The spontaneous appeal by naïve subjects to nonaccidental properties when distinguishing among highly similar members of subspecies of birds closely resembles descriptions produced by experts.

Ori Amir1*, Xiaokun Xu1& Irving Biederman1,2
1Department of Psychology, 2Neuroscience Program, University of Southern California

Problem:
How do humans, expert and novice, distinguish between exemplars of a highly similar subordinate class of natural entities, such as members of a subspecies of birds?

Two Possible Answers:
• They use the same kinds of information – metric and configural – that they use for faces, (e.g., Gauthier & Tarr, 1997), or
• They employ nonaccidental properties at a small scale specifying the location of that feature (Biederman et al., 1999).

Results (continued):

<table>
<thead>
<tr>
<th>Descriptor</th>
<th>Experts</th>
<th>Novices</th>
</tr>
</thead>
<tbody>
<tr>
<td>white</td>
<td>8.04%</td>
<td>10.81%</td>
</tr>
<tr>
<td>black</td>
<td>4.02%</td>
<td>10.65%</td>
</tr>
<tr>
<td>long`</td>
<td>4.25%</td>
<td>9.70%</td>
</tr>
<tr>
<td>dark`</td>
<td>7.81%</td>
<td>3.18%</td>
</tr>
<tr>
<td>gray`</td>
<td>1.22%</td>
<td>2.86%</td>
</tr>
<tr>
<td>feature not present</td>
<td>2.20%</td>
<td>1.75%</td>
</tr>
<tr>
<td>feature present (no ad) descriptors</td>
<td>2.71%</td>
<td>0.40%</td>
</tr>
<tr>
<td>smart`</td>
<td>0.56%</td>
<td>2.46%</td>
</tr>
<tr>
<td>straight`</td>
<td>0.49%</td>
<td>0.35%</td>
</tr>
<tr>
<td>thin`</td>
<td>0.70%</td>
<td>1.91%</td>
</tr>
<tr>
<td>darker`</td>
<td>1.22%</td>
<td>1.35%</td>
</tr>
<tr>
<td>light`</td>
<td>1.40%</td>
<td>1.11%</td>
</tr>
<tr>
<td>large`</td>
<td>1.92%</td>
<td>0.40%</td>
</tr>
<tr>
<td>curved`</td>
<td>0.59%</td>
<td>1.75%</td>
</tr>
</tbody>
</table>

There was a striking similarity in the descriptions of novices and experts, with a significant correlation (r = .66, p < .01) in the frequency of specific adjectives used by both groups.

The table shows the 14 most frequent adjectives and their frequency (as percent of all descriptions in which they appeared) as function of expertise.

Study 1: Online survey
Subjects: 30 bird experts, 27 amateurs, and 25 novices
All subjects performed an online survey in which they were shown sets of 4-5 grey scale images of birds of the same family, all on the same page in the Golden bird guide, e.g., Western US Male Quails.

They were to describe each bird sufficiently so that a “friend on the phone” who has those birds moving around in a cage would know which numbered bird they were referring to.

An example of a set of birds (male US quails, in Robbins, 1983). To distinguish bird 2 from the rest of the birds, many subjects (novices and experts alike) noted its curved head-feather, and dark belly patch.

Results:

As expected, virtually all descriptions could be parsed into:
1. A locative by naming a part (e.g., “beak”), which implied the part’s location/relation specified by a structural description of parts and their relations to other parts, e.g., the beak would be a protuberance emerging from the front of the head.
2. An invariant (i.e., nonaccidental) characteristic of:
   a) of that part’s shape (e.g. curved beak), or
   b) of a surface property of that part (e.g., patch on belly).
3. A comparative metric property because there was a small number of simultaneously presented birds or bird relations (e.g. “smallest of all birds,” “sally is a lighter shade of gray than the chest”).

In all, the proportion of the descriptors that were nonaccidental was 95.5%.

The descriptions from only 6 randomly sampled naïve subjects contained 93.3% of the descriptions in the Golden Bird Guide.

Study 2: Evaluating the Quality of the Descriptions
Naïve subjects were given the description of one of the birds from a subject in Study 1 along with the pictures of the set of birds from which the description was made, and chose the bird that they judged matched the description.

Results:

Mean accuracy was 78.1% (chance = 22.5%), with no significant difference in performance when provided with descriptions by experts vs. novices (although performance with descriptions from amateurs was slightly but significantly (p < .05) higher).

Conclusions:

Both novices and experts use similar descriptors—almost always nonaccidental—in distinguishing highly similar birds.

The authors were readily able to match almost 100% of the descriptions to the intended birds.

The overall level of performance would undoubtedly have been higher if all the subjects from the USC subject pool generating the descriptions and performing the matching were assiduous in doing the task.

Appendix: Automating description parsing

We have developed a program (illustrated below) that can parse subject’s natural language descriptions of the birds into a list of parts/features, their descriptions (general adjectives), and if provided) their location/relation to the rest of the bird. Using a statistical part of speech tagger (Stanford POS Tagger, v 3.0; Toutanova et al., 2003), some learning, and a small set of rules, the program currently extracts 70-90% of the part-descriptor-location triplets correctly (conditioned on description’s grammaticality and complexity).

This result highlights the desirability of computer vision systems that parse images into a similar representation of parts, their nonaccidental properties and relations so the vision system can communicate the results of its processing. Such representation will be necessary for tasks requiring mapping between natural language and vision.

Supported by NSF BCS-04-20794, 05-31177, 06-17699 to I.B.

References