Perceptual simulation in property verification

KAREN OLSETH SOLOMON
Willamette University, Salem, Oregon

and

LAWRENCE W. BARSALOU
Emory University, Atlanta, Georgia

If people represent concepts with perceptual simulations, two predictions follow in the property verification task (e.g., Is *face* a property of GORILLA?). First, perceptual variables such as property size should predict the performance of neutral subjects, because these variables determine the ease of processing properties in perceptual simulations (i.e., perceptual effort). Second, uninstructed neutral subjects should spontaneously construct simulations to verify properties and therefore perform similarly to imagery subjects asked explicitly to use images (i.e., instructional equivalence). As predicted, neutral subjects exhibited both perceptual effort and instructional equivalence, consistent with the assumption that they construct perceptual simulations spontaneously to verify properties. Notably, however, this pattern occurred only when highly associated false properties prevented the use of a word association strategy. In other conditions that used unassociated false properties, the associative strength between concept and property words became a diagnostic cue for true versus false responses, so that associative strength became a better predictor of verification than simulation. This pattern indicates that conceptual tasks engender mixtures of simulation and word association, and that researchers must deter word association strategies when the goal is to assess conceptual knowledge.

Researchers increasingly report that modality-specific systems underlie the representation of knowledge. When people represent an object, they reenact what it looks like visually, what it sounds like auditorally, how they act on it motorically, how they react to it emotionally, and so forth. These reenactments appear related to mental imagery, using representations similar to those in perception and action (e.g., Farah, 2000; Finke, 1989; Grèzes & Decety, 2001; Jeannerod, 1995; Kosslyn, 1980, 1994; Shepard & Cooper, 1982). For theoretical accounts of this perspective, see Barsalou (1999a, 1999b, 2003a), Clark (1997), Damasio (1989), Glenberg (1997), Lakoff (1987), Mandler (1992), and Newton (1996). Increasingly, both behavioral and neural evidence offer empirical support. For recent reviews of this evidence, see Barsalou (2003b); Barsalou, Niedenthal, Barbey, and Ruppert (2003); Martin (2001); Martin and Chao (2001); and Martin, Ungerleider, and Haxby (2000).

Property Verification

The conceptual task examined here—property verification—is a relatively passive, recognition-oriented, and temporally constrained task. On each trial, subjects receive the word for a concept followed by the word for a property. Subjects then specify whether the property is true of the concept as fast as they can. For example, subjects might receive "lawn" then "roots" and respond true; or they might receive "lawn" then "wheels" and respond false.

Semantic network accounts of property verification assume that subjects search through networks of amodal symbols, attempting to link the target concept with the target property (e.g., Collins & Quillian, 1969; Conrad, 1972; Glass & Holyoak, 1975). In feature-matching models, property verification typically requires assessing whether a feature is included in those listed symbolically for a concept (e.g., Rosch & Mervis, 1975; Smith, Shoben, & Rips, 1974). In neither approach are modality-specific representations required or used.

Alternatively, property verification could be performed on sensorimotor simulations of objects. On receiving the word "pony," for example, subjects construct a sensorimotor simulation of a pony. On next receiving the word "mane," subjects construct a sensorimotor simulation of a mane, and then attempt to match it with a region of the PONY simulation. If a match is found, subjects respond true; if not, they respond false. Solomon and Barsalou (2001) have provided further details of this account, along with empirical evidence for it.[1]

244

Predictions for Property Verification from the Simulation View

The experiment reported here tests two predictions that follow from the simulation account of property verification: *perceptual effort* and *instructional equivalence*. Each is addressed in turn.

Perceptual effort. If subjects simulate objects to verify properties, then variables associated with perceptual effort should affect performance. The logic underlying this assumption is as follows. A given perceptual operation varies in the effort it requires. Thus rotating a visual representation a long angular distance takes more time and effort than rotating it less. Similarly, scanning a long distance between two points in a visual representation takes more time and effort than scanning a short distance. When a variable like angular distance or scanned distance affects processing time, this suggests the use of visual representations. Many researchers have used this logic to demonstrate visual images in working memory (for reviews, see Farah, 2000; Finke, 1989; Kosslyn, 1980, 1994; Shepard & Cooper, 1982).

We employ a similar logic here. If subjects simulate objects to verify properties, then perceptual variables should affect verification performance. The more perceptual effort required to process a simulation, the longer verification should take, and the more likely errors should be. In the experiments to follow, we scaled perceptual variables and assessed whether they predicted verification performance. We also scaled other variables that could potentially predict verification. Of primary interest was whether perceptual variables accounted for variance in verification performance after removing variance that nonperceptual variables explain. If so, this would suggest that subjects simulated objects to verify properties.

Instructional equivalence. Assessing instructional equivalence requires two groups of subjects: imagery versus neutral subjects. Imagery subjects are explicitly asked to use the simulation strategy for property verification. Specifically, they are asked to image the object and to respond true only if the property can be found in some region of the image. Much work in the imagery literature indicates that subjects adopt images when asked to do so (e.g., Paivio, 1986).

Neutral subjects, on the other hand, receive no explicit instructions for verifying properties. Of interest is the strategy they adopt spontaneously. If the simulation view is correct, then neutral subjects should perform the task similarly to imagery subjects, and detailed analyses of verification performance should show the same pattern for both groups. Alternatively, if neutral subjects spontaneously use semantic networks, feature lists, or some other amodal representation, detailed analyses of their performance should differ from those of imagery subjects.

Although we predicted that neutral and imagery subjects would both use simulations, we did *not* predict that they would perform identically. In particular, we predicted that imagery subjects would construct richer, more detailed simulations than neutral subjects. This prediction follows from Wu and Barsalou's (2004) finding that imagery subjects consistently produced more properties for concepts than did neutral subjects (and also from Barsalou's [1999b] proposal that simulations vary in componential detail). Thus, we predicted here that imagery subjects would construct richer simulations than neutral subjects while verifying properties. Because imagery subjects were explicitly asked to construct simulations, they should put more effort into constructing simulations than neutral subjects, who adopt simulations implicitly. As a result, imagery subjects should take longer to verify properties than neutral subjects.

Explaining these predictions from amodal views. Amodal views do not appear to predict perceptual effort and instructional equivalence a priori. To our knowledge, no amodal theory has ever predicted them, nor do such predictions follow intuitively from these theories. Nevertheless, amodal theories are so flexible and powerful that they can be configured to explain these results post hoc. We address how amodal theories can explain our results in the final discussion.

Overview of the Experiment

On each trial, subjects received the name of a category followed by the name of a property. On 100 true trials, the property was a part of the category's members (e.g., *sleeve* for BLOUSE). On 100 false trials, the property was not a part of the category's members (e.g., *garage* for CAR). Of primary interest were the factors that predicted the variance in average reaction times (RTs) and error rates for the 100 true pairs. Why were some true properties verified faster and with fewer errors than others?

To answer this question, three types of variables were scaled that might explain this variance: linguistic, perceptual, and expectancy. The linguistic variables included the associative strength from the concept to the property and vice versa, along with word frequency and word length. The perceptual variables included the size and position of properties in images, along with the ease of finding them in images, and sensorimotor aspects of processing them during situated action. The expectancy variables assessed the polysemy of properties. For example, *chimney* is a property of only a few categories, such as HOUSE, whereas *nose* is a property of many categories, such as HUMAN, HORSE, DOG, and AIRPLANE.

Once all the concept–property pairs were scaled on these variables, the average RTs and error rates for the 100 true pairs were regressed onto the scalings. The results of primary interest were the patterns of variables that predicted RTs and errors on true trials. By comparing the patterns of prediction in the imagery and neutral conditions, we assessed whether subjects exhibited perceptual effort and instructional equivalence, as described shortly.

The Word Association Strategy

A potential complication in any property verification study is that subjects can adopt different strategies. In

his classic work, Paivio (1986) consistently found that people have two sources of information available in such tasks, namely, the verbal and perceptual codes of his dual-coding theory. Other work converges on a similar conclusion. After reviewing a variety of word and picture processing paradigms, Glaser (1992) concluded that subjects use word-level and/or conceptual-level information, depending on task conditions. When task conditions do not require conceptual processing, subjects adopt the relatively superficial strategy of only using word-level representations (linguistic forms). Conversely, when task conditions deter this strategy, subjects retrieve conceptual information that produces "deeper" processing. Further evidence for such a distinction comes from studies that manipulate the difficulty of false trials and show that the depth of processing on true trials varies (e.g., James, 1975; McCloskey & Glucksberg, 1979; Smith et al., 1974). When false trials are easy, subjects perform relatively superficial processing on true trials. When false trials are difficult, true trials receive deeper processing.

These findings have implications for property verification, with the associativeness between concepts and properties being of central importance. Imagine that the concepts and properties on true trials have moderate to high associations between them, as would be expected for the standard properties of objects (e.g., HORSE–*mane*). Conversely, imagine that the concepts and properties on false trials have no associations between them (e.g., BICYCLE–*chin*). Under such conditions, subjects can simply attend to the associative strength between concept and property words, responding true when they detect an association and responding false when they do not. Most importantly, subjects need not access conceptual information about either the concept or the property. Simply detecting a superficial association between the words is sufficient for responding correctly. Subjects need not determine whether a conceptual basis underlies the association, such as the property being a *part* of the concept.

If subjects adopt this word association strategy, it is not possible to assess whether conceptual representations are grounded in perceptual simulations or not. Because conceptual representations are not retrieved, they are not subject to examination. No conclusions about the critical issues can be reached. Glaser's (1992) review of related paradigms suggests that subjects are likely to adopt the word association strategy under such task conditions.

To assess the nature of conceptual representations, the word association strategy must be deterred. One approach is to have the concepts and properties on false trials be highly associated, similar to the manipulations in James (1975), McCloskey and Glucksberg (1979), and Smith et al. (1974). Rather than have subjects reject unassociated false properties for concepts (e.g., BICYCLE–*chin*), subjects must reject associated ones (e.g., OWL–*tree*). For associated false properties, subjects cannot respond false on the basis of no association between the concept and

property, because one exists (e.g., people know that owls live in trees). To determine whether the property is *part* of the concept, subjects must access *conceptual* knowledge for OWL and *tree*.

If all of the false properties that subjects receive are associated with their respective concepts, subjects cannot use the word association strategy. Under these conditions, a test of conceptual representations becomes possible. Because conceptual representations must be retrieved, we can assess whether they have perceptual qualities or not. Do effects of perceptual effort occur? Is instructional equivalence observed? To assess these possibilities, subjects received false properties that were either all associated or all unassociated. When subjects received unassociated false properties, we expected to see evidence of the word association strategy. When subjects received associated false properties, we expected to see evidence of perceptual simulations, given our prediction that simulations underlie conceptual processing.

Manipulations and Predictions

Two variables were manipulated orthogonally between subjects: associated versus unassociated false trials, and neutral versus imagery instructions. Manipulating associated versus unassociated false trials assessed whether subjects adopted the word association strategy. Manipulating neutral versus imagery instructions assessed perceptual effort and instructional equivalence. Each manipulation and its predictions are discussed in turn.

The word association strategy. Half the subjects received unassociated false trials, and half received associated ones (crossed with the neutral-imagery manipulation). In the unassociated condition, the false properties were unassociated with the concepts (e.g., *chin* for BICYCLE, *plaster* for HORSE). In the associated condition, the false properties were associated with the concepts but were not parts of the respective objects (e.g., *tree* for OWL, *banana* for MONKEY). The true concept–property pairs were the same in both conditions.

If subjects in the unassociated conditions spontaneously adopt the word association strategy to verify properties, the associative strength between concepts and properties should best explain their performance. As the associative strength between a concept and a property increases, the ability to detect such relations should become faster. When a salient associative relation exists, subjects should be fastest; when the association is not as salient, subjects should be slowest (with errors and RTs correlated positively). Importantly, perceptual and expectancy variables should not be central predictors. Because subjects use "superficial" associations between the concept and property words, they need not retrieve "deeper" conceptual information.

If evidence for the word association strategy appears, then the unassociated condition does not provide a test of whether conceptual knowledge is grounded in simulations. Because the word association strategy allows subjects to bypass conceptual representations, the requisite

conditions for assessing the nature of conceptual representations are not present. To provide an assessment of this issue, we must turn to the associated condition. Because the false properties have strong associative relations to their concepts in this condition, subjects cannot use word association as a basis for responding. Instead subjects must search further and access conceptual representations. Under these latter conditions, the question is whether perceptual variables explain verification performance. If they do, this suggests that perceptual representations are in play. Thus, the key evidence that bears on instructional equivalence and perceptual effort comes from neutral and imagery subjects who received associated-false trials.

Instructional equivalence. Half of the subjects were explicitly asked to use imagery to verify properties. Specifically, they were asked to image the concept object and then to respond true only if they could find the property on the image. Subjects in the neutral condition were simply asked to determine whether the property was true of the concept. Of primary interest were the imagery and neutral subjects who received associated false trials, given that these subjects were most likely to perform conceptual processing. The average RTs and error rates for each of these two conditions were regressed separately onto the linguistic, perceptual, and expectancy predictors described earlier. If the same pattern of variables predicts performance for both imagery and neutral subjects, this provides evidence for instructional equivalence.

Perceptual effort. In performing the regressions, we removed all variance explained by the linguistic and expectancy variables to see whether the perceptual variables explain unique variance. If they do, then this suggests that subjects used perceptual simulations to represent the properties. Of primary interest is whether perceptual variables explain unique variance in the neutral condition. If so, this indicates that neutral subjects spontaneously adopted simulations to perform verification. Of secondary importance is whether perceptual variables explain variance in the imagery condition. If so, this verifies that imagery subjects used images as requested in the instructions.

Method

Design and Subjects. Two between-subjects variables structured the experiment—false properties (unassociated vs. associated) and instructions (neutral vs. imagery)—thereby defining four groups of subjects: unassociated-neutral, unassociated-imagery, associated-neutral, and associated-imagery. Subjects were 96 University of Chicago students who participated for pay, with 24 subjects assigned randomly to each condition. Ten additional subjects were excluded because of error rates that exceeded 20%.

Materials. To obtain enough false properties for the two associated conditions, we limited true properties to the physical parts of objects. For example, *litter* could be liberally construed as a property of CAT, yet it was a false property here, because it is not a physical part. By defining true properties only as parts, a large pool of associated properties remained that could be used as false properties. On the basis of scaling measures described shortly, we selected 100 true pairs from 18 diverse superordinate categories that were

normally distributed on their size (P_area) and their forward associative strength (L_conc_prop). No property or concept was repeated. Examples include ARROW–*shaft*, SALMON–*scales*, BATHTUB–*drain*, and GORILLA–*face*. The largest 50 properties averaged 29.7% of the imaged object; the smallest 50 properties averaged 10.6%. The most-associated 50 properties averaged 7.5 on a scale of 1 to 9; the least-associated 50 properties averaged 4.6. Across properties, size and associative strength were uncorrelated ($r = .13$, $p > .25$).

To form the false materials, 100 false properties were paired with 100 concepts. No false property or concept was a repetition of a true property or concept. In the associated false materials, the false properties were associated to the 100 concepts. These properties included thematically related entities (e.g., BANANA–*monkey*, OTTER–*river*), related taxonomic categories (e.g., DONKEY–*mule*, TABLE–*furniture*), and taxonomically related parts (e.g., GUITAR–*keyboard*, FLASH-LIGHT–*wick*). In the unassociated false materials, the same false properties were reassigned to the same concepts to form unassociated pairs, thereby holding concepts and properties constant. Examples include PLIERS–*river*, ASPARAGUS–*furniture*, and BRIEFCASE–*wick*. The associated false properties averaged 5.0 on L_conc_prop, whereas the unassociated false properties averaged 1.2. Solomon (1997) presents the full set of true and false materials.

Subjects in all four conditions received the same 100 true concept–property pairs. Whereas subjects in the two unassociated conditions received the 100 unassociated false pairs, subjects in the two associated conditions received the 100 associated false pairs. Another 48 pairs were constructed for the practice trials, half true and half false. The practice materials did not share any concepts or properties with the critical materials, although the distributions of concepts and properties were similar. Thus, the 24 true practice pairs had the same types of concepts and properties as the true critical pairs. The 24 false practice pairs were unassociated in the unassociated conditions, but associated in the associated conditions.

Procedure. Macintosh IIci computers running PsyScope controlled all aspects of the experiment. CMU button boxes recorded subjects' responses with millisecond accuracy. Prior to each trial, subjects rested their forefingers on response buttons and a foot on a pedal. To initiate a trial, subjects focused their attention on an asterisk in the middle of the screen and pressed the foot pedal when ready. A concept word appeared immediately for 500 msec, followed by a 1,200-msec interstimulus interval, and then a property word, which remained on the screen until a response occurred. Thus, there was a 1,700-msec SOA between the concept and property words.[2] Verification time was measured from the onset of the property word to the detection of a response. Subjects were instructed to respond true only if a property was a "physical part" of the concept object; otherwise, they were to respond false. Subjects used their dominant hand to make true responses and were instructed to respond as quickly as possible but to avoid making errors. On making an error, subjects received feedback on the computer screen ("Incorrect Response").

The two imagery groups received explicit instructions to form an image of each concept and then to look for the property in the image. If they found the property, they were to respond true; otherwise, they were to respond false. Imagery subjects were explicitly asked to base each response on their image, even if they knew the answer without consulting it. The two neutral groups received no explicit instructions about strategies for performing the task. Solomon (1997) presented the full set of instructions for imagery and neutral subjects.

To ensure optimal performance on the critical trials, subjects were led to believe that only the first 16 practice trials were practice and that the last 32 practice trials were critical. The 48 practice trials continued seamlessly into the 200 critical trials. Subjects received eight short breaks distributed evenly over the 248 total trials. Each subject received different random orders of the practice trials and of the critical trials. The experiment lasted 40 to 45 min.

Scaling overview. To perform the planned regression analyses, it was first necessary to scale the concept–property pairs on vari-

ables capable of predicting RTs and errors. As described earlier, two sets of predictors were of primary interest: *linguistic predictors* and *perceptual predictors*. To develop as complete a predictor set as possible, we also scaled *expectancy predictors*. Table 1 lists the 14 predictors included. Solomon (1997) has provided their means and standard deviations, as well as principal component analyses of them.

Scaling the linguistic predictors. The linguistic predictors assessed the presence of linguistic processing during property verification. If subjects use the word association strategy to verify properties, then the associative strength between a concept word and its property word should predict performance. When concept and property words are highly associated, subjects should detect their associativeness faster and produce faster RTs. Thus, the forward associative strength from each concept word to its respective property word was scaled (i.e., L_conc_prop). To rate L_conc_prop, 20 new subjects first read a concept word on one side of a card, then read the respective property word on the other side, and finally rated on a 1 to 9 scale how much the property word had come to mind while they were reading the concept word. The average rating across subjects constituted the value for the concept–property pair. L_conc_prop was also scaled in the same manner for the false materials, but with 12 new subjects. No other predictor was scaled for the false materials.

To further assess the word association strategy, the backward strength from the property word to the concept word was scaled (L_prop_conc). Sixteen new subjects heard each property word and named the first three objects having the property that came to mind. The objects named first, second, and third were scored 4, 3, and 2, respectively, to indicate decreasing ease of production. If the object used in the true materials was not produced, it was scored 1, indicating a lower associative strength than the three produced objects. The scaled value for each property was the average output position of its concept across subjects.[3]

Two other linguistic variables were also scaled that could potentially predict performance: the frequency of each property word (L_freq), and its length in letters (L_letters). L_freq was the property word's frequency in Kučera and Francis (1967). L_letters was the number of letters in the property word.

Scaling the perceptual predictors. The perceptual predictors assessed the presence of perceptual simulation during property verification. If subjects use perceptual simulation to verify properties,

then the perceptual qualities of the properties should predict RTs and errors.

If processing large properties takes more perceptual effort than processing small ones (see Kosslyn, Cave, Provost, & von Gierke, 1988; Kosslyn, Reiser, Farah, & Fliegal, 1983), RTs and errors should increase with property size. Thus, size was scaled for each property by having subjects estimate its area in the respective object (P_area). To assess size, 12 new subjects imagined each object and estimated the percentage of the object that the property occupied. For example, subjects were asked, "What percentage of the total volume of a fox is its nose?" The average percentage across subjects constituted P_area.

Because the distance that a subject scans from some starting point in an image to the property could also be important (Kosslyn, Ball, & Reiser, 1978), two position measures were scaled for each property, one from each of two possible origins: the property's distance from the bottom edge of the object (P_bottom), and its distance from the object's left edge (P_left). To rate P_bottom and P_left, the previous 12 subjects imagined each object in a "minimal" box that just fit the object, touching it on all four sides. For P_bottom, subjects were asked to find the minimal vertical distance from the bottom of the box to the nearest point on the property, and then to express this distance as a percentage of the box's height. For example, *paw* for DOG typically received 0% because it lies on the box's bottom edge. Subjects similarly estimated P_left, using the box's left edge as the reference point. Solomon (1997) scaled property position and size in drawings of objects and found the same results as for ratings here.

Three more general predictors of perceptual effort were also scaled. P_initial assessed whether each property was in the initial image of its respective concept or not. P_sal assessed the salience of the properties in the images of the respective concepts. P_find assessed subjects' intuitive sense of how easy it was to locate the properties in these images. Three new groups of 16 subjects scaled these three measures. For P_initial, subjects indicated whether the property was in their initial image (1) or not (0), with the scaled value being the proportion of 1s across subjects. For P_sal, an experimenter stated an object's name to subjects, who closed their eyes and imaged it. The experimenter then named a property of the object, and subjects rated the property's salience in their image using a scale that ranged from *extremely nonsalient* (1) to *extremely*

Table 1
Predictors and Their Reliabilities for the True Concept–Property Pairs

Predictor	Description	Reliability
Linguistic predictors		
L_conc_prop	associative strength from concept word to property word	.97
L_prop_conc	associative strength from property word to concept word	.96
L_freq	word frequency of the property word	–
L_letters	number of letters in the property word	–
Perceptual predictors		
P_area	area of the property in the image	.97
P_bottom	distance from the image's top/bottom edge to the property	.95
P_left	distance from the image's left edge to the property	.92
P_initial	whether the property is in the initial image of the concept	.87
P_sal	salience of the property in the image	.92
P_find	ease of finding the property in the image	.81
P_attn	likelihood of attending to the property on seeing the object	.88
P_handle	likelihood of touching the property on seeing the object	.91
Expectancy predictors		
E_num_obj	how many different objects have the property	.85
E_object	whether the property can refer to an object	.84

Note—Predictors without reliabilities were collected from sources other than subjects in this experiment.

salient (9), with the scaled value being the average across subjects. For P_find, subjects received similar instructions to image objects. This time, however, the experimenter stated, "Once you have had time to form the image, I will read you the name of a property. At this point, try to find the property on your image. Once you have found the property on the image, note how easy or difficult it was to find, and indicate the difficulty on the scale in front of you." Subjects used a scale that ranged from *very difficult* (1) to *very easy* (9), with ratings averaged across subjects to form the scaled values.

Finally, two sensorimotor aspects of interacting with properties were scaled. P_attn assessed the likelihood of attending to a property on seeing the respective object in the world. P_handle assessed the likelihood of touching the property while interacting with the object. For P_attn, 16 new subjects were asked to imagine encountering an object in the world and then were asked, "How likely would you be to focus your attention on the following property?" Subjects provided a rating of *extremely unlikely* (1) to *extremely likely* (9), with the scaled value being the average across subjects. For P_handle, 16 new subjects were similarly asked to imagine encountering an object in the world and then were asked, "How likely would you be to touch this particular property upon encountering the object?" Subjects provided a rating of *extremely unlikely* (1) to *extremely likely* (9), with the average across subjects being the scaled value.

Scaling the expectancy predictors. A third set of predictors attempted to capture people's expectancies about the different forms that a property can take. In developing concept–property pairs, it became clear that property words are highly polysemous, typically referring to distinctly different properties across objects. For example, "nose" refers to distinctly different properties in HUMAN, HORSE, MOUSE, and AIRPLANE. Similarly, "handle" refers to different properties in DRAWER, SHOVEL, SUITCASE, and CAR. Halff, Ortony, and Anderson (1976) document the polysemy of *red* across concepts such as KETCHUP, HAIR, BRICK, and WINE (also see Solomon & Barsalou, 2001; Wisniewski, 1998).

Because the amount of polysemy a property exhibits might affect verification, a measure of polysemy was scaled (E_num_obj). Sixteen new subjects were asked to think of all the different types of objects that have a particular property (e.g., different objects with *handles*). Subjects then rated the number of objects sharing the property on a 5-point scale: *1 object* (1), *2–3 objects* (2), *4–10 objects* (3), *11–30 objects* (4), or *31–100 objects* (5). Average values of these ratings were computed for each concept–property pair.

Some property words refer to both properties and objects, thereby making them not as "good" as property words that refer only to properties. For example, "barrel" refers both to an object (e.g., a rain barrel) and to a property (e.g., a gun barrel), thereby making it a poorer property word than "nose," which never refers to an object. In another study, Solomon (1997) found that such properties were difficult to verify. To remove this irrelevant source of noise, a second expectancy measure, E_obj, was scaled to help identify such properties and to remove them from the materials. Sixteen new subjects heard each property name and rated whether their first impression of its meaning was a property (0) or an object (1), with proportions computed across subjects.

Reliability of the scalings. To establish the stability of each predictor, the reliability of its mean values was computed across individual concept–property pairs. Reliabilities were not computed for L_freq and L_letters, because these measures were not obtained from subjects. Reliable measures are important in regression, given that an unreliable predictor can attenuate correlations between the predictor and other variables, such that observed correlations underestimate their true values in the population. As Table 1 illustrates, group reliabilities ranged from .81 to .97, indicating that attenuation will not be a major factor in explaining RTs and errors.

Results

We summarize the results before presenting them in detail. As predicted, imagery subjects took longer to ver-

ify properties than did neutral subjects, suggesting that imagery subjects constructed richer simulations. Also as predicted, the associated false properties slowed RTs and increased errors, relative to unassociated false properties, consistent with the hypothesis that subjects adopted a fast word association strategy when false trials were unassociated. Further supporting this conclusion, regression analyses demonstrated that associative strength best predicted verification performance in the unassociated-neutral and unassociated-imagery conditions. When subjects in the associated-neutral and associated-imagery conditions were prevented from using this strategy, they used perceptual simulation instead. Under these conditions, perceptual variables predicted verification performance in both conditions, providing evidence of perceptual effort. Furthermore, instructional equivalence held between associated-neutral and associated-imagery subjects, with the same ordinal pattern of predictors explaining performance for each group.

Average RTs and errors. In all RT analyses, incorrect RTs were removed, as were outliers (i.e., correct true trials 2 *SD* above or below a subject's average true RT, and correct false trials 2 *SD* above or below the subject's average false RT; 4% of the data overall). The averages in Table 2 were computed across concept–property pairs (not across subjects), because all later analyses use these averages. However, statistical tests were performed both for concept–property pairs and for subjects.

As predicted, true trials were 133 msec slower when false trials were associated than when false trials were unassociated [items, $F(1,396) = 148.35, MS_e = 11,923, p < .001$; subjects, $F(1,92) = 12.67, MS_e = 32,155, p < .001$]. As Table 2 illustrates, sizable effects occurred for both imagery subjects (156 msec) and for neutral subjects (111 msec). As in previous research, increasing the relatedness of false materials slowed the processing of true materials (e.g., James, 1975; McCloskey & Glucksberg, 1979; Smith et al., 1974).

Relatedness interacted with instructions for true trials [items, $F(1,396) = 4.37, MS_e = 11,923, p < .05$; subjects, $F(1,92) = .22, MS_e = 32,155, p > .25$]. In the associated conditions, imagery subjects were 49 msec slower than neutral subjects [items, $F(1,396) = 78.71, MS_e = 11,923, p < .001$; subjects, $F(1,92) = .90, MS_e = 32,155, p > .25$].[4] As predicted, subjects appeared to construct richer representations under imagery instructions than under neutral instructions. In the unassociated conditions, imagery subjects did not differ from neutral subjects, where the difference was only 4 msec [items, $F(1,396) = .07$,

Table 2
Average Verification Times and Error Rates

Condition	Verification Times		Error Rates	
	True	False	True	False
Unassociated neutral	731	748	.05	.02
Unassociated imagery	735	760	.05	.02
Associated neutral	842	979	.07	.10
Associated imagery	891	1,001	.07	.07

$MS_e = 11,923$, $p > .25$; subjects, $F(1,92) = .01$, $MS_e = 32,155$, $p > .25$]. Imagery instructions may have had no effect in the unassociated condition because the difference in associative strength between true and false trials was so salient that imagery subjects adopted the word association strategy. The regression analyses strongly support this interpretation.

Not surprisingly, associativeness affected subjects' ability to reject false trials. Subjects in the associated conditions were 236 msec slower than subjects in the unassociated conditions [items, $F(1,396) = 509.87$, $MS_e = 10,894$, $p < .001$; subjects, $F(1,92) = 22.37$, $MS_e = 54,455$, $p < .001$], and they made .065 more errors [items, $F(1,396) = 55.97$, $MS_e = .13$ arcsin units, $p < .001$; subjects, $F(1,92) = 200.62$, $MS_e = .012$ arcsin units, $p < .001$]. Imagery versus neutral instructions had no effects on the false trials and exhibited no interactions with associativeness.

Error rates for the true trials did not differ across the four conditions, indicating no speed–accuracy tradeoff [for instructions: items, $F(1,396) = .00$, $MS_e = .13$ arcsin units, $p > .25$; subjects, $F(1,92) = .01$, $MS_e = .02$ arcsin units, $p > .25$; for associativeness: items, $F(1,396) = 2.09$, $MS_e = .13$ arcsin units, $p > .10$; subjects, $F(1,92) = 2.98$, $MS_e = .02$ arcsin units, $p < .10$].

Overview of the regressions. Although the average RTs and error rates provide a rough summary of subjects' performance, they mask critical patterns in the data. Regressions were performed to identify these patterns and to assess perceptual work and instructional equivalence.

As described earlier, there were more perceptual predictors (8) than linguistic (4) or expectancy predictors (2). However, our primary interest was not in the absolute amount of variance that a set of predictors captured. Instead, ordinal relations between predictor sets were critical. If the word association strategy is a factor, for example, linguistic predictors should dominate in the unassociated conditions, whereas perceptual predictors should dominate in the associated conditions (i.e., an ordinal interaction). Similarly, perceptual effort predicts that perceptual variables should account for more variance in the neutral condition when the word association strategy is deterred than when it is not (another ordinal prediction). Finally, instructional equivalence predicts that the overall pattern of prediction should be ordinally the same for imagery and neutral subjects, when the word association strategy is deterred. Because the predictions focus on ordinal relations, the sizes of the predictor sets are irrelevant.

Regressions for RTs: Screening predictors. The analysis of average RTs took place in three stages: screening ineffective predictors from later stages of analysis (Stage 1); assessing unique group variance (Stage 2); and assessing unique predictor variance (Stage 3).

The initial screening stage attempted to identify predictors that explained RTs effectively and to exclude predictors that did not. Decreasing the number of pre-

dictors had the effect of concentrating the analysis on the most important predictors and increasing the power of the statistical tests. Three regressions were performed during screening for each of the four between-subjects conditions: one on the linguistic predictors, one on the perceptual predictors, and one on the expectancy predictors. In each regression, the 100 average RTs in a condition were regressed onto the 100 scale values for each predictor within a group (i.e., linguistic, perceptual, or expectancy). The predictors were allowed to enter freely, thereby maximizing the chances that any particular predictor in a group would account for significant variance. Table A1 in Appendix A presents the standardized betas from these analyses.

As Table A1 illustrates, seven predictors were significant in at least one of the four conditions: L_conc_prop, P_area, P_left, P_initial, P_find, P_handle, and E_num_obj. In the remaining two stages of analyzing the average RTs, only these seven predictors were included.

Due to extensive shared variance between predictors, care is essential in interpreting the significance of individual predictors. In general, the predictors formed intercorrelated clusters, with one predictor in each cluster typically winning the competition during screening. Thus a predictor's lack of significance does not necessarily imply no role in the verification process.

Regressions for RTs: Assessing unique group variance. The second stage of analysis established how much unique variance each group of predictors—linguistic, perceptual, and expectancy—explained as a whole in the RTs. To establish how much unique variance a group explained, it was first necessary to establish how much variance all three groups explained together. Thus, the seven significant predictors from the screening analysis were allowed to enter freely into a regression of each condition's average RTs. Table 3 presents the total variance explained in each of these four regressions (i.e., the row for RTs under Total Variance). As can be seen, the seven predictors significant in the screening analysis explained 34% to 51% of the variance across the four conditions. Note that these percentages include shared variance between the predictors, as well as variance that each predictor explains uniquely.

Once the total variance was established for each condition, it was possible to establish how much unique variance each group of predictors explained. Assessing a group's unique variance provides the clearest means of establishing the importance of its predictors in explaining RTs. Specifically, a group's unique variance represents the explanatory value of its predictors independently of the other predictors. It provides a conservative assessment of a group's importance, excluding variance that the group shares with the other two groups.

To establish unique variance, three further regressions were performed in each condition. In each, the predictor group of interest was excluded from a regression that allowed the significant predictors from the other two groups to enter freely. For example, to establish unique

Table 3
Percentage of Unique Regression Variance Explained for RTs, for Errors,
and for RTs and Errors Combined

Predictors/Measure	Unassociated		Associated	
	Neutral	Imagery	Neutral	Imagery
Linguistic				
RTs	17*	17*	8*	6*
Errors	14*	10*	3*	4*
RTs and errors combined	14*	16*	5*	4*
Perceptual Work				
RTs	9*	9*	16*	27*
Errors	17*	6	21*	14*
RTs and errors combined	14*	10*	24*	26*
Expectancy				
RTs	2	1	7*	9*
Errors	2	1	9*	10*
RTs and errors combined	2	2	9*	10*
Total Variance (unique and shared)				
RTs	46*	51*	34*	44*
Errors	40*	33*	38*	29*
RTs and errors combined	46*	50*	41*	42*

*$p < .05$.

variance for the linguistic predictors in the unassociated-neutral condition, the average RTs in this condition were regressed onto the perceptual and expectancy predictors. The total variance explained in this regression was then subtracted from the initial regression in the unassociated-neutral condition that included all seven significant predictors from all three groups, leaving the unique linguistic variance. In other words, the unique linguistic variance was the difference between the regression with all seven predictors and the regression with only the perceptual and expectancy predictors. It is the variance that only the linguistic predictors explain.

The first row of Table 3 illustrates that the unique linguistic variance ranged from 6% to 17%. As predicted, the linguistic predictors explained more unique variance in the two unassociated conditions (17%) than in the two associated conditions (7%). When the false properties were unassociated, neutral subjects used the word association strategy to verify properties. Because associative strength predicted correct responses, these subjects capitalized on this strategy.

Interestingly, unassociated-imagery subjects used the word association strategy as much as unassociated-neutral subjects, consistent with the earlier lack of an imagery effect on true RTs in these two conditions. Even though these subjects were asked to use imagery, they adopted the word association strategy because of the salient difference between true and false trials in associativeness.

Analogous to the regressions that established the unique linguistic variance in the average RTs, further regressions established the unique perceptual variance, as well as the unique expectancy variance. Table 3 presents the results of these analyses. As predicted, the perceptual predictors explained more unique variance in the two associated conditions (21.5%) than in the two unassociated conditions (9%). Because subjects in the two as-

sociated conditions could not use the word association strategy, they had to adopt some other strategy, which appeared to be simulation.

As Table 3 illustrates, perceptual predictors captured more variance for associated-imagery subjects (27%) than for neutral subjects (16%). It might initially appear that instructional equivalence does not hold. However, two considerations suggest otherwise. First, we are primarily concerned with ordinal patterns of prediction. As Table 3 illustrates, the same ordinal pattern held for the associated-neutral and associated-imagery conditions. For both, the perceptual predictors captured the greatest unique variance, followed by the expectancy predictors, and finally the linguistic predictors. At the ordinal level, instructional equivalence occurred. Second, we will see shortly that the perceptual predictors captured more variance for *errors* in the associated-neutral condition than in the associated-imagery condition. Furthermore, when later regressions were performed for RTs and errors combined, the perceptual predictors captured *equal* variance in the associated-neutral and associated-imagery conditions. For the combined measure, instructional equivalence held not just at the ordinal level, but more precisely as well.

Finally, Table 3 presents the unique variance captured by the expectancy predictors. As can be seen, the expectancy predictors played a role in the two associated conditions but not in the two unassociated conditions. Because one particular expectancy predictor was responsible for this pattern, E_num_obj, we address it in later sections on individual predictors.

Regressions for RTs: Assessing unique predictor variance. The third and final stage of analysis established the unique variance for each predictor significant in the screening analysis. The same procedure used to establish unique group variance was used to establish

unique predictor variance. For each significant predictor, one regression was performed on the other six significant predictors, excluding the predictor of interest (once in each condition). The variance explained by each of these regressions was subtracted from the regression on all seven significant predictors in the same condition, leaving the unique variance for the predictor of interest. For example, to compute the unique variance for L_conc_prop in the unassociated-neutral condition, the average RTs in this condition were regressed onto the other six significant predictors from the screening analysis, excluding L_conc_prop. The variance in this regression was then subtracted from the regression on all seven significant predictors in the unassociated-neutral condition, leaving the unique L_conc_prop variance.

Table B1 in Appendix B presents the results of this analysis for the RTs. As can be seen, L_conc_prop was responsible for the importance of the linguistic variables as a group. The ease of generating a property from a concept appeared critical to the word association strategy. As associative strength increased, RTs decreased.

As Table B1 further illustrates, P_area was primarily responsible for the importance of perceptual predictors. As properties became larger, they took longer to verify. One interpretation of this effect is that subjects must simulate a property before they can search for it in an image, with large properties taking longer to simulate than small ones. Previous imagery research is consistent with this conclusion. When subjects are explicitly asked to image objects, the time to image an object increases with its complexity (e.g., Kosslyn et al., 1988; Kosslyn et al., 1983). Analogously, when subjects verify properties, large properties take longer to simulate than small ones, thereby increasing the duration of the verification process.

In summary, the interaction of the linguistic and perceptual predictors with the unassociated and associated conditions in Table 3 illustrates a major shift in verification strategies. When false trials were unassociated, the word association strategy dominated. When the false trials were associated, perceptual simulation dominated.

Because regressions on group data may not provide an accurate account of individual subjects' performance, regressions on individual subjects are necessary (e.g., Lorch & Myers, 1986). In such regressions, Solomon (1997) again found that linguistic variables captured the most variance in the unassociated conditions, whereas perceptual variables captured the most variance in the associated conditions. Thus the group regressions are representative of individual subjects.

P_handle and analogical verification. As Table B1 in Appendix B illustrates, P_handle was the second most important predictor of the RTs after P_area. Counterintuitively, however, the more likely a part was to be handled, the longer it took to verify. One might think that properties handled frequently should be verified quickly. Another experiment in Solomon (1997) showed the same correlation, indicating that this is not a spurious result.

Post hoc analyses suggest an interpretation of this result that begins with the phenomenon of *analogical verification*, demonstrated in Solomon and Barsalou (2001). In this phenomenon, subjects use a property in one concept (e.g., *mane* for HORSE) to verify an analogous property in another concept (e.g., *mane* for PONY). When verifying *mane* for PONY, the representation of *mane* accessed initially is *horse mane*, because it is the most dominant sense of *mane* available in memory. Once retrieved, *horse mane* is used analogically to verify that *mane* is true of PONY. Because *horse mane* matches the contents of the analogous region in a PONY simulation reasonably well, a true response follows. Note, however, that a true response has been reached analogically via *horse mane*, not directly via *pony mane*.

The unusual effects of P_handle appear to reflect analogical verification. From examining individual concept–property pairs, it became clear that pairs low on P_handle had properties that consistently occurred in the same location across objects. For example, *tail* is handled infrequently but occurs fairly consistently in the same location across objects that have tails (e.g., HORSE, CAT, ROBIN). Conversely, pairs high on P_handle had properties that tend to occur in different locations across objects. Thus, *button* is handled frequently but occurs in different locations across objects that have buttons (e.g., SHIRTS, CALCULATORS, DOORBELLS).

To confirm this relationship, two judges rated each property for the consistency of its location across objects on a 1 to 5 scale (1 = *a single location*, 3 = *a few*, and 5 = *many*). The judges' ratings correlated .82. When disagreements arose, the average of their two ratings for a property was taken. The correlation of this measure with P_handle was .34 ($p < .01$). When a property had a low P_handle rating, it tended to occur consistently in the same location across objects and could therefore benefit from analogical verification. If an analogous property in a different concept is retrieved initially, it can be used to verify the property in the target concept. Because it is not necessary to retrieve the target property itself—which might be less accessible—verification is speeded. Conversely, properties rated high on P_handle are more likely to vary in location, making analogical verification less available and increasing response time.

An alternative hypothesis about the importance of P_handle is that it simply reflects the distinction between artifacts and natural kinds. If people handle artifacts more often than natural kinds, and if artifacts are verified more slowly than natural kinds, the effects of P_handle could reflect this difference. Because subjects did verify artifacts more slowly than natural kinds (822 vs. 780 msec), we assessed this relationship by creating another new predictor, namely, whether each concept was an artifact or a natural kind.

The regression analyses of unique predictor variance were redone with the addition of the location predictor and the artifact/natural kind predictor. Across the four conditions, adding these two predictors increased the

total explained variance by 6% (from 44% to 50%). Thus, these two predictors captured significant variance not explained by any of the preexisting predictors. However, the location predictor—not the artifact/natural kind predictor—was primarily responsible. Across the unassociated-neutral, unassociated-imagery, associated-neutral, and associated-imagery conditions, the location predictor explained 2%, 3%, 5%, and 4% of the unique variance, respectively (significant in the unassociated-imagery condition and in the two associated conditions). In contrast, the artifact/natural kind predictor explained 2%, 0%, 0%, and 0% of the unique variance (none significant). This pattern indicates that increasing consistency in a property's location speeded its verification, consistent with the hypothesis that subjects used analogical verification when possible. Whether a concept was an artifact or natural kind was not important. Finally, the unique variance for P_handle dropped from 2%, 2%, 4%, and 4% across conditions to 0%, 0%, 1%, and 2%, all now insignificant. Thus, the effects of P_handle appeared to reflect the ease of analogical verification. To the extent that a property's location was consistent across objects, analogical verification was available and speeded verification time.

E_num_obj and opportunities for analogical verification. As Table 3 illustrates, the expectancy predictors accounted for significant unique variance in the associated-neutral and associated-imagery conditions (7% and 9%). As Table B1 in Appendix B shows, E_num_obj was responsible for this effect. Surprisingly, however, increases in E_num_obj were associated with *decreases* in RTs. Intuitively, the more objects sharing a property, the more interference they should produce in verifying it for any one concept. If so, then E_num_obj should have correlated positively with RTs, not negatively.

Again, analogical verification provides a natural explanation of this finding. As more objects share a property, the opportunity for analogical verification increases, and the time to retrieve a relevant property sense decreases, thereby speeding verification (see Logan, 1988). Notably, this account further explains why E_num_obj did not capture unique variance in the unassociated conditions (Table 3). Because subjects in these conditions primarily used the word association strategy—not perceptual simulation—analogical verification did not occur, and the number of concepts sharing a property was irrelevant.

Regressions for group error rates. The error rates for individual concept–property pairs provide another means of assessing the critical issues. Just as RTs reflect the difficulty of verifying particular properties, so do errors. Thus, the procedure for performing regressions on RTs was applied to the error rates. In the screening analysis, eight predictors were significant in at least one condition, and were therefore used to compute unique variance later: L_conc_prop, P_area, P_left, P_initial, P_sal, P_find, P_handle, and E_num_obj. Together these predictors explained 29% to 40% of the variance across conditions (Table 3). Table A1 in Appendix A presents the betas from the screening analyses. Table B2 in Appendix B pre-

sents the unique variance explained by individual predictors that passed the screening analysis.

As Table 3 illustrates, the pattern of prediction for errors was similar to the pattern for RTs. Whereas the linguistic predictors were more important in the unassociated conditions than in the associated conditions (12% vs. 3.5%), the perceptual predictors were more important in the associated conditions than in the unassociated conditions (17.5% vs. 11.5%). Again this confirms the prediction that unassociated subjects use word association, whereas associated subjects use perceptual simulation. The finding that the expectancy predictor, E_num_obj, only predicted errors in the associated conditions further supports the conclusion that analogical verification was only used in conjunction with perceptual simulation.

Perceptual simulation in the unassociated-neutral condition. As Table 3 illustrates, 17% of the error variance in the unassociated-neutral condition was explained by perceptual variables. If these subjects were using the word association strategy, why would perceptual variables explain their error variance?

One hypothesis is that unassociated-neutral subjects used perceptual simulation when the diagnosticity of word association as a response cue was low. When a concept and property were weakly associated, subjects looked for a more diagnostic source of information, namely, searching an object simulation for the property. To test this hypothesis, the 100 true pairs were divided into two halves, based on a median split of L_conc_prop (i.e., associative strength). In the unassociated-neutral condition only, average correct RTs for the 50 pairs in each half were then regressed onto the three predictor sets. If unassociated-neutral subjects relied on perceptual simulation as associative strength decreased, then perceptual predictors should not explain RTs in the highly associated half of the true materials, but should explain RTs in the weakly associated half.

This pattern did indeed occur. In the highly associated half of the true materials, the perceptual predictors captured only 4% of the unique RT variance. In the weakly associated half, however, the perceptual predictors captured 21%. For unassociated-neutral subjects, perceptual predictors were roughly five times more important for weakly associated properties than for strongly associated ones. When associativeness was nondiagnostic, these subjects fell back on simulation.

Further evidence for this conclusion can be found in Solomon (1997). In another experiment that had a much broader range of values for L_conc_prop, the true materials were divided into four quartiles based on L_conc_prop. From the highest associated quartile to the lowest, the perceptual variables captured 4%, 11%, 18%, and 36% of the unique RT variance, respectively, in the unassociated-neutral condition. In both experiments, unassociated-neutral subjects fell back on simulation as word association became less diagnostic.

Differential roles of perceptual variables for neutral and imagery subjects. As Table 3 illustrates, the following interaction occurred in the two associated con-

ditions. For RTs, the perceptual variables explained more variance for associated imagery subjects (27%) than for associated neutral subjects (16%). Conversely, for errors, the perceptual variables explained more variance for associated neutral subjects (21%) than for associated imagery subjects (14%). Two factors appear responsible for this interaction: verification speed and property size. First, imagery subjects verified properties more slowly than did neutral subjects, suggesting that imagery subjects constructed more complete simulations (Table 2). Second, large properties generally had longer RTs than smaller ones, and also produced more errors (see the positive correlations of P_area with RTs and errors in Appendix B). The combined effects of speed and size may have led perceptual variables to play different roles for neutral and imagery subjects. Because imagery subjects took longer to verify properties, they had extra time to simulate large properties (which take longer to simulate than small properties). Because neutral subjects responded faster, they were more likely to omit large properties in their simulations, thereby erring on them more.

Evidence for this interpretation can be found in the measure for size (P_area). As Tables B1 and B2 of Appendix B illustrate, P_area predicted RTs better for imagery subjects (16%) than for neutral subjects (5%). Conversely, P_area predicted error rates better for neutral subjects (10%) than for imagery subjects (7%). Because imagery subjects were likely to include large properties in their object simulations (not just smaller ones), their simulations contained properties that varied widely in size. As a result, P_area had much potential to explain the time required to process different-sized properties. Conversely, because neutral subjects were less likely to include large properties in their object simulations, size variability was reduced, and P_area was not as effective.

A similar account explains why P_area predicted errors better for neutral than for imagery subjects. Because neutral subjects simulated large properties less often than smaller ones, P_area predicted their errors, with errors occurring more often for large properties. Conversely, because imagery subjects were more likely to simulate properties of all sizes, they were less biased to err on large properties, so that P_area did not predict errors as well.

Further evidence for this conclusion is that P_initial (the probability than an initial object simulation contained the target property) was more important for neutral than for imagery subjects. As Tables B1 and B2 illustrate, P_initial explained 4% of the RT variance for neutral subjects and 7% of the error variance. For imagery subjects, P_initial only explained 2% of the variance for both RTs and errors. Furthermore, P_initial was more important for explaining the error variance of neutral subjects than for explaining their RT variance (7% vs. 4%). This pattern suggests that neutral subjects omitted properties in their simulations more often than imagery subjects, with these omissions having more impact on errors than RTs.

To assess this hypothesis further, we identified large properties that were absent in subjects' initial images for further analysis (i.e., properties above the median on

P_area but below the median on P_initial). Examples of such properties include CAR–*hood*, OAK–*bark*, SALMON–*scales*, and OWL–*chest*. As predicted, the error rates for the 15 properties meeting these criteria were twice as high in the associated-neutral condition (.08) as in the associated-imagery condition (.04). Because the number of relevant properties was small, this difference only approached significance [$F(1,14) = 2.34, MS_e = .043, p < .15$]. Nevertheless, the direction of this difference is consistent with the larger pattern of findings indicating that response speed underlies the ability of perceptual predictors to explain errors better for neutral subjects than for imagery subjects. Large properties took relatively longer to simulate, occurred less often in neutral subjects' sketchier simulations, and therefore produced more errors than could be explained by perceptual variables.

Regressions for combined RTs and error rates. As we just saw, perceptual predictors explained more variance for RTs in the associated-imagery condition, but explained more variance for error rates in the associated-neutral condition. This suggests that RTs and error rates tap different aspects of the verification process, and that both must be assessed to explain it. To assess this possibility, the average verification time for each concept–property pair was transformed to a z score, as was the error rate for the pair. The two z scores for a pair were then averaged to establish an overall measure of the difficulty in verifying it. As these average z scores increased, the difficulty of verifying a pair increased, as reflected in both verification time and errors.

The average z scores were submitted to regressions. In the screening analysis, seven predictors were significant in at least one condition, and were therefore used to compute unique variance: L_conc_prop, P_area, P_left, P_initial, P_find, P_handle, and E_num_obj. Together these predictors explained 41% to 50% of the variance across conditions (Table 3). Table A1 in Appendix A presents the betas from the screening analyses. Table B3 in Appendix B presents the unique variance explained by individual predictors that passed the screening analysis.

As Table 3 illustrates, the pattern of prediction unifies the patterns for RTs and error rates. The linguistic predictors explained more unique variance in the unassociated conditions than in the associated conditions (15% vs. 4.5%), indicating the prevalence of the word association strategy when the false trials allowed it. Conversely, the perceptual predictors explained more unique variance in the associated conditions than in the unassociated conditions (25% vs. 12%), indicating the prevalence of perceptual simulation when the false trials deterred word association. However, the perceptual predictors also explained unique variance in the unassociated conditions because unassociated subjects fell back on perceptual simulation for weakly associated properties when word association was not a diagnostic cue (as described earlier). Finally, the expectancy predictors again explained significant unique variance in the associated conditions, because these subjects used analogical verification together with perceptual simulation.

These results provide an even stronger case for instructional equivalence. When RTs and errors are considered together, the overall pattern of prediction is nearly identical for the associated-neutral and associated-imagery conditions. Clearly, these two conditions differ somewhat in more detailed analyses of performance; namely, perceptual factors contribute more to RTs in the associated-imagery condition but contribute more to error rates in the associated-neutral condition. Regardless, perceptual predictors enter into both conditions overall at comparable levels.

Discussion

These results support the a priori predictions of perceptual views. When the word association strategy was available in the unassociated conditions, subjects adopted it, as indicated by the importance of linguistic predictors. Conversely, when the word association strategy was deterred in the associated conditions, linguistic predictors became less important, and perceptual predictors became more important. Consistent with the predictions of perceptual views, subjects used perceptual simulation to verify properties when they could not use associative strength. These results support the hypothesis that perceptual effort should occur not only in the associated-imagery condition, but most importantly, in the associated- neutral condition.

Ordinally, the pattern of prediction was the same for the associated-neutral and associated-imagery conditions, consistent with the prediction of instructional equivalence. For both conditions, perceptual predictors were most important, followed by expectancy predictors, and then linguistic predictors. At a more detailed level, the perceptual predictors explained more variance for RTs in the associated-imagery condition, but explained more variance for error rates in the associated-neutral condition. Because neutral subjects responded faster than imagery subjects, they appeared less likely to simulate large properties, thereby being more likely to err on them. When both RTs and errors were combined into a single measure, the perceptual predictors explained nearly identical variance in the associated-neutral and associated-imagery conditions. Thus, when the larger process of verification was considered as a whole— including both RTs and errors—instructional equivalence held at a precise level, not just ordinally.

Perceptual simulation even played a role in the unassociated-neutral condition. Although these subjects primarily used word association, they increasingly constructed simulations when associative strength was not diagnostic for correct responses. As a result, perceptual predictors increasingly explained these subjects' performance as verifying properties became more difficult.

Replication of these findings comes from two sources. First, Solomon (1997) performed a similar experiment with different materials and found the same pattern of results. When unassociated false trials allowed the word association strategy, subjects adopted it. When associated false trials deterred this strategy, perceptual effort

and instructional equivalence were both observed. Regressions for associated-neutral subjects were ordinally the same as those for associated-imagery subjects, and perceptual variables accounted for substantial unique variance in both conditions.

Second, Kan, Barsalou, Solomon, Minor, and Thompson-Schill (2003) ran the unassociated-neutral and associated-neutral conditions in an fMRI scanner. Of primary interest, the left fusiform gyrus was active in associated-neutral subjects but not in unassociated-neutral subjects. Much previous work has shown that the left fusiform plays central roles in the generation of mental images. Thus, fusiform activation in Kan et al.'s associated-neutral subjects corroborates our conclusion here that associated-neutral subjects used simulation to verify properties. Conversely, the lack of fusiform activation in Kan et al.'s unassociated-neutral subjects corroborates our conclusion here that unassociated-neutral subjects relied heavily on word association.

Amodal accounts of perceptual effects in conceptual processing. Theorists have argued that amodal theories have the expressive power to represent any information, including the information in images (e.g., Anderson, 1978; Pylyshyn, 1973, 1981). On the basis of these arguments, many researchers believe that amodal views can explain all imagery phenomena. These researchers might further believe that amodal theories can explain all perceptual effects in conceptual processing. Such accounts, however, should be placed in the proper perspective: They are post hoc, not a priori, and they are possible because amodal theories are unfalsifiable. From the perspectives of inferential statistics, Bayesian reasoning, and philosophy of science, such accounts should be viewed with caution. If a priori prediction is favored over post hoc prediction in these other areas, should it not also be favored here? If perceptual theories motivate the experiments that produce perceptual effects in conceptual processing, why should explanatory equivalence between perceptual and amodal theories be viewed as a tie?

We realize that many researchers are deeply invested in amodal views (as we once were), and that they will want to explore amodal accounts of our findings, even if post hoc. Thus, we present two amodal accounts of our results: *tacit knowledge* and *perceptual carryover in transduction*. Both apply not only to our results here but to perceptual effects more generally.

Tacit amodal knowledge. According to the tacit knowledge position, subjects know how to perform *as if* they are using perceptual simulation to verify properties, while really using amodal knowledge (Goldston, Hinrichs, & Richman, 1985; Mitchell & Richman, 1980; Pylyshyn, 1981). In other words, subjects possess amodal instructions for how to act perceptually, and then execute these instructions during performance when task demands suggest that doing so is appropriate.

Several problems exist for this account. First, nothing in the instructions to associated-neutral subjects should have motivated them to exhibit perceptual effects, yet

these subjects did so. Indeed, the tacit knowledge position predicts a priori that associated-neutral subjects should not have exhibited perceptual effects, given that the instructions were nonperceptual.

Second, unassociated-neutral subjects received exactly the same instructions as associated-neutral subjects, yet the two groups performed quite differently. Whereas linguistic predictors best explained the performance of unassociated-neutral subjects, perceptual predictors best explained the performance of associated-neutral subjects. If implicit task demands from the instructions had controlled subjects' use of tacit knowledge, these two groups should have performed similarly. One could argue that perceived differences in the difficulty of the false trials created different demands. However, it is implausible that subjects would know that they should act perceptually for associated false properties but linguistically for unassociated false trials.

Finally, associated-imagery subjects received very different instructions than associated-neutral subjects but performed quite similarly. If imagery instructions had generated different task demands, the two groups should have diverged. To the contrary, their similar performance again suggests that tacit knowledge was not a factor.

Perceptual carryover in transduction. Amodal views assume that perceptual states are transduced into amodal representations that support cognitive processes (Barsalou, 1999b). Perhaps the nature of this transduction process causes perceptual variables to become established implicitly in amodal representations. For example, occlusion effects could arise because unoccluded properties are stored earlier in amodal sequences of symbols than occluded properties. Similarly, size effects could arise because small properties are stored earlier than large properties. As these examples illustrate, it is conceivable that the transduction process preserves perceptual structure in amodal knowledge implicitly. During the processing of properties, this implicit structure enters into processing, appearing to give it a perceptual character.

Problematically, this view does not reflect the a priori spirit of amodal theories, which typically assume that conceptual knowledge abstracts over perceptual details. No amodal view to our knowledge has ever predicted that perceptual effects like those observed here should occur. Indeed, amodal views generally assume that fundamentally different factors structure perception and conception, not the same. Thus, the perceptual carryover view is post hoc, motivated solely by results that perceptual views predict a priori.

We hasten to add, however, that our experiments do not rule out the possibility that amodal symbols reside in the cognitive system (see Simmons & Barsalou, 2003). They may play roles in concepts and processes that we have not studied, and they may enter into the concepts that we have studied under other circumstances. Our primary argument is simply that perceptual simulation explains the results here in a more natural and a priori man-

ner than do amodal accounts. These results suggest to us, at least, that simulation plays a central role in representing these concepts.

REFERENCES

ANDERSON, J. R. (1978). Arguments concerning representations for mental imagery. *Psychological Review*, **85**, 249-277.

BARSALOU, L. W. (1999a). Language comprehension: Archival memory or preparation for situated action. *Discourse Processes*, **28**, 61-80.

BARSALOU, L. W. (1999b). Perceptual symbol systems. *Behavioral & Brain Sciences*, **22**, 577-660.

BARSALOU, L. W. (2003a). Abstraction in perceptual symbol systems. *Philosophical Transactions of the Royal Society of London: Series B*, **358**, 1177-1187.

BARSALOU, L. W. (2003b). Situated simulation in the human conceptual system. *Language & Cognitive Processes*, **18**, 513-562.

BARSALOU, L. W., NIEDENTHAL, P. M., BARBEY, A., & RUPPERT, J. (2003). Social embodiment. In B. Ross (Ed.), *The psychology of learning and motivation* (Vol. 43, pp. 43-92). San Diego: Academic Press.

CLARK, A. (1997). *Being there*: *Putting brain, body, and world together again.* Cambridge, MA: MIT Press.

COHEN, J., & COHEN, P. (1983). *Applied multiple regression/correlation analysis for the behavioral sciences*. Mahwah, NJ: Erlbaum.

COLLINS, A. M., & QUILLIAN, M. R. (1969). Retrieval time from semantic memory. *Journal of Verbal Learning & Verbal Behavior*, **8**, 240-248.

CONRAD, C. (1972). Cognitive economy in semantic memory. *Journal of Experimental Psychology*, **92**, 149-154.

DAMASIO, A. R. (1989). Time-locked multiregional retroactivation: A systems-level proposal for the neural substrates of recall and recognition. *Cognition*, **33**, 25-62.

FARAH, M. J. (2000). The neural bases of mental imagery. In M. S. Gazzaniga (Ed.), *The cognitive neurosciences* (2nd ed., pp. 965-974). Cambridge, MA: MIT Press.

FINKE, R. A. (1989). *Principles of mental imagery*. Cambridge, MA: MIT Press.

GLASER, W. R. (1992). Picture naming. *Cognition*, **42**, 61-106.

GLASS, A. L., & HOLYOAK, K. J. (1975). Alternative conceptions of semantic memory. *Cognition*, **3**, 313-339.

GLENBERG, A. M. (1997). What memory is for. *Behavioral & Brain Sciences*, **20**, 1-55.

GOLDSTON, D. B., HINRICHS, J. V., & RICHMAN, C. L. (1985). Subjects' expectations, individual variability, and the scanning of mental images. *Memory & Cognition*, **13**, 365-370.

GRÈZES, J., & DECETY, J. (2001). Functional anatomy of execution, mental simulation, observation, and verb generation of actions: A meta-analysis. *Human Brain Mapping*, **12**, 1-19.

HALFF, H. M., ORTONY, A., & ANDERSON, R. C. (1976). A context-sensitive representation of word meanings. *Memory & Cognition*, **4**, 378-383.

JAMES, C. (1975). The role of semantic information in lexical decisions. *Journal of Experimental Psychology: Human Perception & Performance*, **104**, 130-136.

JEANNEROD, M. (1995). Mental imagery in the motor context. *Neuropsychologia*, **33**, 1419-1432.

KAN, I. P., BARSALOU, L. W., SOLOMON, K. O., MINOR, J. K., & THOMPSON-SCHILL, S. L. (2003). Role of mental imagery in a property verification task: fMRI evidence for perceptual representations of conceptual knowledge. *Cognitive Neuropsychology*, **20**, 525-540.

KOSSLYN, S. M. (1980). *Image and mind*. Cambridge, MA: Harvard University Press.

KOSSLYN, S. M. (1994). *Image and brain: The resolution of the imagery debate*. Cambridge, MA: MIT Press.

KOSSLYN, S. M., BALL, T., & REISER, B. J. (1978). Visual images preserve metric spatial information: Evidence from studies of image scanning. *Journal of Experimental Psychology: Human Perception & Performance*, **4**, 47-60.

KOSSLYN, S. M., CAVE, C. B., PROVOST, D. A., & VON GIERKE, S. M. (1988). Sequential processes in image generation. *Cognitive Psychology*, **20**, 319-343.

KOSSLYN, S. M., REISER, B. J., FARAH, M. J., & FLIEGAL, S. L. (1983). Generating visual images: Units and relations. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **112**, 278-303.

KUCĚRA, H., & FRANCIS, W. N. (1967). *Computational analysis of present-day American English*. Providence, RI: Brown University Press.

LAKOFF, G. (1987). *Women, fire, and dangerous things: What categories reveal about the mind*. Chicago: University of Chicago Press.

LOGAN, G. D. (1988). Toward an instance theory of automatization. *Psychological Review*, **95**, 492-527.

LORCH, R. F., & MYERS, J. L. (1986). Regression analyses of repeated measures data in cognitive research. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **16**, 149-157.

MANDLER, J. M. (1992). How to build a baby: II. Conceptual primitives. *Psychological Review*, **99**, 587-604.

MARTIN, A. (2001). Functional neuroimaging of semantic memory. In R. Cabeza & A. Kingstone (Eds.), *Handbook of functional neuroimaging of cognition* (pp. 153-186). Cambridge, MA: MIT Press.

MARTIN, A., & CHAO, L. (2001). Semantic memory and the brain: Structure and process. *Current Opinion in Neurobiology*, **11**, 194-201.

MARTIN, A., UNGERLEIDER, L. G., & HAXBY, J. V. (2000). Category-specificity and the brain: The sensory-motor model of semantic representations of objects. In M. S. Gazzaniga (Ed.), *The new cognitive neurosciences* (2nd ed., pp. 1023-1036). Cambridge, MA: MIT Press.

MCCLOSKEY, M., & GLUCKSBERG, S. (1979). Decision processes in verifying category membership statements: Implications for models of semantic memory. *Cognitive Psychology*, **11**, 1-37.

MITCHELL, D. B., & RICHMAN, C. L. (1980). Confirmed reservations: Mental travel. *Journal of Experimental Psychology: Human Perception & Performance*, **6**, 58-66.

NEWTON, N. (1996). *Foundations of understanding*. Philadelphia: John Benjamins.

PAIVIO, A. (1986). *Mental representations: A dual coding approach*. New York: Oxford University Press.

PYLYSHYN, Z. W. (1973). What the mind's eye tells the mind's brain: A critique of mental imagery. *Psychological Bulletin*, **80**, 1-24.

PYLYSHYN, Z. W. (1981). The imagery debate: Analogue media versus tacit knowledge. *Psychological Review*, **88**, 16-45.

ROSCH, E., & MERVIS, C. B. (1975). Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology*, **7**, 573-605.

SHEPARD, R. N., & COOPER, L. A. (1982). *Mental images and their transformations*. New York: Cambridge University Press.

SIMMONS, K., & BARSALOU, L. W. (2003). The similarity-in-topography principle: Reconciling theories of conceptual deficits. *Cognitive Neuropsychology*, **20**, 451-486.

SMITH, E. E., SHOBEN, E. J., & RIPS, L. J. (1974). Structure and process in semantic memory: A featural model for semantic decisions. *Psychological Review*, **81**, 214-241.

SOLOMON, K. O. (1997). *The spontaneous use of perceptual representations during conceptual processing*. Unpublished doctoral dissertation, University of Chicago.

SOLOMON, K. O., & BARSALOU, L. W. (2001). Representing properties locally. *Cognitive Psychology*, **43**, 129-169.

WISNIEWSKI, E. J. (1998). Property instantiation in conceptual combination. *Memory & Cognition*, **26**, 1330-1347.

WU, L., & BARSALOU, L. W. (2004). *Perceptual simulation in property generation*. Manuscript submitted for publication.

NOTES

1. We use quotes to indicate words ("pony," "mane"), small capital letters to indicate object concepts (PONY), and lowercase italics to represent property concepts (*mane*).

2. Solomon and Barsalou (2001) showed that the simulation effects found here for a 1,700-msec stimulus onset asynchrony (SOA) also occur for a 250-msec SOA, suggesting that they reflect automatic processing at least to some extent.

3. Somewhat different measures were used to scale L_conc_prop and L_prop_conc for the following reason. Optimally, we preferred generation measures over rating measures, given that generation seemed closer than rating to the mechanism that we were trying to measure. For L_conc_prop, however, the likelihood of generating a particular property was very low, which was why we used the rating task. For L_prop_conc, the likelihood of generating the concept is much higher, which enabled use of the generation measure for it.

4. The subjects *F* is not significant because the imagery versus neutral manipulation was between subjects, who varied considerably in overall speed, ranging from very fast to very slow. The items ANOVAs have more power given that, for items, imagery versus neutral instructions was a repeated measure, thereby neutralizing item variability. In another experiment, Solomon (1997) similarly found a 134-msec effect of imagery instructions that, for the same reasons, was highly significant in the items analysis ($p < .0001$) but not in the subjects analysis.

APPENDIX A
Standardized Betas for Initial Variance Explained in the Regressions

The standardized betas in Table A1 below were obtained from regressions in which all potential predictors from a group of predictors were allowed to enter freely.

Table A1
Standardized Betas for Predictors' Ability to Explain Variance in Group Verification Times, Error Rates, and Combined Verification Times and Error Rates

	Unassociated						Associated					
	Neutral			Imagery			Imagery			Neutral		
Predictor	RTs	Errors	Both	RTs	Errors	Both	RTs	Errors	Both	RTs	Errors	Both
Linguistic												
L_conc_prop	−.55*	−.38*	−.52*	−.64*	−.57*	−.66*	−.38*	−.18	−.31*	−.27*	−.19	−.24
L_prop_conc	−.04	−.10	−.04	−.02	.03	.00	.05	.15	.11	.07	.14	.12
L_freq	.05	−.05	−.02	.01	−.13	−.09	−.15	−.07	−.14	−.12	−.15	−.13
L_letters	.02	−.08	−.01	.10	.04	.10	.11	.00	.07	.16	−.05	.09
Perceptual												
P_area	.20	.16	.18	.29*	.20	.23*	.29*	.36*	.40*	.53*	.34*	.49*
P_bottom	−.03	−.17	−.09	.04	.02	.01	−.10	−.09	−.07	.20	.10	.17
P_left	.26*	.24*	.30*	.15	.04	.14	.12	.09	.14	−.01	.03	.02
P_initial	−.10	−.23	−.19	−.10	−.25	−.21	−.26*	−.42*	−.33*	−.32*	−.27*	−.34*
P_sal	−.05	.15	.11	−.25	−.23	−.16	.09	.34*	.27	.15	.24	.26
P_find	−.22	−.48*	−.36*	−.29*	−.31*	−.36*	−.12	−.32*	−.26*	−.33*	−.25	−.36*
P_attn	−.24	.07	−.16	−.03	.24	.03	−.26	−.04	−.22	−.01	−.07	−.04
P_handle	.15	.22*	.21	.13	.10	.15	.21	.16	.23*	.30*	.22	.30*
Expectancy												
E_num_obj	.26*	.13	.21*	.30*	.21*	.25*	−.07	−.27*	−.17	−.17	−.25*	−.21*
E_object	−.05	−.06	−.07	.06	.02	.04	.01	−.04	−.02	−.05	−.11	−.08

*$p < .05$.

APPENDIX B
Semipartial Correlation Coefficients and Percentages of Unique Variance in the Regressions

Each of the three tables to follow shows the semipartial correlation coefficient (sr) and the percentage of unique variance explained (%Var) for one predicted measure (RTs, errors, or combined RTs and errors). The predictors in each table are those that passed the screening stage. The unique variance for each predictor was established using the procedure described in the text. The semipartial correlation is the correlation between a predicted measure (RTs or errors) and the portion of a predictor's variance that is uncorrelated with the remaining predictors. It was obtained by taking the square root of a predictor's unique variance and giving it the sign of its beta in the regression on all predictors passing the screening stage. Significant predictors were those whose contributions differed significantly ($p < .05$) from zero using the F statistic, as suggested by Cohen and Cohen (1983).

Table B1
Semipartial Correlation Coefficients (sr) and Percent Variance Explained in Regressions That Assess Unique Contributions to Reaction Times

	Unassociated				Associated			
	Neutral		Imagery		Neutral		Imagery	
Predictor	sr	%Var	sr	%Var	sr	%Var	sr	%Var
Linguistic								
L_conc_prop	−.41*	17*	−.41*	17*	−.28*	8*	−.24*	6*
Perceptual								
P_area	.12	1	.18*	3*	.23*	5*	.40*	16*
P_left	.23*	5*	.15	2	.09	1	.11	1
P_initial	−.07	1	−.08	1	−.20*	4*	−.13	2
P_find	−.11	2	−.14	2	−.06	0	−.13	2
P_handle	.16*	3*	.14	2	.20*	4*	.19*	4*
Expectancy								
E_num_obj	−.14	2	−.10	1	−.26*	7*	−.31*	9*

*$p < .05$.

APPENDIX B (Continued)

Table B2
Semipartial Correlation Coefficients (sr) and Percent Variance Explained in Regressions
That Assess Unique Contributions to Error Rates

| | Unassociated | | | | Associated | | | |
| | Neutral | | Imagery | | Neutral | | Imagery | |
Predictor	sr	%Var	sr	%Var	sr	%Var	sr	%Var
Linguistic								
L_conc_prop	−.36*	14*	−.32*	10*	−.18*	3*	−.20*	4*
Perceptual								
P_area	.12	1	.12	2	.31*	10*	.26*	7*
P_left	.16*	3*	.06	0	.09	1	.10	1
P_initial	−.17*	3*	−.16	2	−.26*	7*	−.12	2
P_sal	.22*	5*	.06	0	.12	1	.06	0
P_find	−.22*	5*	−.09	1	−.14	2	−.10	1
P_handle	.25*	6*	.11	1	.12	2	.14	2
Expectancy								
E_num_obj	−.13	2	−.08	1	−.30*	9*	−.31*	10*

*$p < .05$.

Table B3
Semipartial Correlation Coefficients (sr) and Percent Variance Explained in Regressions
That Assess Unique Contributions to Combined RTs and Error Rates

| | Unassociated | | | | Associated | | | |
| | Neutral | | Imagery | | Neutral | | Imagery | |
Predictor	sr	%Var	sr	%Var	sr	%Var	sr	%Var
Linguistic								
L_conc_prop	−.38*	14*	−.39*	16*	−.21*	5*	−.21*	4*
Perceptual								
P_area	.13	2	.15*	2*	.34*	11*	.38*	14*
P_left	.25*	7*	.14	2	.13	2	.12	1
P_initial	−.08	1	−.12	1	−.20*	4*	−.13	2
P_find	−.13	2	−.13	2	−.13	2	−.13	2
P_handle	.21*	4*	.15*	2*	.19*	4*	.19*	4*
Expectancy								
E_num_obj	−.15*	2*	−.13	2	−.30*	9*	−.31*	10*

*$p < .05$.