



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.



*(oamir@usc.edu)
http://geon.usc.edu/~ori/

Ori Amir¹, Xiaokun Xu¹ & Irving Biederman^{1,2}

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:

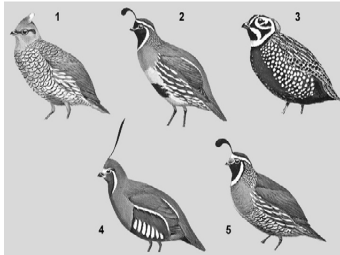
- They use the same kinds of information – metric and configural – that they use for faces, (e.g. Gauthier & Tarr, 1997), or
- They employ nonaccidental properties at a small scale specifying the location of that feature (Biederman et al., 1999).

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.



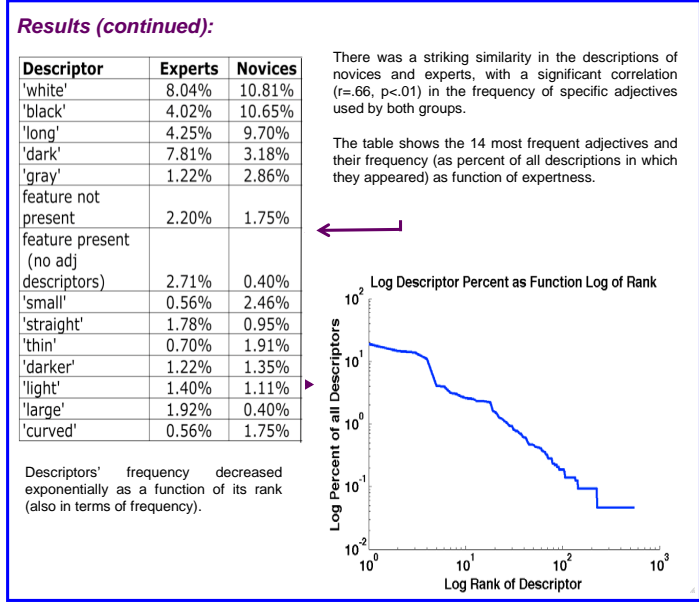
An example of a set of birds (male US quails, in Robbins, 1983). To distinguish bird 2 from the rest of the birds, many subjects (novices and experts alike) noted its curved head-feather, and dark belly patch.

Results:

Virtually all descriptions could be parsed into:

1. A locative by naming a part (e.g., "beak"), which implied the part's location/relationship 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 possible because there was a small number of simultaneously presented birds or bird relations (e.g. "smallest of all birds", "belly is a lighter shade of gray than the chest"). If the birds or bird parts were not presented simultaneously, then it would have been absolute (rather than comparative) *metric* (rather than nonaccidental) judgment.

- ✓ 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.



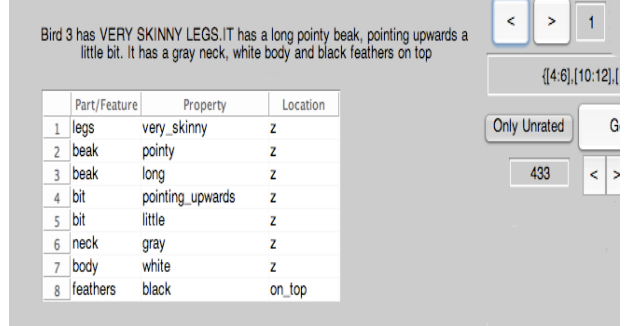
Conclusions:

- ✓ Both novices and experts use similar descriptors—almost always nonaccidental—in distinguishing highly similar birds.
- ✓ The descriptions from individual subjects were sufficient to allow accurate matching to the intended birds by naïve subjects.
- ✓ The descriptions of naïve subjects tended to match the descriptors in the Bird Guide. From a handful of such subjects, the complete descriptions in the bird guide could be generated.

Appendix: Automating description parsing

We have developed a program (illustrated below) that can parse subjects' natural language descriptions of the birds into a list of parts/features, their descriptors (generally adjectives), and (if provided) their location/relationship 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.



Part/Feature	Property	Location
1 legs	very_skinny	Z
2 beak	pointy	Z
3 beak	long	Z
4 bit	pointing_upwards	Z
5 bit	little	Z
6 neck	gray	Z
7 body	white	Z
8 feathers	black	on_top

A typical parsing returned by the program. 6/8 of the triplets were extracted correctly, triplets 4 and 5 are errors that need to be corrected by a human user. Further improvements and learning may eliminate some of these errors.

References:
Biederman, I., Subramaniam, S., Bar, M., Kalocsai, P., & Fiser, J. (1999). Subordinate-level object classification reexamined. *Psychological Research*, 62, 131-153.
Gauthier, I., & Tarr, M. J. (1997). Becoming a "Greeble" expert: Exploring mechanisms for face recognition. *Spatial Vision*, 37, 1673-1682.
Toutanova, K., Klein, D., Manning, C., & Singer, Y. (2003). Feature-rich part-of-speech tagging with a cyclic dependency network. In *Proceedings of HLT-NAACL 2003*, 252-259.
Robbins, C.S., Bruun, B., & Zim, H.S. (1983). *Birds of North America: A Guide to Field Identification*. Golden Press, New York.

Supported by NSF BCS 04-20794, 05-31177, 06-17699 to I.B.
We would like to thank, June Wang for her help in analyzing the surveys, and to Kimball L. Garrett for his help in recruiting bird experts for our experiment. Special thank for our participants from the Yahoo! groups: LACoBirds, CALBIRDS, SDBIRDS, SFBIRDS and venturacobirding.
"Geon-Brain" symbol of the Image Understanding Lab (top right) is the work of Mark Lescoart.

