I 202: Information Organization & Retrieval Fall 2025

Categorization, part 2

This week's focus

Data / Information Collections

Categories

- **Types of categories**
- **Cognitive / language aspects**
  - Naming / Lexical similarity
- Structure
  - Hierarchical / Taxonomy
  - Faceted
  - Overlapping / Clustering
  - Network / Ontology
- Use in Navigation & Search
  - Information Architecture
  - Faceted Navigation

Technology Support for Info Org

- Identifiers
- Metadata
- Markup
- Schema / Databases
- Search Ranking / Evaluation
- Automated category creation
- Automated similarity

Social / Ethical Aspects

- Cultural Bias
- Intellectual Property
- Standards Process

Today's Outline

Let's Play a Game!

Prototypes vs Classical Categories

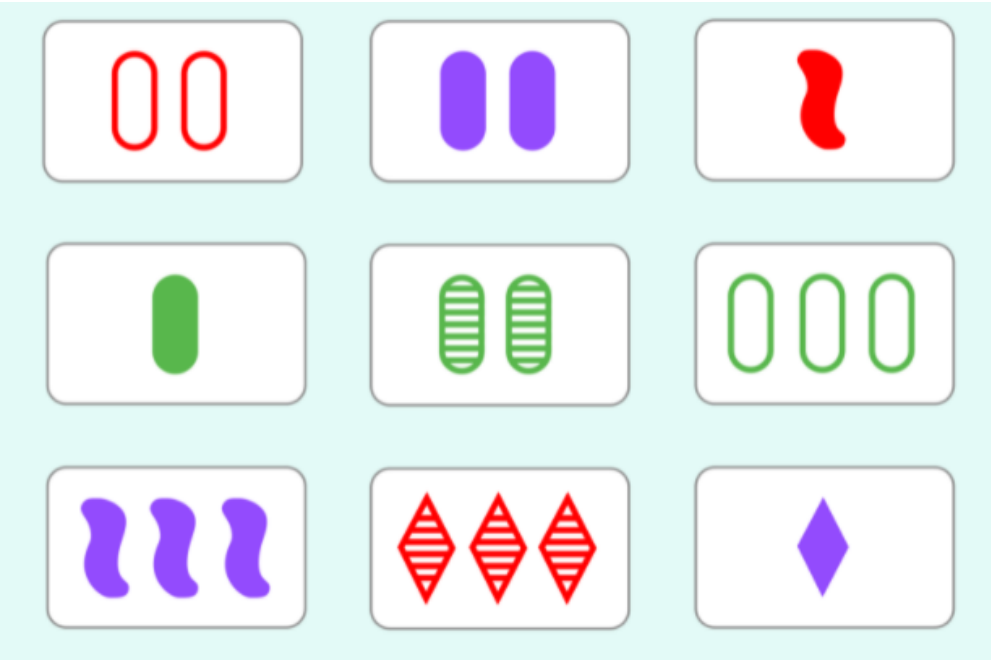
Basic-Level Categories

Language, Concepts, & Categories

Genre as a Category

Let's Play another Game!

The Game of Set



https://smart-games.org/en/set/start

THE GAME OF SET

A SET consists of 3 cards

Each card has a color, shape, number, and fill

For **each** of the 4 categories, the 3 cards must display that category as either

a) all the same, or

b) all different.

Put another way: For each category, the 3 cards must avoid having 2 cards showing one version of the category and the remaining card showing a different version.

Valid SET:

Invalid SET:

two cards have the same shape and one is different

https://smart-games.org/en/set/start

The Game of Set

What kind of category system is this?

The Set Game and Categorization

- **Attributes**: color, shape, number, texture
- **Rule**: a set must consist of exactly three cards; all 3 cards must be the same for an attribute or all different
- These rules define:
  - *What IS a set*
  - *What IS NOT a set*
- The game's rules define **a classic category**

Classical vs Cognitive Science-Based Categories

Culturally / Cognitively-Created Categories

- Learned implicitly through development via parent-child interactions, language, and experience
- Learned explicitly through education, reading
- Informal and contextual acquisition makes categories flexible, creative, generative, and biased

What Are **Necessary and Sufficient** Conditions for Something to be a Game?

A list of properties shared by **all elements?**

Defining **Game** with Necessary & Sufficient Conditions

- All activities that are done for amusement?
  - *Would **include** movies, hanging with friends …*
- All activities that have rules?
  - *Would **include** school, filling out forms*
- All activities that involve competition?
  - *Would **include** non-game activities like racing, competing for jobs*
- All activities done for amusement that have rules and involve competition?
  - *Would **exclude** jigsaw puzzles – not competitive*
- All activities done for amusement that have rules?
  - *Would **exclude** a child tossing a ball and seeing how many times she can catch it*

Merriam-Webster: Game

Splits it up into different concepts:

- “A physical or mental competition conducted according to rules with the participants in direct opposition to each other
- Activity engaged in for diversion or amusement : PLAY
- A procedure or strategy for gaining an end : TACTIC”

Different definitions signal a category whose members are not described by all of the same properties

Oxford Dictionary: GAme

"**Game**: form of play or sport, especially a competitive one played according to rules and decided by skill, strength, or luck."

"or" and "especially" signal a category whose members are not described by all of the same properties

Game Example: The Point

- There is no single set of features all members have in common -- no single all-encompassing definition
- Boundaries of categories are (usually) **not accurately described by one rule.**

Problems with the Classical View

It does not reflect how people categorize:

- People do **not** rely on abstract definitions or lists of shared properties (Rosch 1973)

  Example: Are curtains furniture?
- Some members are **more typical** than others

  Example: Chicken as a bird vs eagle as a bird
- At least some aspects of categorization seem to reflect the human body and mind

  Examples: Color categories, emotion categories

Prototypes: Eleanor Rosch

- UC Berkeley linguistics professor (emeritus)
- Influential work on mental categorization and prototypes
- Prototype theory: People rely less on abstract definitions of categories than on a comparison of the given object or experience with what they deem to be the object or experience best representing a category ("prototype")

https://en.wikipedia.org/wiki/Eleanor_Rosch

PROTOTYPES

- A prototype is a single best example that captures the “central tendency” of a category
- The prototype can be thought of as an average of the features of the category members.
- Gradience in category membership: some members of the category are “better” examples than others

Gradience in Category Membership

Which bird is most typical of the category bird?
Least typical? What does it mean that you can answer this question?

PROTOTYPES

- Which dog breed is central?
- Which are “better” or “worse” examples?

https://articles.hepper.com/great-dane-chihuahua-mix/

https://www.pageant.dog/breeds/dog/1120-Labrador-Retriever

Evidence For Prototypes

Typicality ratings

Order in which members are named

Time needed to verify category membership

Istockphoto.com

What is the difference between a Prototype and a Stereotype?

Prototype vs. Classical View

- **Prototype view**: we start by having prototypes pointed out to us and then cluster other things around them
- **Definitional view** (classical): we start with criteria and then find some good examples
- **Prototype view**: the boundaries of a category are fuzzy
- **Definitional view**: draws sharp boundary lines

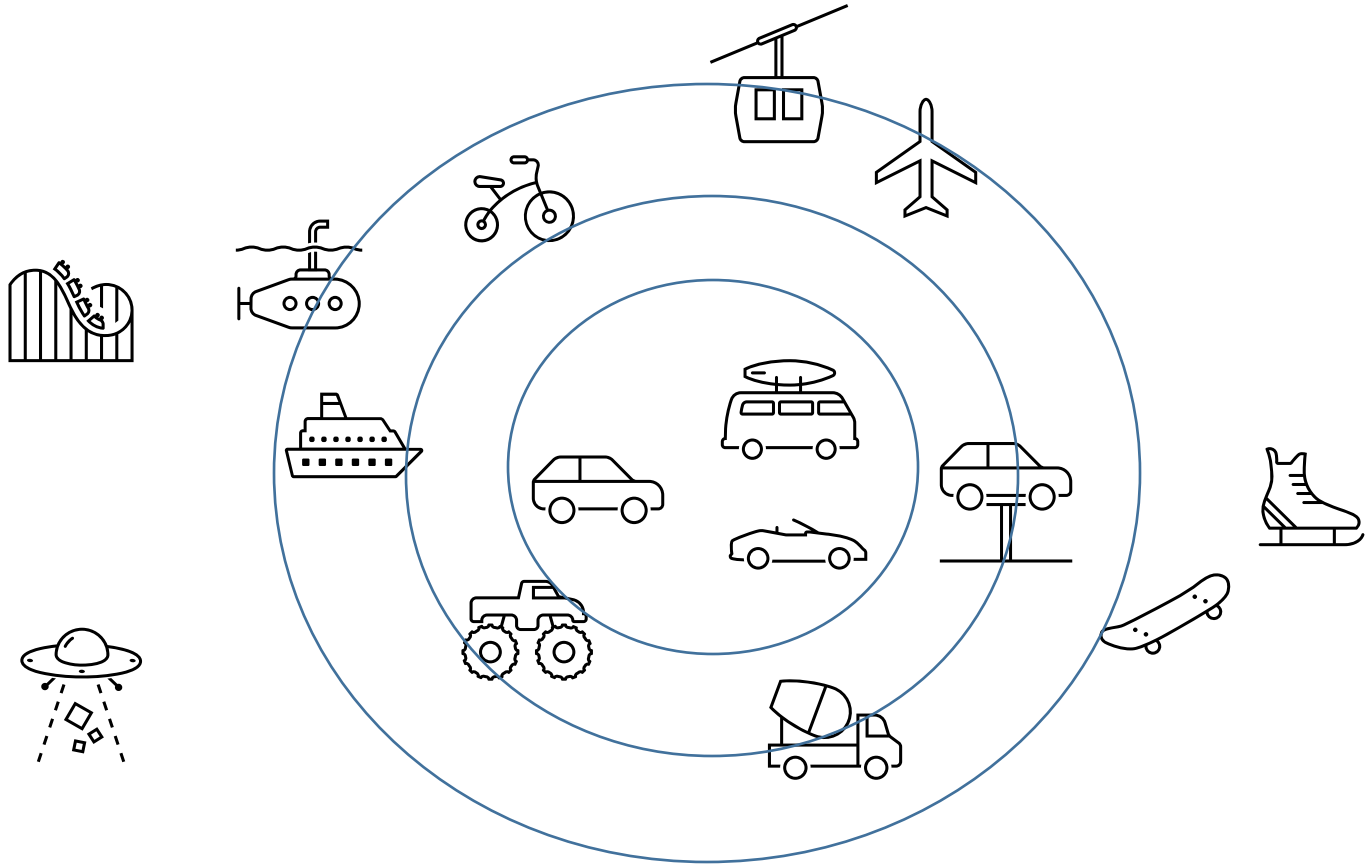
What is the first Example That Comes To Mind of A …

Vehicle

Example: Vehicle

More **Prototypical** is closer to the center

There is a **gradience** in category membership



What Determines Category Prototypes?

- **Purpose categories** – the key concepts center around what is most needed for that purpose
  - *Things you need for international travel: suitcase, passport*
  - *Things you need for camping: tent, sleeping bag*
- **Natural objects categories** – key concepts center around:
  - *their appearance, what they might do to us (snakes)*
- **Human made object categories** – key concepts center around:
  - *how they are used, how we physically interact with them*
- **Human psychological categories** – key concepts center around:
  - *how they make us feel, how they impact our social interactions*

CATEGORY LEVELS

- **Basic-level categories:** those categories at the highest level of generality that still share many common attributes and have fewer distinctive attributes (shape, color, size, etc.)
- **Super-ordinate**: more general categories such as *animal* will share fewer attributes and demonstrate more variability.
- **Subordinate**: more specific categories, such as *American Robin* will share even more attributes among themselves

Rosch. Principles of categorization. *Cognition and Categorization*, pages 27–48, 1978.

Ordonez, et al. "Predicting entry-level categories." *IJCV* 115.1 (2015): 29-43.

Superordinate and Subordinate Levels

SUPERORDINATE	animal	furniture	emotion
BASIC LEVEL	dog	chair	happy
SUBORDINATE	terrier	rocker	joy

What are other examples?

- Children take longer to learn superordinate
- Superordinate not associated with mental images or motor actions

What is special about basic-level categories?

- They are typically learned first by children
- They are the most commonly used in language
  - However, this can differ by culture
- People tend to recognize basic level categories more quickly than super- or sub-ordinate categories
- Related to, but not the same as, prototypes

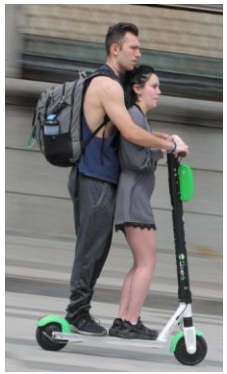
How Do Children Learn Categories?

- **Prototype Formation**: A process in which a single best example (a prototype) is abstracted from experience with different category members and stored in memory
- Infants shown images of specific animals from the same category that varied on size of features responded to a novel prototype animal as if it was more familiar than the detailed examples they had seen.
- As children age, they become able to use other methods of categorization, **including using rules** to learn new categories. However, if the rules are complex, young children cannot form the categories without prototype examples

Klinger & Dawson, Prototype formation in autism, Development & Psychopathy, 2001

Prototype vs. Classical In Practice

- Although we think more naturally with fuzzy boundaries, we sometimes are forced to make sharp distinctions
- Example: Dept of Motor Vehicles has to classify which vehicles require licenses
- Answers questions like:
  - *Licenses?*
  - *Helmets?*
  - *Sidewalk?*

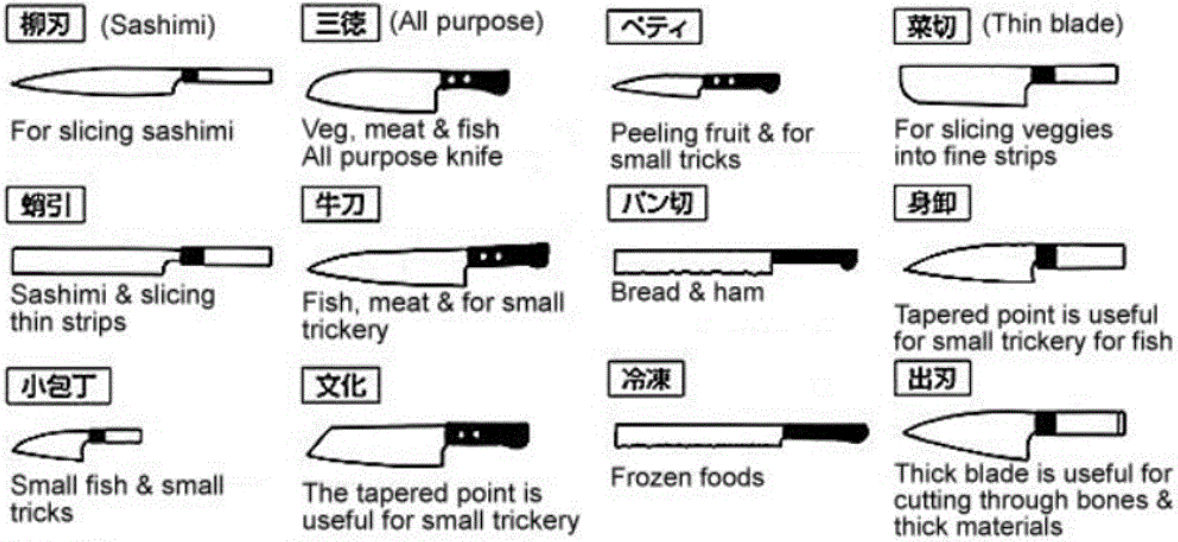


What are the California DMV Classes of E-Bikes? What are NOT E-Bikes?

How do licenses, rules for use, vary among them?

Language? Concepts?
Which determines which?

Whorf's Hypothesis: Does Language Constrain concepts?



Does your language have different words for each of these kinds of knives? If it doesn't, can you still distinguish them?

Language and Categories

- Languages differ a great deal in the words they contain (which concepts are lexicalized)
- The **Whorfian** hypothesis: language influences the speaker's worldview
- But we can understand concepts that we don't have words for.
- The current consensus is that the strong Whorf hypothesis is incorrect, but language can influence rather than determine thought.

Color Naming Research

- Berlin & Kay: there seem to be 11 basic color categories across 110 cultures
- Cultures disagreed about which subset of 11 colors count as basic
  - *Russian: no single word for blue*
  - *French: no word for brown*
- Rosch studied a tribe in New Guinea called the Dani that had only 2. Nonetheless, the Dani recognized the same 11 basic colors more quickly than non-basic colors.
- In other words, universally, some reds are better examples than other reds.
- This also suggests that something embodied in determining what is a "good" version of a color

Color Naming Research

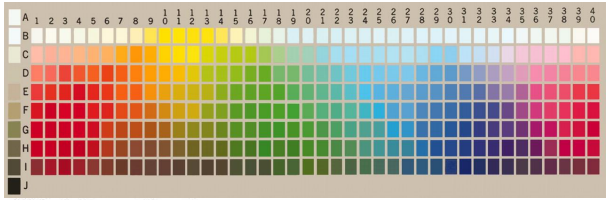
- World color survey: 110 languages of non-industrialized societies, 330 color chips
- Findings: the best-example choices for color terms in these languages cluster near the prototypes for English *white*, *black*, *red*, *green*, *yellow*, and *blue*

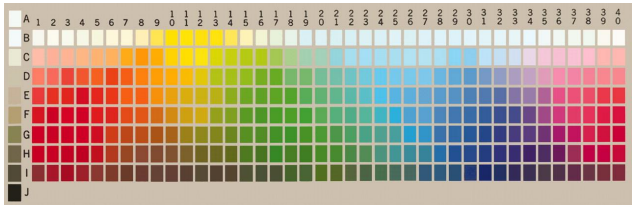
Terry Regier
Prof, UCB Linguistics

Focal colors are universal after all, Regier, Kay, Cook, PNAS 2005

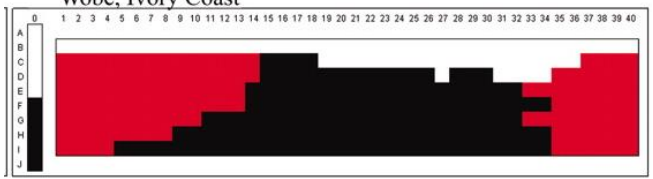
Paul Kay
Prof, UCB Linguistics

Example Color Naming Systems

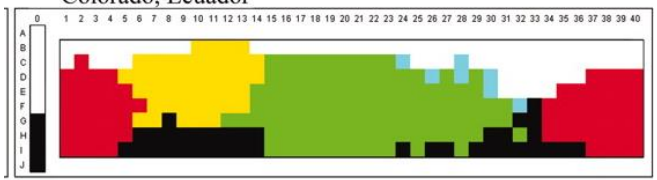




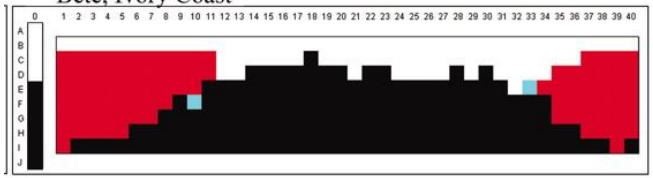
Wobé, Ivory Coast



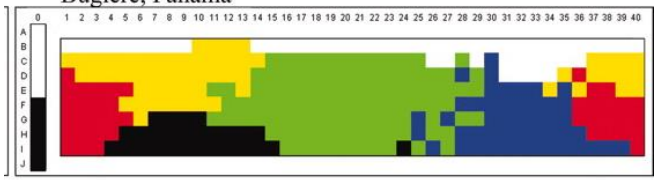
Colorado, Ecuador



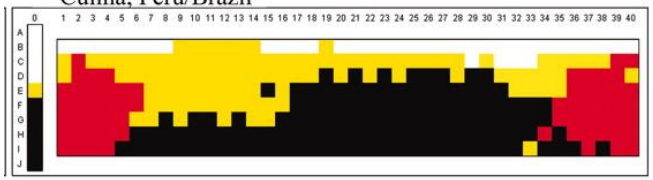
Bété, Ivory Coast



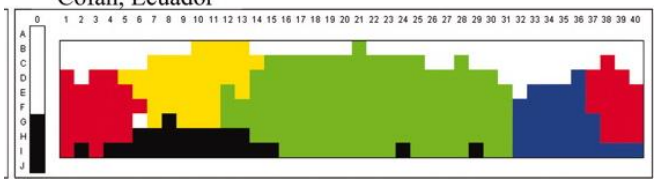
Buglere, Panama



Culina, Peru/Brazil



Cofán, Ecuador



Color naming reflects optimal partitions of color space, Regier, Kay, Khetarpal, PNAS 2007

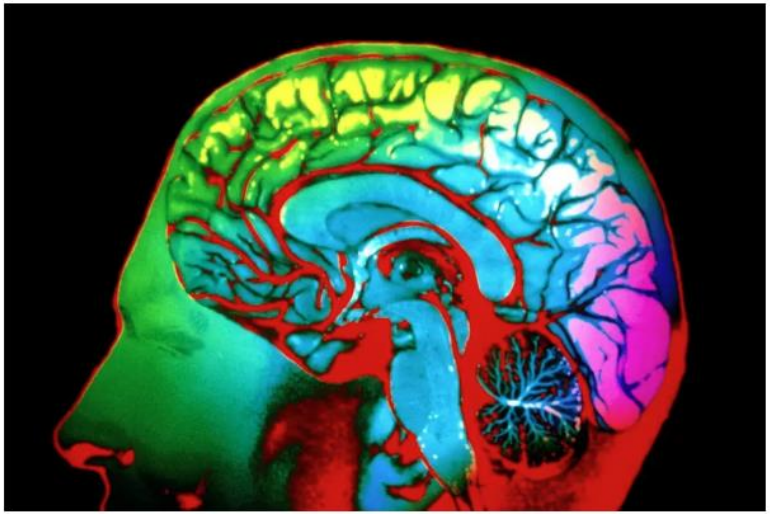
The New York Times

Do You See the Same Colors That I Do?

Scientists cannot say for certain, but new research suggests that different people's brains respond similarly when looking at a particular hue.

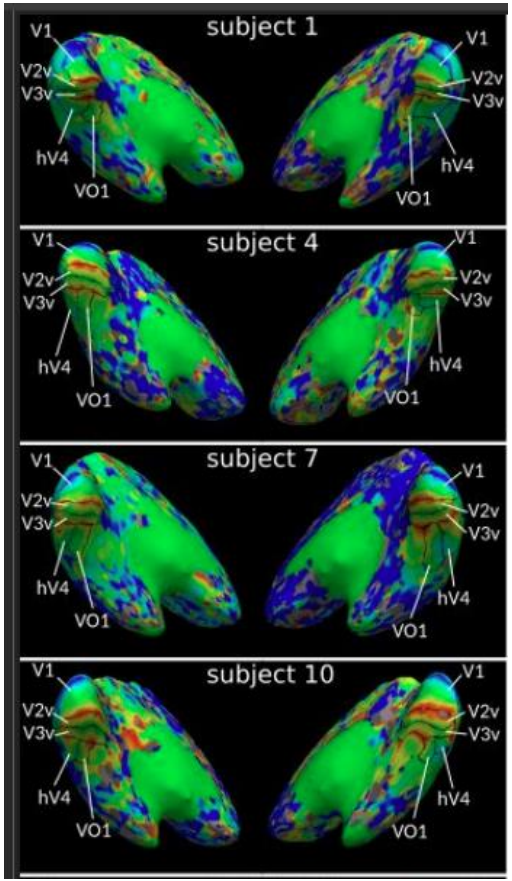
Listen to this article · 4:49 min Learn more

Share full article



GJLP/Science Source

Chang, NYTimes, 9/8/25



Bannert, Bartels, Journal of Neuroscience 8 September 2025,

Color Naming: The Point

- Categories develop as a combination of cultural, cognitive, and experiential factors
- Color naming provides evidence that this category has a prototype structural across all cultures

Concept Acquisition and Blindness

- Recent studies show that blind people acquire most of the same concepts as sighted people in the same culture
- Much of this is conveyed through language and concept formation
- For example, blind and sighted adults are
  - Equally likely to infer that two bananas (natural kinds) and two stop-signs (artifacts with functional colors) are more likely to have the same color than two cars (artifacts with non-functional colors)
  - Make similar inferences about novel objects' colors
- There are some differences in how blind vs sighted people make similarity judgements about fruits and vegetables, but not about human made artifacts.

Kim, et al. "Shared understanding of color among sighted and blind adults." *PNAS* (2021).

Connolly et al. "Effect of congenital blindness on the semantic representation of some everyday concepts." *PNAS 2007*

How do we deal with a new concept?

grampus griseus

Ordonez, et al. "Predicting entry-level categories." *IJCV* 115.1 (2015): 29-43.

WHAT IS THE RELATIONSHIP BETWEEN NEW **CONCEPTS** AND **CATEGORIES**?

- We learn new concepts all the time
  - *Being "cancelled"*
  - *The blockchain*
  - *Fire-induced storms*
  - *E-sports*

- We can
  - *Slot it into an existing category, perhaps expanding its boundaries*
  - *Start to create a new category with this concept as a prototype*
  - *Shift boundaries of multiple existing boundaries*

GENRE

Definition (Oxford Languages): “A category of artistic composition, as in music or literature, characterized by similarities in form, style, or subject matter”

FILM GENRE POPULARITY 1910-2018

CLICK TO STANDARDIZE AXIS RANGE

BO MCCREADY @BOKNOWSDATA

This graphic shows film genre popularity over time, represented as the percentage of all films released that year with the specified genre tagged on IMDB. Each genre has a different axis range, so these lines show popularity relative to other years, not necessarily relative to other genres.

tableau

Film Genre Popularity - 1910-2018 by Bo McCready

What is Genre?

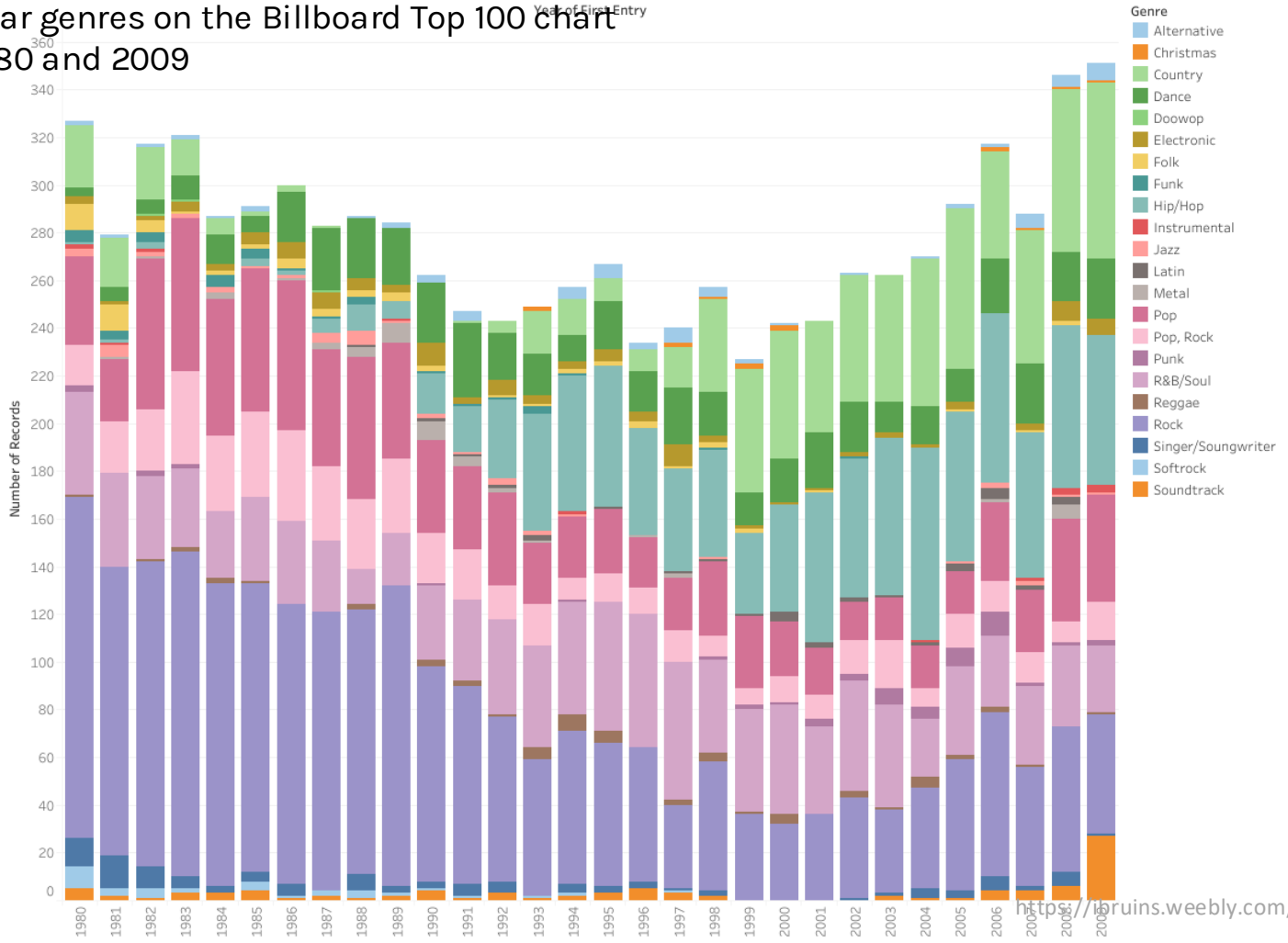
- "Any widely recognized class of texts defined by some common communicative purpose or other functional traits, provided the function is connected to some formal cues… and is extensible."
- Genre is a **heterogenous** classificatory system, based on
  - How the text was created
  - The way it is distributed
  - The register of language it uses
  - The kind of audience it is addressed to
- Examples: reporting, legal, scientific articles, fiction, editorial

Register corresponds to context of use: a formal speech, a broadcaster, talking with friends all have different registers.

Kessler, Nunberg, Schuetze. "Automatic Detection of Text Genre." *EACL 1997*

Popularity of Genres 1980-2010

most popular genres on the Billboard Top 100 chart between 1980 and 2009



https://bruins.weebly.com/visualizations.html

Petrusich Genre Reading Raises Questions about the Role of Genre

- Do the changing patterns of consumption of commercial music make genre irrelevant / antiquated?
- What is the link between identity and musical genre?
- Are there fixed attributes for defining musical genre?

About Discussion of Sensitive Topics

Practice Active Listening: listening to learn, understand, integrate and contribute

Carefully attend to what is said

Not just how it relates to what you want to say

Respect others' knowledge

Give people the **benefit of the doubt**

We have different experiences and priors, and come with different knowledge and sensitivities.

Ask other parties to **explain what is meant**

*Ask for clarifications or further explanation of **ambiguous ideas or statements***
No question is a bad question.

Address disagreement purposely and openly

Acknowledge and clarify differences
Don't impute justifications or intent, solicit more information
*Critique the **idea**, not the person*

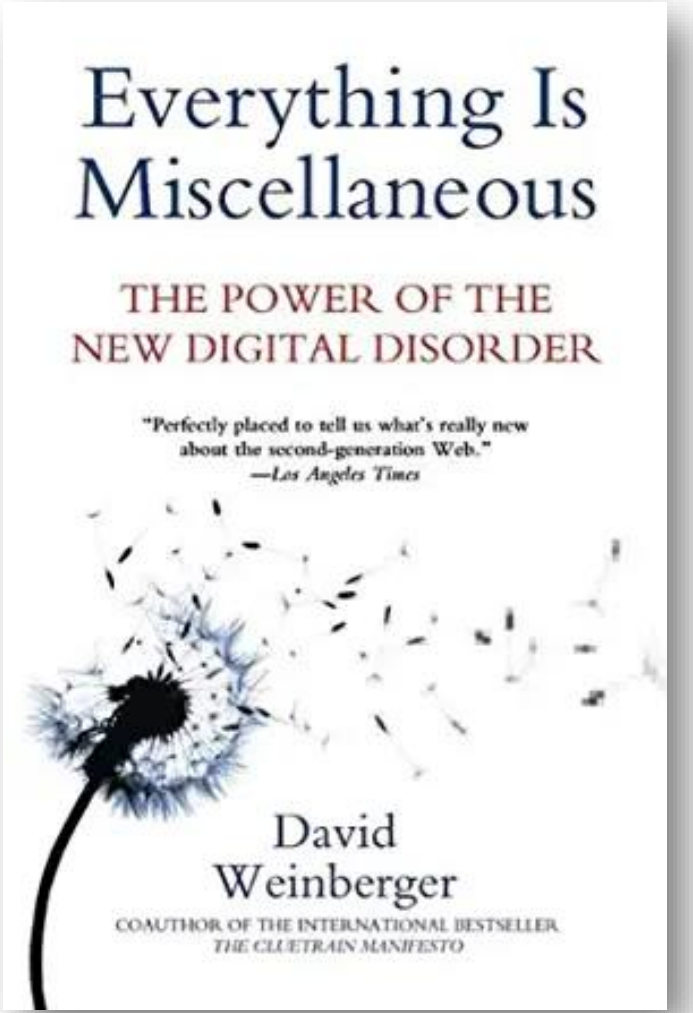
SUMMARY: CATEGORIZATION

Categorization is central to how we organize information and the world

Categories are involved whenever we communicate, analyze, predict, or classify

Categorization is much messier than our computer systems and applications would like

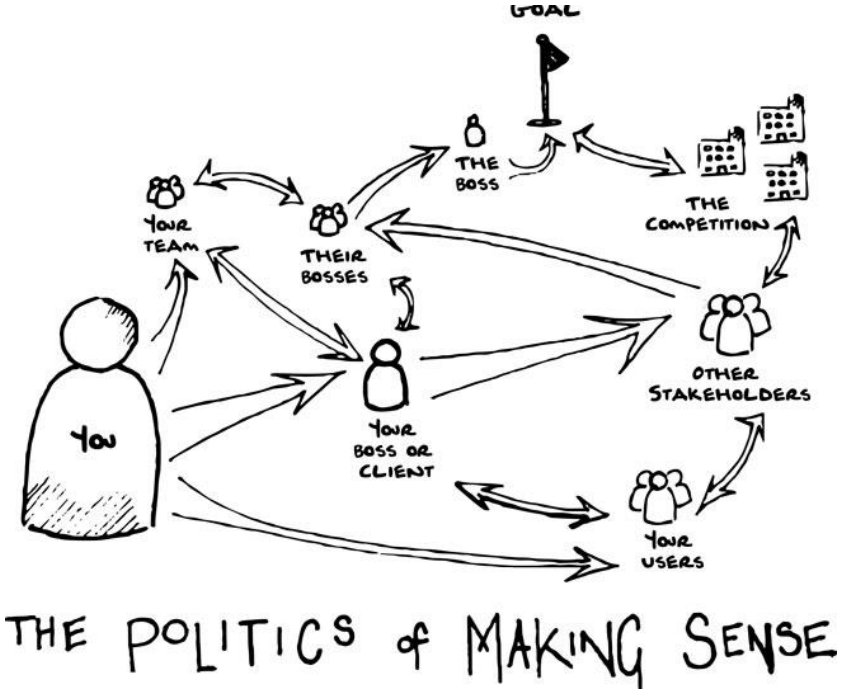
But understanding how people (and each of us) categorize can help us design better systems and interfaces



- Terrific and fun book on categorization and organization
- Written by an entertaining author
- His other books are great too

How to Make Sense of Any Mess by Abby Covert

Table of Contents | Indexed Lexicon | IA Resources | Buy the Book



Fun online book

This week's focus

Data / Information Collections

Categories

- **Types of categories**
- **Cognitive / language aspects**
  - Naming / Lexical similarity
- Structure
  - Hierarchical / Taxonomy
  - Faceted
  - Overlapping / Clustering
  - Network / Ontology
- Use in Navigation & Search
  - Information Architecture
  - Faceted Navigation

Technology Support for Info Org

- Identifiers
- Metadata
- Markup
- Schema / Databases
- Search Ranking / Evaluation
- Automated category creation
- Automated similarity

Social / Ethical Aspects

- Cultural Bias
- Intellectual Property
- Standards Process

Next week's focus

Data / Information
Collections
Categories
- Types of categories
- Cognitive / language aspects
    - Naming / Lexical similarity
- **Structure**
    - **Hierarchical / Taxonomy**
    - **Faceted**
    - **Overlapping / Clustering**
    - **Network / Ontology**
- Use in Navigation & Search
    - Information Architecture
    - Faceted Navigation

Technology Support for Info Org
- Identifiers
- Metadata
- Markup
- Schema / Databases
- Search Ranking / Evaluation
- Automated category creation
- Automated similarity

Social / Ethical Aspects
- Cultural Bias
- Intellectual Property
- Standards Process