OBSERVATIONS
Producing Biased Diagnoses With Unambiguous Stimuli:
The Importance of Feature Instantiations
Samuel D. Hannah and Lee R. Brooks
McMaster University

In this article, the authors demonstrate a laboratory analogue of medical diagnostic biasing (V. R. LeBlanc, G. R. Norman, & L. R. Brooks, 2001) in 2 experiments and explore the basis of this effect. Before categorizing novel exemplars, participants first evaluated the likelihood that the item was a member of the category suggested on that trial: either the correct category or a plausible alternative category. This was sufficient to produce a substantial bias toward the suggested category despite the use of unambiguous stimuli, explicit rules, and unhurried conditions—each of which would be likely to limit diagnostic bias. The authors argue that the production of this effect requires distinguishing between particular feature instantiations and more abstract representations of those features as well as allowing people to adopt a particular decision strategy mediating the use of instantiated features: a feature-recognition heuristic.

Keywords: concept formation, feature processing, categorization, instantiated features, informational features

LeBlanc, Norman, and Brooks (2001) found that medical students and residents could be biased toward a correct diagnosis or a plausible, alternative diagnosis simply by having them first evaluate the plausibility of either the correct or the alternative diagnosis. This biasing effect was large: Providing a diagnosis to evaluate shifted the probability of accepting it by 20%–70%. The biasing suggestion also affected the reporting of features, increasing the likelihood that reported features were consistent with the suggested diagnosis. This diagnostic biasing effect arose despite the diseases being fairly well known (e.g., stomach cancer, lupus, Cushing’s disease) and despite the photographs being taken from medical textbooks, and thus presumably representative of the disorders. A related study by Brooks, LeBlanc, and Norman (2000) that used the same photographs found that expert physicians also missed features central to the diagnosis. Prior to the study, another group of physicians rated these “cardinal features” as being unambiguous. However, this still leaves the question as to whether the rating physicians were themselves biased by knowing or generating the diagnosis when they rated the features. The biasing at both levels may be dependent on complexities of medicine, such as complex and ambiguous stimuli and a vast number of potential diagnoses.

In this article, we explore the basis and conditions of diagnostic biasing by attempting to reproduce it with a limited number of categories and with simple, artificial stimuli consisting of clear and unambiguous features. As with medical instruction, the categories in this article are acquired through formal instruction and are defined in part by explicit diagnostic rules of similar form to those used in medicine. Further, the presentation conditions, as with the medical experiments, are clear and unhurried. Our experiments show diagnostic biasing under these conditions and further show that it is dependent on knowledge of the appearance of features and the hierarchical organization of the categories. We also provide evidence suggesting that biasing diagnostic decisions may involve the use of a heuristic in which features are evaluated by their perceptual similarity to familiar feature instantiations. We argue that producing these biasing effects will be extremely difficult using the materials and training methods commonly used in laboratory studies, as these do not implement the distinction between instantiated and informational features (Brooks & Hannah, 2006).

Diagnostic Biasing and Use of Instantiated Knowledge

We argue that diagnostic biasing is strongly influenced by reliance on instantiated, rather than informational, features. By instantiated features, we mean representations of specific previously encountered features; by informational feature, we mean a
more general description of those features. In these experiments, the informational features that we consider are parts of a rule initially given to the participants; in LeBlanc et al.’s (2001) study, the informational features are those referred to in standard medical rules. If an item contains a feature whose particular manifestation (instantiation) normally occurs in another category, then suggesting that other category is likely to be substantially more seductive than if that feature had a novel appearance, even if the informational content (e.g., “two legs”) is the same. In addition, any misleading feature will be more seductive if it occurs in a location that is normally important for the suggested category. If a physician incorrectly suspects a dermatological disorder, then any misleading evidence on the skin will be more seductive than misleading evidence in posture.

This perceptual information would not be important if the person were attending only to features on an informational level, that is, attending only to a general description of the features. In Experiment 1, we show that participants differ markedly in the extent to which they rely on informational or instantiated feature representations. The effect of feature instantiations is shown to apply differentially to those participants for whom the appearance of features is heavily weighted, a strategy we characterize as a feature-recognition heuristic.

Experiment 1

The biasing manipulation in all of the experiments in this article, as with those in LeBlanc et al.’s (2001) study, was to ask participants to rate the likelihood that the item was a member of a given category (e.g., “How likely is it that this item is a ramus?”) and then to indicate which category they thought it belonged. The biasing suggestions were intended to induce the participants to consider either the correct or a plausible alternate category before making their classification.

We designed this experiment to capture the role of stimulus-specific feature manifestations. All the training items (see left and middle columns, Figure 1) except for prototypes were characterized by an overlap of informational features across categories (e.g., the term “two legs” could apply to members of both prin and ramus category) but no overlap of instantiated features (e.g., members of both prin and ramus categories have different looking legs, even when they have the same number). Test stimuli (see right column, Figure 1) embodied the same informational structure as the training exception items. For high-familiarity participants, the informational overlap in training was replaced with a perceptual overlap so that the exact legs encountered in one training group appeared in test exemplars of another, making the overlap feature the most familiar feature. For the reduced-familiarity group, the overlap feature was also taken from the overlap category’s training set, but was itself skewed to the same degree as the rule-consistent features, so that all features were equally familiar.

We embedded feature instruction within causal stories regarding the adaptiveness of each characteristic, or common, feature. These stories were intended to create distinct semantic contexts for each imaginary animal and were meant to parallel the background knowledge the medical students had of diseases. The characteristic features associated with a category were explicitly taught to participants, as is done in medical education. This also ensured that the features extracted as relevant by participants would map onto those found in the test set. Without this constraint, biasing suggestions could work because the test set did not match the feature descriptions generated by the participants, leaving them confused and following the biasing suggestion out of desperation.

Last, we did a post hoc separation of participants based on their reported decision strategies. This is tied to one of the key points raised by Brooks and Hannah (2006). There we suggested that the feature-list rules that are characteristic of everyday classification rules, including most diagnostic rules in medicine, point to the particular instantiations to be learned rather than acting as a definition or the terms of an implicit multiple regression. In this article, we suggest that people providing feature lists as their classification strategy are indicating a reliance on instantiated features. In contrast, many of our participants have provided rules also containing an explicit decision procedure to resolve conflicts among features (e.g., “I based my decisions on which species had the most features present”). As the conflicts in training exist only at an informational level, such rules are suggestive of a reliance on informational features.

Thus, we hypothesized that different strategies (feature listing vs. feature counting) would point to a differential reliance on instantiated or information features. This raises the possibility that these different strategy reports reflect critical differences in decision processes. People giving feature-counting strategies have a strong decision rule to resolve feature conflicts; people producing features listing strategies must find some other recourse to resolve such conflicts. We suggest that they are doing so by weighting features by their ease of recognition and would thus weight famil-
iar more than less familiar features. More important, the feature-listing group’s bias effect should be sensitive to the familiarity of the overlap feature, whereas any bias effect elicited among feature-counting participants should be constant regardless of feature familiarity.

**Method**

Participants. A total of 139 participants contributed data in this study, with 11 people dropped, 6 in the reduced-familiarity test condition and 5 in the high-familiarity test condition. This left 128 participants who provided useful data in the high- and reduced-familiarity conditions. The last 2 participants in the high-familiarity counting condition were dropped to produce a balanced design, resulting in data from 126 participants being analyzed. An additional 20 participants were run in an earlier pilot study to ensure no stimulus bias existed that could account for our current results.

The number of participants was large because of the need for a minimum of 20 participants in each strategy condition (listing, counting) within each overlap condition (reduced-familiarity, high-familiarity). We ended up with 42 participants who produced a listing strategy, 21 in each familiarity condition. Another 86 participants produced a counting strategy (44 in the high-familiarity condition, truncated to 42 as noted above, and 42 in the reduced-familiarity condition). Ten other participants produced a strategy other than listing or counting (6 in the perceptual overlap condition, 4 in the modified overlap condition). Participants were run in cohorts ranging in size from 1 to 10.

Materials. Stimuli were presented on an overhead projector and were line drawings of four species of imaginary animals: bleeb, ramus, croom, and prin. Each category was created around a family-resemblance structure based on three features: tail type, torso shape, and number of feet. For example, as shown in Figure 1, the characteristic feature values for ramuses (top row) are pentagonal torsos, spotted tails, and four legs. Perfect identification was possible with a two-out-of-three features rule.

The training set for each category was one animal with all the characteristic features (prototype) and three exemplars that differed by one characteristic feature (exception exemplars). For all exception exemplars, the informational value of the differing feature was identical to the characteristic value of that feature for one of the other three categories, but it had a unique perceptual manifestation (informational, but not perceptual, overlap). For example, the ramus legs-deviant exception item had two legs, which was the characteristic value for prin legs, but the legs of the ramus item were perceptually distinct from the prin four legs. We refer to the differing feature in an exception item as the overlap feature, the category from which the overlap feature came as the overlap category, and the category indicated by the two-out-of-three rule as the correct category.

The 24 test items were all exception items and were generated by skewing each training feature $+20^\circ$ and $-20^\circ$. This generated two versions of each feature, allowing us to assemble two training items matched exactly for informational content and matched approximately for overall similarity. For the high-familiarity condition, the original overlap feature in each item was replaced with the exact manifestation that feature took in the overlap category. In Figure 1 (right column), the instantiation of the four legs that was part of the bleeb training item was replaced with the exact four legs previously occurring only in ramuses. For the reduced-familiarity participants, these perceptual-overlap features were themselves skewed $+20^\circ$ or $-20^\circ$.

Procedure. The participants were told at the start of training that they were to learn four species of imaginary animals and to later use that knowledge to classify new exemplars into one of the categories. Training began with the experimenter pointing out the correct and the overlap features, naming the correct features, and explaining their adaptive functions. In this training round, a category label accompanied items. Although the overlap feature was pointed out, its overlapping nature was not, and the two-out-of-three rule was not given.

Participants then saw eight presentations of each training item, spread over three blocks. Items were presented 3 times as quartets (one item from each of the four categories) in the first block, 3 times as pairs in the second block, and 2 times as single items in the final block. At different points in the experiment, participants were required to (a) silently identify the consistent features of each displayed exemplar, followed immediately with feedback to the whole cohort, (b) silently categorize exemplars, with feedback, (c) write down the classification of exemplars, with feedback, and (d) study items as they wished (free study).

We used performance on the final presentation of the individual training items to assess learning of the training set. Identification of the same items was also assessed at the end of test to ensure that learning was sufficiently robust for relevant knowledge to be available throughout test. Only participants whose performance on each of these rounds exceeded 70% had their test data included in analyses. This criterion represents 60% of the distance between perfect and chance performance ($6(1 – \text{chance})$) and was used for both experiments in this article.

Before presenting each test item, the experimenter asked participants in the suggestion condition to rate the likelihood it was a member of a given category; for example, “How likely is it that [trial] number ten is a ramus?” After viewing each item, participants rated the likelihood that the suggested category was correct on a scale ranging from 0% to 100%. After rating the likelihood that the item was in the suggested category, participants identified it. The experimenter suggested the correct category for one member of each skewed pair ($+20^\circ$ or $-20^\circ$) when that item appeared and suggested the overlap category for the second member when it appeared. Twelve items (3 exception items $\times$ 4 categories) were cued to the correct category, and 12 were cued to the overlap category. Participants had a maximum of 30 s to respond to each item, or until everyone was finished. All participants finished before 30 s for the overwhelming majority of trials.

As the nature of the display precluded randomizing the order of stimuli for each participant, a single randomized order was used. To control for stimulus bias, we ran a no-suggestion control group using the high-familiarity test materials in a pilot study. Participants in the no-suggestion control condition simply identified the items, and then they rated their confidence on the same scale used by suggestion participants to make the two conditions roughly equivalent in decision complexity. An item analysis identified one problematic item, and after removing this item and its informationally matched partner, there were no reliable differences in any response between dummy-coded cueing conditions ($F < 1.0$). The problematic item and its partner were replaced for the high-familiarity participants in this experiment with two items modified to be unambiguous.

These new items were further pilot tested among a small no-suggestion group that uniformly identified them correctly.

Analysis. For both experiments, we scored responses as correct, overlap, or other, and we computed the mean response level for each response type for each suggestion condition. If a suggestion biases responding, we should see changes as a function of the suggestion, with overlap responses, for example, increasing when the overlap category rather than the correct category is suggested. Because biasing suggestions should affect both

---

1 The 11 participants were dropped for the following reasons: failing learning criterion (4 participants), failing to follow directions (2 participants), being an extreme outlier (biasing effects three standard deviations above the mean; 3 participants), having English as a second language (1 participant), or using a strategy in training that rendered interpretation of his or her training performance impossible (1 participant).

2 The designation of the correct category as correct is somewhat arbitrary, as it assumes the use of our counting rule to classify items. Such rules may be unusable in ordinary categorization; therefore, it may be argued that it is unrealistic to expect people to apply such a rule. Nonetheless, using the term correct to refer to rule application is useful for communication within this article.
correct and overlap responses, we conducted separate analyses on each response type. However, as the effects of suggestions on overlap responses are smaller, and thus analyses of the overlap responses more conservative, we only report analysis of the overlap responses, which were analyzed by a 2 × 2 mixed design analyses of variance (ANOVAs), with familiarity condition (high, reduced) as the between-subjects factor and cue direction (cued to overlap, cued to correct) as the within-subjects factor. Similar analyses were run on the identification of training items before and after test.

This asymmetry in size of biasing on response type likely resulted from the apparent tendency of incorrect suggestions to produce both a bias and a small confusion effect. Suggesting the overlap category may have interfered with retrieval of the correct category, leading people to decide on an answer other than the correct or overlap category, whereas suggesting the correct category only interfered with incorrect classification responses. Although other response levels did not reliably differ across cueing suggestions in either study, and nominal differences between cue direction conditions were quite small—justifying our treatment of responses as including strategies or rules developed. These latter between-subjects differences were due to the coding of the overlap responses fewer other responses relative to correct suggestions.

This asymmetry across response type is one reason why it makes more sense to compare responses within subjects (between cue conditions) rather than to compare between suggestion and no-suggestion participants (within cue conditions), as this would lead to misleading asymmetries in the magnitude of between-subjects cueing effects. The misleading nature of between-subjects comparisons is heightened by the conclusion of this study and that of Brooks and Hannah (2006): How conceptual information is used depends on the type of feature representations relied on and the consequent strategies or rules developed. These latter between-subjects factors would confound attempts to untangle between-subjects cue effects.

**Results and Discussion**

There were no significant effects found for performance on training items, and for all groups, accuracy was above 90%, and usually 95%. Prior to analysis, suggestion participants were segregated into listing and counting participants on the basis of their postexperimental description of decision strategy.

**Listing participants.** There was a reliable main effect of feature familiarity, \(F(1, 40) = 5.79, \text{MSE} = 4.35, \omega^2 = .091\). Participants receiving the high-familiarity materials had a higher rate of mean overlap responses (3.19) than participants tested on the reduced-familiarity materials (2.19), a difference of 9.2%. There was also a main effect of cue direction, \(F(1, 40) = 28.27, \text{MSE} = 0.97, \omega^2 = .098\). Suggesting the overlap category increased mean overlap responses compared with suggesting the correct category from 2.17 responses to 3.31 overlap, a change of 9.5%. Critically, we found a reliable Cue Direction × Familiarity interaction, \(F(1, 40) = 4.91, \text{MSE} = 0.97, \omega^2 = .018\); the effect of suggesting the overlap category was smaller for reduced-familiarity participants than for high-familiarity participants. For high-familiarity participants, suggesting the overlap category increased mean overlap responses (4.09) compared with suggesting the correct category (2.48), a change of 13.5%. For the reduced-familiarity participants, suggesting the overlap category increased mean overlap responses from 1.86 overlap response to 2.52 overlap response, a change of 5.6%. Repeated measure ANOVAs done within each familiarity condition revealed a reliable main effect of cue direction for the high-familiarity participants, \(F(1, 20) = 33.41, \text{MSE} = 0.82, \omega^2 = .203\). The main effect of cue direction was nearly significant for the reduced-familiarity participants, \(F(1, 20) = 4.67, \text{MSE} = 1.12, p = .054, \omega^2 = .031\).

**Counting participants.** The only reliable effect was a small main effect of cue direction, \(F(1, 82) = 13.79, \text{MSE} = 0.50, \omega^2 = .003\). Suggesting the overlap category increased mean overlap responses from 0.62 to 1.02, a change of 3.3%. Importantly, there was no Cue Direction × Familiarity interaction for counting participants, \(F(1, 82) \leq 1.00, \text{MSE} = 0.50\), despite a larger sample size and smaller error variance for the counting participants’ interaction than for listing participants. Had there been a Cue Direction × Familiarity effect for counting participants that was as large as that of listing participants, our test of the counting participants’ interaction would have had a power close to .80. Although suggestions slightly biased the counting participants, this bias was constant across the overlap feature’s familiarity.

We were able to bias people’s categorization decisions even with simple, unambiguous materials. Participants were more likely to classify an item as being in the overlap category after first estimating the likelihood that it was a member of the overlap category than after first estimating the likelihood that it was a member of the correct category. High accuracy on training items by the no-suggestion participants rules out a role for confusion stemming from poor knowledge. Diagnostic biasing, therefore, does not require vague criteria or either the level of training or the complexity of materials found in medicine.

The crucial finding for use of strategies is that reducing the perceptual familiarity of the overlap feature reduced the diagnostic biasing effect for people reporting a listing strategy; for those reporting a counting strategy, however, the effect of a suggestion was small and constant regardless of the familiarity of the overlap feature. This confirms our conjecture that people reporting feature-listing strategies are more sensitive to the perceptual manifestations of the features than are those people reporting counting. This interaction of feature familiarity and bias suggestions implies a common mechanism, which we suggest is a feature-recognition heuristic. We suggest it is reasonable that a suggestion primes processing for features consistent with the suggestion, making them more readily recognizable than rivals. If instantiated-oriented people are weighting features by their ease of recognition, then suggestion-consistent features should be treated as more important.

We further suggest that the large number of our participants who reported a counting procedure were reacting to the special circumstances that each category had only three relevant features and that (with the exception of the prototypes) every item had exactly the same number of these relevant features. Because this is likely to occur only in artificial materials, we suspect that most people would be like our listing-strategy participants under everyday identification conditions.

**Experiment 2: Superordinate Structure and Biasing**

Hierarchical organizations based on semantic or structural factors are a common aspect to natural categories and may establish separate contexts that influence how individuals attend to information. In medicine, there are superordinate classes based on causal mechanisms, such as genetic disorders or infectious diseases. If hepatitis is suggested by a colleague, this may limit the chance of stomach cancer coming to mind for consideration. Different systems (endocrine, cardiac, etc.) form superordinates...
with particular spatial arrangements. Part of the concept of lung
disease may be knowledge about where to look for the signs of
lung diseases. If this knowledge includes preserved records of
attention shifts, then a suggestion could bias attention by priming
previous attentional patterns and, through biasing the selection of
evidence, bias the final decision.

In Experiment 2, we explored the role of superordinate structure
by putting our four categories into two superordinates: zoots
(bleebs and crooms) and soots (ramus and prins), which were
distinguished structurally and semantically. To distinguish them
structurally, we placed the relevant features for zoots in the upper
half of the body; for soots, they remained in the lower half of the
body. To distinguish the classes semantically, we gave different
types of evolutionary stories. The evolutionary stories for zoots
were based on social structures, whereas the stories for soots
involved adaptations to terrain and climate. If either semantic or
physical organization influences diagnostic biasing, then the size
of the diagnostic biasing effect will vary depending on whether
feature competition occurs among members of the same superor-
dinate or among members of different superordinates. That is,
there will be a Cue Direction × Superordinate interaction.

Method

Participants. Forty-four McMaster University undergraduates partici-
pated, with 4 dropped for failing to meet learning criterion—3 in the
suggestion and 1 in the no-suggestion condition—leaving us with 20
participants in each condition.

Materials. We modified the stimuli from Experiment 1 to create two
genera, each consisting of two species. Examples of training (top two rows
of each panel) and test stimuli (bottom row of each panel) are in Figure 2.
The zoot genus consisted of bleebs and crooms (see top panel, Figure 2),
and the soot genus (see bottom panel, Figure 2) consisted of prins and
ramuses. A schematic description of the bleeb (zoot) and ramus (soot)
categories is given in Table 1.

For zoots, the diagnostic features were horn curvature, head shape, and
neck length. For soots, the diagnostic features remained tail type, torso
shape, and number of feet, with only their nondiagnostic features (horns,
head shape, and neck length) modified from Experiment 1. For exception
training items, there was overlap at an informational level on both diag-
nostic and nondiagnostic features. Two nondiagnostic features informa-
tionally overlapped with another category, with the third nondiagnostic
feature being both informationally and perceptually novel (indicated
in Table 1 by Xs). For example, when an exception bleeb’s tail descriptively
matched that of the prin, its feet took the number of ramus feet, whereas its
torso was a novel value. For prototypes, all three nondiagnostic features
had novel values. No perceptual overlap occurred across categories.

In addition, the stories involving the relevant features were amended.
For the soots, the features were explained as adaptations to different social
structures (solitary, highly territorial animals for crooms vs. gregarious,
social animals for bleebs). For the zoots, we preserved the stories from
Experiment 1, which described features as adaptations to terrain and
climate.

Test items were created as before, with changes made to allow for
perceptual overlap either within the same genus or across genera. For each
exception training item, we made four test versions. First, rule-consistent
features were skewed 20° clockwise or counterclockwise and reassembled
into new items. For two of these, we made a perceptual overlap by
replacing the informationally overlapping diagnostic feature with a training
feature from the other category in the same genus (same superordinate
overlap; see Figure 2, third rows of each panel). For the other two test
items, we created a perceptual overlap by replacing one of the informa-
tionally overlapping nondiagnostic features with a training feature from

Results and Discussion

Mean overlap responses are summarized in Table 2. We tested
for the biasing of mean overlap responses with a $2 \times 2 \times 2$ mixed
design ANOVA, with biasing suggestion (suggestion, no-suggestion) as a between-subjects factor, and superordinate (same-superordinate items, different-superordinate items) and cue direction (cued correct, cued overlap) as within-subjects factors. For no-suggestion participants, items were dummy coded for cue direction on the basis of how they were cued for the suggestion participants. As this experiment was actually run before Experiment 1, we did not ask for decision rules.

Overall, suggesting the overlap category increased mean overlap responses (2.82) over suggesting the correct category (2.82), yielding a main effect of cue direction, F(1, 38) = 8.09, MSE = 3.78, ω² = .0.80. The effect of cue direction was larger for suggestion than for no-suggestion participants, yielding a significant Cue Direction × Suggestion interaction, F(1, 38) = 15.22, MSE = 3.78, ω² = .076. For suggestion participants, mean overlap responses increased after the experimenter suggested the overlap category compared with when the experimenter suggested the correct category. For no-suggestion participants, items dummy coded as cued to the overlap category produced fewer mean overlap responses than items dummy coded as cued to the correct category. Critically, the effect of superordinate structure on the diagnostic bias effect varied across suggestion conditions, producing a Cue Direction × Suggestion × Superordinate interaction, F(1, 38) = 13.50, MSE = 1.45, ω² = .027. For suggestion participants, suggesting the overlap category seemed to have an even bigger effect for different-superordinate items than for same-superordinate items, whereas no-suggestion participants seemed to show a negative cueing effect for different-superordinate items.

To clarify the three-way interaction, we performed analyses consisting of 2 (cue direction) × 2 (superordinate) repeated measures ANOVAs within each suggestion condition (suggestion, no-suggestion). For suggestion participants, the main effect of cue direction was significant, F(1, 19) = 13.68, MSE = 6.30, ω² = .200. Suggesting the overlap category increased overlap responses compared with suggesting the correct category. The Cue Direction × Superordinate interaction was virtually significant, F(1, 19) = 4.13, MSE = 1.89, p = .056, ω² = .015. When perceptual overlap occurred within a superordinate, the bias effect was smaller than when perceptual overlap involved members of different superordinate classes.

For no-suggestion participants, the only main effect was that of superordinate, F(1, 19) = 6.45, MSE = 0.56, ω² = .013; items with a perceptual overlap feature from a different superordinate produced fewer overlap responses than items in which the overlap occurred among members of the same superordinate. The Cue Direction × Superordinate interaction was also significant, F(1, 19) = 11.86, MSE = 1.01, ω² = .046. For the same-superordinate items, there was little difference in overlap responses across items dummy coded as cued to overlap and those dummy coded as cued to correct. However, for the different-superordinate items, items dummy-coded as cued to overlap category elicited fewer mean overlap responses than items dummy coded as cued to correct. This interaction shows a stimulus bias opposite to the experimental manipulation, suggesting the actual magnitude of biasing for suggestion participants may be underestimated by the data.

These differences between suggestion groups were not due to different levels of pretest knowledge. A 2 × 2 mixed design ANOVA on the identification accuracy of the training items, with suggestion (suggestion, no suggestion) as a between-subjects factor and assessment round (before test, after test) as a within-subjects factor, found only a significant main effect of assessment round, F(1, 38) = 4.59, MSE = 0.61, ω² = .027. For both groups, accuracy again dropped slightly over time (after training = 94% correct, after test = 91% correct). Despite the greater complexity

Table 1
Informational Structure of Bleeb (Zoot) and Ramus (Soot)
Training Items, Experiment 2

<table>
<thead>
<tr>
<th>Zoot diagnostic</th>
<th>Soot diagnostic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Head Horns Neck</td>
<td>Legs Torso Tail</td>
</tr>
<tr>
<td>Bleeb prototype</td>
<td></td>
</tr>
<tr>
<td>1-away 1</td>
<td>1 1 1 X X X</td>
</tr>
<tr>
<td>1-away 2</td>
<td>1 1 2 3 X</td>
</tr>
<tr>
<td>1-away 3</td>
<td>2 1 1 4 3 X</td>
</tr>
<tr>
<td>Ramus prototype</td>
<td></td>
</tr>
<tr>
<td>1-away 1</td>
<td>X X X 4 4 4 4</td>
</tr>
<tr>
<td>1-away 2</td>
<td>1 1 1 4 4 3 X</td>
</tr>
<tr>
<td>1-away 3</td>
<td>2 1 1 2 3 X 4</td>
</tr>
</tbody>
</table>

Note. Values characteristic of bleebs are denoted with 1, prins with 2, crooms with 3, and ramuses with 4. X values are not associated with any category.

Table 2
Mean Overlap Responses (and Standard Deviations) Within Cueing and Superordinate Conditions by Suggestion Condition, Experiment 2

<table>
<thead>
<tr>
<th>Condition</th>
<th>Cued to</th>
<th>Same superordinate</th>
<th>Different superordinate</th>
<th>Cue M</th>
</tr>
</thead>
<tbody>
<tr>
<td>Suggestion (n = 20)</td>
<td>Overlap</td>
<td>3.60 (2.64)</td>
<td>3.95 (2.09)</td>
<td>3.78 (2.36)</td>
</tr>
<tr>
<td></td>
<td>Correct</td>
<td>2.15 (1.39)</td>
<td>1.25 (1.55)</td>
<td>1.70 (1.52)</td>
</tr>
<tr>
<td></td>
<td>Bias effect</td>
<td>1.55/12.9%</td>
<td>2.70/22.5%</td>
<td>2.08/17.3%</td>
</tr>
<tr>
<td></td>
<td>Superior</td>
<td>2.87 (2.21)</td>
<td>2.60 (2.67)</td>
<td></td>
</tr>
<tr>
<td>No suggestion (n = 20)</td>
<td>Overlap</td>
<td>2.55 (1.96)</td>
<td>1.35 (1.57)</td>
<td>1.95 (1.63)</td>
</tr>
<tr>
<td></td>
<td>Correct</td>
<td>2.10 (1.65)</td>
<td>2.45 (1.64)</td>
<td>2.28 (1.85)</td>
</tr>
<tr>
<td></td>
<td>Bias effect</td>
<td>0.45/3.7%</td>
<td>-1.1/-9.2%</td>
<td>-0.33/-2.7%</td>
</tr>
<tr>
<td></td>
<td>Superior</td>
<td>2.33 (1.80)</td>
<td>1.90 (1.68)</td>
<td></td>
</tr>
</tbody>
</table>

Note. Bias effect = (cued to overlap − cued to correct). Numbers before the backslash represent the bias effect in terms of raw responses, and numbers after the backslash represent the effect as a percentage of responses.
of the materials, both groups performed well, and roughly equally, on the training items. The suggestion group was 92.5% correct after training and 90.6% correct on the same items after test; the no-suggestion group was 95.3% correct after training and 91.3% correct after test.

The diagnostic biasing effect was much larger when the correct and alternative categories were in different superordinates than when rivals were in the same superordinate. For different-superordinate items, the diagnostic biasing effect was equivalent in size to some of the effects seen in the medical literature (LeBlanc et al., 2001). This increase due to hierarchical structure could be due to the different semantic contexts being established by the different kinds of evolutionary stories accompanying feature instruction. Or, this superordinate bias effect could happen because of the different spatial relations in the same- and different-superordinate conditions. Features of rival categories were adjacent in the same-superordinate condition but spatially segregated in the different-superordinate condition. This spatial segregation, combined with a suggestion that would imply a particular superordinate, and thus a particular spatial arrangement, could encourage a neglect of suggestion-inconsistent features.

However, a study published in Samuel D. Hannah’s dissertation eliminated the evolutionary cover stories and replicated the effect of superordinate structure with no reliable diminution of biasing for different-superordinate overlap (Hannah, 2004). Thus, physical segregation of features seems to be sufficient to increase the diagnostic biasing effect. If our future research confirms that it is the physical segregation that enhances the diagnostic biasing effect, then this would suggest that suggesting a category name biases not only semantic processing involved in diagnostic decisions but more basic processes, such as the distribution of attention across features. This in turn would suggest that attentional processing can be concept specific, reminiscent of findings of concept-specific feature parsing (Schyns, Goldstone, & Thibaut, 1998; Schyns & Murphy, 1994; Schyns & Rodet, 1997). This would also imply that concept representations include not only semantic and perceptual information but also processing responses, such as attentional patterns, consistent with situated accounts of concepts (e.g., Barsalou, 1999; Glenberg & Robertson, 2000; Lakoff, 1987).

General Discussion

We have demonstrated a diagnostic biasing effect using materials consisting of a small number of categories with well-known, unambiguous features and an explicit feature list, which were run under unhurried decision conditions. Although the magnitude was smaller than that found in medical diagnosis by LeBlanc et al. (2001), this is likely due to the elimination of factors that may well contribute to actual medical classification, such as ambiguous features. Even so, in some conditions, our biasing effect approached the bottom of the range found by LeBlanc et al. (different-superordinate condition, Experiment 2). Diagnostic biasing is not, therefore, some special phenomenon arising out of the particulars of medical cognition (e.g., extensive training, vast numbers of categories) or medical materials (e.g., ambiguous or complex features).

We showed that diagnostic biasing was strongly affected by perceptual familiarity for those reliant on instantiated features but not for those reliant on informational features (Experiment 1). We demonstrated that one salient feature of ordinary concepts, their hierarchical organization, also influenced diagnostic biasing. This influence could be due to the preservation of learned attentional patterns cued by a suggestion. When rival features were physically segregated, the diagnostic biasing effect increased compared with when rival features were physically adjacent. Work published in Samuel D. Hannah’s dissertation (Hannah, 2004) revealed that eliminating the semantic contexts had no effect on the results and neither did reducing the number of categories to two. A biasing of attention may explain both the small familiarity-independent bias effect shown by the counting participants in Experiment 1 and the influence of suggestions on feature reporting in LeBlanc et al.’s (2001) work.

We suggest that at least two themes are necessary to account for these biasing effects and, presumably, many other aspects of real-world categorization and concept use.

1. Instantiated features. We believe that our training materials better capture the relations between the appearance of features and category identity than has been done by most prior research. Our training materials had a high degree of similarity of feature appearance within categories and a very low similarity of features across categories. Such an association between the appearance of features and category is a normal aspect of most everyday categories, at least at the basic level. In the laboratory, however, features usually have little perceptual distinctiveness across categories because of perceptual overlap across categories. Because the physical arrangement of features is typically held constant across categories, expecting a given category does not change the distribution of attention or search patterns. The only useful information left is the relation between imperfect informational features and category identity.

In their highly influential article, Shepard, Hovland, and Jenkins (1961) concluded that stimulus generalization was not enough to account for classification learning in their paradigm because people made fewer errors than would be accounted for by the stimulus confusions such generalization would generate. They concluded that concept formation is a process of abstraction of relational knowledge by selective attention, directed by hypotheses expressed in verbal rules. However, as in most classification experiments, their features overlapped across categories at a perceptual level. This eliminated the diagnosticity of individual feature instantiations and forced participants to learn about the general relations of the categories because this was the only diagnostic information available. This real-world phenomenon surrounding the use of concepts (such as diagnostic biasing), which is reliant on instantiated knowledge, suggests that Shepard et al.’s conclusion is correct as a description of what occurs in experiments such as theirs but not as a description of everyday concept learning.

2. Feature recognition as a judgment heuristic. Participants who gave a feature-list strategy showed not only a diagnostic biasing effect but one that interacted with feature familiarity. This is readily understandable if such people are relying on instantiated features and using the ease of feature recognition to judge the significance of each feature. The priming arising from a biasing suggestion may aid feature recognition/saliency, enhancing the perceived significance or goodness of a suggestion-consistent feature at the expense of suggestion-inconsistent features.
References


Received December 16, 2005
Revision received April 28, 2006
Accepted May 8, 2006

\[1423\]