our description-picture task in a completely different manner than in Experiment 2. In this approach, we sought to uncover processes common to word-picture matching and object naming by testing for intertask transfer-that is, for changes in naming performance resulting from wordpicture verification. The following points will help to clarify and motivate this use of transfer. When rotated objects are repeated in a naming experiment, there are two prominent markers that distinguish performance on Block 2 from that on Block 1 (see Jolicœur, 1985; Maki, 1986). First, there is object priming: Repeated objects are named faster in Block 2 than in Block 1. Second, there is orien*tation priming*: The size of the orientation effect is smaller in Block 2 than in Block 1, yielding a much flatter naming function. These two markers show priming, rather than mere practice, because they disappear when new objects are introduced in later blocks of trials (Jolicœur, 1985).

In principle, transfer of these two priming effects can be used to test convergence between word-picture matching and object naming. Objects are typically named with entry-level terms (Snodgrass & Vanderwart, 1980), the same level as that used in our descriptions. If objects are *verified* at the entry level, priming should transfer from matching to naming, manifesting itself in the very first, rather than the second, block of naming trials. Object priming is not expected when repeated objects are not identified in Block 1 (Murray, 1995), and orientation priming is not expected when repeated objects are exclusively upright in Block 1 (Jolicœur & Milliken, 1989). The assumption that objects are verified at the entry level in the description–picture task and the fact that objects are to be depicted at several orientations together predict both objectpriming and orientation-priming effects on naming times in Block 1.

Method

Subjects. The subjects were 32 undergraduate students, 16 of whom were the same as those in Experiment 3A. The other 16 subjects were *controls*, having not participated in any of the previous experiments.

Design. The control subjects completed two blocks of naming trials (denoted B1-control, and B2-control, respectively), whereas the test subjects from Experiment 3A completed only a single block of naming trials (B1-test). We expected transfer from Block 1 naming to Block 2 naming with the control subjects, owing to the two documented priming effects. The purpose of this design was to discover whether priming would also transfer from the matching task to naming with the test subjects. If so, naming times in the B1-test condition should more closely follow the naming times in the B2-control condition.

The items used on B1-control trials were identical to those used on the description-picture trials in Experiment 3A. Each control subject was paired with one data file from Experiment 3A. These files contained records of the objects and views that were used on each description-picture trial. A computer program read this information and then generated equivalent trial schedules for the control subjects. In addition, the items used on B2-control trials were identical to those used on B1-test trials. These two critical blocks comprised only the previously rotated objects (for the test subjects, these were all the rotated objects presented in the description-picture task). It was necessary to exclude the 50% of the objects previewed at 0° because repeated objects generally do not produce orientation priming in Block 2 if they were upright in Block 1 (Jolicœur & Milliken, 1989). Each of the 64 previously rotated objects was randomly assigned to appear at a new orientation (including 0°), with the constraint that an equal number of objects was assigned to the five views. A different randomization sequence was used for each test subject and then duplicated for the corresponding control subject.

Procedure. The equipment used to present objects on the naming trials was the same as that described in Experiment 1. A black ring appeared at the center of the display to signal the start of each trial. The subject then pressed a mouse button, which blanked the display and caused the picture to appear (until named) 500 msec later. The subjects were instructed to name each object as quickly and clearly as possible. A microphone, voice-activate d relay, and software timer measured naming latencies to the nearest millisecond. The sensitivity of the voice-activated relay was calibrated to each subject during an initial practice block, while the subject named a series of alphanumeric characters. The test subjects were run approximately 5 min after completing the description–picture experiment (the time needed to complete the practice trials and voice calibration).

Results and Discussion

Naming times longer than 3 sec were discarded as obvious outliers, which affected 0.7% of the data. Incorrect naming responses and spoiled trials were also discarded, which excluded 8.0% of the data. A trial was considered spoiled if the naming response failed to activate the voice key or if a vocal hesitation activated the voice key prematurely.

 Table 2

 Mean Correct Naming Time (in Milliseconds) and Standard Deviation for Each Type of Block in Experiment 3B

					B1 Test						
B1 Control		B2 Control		I+/O+		I+/O-		I-/O+		I-/0-	
М	SD	М	SD	М	SD	М	SD	М	SD	М	SD
1,070	130	888	111	874	144	874	141	890	140	922	187

Note—The data from test subjects who completed the description—picture task prior to naming objects are broken down to show in which match condition the objects were first presented: I and O denote identity and orientation, respectively, and + and - denote match and mismatch, respectively.

Control priming. The data from the control subjects were submitted to a two-way repeated measures ANOVA to measure the effects of block and view and the interaction. The main effect of block was significant [F(1,15) = $150.18, MS_e = 8,833.79, p < .001$]. Naming times were faster in Block 2 (M = 888 msec) than in Block 1 (M =1,070 msec). The main effect of view was also significant $[F(4,60) = 10.75, MS_e = 5,696.24, p < .005]$. Naming times increased with rotation from the upright, but with a departure in linearity at 180° that is typical in rotated object naming (e.g., Jolicœur, 1985). The overall block \times view interaction was not reliable in this case [F(4,60) = $2.37, MS_e = 9,442.23, p = .14$]. However, the slope of the orientation effect was smaller in Block 2 than in Block 1, as was shown by a significant linear interaction component $[F(1,60) = 7.31, MS_e = 9,442.23, p < .05]$. Thus, these data are consistent with previous object-naming experiments showing both object priming and orientation priming in Block 2.

Test priming. Table 2 shows the mean correct naming times for each type of block. The B1-test trials are broken down to show the specific match condition in which the objects were previewed (in Experiment 3A) prior to naming. It is clear from these data that naming times were shorter on B1-test trials (M = 890 msec) than on B1-control trials [M = 1,070 msec; t(30) = 3.76, p < .001]. Table 2 also shows that the four match conditions in the description–picture task primed equally well [F(3,45) = 1.53, $MS_e = 5,450.59$, p = .23].

Next, data from the B1-test and B2-control conditions were submitted to a mixed two-way ANOVA with block and view as the between- and within-subjects factors, respectively. The two groups of subjects showed equivalent priming: There was an effect of view $[F(4,120) = 5.10, MS_e = 7,370.62, p < .05]$, but neither the effect of block $[F(1,30) = 0.03, MS_e = 82,745.61]$ nor the block × view interaction $[F(4,120) = 0.64, MS_e = 7,370.62]$ was significant (ps > .40). Trend analysis showed that any linear interaction component also was unreliable (p = .19).

The substantive importance of these results is obvious from Figure 5: Verifying objects in the description–picture task produced significant priming effects in a subsequent object-naming task. We obtained both object priming and orientation priming effects in Block 1 that were comparable to the effects normally seen in Block 2 with other subjects. The findings of Experiment 3B thus strengthen the interpretation of results from the surprise recognition test in Experiment 2. Taken together, a more convincing argument can be made that match and mismatch objects were verified at a semantic level consistent with entry-level classification.

GENERAL DISCUSSION

In three experiments, we have attempted again to demonstrate that object class is determined even before general orientation (upright vs. misoriented) is known (DeCaro, 1998) and that the recognition of plane-rotated views of common objects does not require a prior stage of mental rotation (Jolicœur, Corballis, & Lawson, 1998) or an incremental normalization process that mimics mental rotation. The psychometric functions in Experiment 1 show a clear shift and indicate that, using RT this time, orientation is not merely encoded later than identity at all levels of performance: It accrues more slowly. We have suggested that a mental rotation-like process is used to determine (or



Figure 5. Mean correct naming times as a function of spatial orientation for each block in Experiment 3B.

check) orientation in the picture plane (DeCaro & Reeves, 2000), but our data in Figure 3 imply a rotation speed of 1,538 deg/sec (or 0.65 msec/deg), much too fast for mental rotation. The very rapid transformations implied by these data might reflect the fact that our task required orientation judgments at only the lowest level of specificity: Simply determining that the picture is not upright was sufficient classification. If we had required subjects to classify specific views (such as 45° vs. 90° vs. 135°), we might have observed much slower rates, consistent with earlier mental rotation studies (e.g., Cooper & Shepard, 1973). The orientation task was designed to be overly simple (to bolster the identity-before-orientationcase), and so the rotation curves may reflect this ease.

Perhaps more important, in light of the results already established, we showed how an unconventional wordpicture matching task could be used to probe visual object recognition. DeCaro's (1998) method implies a high proportion of mismatch trials (here, 75%) so that latencies for the same response (no) can be used as a probe across different conditions. This has the enormous advantage that incidental factors that can influence responses in samedifferent paradigms (such as handedness, frequency of response, and number of matching elements) cannot affect the conclusions and, so, need not be controlled (see Krueger, 1973; Nickerson, 1967). Still, a possible criticism is that mismatches can be determined on open-ended, negative grounds ("not an X"), and if so, they would not address object recognition at the entry level. A critical finding, however, is that most of the subjects demonstrated intact memory for mismatch object names in a surprise recognition test following the description-picture trials, showing that the subjects had classified the objects in a positive manner. In early pilot work, we obtained similar recognition scores when picture duration in the descriptionpicture task was limited to 55 msec with backward masking, which makes it less likely that memory performance reflects any encoding beyond that needed to make a speeded response in the description-picture task.

Another possible criticism is that the task set to subjects in word–picture verification, a simple *same–different* judgment, is so different from naming that word–picture matching performance cannot bear on naming RTs. Thus, the finding by Hamm and McMullen (1998), that basiclevel judgments showed weak effects of orientation, and our second critical finding, that times to verify identity were flat across 180° of rotation, may not contradict the classic naming experiments of Jolicœur (1985) and Maki (1986), which do show a clear, systematic effect of orientation on times to name similar objects. However, in Experiment 3B, we showed direct transfer of priming from word–picture matching to naming, demonstrating an important convergence in the encoding and recognition processes brought to bear on the two tasks.

An argument can always be made that naming is special, but we have also found direct evidence that rotation in the plane has systematic effects on the encoding of orientation, rather than identity, using a backward-masking procedure and *picture–word* matching (DeCaro & Reeves, 2000). Thus, we have argued, like Corballis (1988), that the effects of orientation on naming RTs occur at a *postrecognition* stage of double-checking. Familiarity with the misoriented material is sufficient to virtually eliminate orientation effects on naming RTs for common objects—unlike the effects on times to judge orientation, which are linear and robust across practice (e.g., Cooper & Shepard, 1973; Corballis & Cullen, 1986; Jolicœur, Ingleton, Bartram, & Booth, 1993; McMullen & Jolicœur, 1990, 1992). We therefore argued that the increase in RTs associated with rotations in the picture plane can be attributed quite reasonably to the use of view-dependent mechanisms to check or determine the *orientation* of newly rotated objects (DeCaro & Reeves, 2000).

Our claims regarding view dependency in visual object recognition apply to line-drawing depictions of common objects. That such drawings can support recognition almost as well as full-color representations (Biederman & Ju, 1988) implies a powerful perceptual engine that can abstract critical features for recognition from line segments, intersections, junctions, and so forth. Such critical features are heavily dependent on correct analysis of relative orientation (e.g., that one line segment is clockwise 15° relative to another, or that two lines meet at a right angle). A rotation in the picture plane, of course, changes none of these geometric relations, but such transformations can have a strong perceptual effect on global shape (Rock, 1973, 1974) and may affect particular features of the image. Our present conclusion is that double-checking (for orientation) is still a valid hypothesis to explain how early, orientation-independentrecognition might deal with the perceptual consequences of common objects depicted at uncommon views.

REFERENCES

- BIEDERMAN, I., & JU, G. (1988). Surface versus edge-based determinants of visual recognition. *Cognitive Psychology*, 20, 38-64.
- COOPER, L. A., & SHEPARD, R. N. (1973). The time required to prepare for a rotated stimulus. *Memory & Cognition*, 1, 246-250.
- CORBALLIS, M. C. (1988). Recognition of disoriented shapes. *Psychological Review*, 95, 115-123.
- CORBALLIS, M. C., & CULLEN, S. (1986). Decisions about the axes of disoriented shapes. *Memory & Cognition*, 14, 27-38.
- CORBALLIS, M. C., ZBRODOFF, N. J., SHETZER, L. I., & BUTLER, P. B. (1978). Decisions about identity and orientation of rotated letters and digits. *Memory & Cognition*, 6, 98-107.
- DECARO, S. A. (1998). On the perception of objects and their orientations. Spatial Vision, 11, 385-400.
- DECARO, S. A., & REEVES, A. (2000). Rotating objects to determine orientation, not identity: Evidence from a backward-masking/dual-task procedure. *Perception & Psychophysics*, 62, 1356-1366.
- GRAVES, R., & BRADLEY, R. (1987). Millisecond interval timer and auditory reaction time programs for the IBM PC. *Behavior Research Methods, Instruments, & Computers*, 19, 30-35.
- GRAVES, R., & BRADLEY, R. (1991). Millisecond timing on the IBM PC/ XT/AT and PS/2: A review of the options and corrections for the Graves and Bradley algorithm. *Behavior Research Methods, Instruments, & Computers*, 23, 377-379.
- HAMM, J. P., & MCMULLEN, P. A. (1998). Effects of orientation on the identification of rotated objects depend on the level of identity. *Journal of Experiment Psychology: Human Perception & Performance*, 24, 413-426.

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

- JOLICEUR, P. (1990). Identification of disoriented objects: A dualsystems theory. *Mind & Language*, 5, 387-410.
- JOLICŒUR, P., CORBALLIS, M. C., & LAWSON, R. (1998). The influence of perceived rotary motion on the recognition of rotated objects. *Psychonomic Bulletin & Review*, 5, 140-146.
- JOLICŒUR, P., GLUCK, M. A., & KOSSLYN, S. M. (1984). Pictures and names: Making the connection. *Cognitive Psychology*, 16, 243-275.
- JOLICŒUR, P., INGLETON, M., BARTRAM, L., & BOOTH, K. S. (1993). Top-bottom and front-behind decisions on rotated objects. *Canadian Journal of Experimental Psychology*, 47, 657-677.
- JOLICŒUR, P., & MILLIKEN, B. (1989). Identification of disoriented objects: Effects of context of prior presentation. *Journal of Experimental Psychology: Learning, Memory*, & Cognition, 15, 200-210.
- KRUEGER, L. E. (1973). Effect of stimulus frequency on speed of "same"-"different" judgments. In S. Kornblum (Ed.), Attention and performance IV (pp. 497-506). New York: Academic Press.
- LEWIS, C. (1993). Analyzing means from repeated measures data. In G. Keren & C. Lewis (Eds.), A handbook for data analysis in the behavioral sciences: Statistical issues (pp. 73-94). Hillsdale, NJ: Erlbaum.
- MACMILLAN, N. A., & CREELMAN, C. D. (1991). Detection theory: A user's guide. New York: Cambridge University Press.
- MAKI, R. H. (1986). Naming and locating the tops of rotated pictures. *Canadian Journal of Psychology*, **40**, 368-387.
- MAKI, R. H., & BRAINE, L. G. (1985). The role of verbal labels in the judgment of orientation and location. *Perception*, 14, 67-80.
- MCMULLEN, P. A., & JOLICEUR, P. (1990). The spatial frame of reference in object naming and discrimination of left–right reflections. *Memory* & Cognition, 18, 99-115.
- MCMULLEN, P. A., & JOLICEUR, P. (1992). Reference frame and effects of orientation on finding the tops of rotated objects. *Journal of Experimental Psychology: Human Perception & Performance*, 18, 807-820.
- MURRAY, J. E. (1995). The role of attention in the shift from orientationdependent to orientation-invariant identification of disoriented objects. *Memory & Cognition*, 23, 49-58.

- NICKERSON, R. S. (1967). "Same"-"different" response times with multi-attribute stimulus differences. *Perceptual & Motor Skills*, 24, 543-554.
- ROCK, I. (1973). Orientation and form. New York: Academic Press. ROCK, I. (1974, January). The perception of disoriented figures. Scientific American, 230, 78-85.
- 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.
- SEGALOWITZ, S. J., & GRAVES, R. E. (1990). Suitability of the IBM XT, AT, and PS/2 keyboard, mouse, and game port as response devices in reaction time paradigms. *Behavior Research Methods, Instruments*, & Computers, 22, 283-289.
- 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., & PINKER, S. (1991). Orientation-dependent mechanisms in shape recognition: Further issues. *Psychological Science*, 2, 207-209.

NOTES

1. There were minor variations in the lists of match and mismatch names used throughout this study (for instance, the lists were expanded in Experiment 3), but most of the items are described in DeCaro (1998). Alternatively, the reader can e-mail the first author for a complete list of the match and mismatch names used in the present study.

2. This curve is a simplified form of a more general three-parameter equation, in which x = (RT - c)/a permits scaling and shifting of the RTs with two parameters, and then the logistic gives *cumulative proportion correct* = $x^b/(1 + x^b)$, where *b* defines the curve shape. Our data were satisfactorily fit with c = 0 (no shift), perhaps because our RTs were sufficiently long. Had they been shorter, removal of the residual RT (namely, *c*) might have been necessary.

(Manuscript received November 22, 2001; revision accepted for publication February 25, 2002.)