Exploring the contributions of environmental factors in folk categorization

Anonymous CogSci submission

Abstract

How do we name things? What role does frequency of observation and physical size play in categorization of animals? Here we explore these questions using ideas from anthropology and ethnobiology, and utilizing large-scale citizen science datasets.

Keywords: ethnobiology; categorization; bird naming; citizen science

Introduction

What role does frequency of exposure have in being able to determine the name for a bird or its role in a taxonomy? What about the size of the bird? BIG QUESTION: What contributes to naming data?

More generally, we're interested in exploring computational principles that guide the naming and structuring of categories. Following a tradition of ethnobiological classification (Berlin, Breedlove, & Raven, 1973; Berlin, 2014), we aim to explore what kinds of features influence these behaviors.

These questions are interesting because we typically take for granted the categories of natural kinds. However, scientific taxonomies are just another human-constructed category system. When considering the set of birds in particular, it has been difficult for biologists to agree on a standardized taxonomy, which has been shown to severely impact decisions on conservation policy (Peterson, 2006; Garnett & Christidis, 2017).

What ebird can contribute to the folk biology literature: frequency data.

This connects with the theoretical debate about cognitive vs utilitarian view of classification (E. Hunn, 1982; Lopez, Atran, Coley, Medin, & Smith, 1997). Frequency data provide some new ways to test the utilitarian view. For example, we could use frequencies to address questions like:

1) do unnamed species tend to have low frequency?

2) if a species is lumped in with another species under the same label – does one of these species have low frequency?

3) in cases where nomenclature reveals prototype effects – is the prototype highest in frequency?

For more on point 3 see (Berlin, 1972): "A highly regular labeling process can be described for the encoding of specific taxa, given the primarily binary partition of a generic taxon. In general, one specific category, because it is most widespread, larger, best known, or the like, will always be recognized as the typical species of the folk genus. This taxon can be referred to as the type-specifict, he archetype, or the ideal type.... As Wyman and Harris have said in referring to Navaho ethnobotany, 'The situation is as if in our binomial system the generic name were used alone for the best known species of a genus, while binomial terms were used for all other members of the genus'"

Environmental factors in ethnobiological classification

Environmental factors have played a major role in scientific classification of species (Amadon, 1943).

Here we talk about the data we use. We show how to utilize publicly available digitized information to explore these questions.

For an official taxonomic system, we follow the Clements taxonomy (Clements, 2007).

Language naming data

We focus on a single language for brevity. We use bird-naming data from (E. S. Hunn, 2008), also found online[1], a Zapotec language spoken in a small village in San Juan Gbëë, Oaxaca Mexico.

Frequency data

We utilize a citizen-based bird observation network, eBird (Sullivan et al., 2009). We sampled bird observational data from just the region containing the state of Oaxaca, Mexico, for the timespan between Jan 1, 2018 and Jan 1, 2020. This resulted in XXX entries.

Physical Size

Bird weights as an aid in taxonomy (Amadon, 1943).

We'll also look at bird size as a factor, following (E. Hunn, 1999) and using data from EltonTraits (Wilman et al., 2014). This data set provides information on key attributes for all 9993 and 5400 extant bird and mammal species, derived from key literature sources. Variables include relevance of select diet types and foraging strata, body size, and activity time.

We focus on the body mass data, separately sourced from (Dunning Jr, 2007), which is measured as the geometric mean of average values provided for both sexes.

[1] http://faculty.washington.edu/hunn/zapotec/z5.html

What categories are given a name?

Frequency

Here we analyze the frequencies of birds named in Zapotec. We plot the densities of birds named and unamed birds, along with all birds observed in the state of Oaxaca in Figure 1.

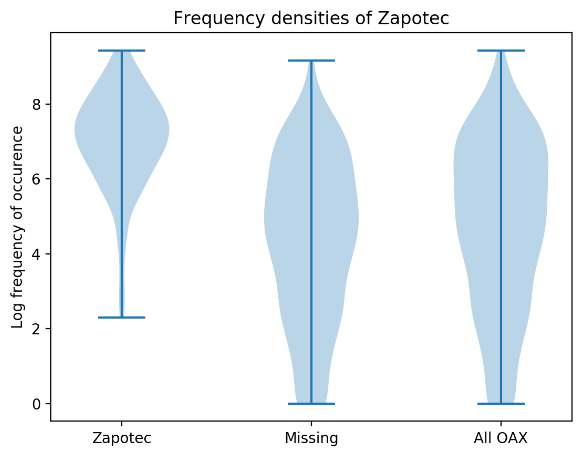

Figure 1: Frequency densities of birds named in Zapotec and those observed in the state of OAX.

Size

SSRRs.

Analysis of name-forms

Name-length

Here we analyze the frequencies of birds named in Zapotec in relation to the name length of the bird. See Figure 2

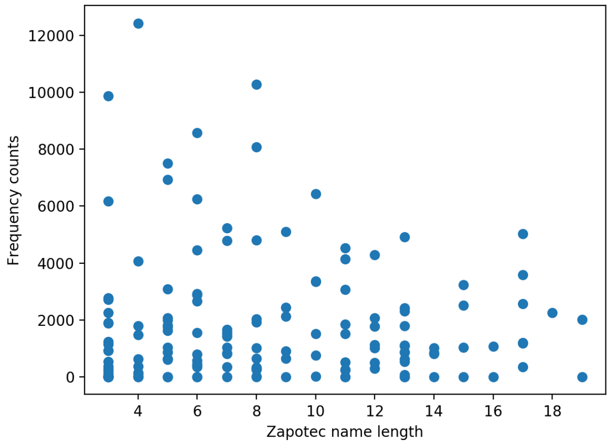

Figure 2: Frequency densities of birds named in Zapotec as a function of name length.

Compound names

Here we further examine names based on whether the Zapotec label is a single word (a monomial) or a compound of multiple words. First we examine frequencies, in Figure 3. Here we see that the monomials tend to be more frequently observed than compounds. The raw mean frequency counts are mono = 2465 and compound = 1715, for monomials and compound names, respectively.

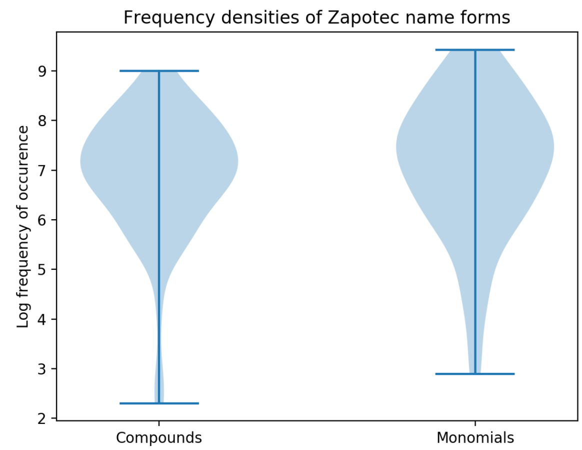

Figure 3: Frequency densities of birds named in Zapotec as a function of name form.

We also explore how masses are distributed based on name form. See Figure 4. Here we see a similar trend as before, with raw mean masses of mono = 375g and compound = 152g.

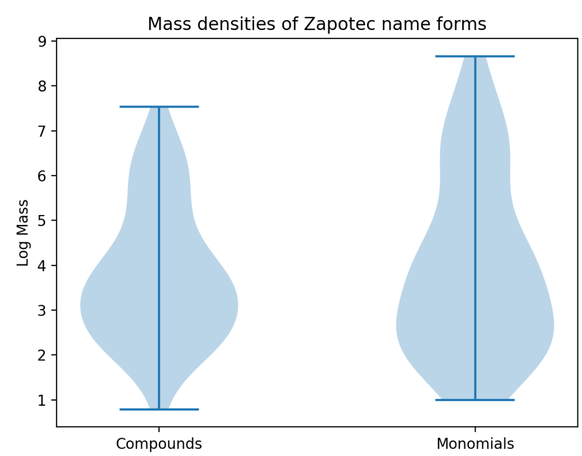

Figure 4: Mass densities of birds named in Zapotec as a function of name form.

Prototypes

Here we look at unmarked-prototypes in Hunn's data on Zapotec bird-naming.

Discussion

This work extends the investigation of language categories using large-scale datasets online (Kemp & Regier, 2012; Regier, Carstensen, & Kemp, 2016)

Future Directions

The next step would be to expand these analyses to more languages. To do this one needs to find trustworthy ethnographies similar to the Zapotec naming data we used here from E. S. Hunn (2008), and one needs decent coverage in eBird over the geographic region in question. Clear next steps would be to analyze the Tzeltal language from Chiapas, Mexico, and the Tlingit language from the south-east Alaska, both published by Hunn as well (E. S. Hunn, 1977; E. S. Hunn & Thornton, 2012), which have decent coverage within their respective geographics regions in eBird observational data.

That said, it can be difficult to find languages with both expert ethnographries of the folk biological naming systems which also have good coverage in eBird. This has prohibited us from exploring bird naming data from known experts in regions with low coverage in eBird (e.g., naming data summarized in (Holman, 2002), including the Tobelo language from Indonesia (Taylor, 1990) and the Anindilyakwa language from Australia (Waddy et al., 1988), which do not have coverage in eBird currently).

Conclusion

References

Amadon, D. (1943). Bird weights as an aid in taxonomy. *Wilson Bull*, *55*(3), 164–177.

Berlin, B. (1972). Speculations on the growth of ethnobotanical nomenclature. *Language in society*, *1*(1), 51–86.

Berlin, B. (2014). *Ethnobiological classification: Principles of categorization of plants and animals in traditional societies* (Vol. 185). Princeton University Press.

Berlin, B., Breedlove, D. E., & Raven, P. H. (1973). General principles of classification and nomenclature in folk biology. *American anthropologist*, *75*(1), 214–242.

Clements, J. F. (2007). *Clements checklist of birds of the world*. Comstock Pub. Associates/Cornell University Press.

Dunning Jr, J. B. (2007). *CRC handbook of avian body masses*. CRC press.

Garnett, S. T., & Christidis, L. (2017). Taxonomy anarchy hampers conservation. *Nature News*, *546*(7656), 25–27.

Holman, E. W. (2002). The relation between folk and scientific classification of plants and animals. *Journal of Classification*, *19*(1), 131–159.

Hunn, E. (1982). The utilitarian factor in folk biological classification. *American Anthropologist*, *84*(4), 830–847.

Hunn, E. (1999). Size as limiting the recognition of biodiversity in folkbiological classifications: One of four factors governing the cultural recognition of biological taxa. *Folkbiology*, *47*, 47–69.

Hunn, E. S. (1977). *Tzeltal folk zoology: The classification of discontinuities in nature*. New York: Academic Press.

Hunn, E. S. (2008). *A Zapotec natural history: Trees, herbs, and flowers, birds, beasts, and bugs in the life of San Juan Gbëë*. University of Arizona Press.

Hunn, E. S., & Thornton, T. F. (2012). Tlingit birds: An annotated list with a statistical comparative analysis. In *Ethno-ornithology* (pp. 211–240). Routledge.

Kemp, C., & Regier, T. (2012). Kinship categories across languages reflect general communicative principles. *Science*, *336*(6084), 1049–1054.

Lopez, A., Atran, S., Coley, J. D., Medin, D. L., & Smith, E. E. (1997). The tree of life: Universal and cultural features of folkbiological taxonomies and inductions. *Cognitive Psychology*, *32*(3), 251–295.

Peterson, A. T. (2006). Taxonomy is important in conservation: a preliminary reassessment of philippine species-level bird taxonomy. *Bird Conservation International*, *16*(2), 155–173.

Regier, T., Carstensen, A., & Kemp, C. (2016). Languages support efficient communication about the environment: Words for snow revisited. *PloS one*, *11*(4).

Sullivan, B. L., Wood, C. L., Iliff, M. J., Bonney, R. E., Fink, D., & Kelling, S. (2009). ebird: A citizen-based bird observation network in the biological sciences. *Biological Conservation*, *142*(10), 2282–2292.

Taylor, P. M. (1990). Folk biology of the tobelo people: A study in folk classification. *Smithsonian Contributions to Anthropology*.

Waddy, J. A., et al. (1988). *Classification of plants and animals from a groote eylandt aboriginal point of view*. The Australian National University.

Wilman, H., Belmaker, J., Simpson, J., de la Rosa, C., Rivadeneira, M. M., & Jetz, W. (2014). Eltontraits 1.0: Species-level foraging attributes of the world's birds and mammals: Ecological archives e095-178. *Ecology*, *95*(7), 2027–2027.