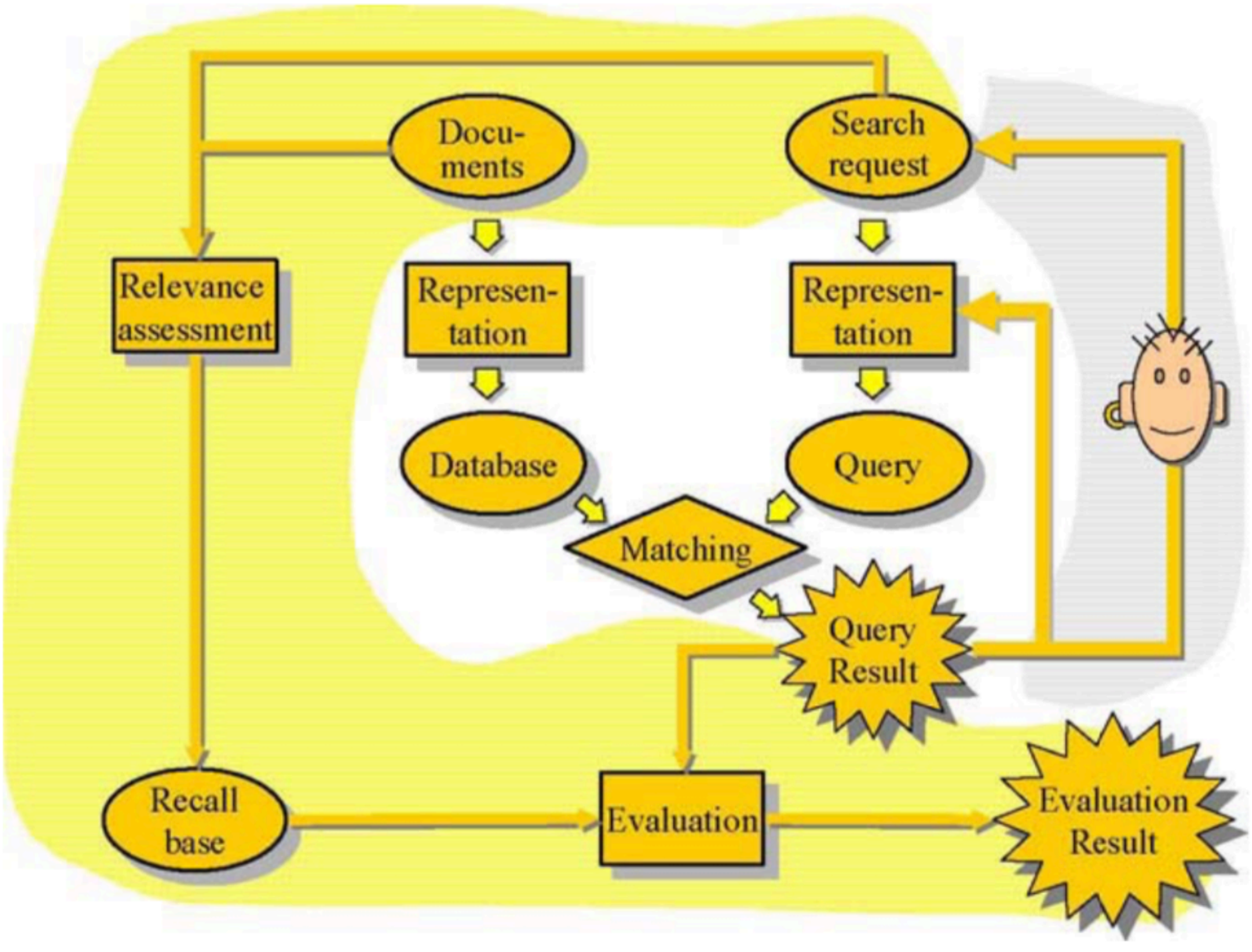
Topics in ~~User~~ Cognitive Modelling in Interactive Information Retrieval

Alan Medlar (alan.j.medlar@helsinki.fi)
Dorota Glowacka (glowacka@cs.helsinki.fi)

Schedule

- **Changes will appear on the course webpage**
- ~~04.09.19 Lecture 1: Introduction to IR and IIR~~
- **11.09.19 Lecture 2: Cognitive modelling**
- 25.09.19 Deadline for topic selection (title + 3 papers min.)
- 09.10.19 Presentation of chosen topic (5 mins, 5 slides)
- 30.10.19 Feedback session
- 20.11.19 Final presentations (20 mins, 20 slides) - if necessary
- 27.11.19 Final presentations (20 mins, 20 slides)
- 11.12.19 Deadline for final paper submission

Recap 1: IR evaluation



Recap 2: IIR evaluation

Research Questions, Hypotheses, and Theory. Explicit research questions were found in 19.3% of the studies ($n = 29$), explicit hypothesis were found in 10.7% ($n = 16$) of the studies, and both a research question and a hypothesis were found in 4.7% of the studies ($n = 7$). In 65.3% ($n = 98$) of the studies, there was neither an explicitly stated research question nor hypothesis.

[objectives are to] **"compare two search systems,"** which suggests an implicit research question focused on basic evaluation.

D. Kelly and C. Sugimoto (2013). "**A systematic review of interactive information retrieval evaluation studies, 1967–2006**". JASIST 64.4, pp. 745–770.

Recap 2: IIR evaluation

statistics, while 9% ($n = 11$) did not provide any indication of which type of analysis was used, despite claiming statistically significant results or presenting probability values. Almost all the analyses were performed variable-by-variable and were conducted to compare the systems. Only a small percentage of articles described statistical analyses that attempted to model performance using multiple input variables ($n = 6$, 5%).

Why cognitive models?

- Cognitive models relate models of **psychological processes** to behavioral data

- A cognitive model should be viewed as a **hypothesis** for an experiment

- We use **data** to identify which model has the best empirical support (model selection)

- We can perform evaluation on the basis of **model parameters** instead of raw data

Essay structure

- Essay will have 3 sections:

  - **An IIR component (or search task)** (e.g. ranking, relevance feedback, implicit relevance feedback) - what does it do? how is it implemented? how is its effectiveness validated?

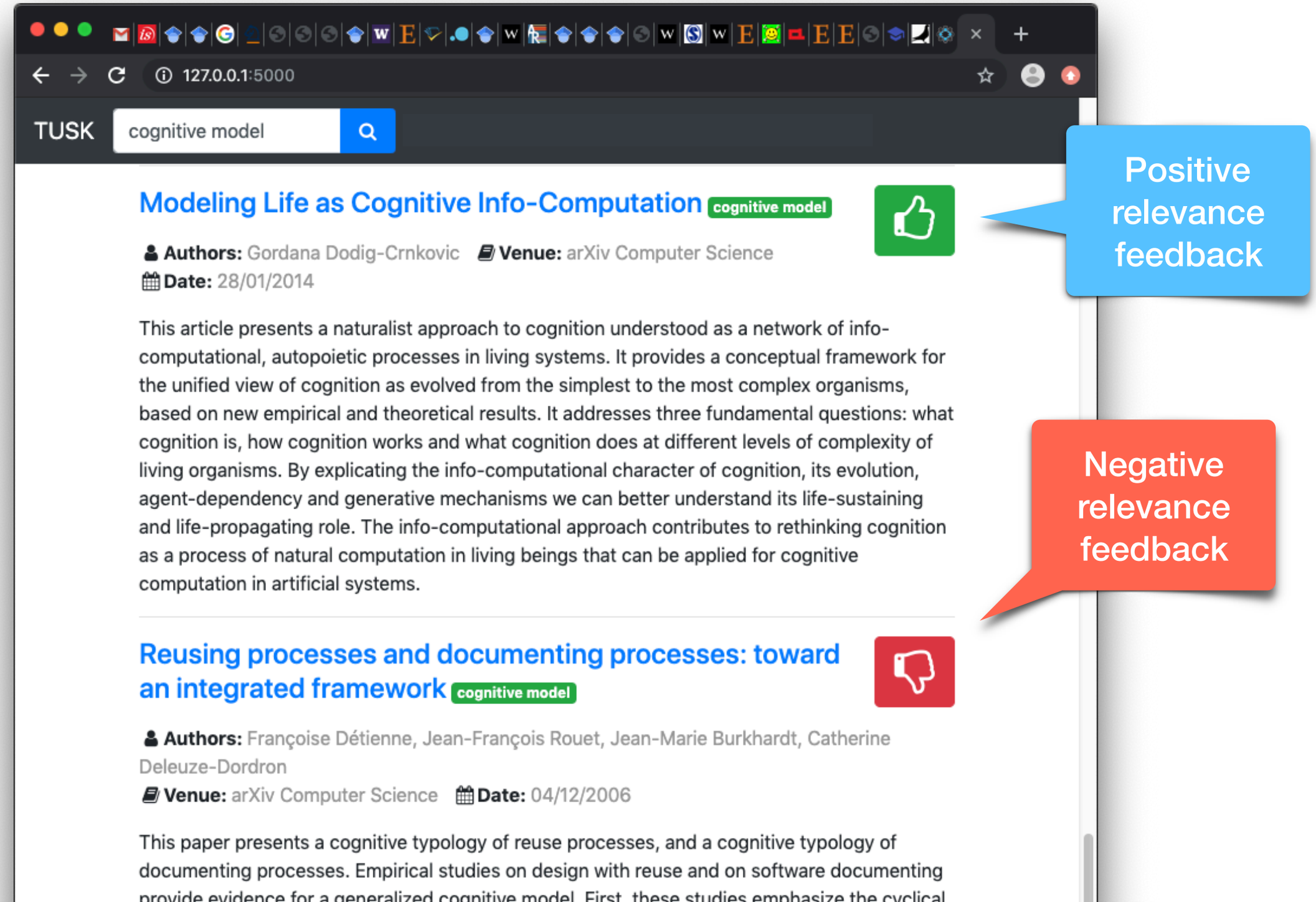
  - **A cognitive process** (e.g. categorisation, decision making, implicit learning) - what does it study? describe the model, what type of experiment is used to gather data?

  - **Cognitive modelling in IIR** (e.g. modelling relevance feedback as a categorisation process) - sketch an experimental design, what old results can be replicated? what new results will we get?

Essay structure

- Two approaches:
  - **Correlation:** Gain an understanding of IIR component/ search task + cognitive model ⇒ propose hypothesis relating the two + experiment to test hypothesis
  - **Integrative:** Gain a deep understanding of IIR component/search task + cognitive model ⇒ propose an experiment to fit the model itself

Relevance feedback



Theory of Categorization

- Can do we determine if something is the member of a category?
  - Aristotelian model
  - Prototype models
  - Exemplar models

Aristotelian categorization model

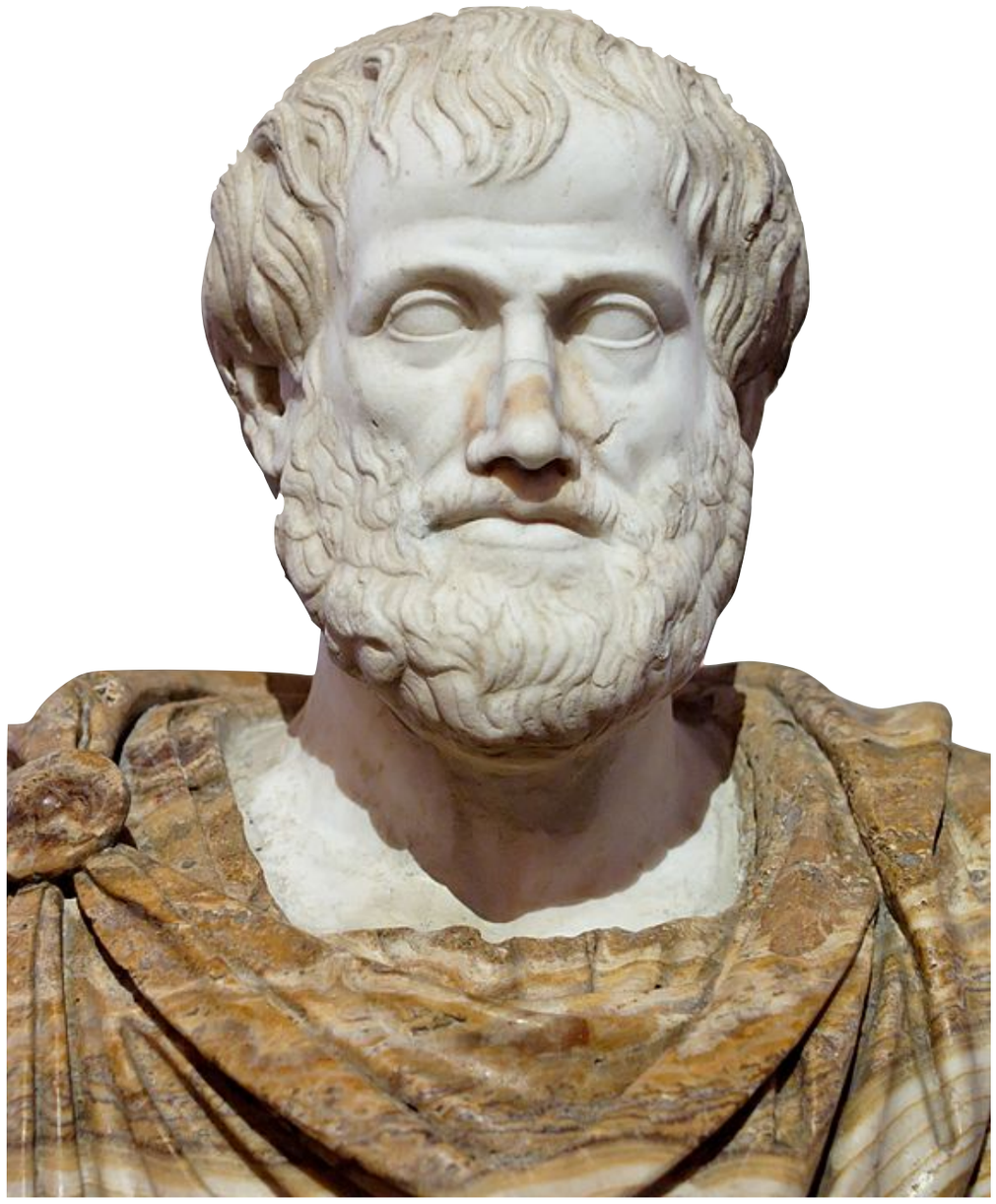
- Classical view of categories: Plato + **Aristotle** (The Categories)

- Categories defined by list of features shared by all category members

- Properties are necessary conditions of category membership: entity must have **all features** to be a member of category

- Categories are strictly defined, mutually exclusive and collectively exhaustive

- Members are equal; no entity is more of that category than another



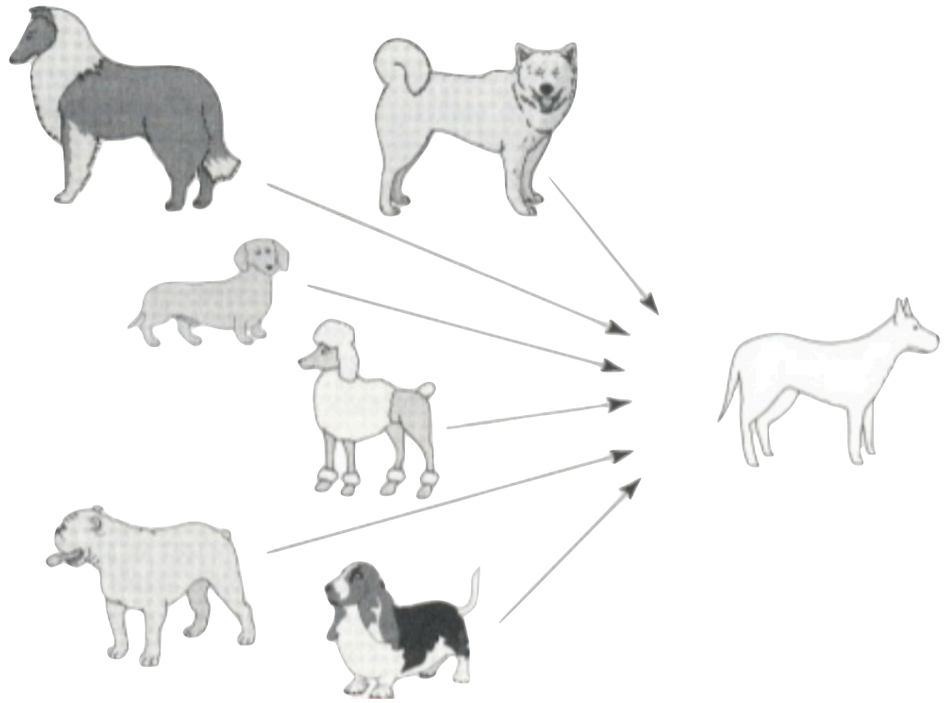
https://plato.stanford.edu/entries/aristotle-categories/

Aristotelian categorizatio[illegible]l

- Criticisms of Aristotelian categorization:
  - Entities can belong to a category, without sharing all features
  - Some entities are better exemplars of categories than others (degrees of membership)
  - Membership not strictly binary

Prototype models

- Category judgments are made by comparing an object to a **prototype** (summary, average) in terms of **similarity**

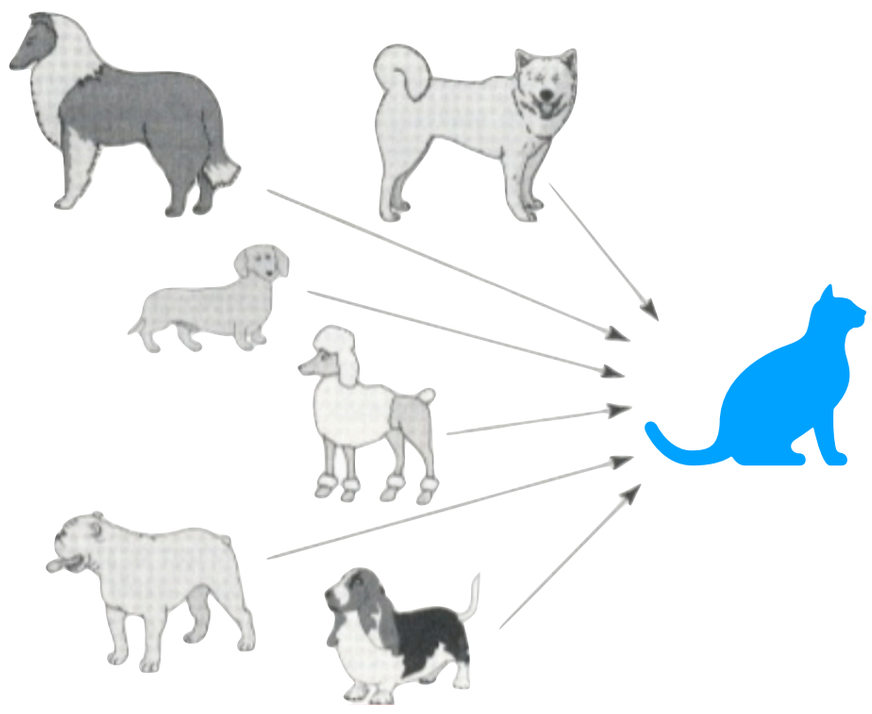


- Prototype need not be real, purpose is to define membership by similarity - could emphasise features that distinguish between categories

- Categorises are not "out there" in the world, but rooted in human experience

Prototype models

- Criticisms of Prototype models:
  - Information can be lost, i.e. it is difficult to model:
    - differences in category size
    - differences in category variability
    - correlations between features
    - multimodal distributions
    - outliers!



Exemplar models

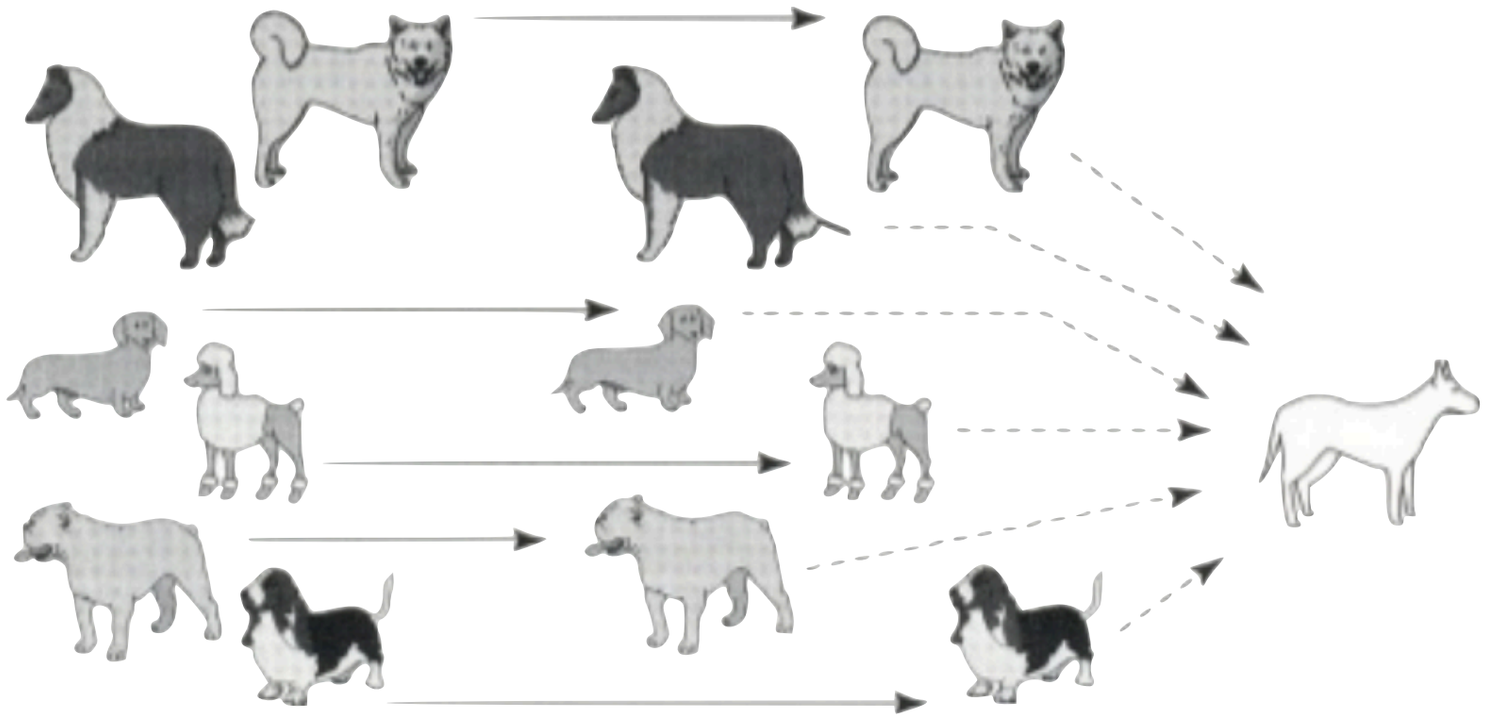
- Category judgments are made by comparing an object to **all members of category** (so-called **exemplars**) in terms of similarity

- Similarities are aggregated to make a categorization decision

- **Note similarities:**

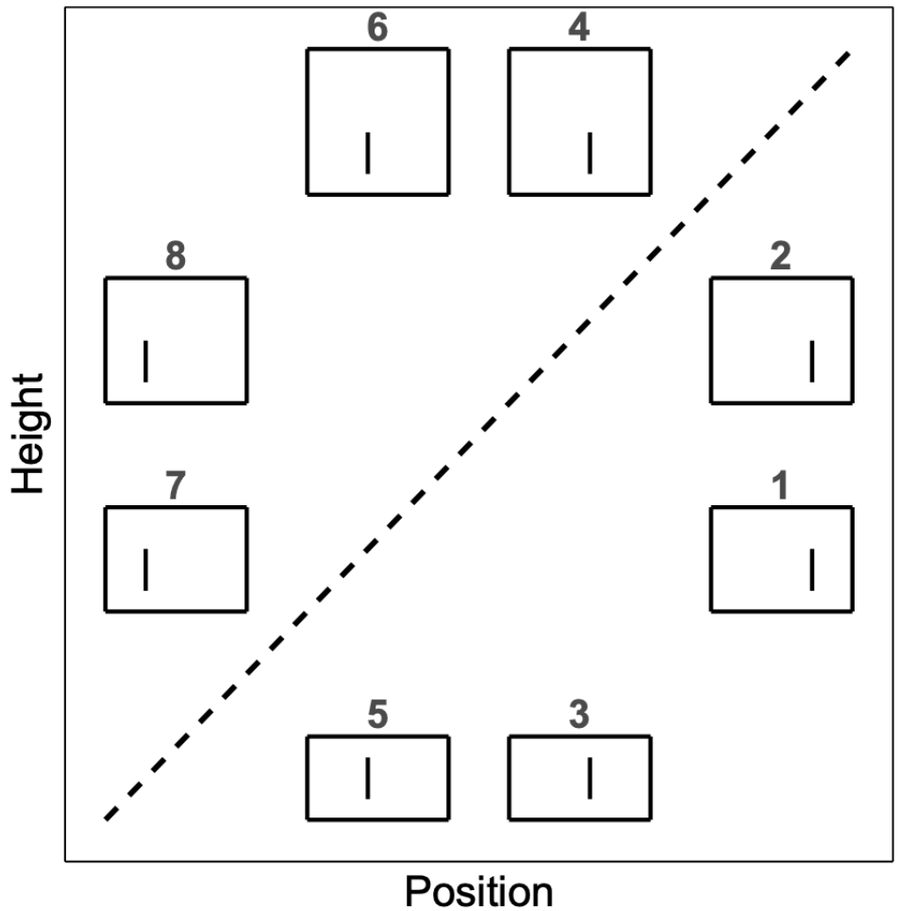
  - 1 exemplar/category = prototype

  - low variability of exemplars = little information loss with prototype



Categorization Example

- Categories A and B (above and below x=y)
- Two features:
  - position of line
  - height of box
- **What is the relative importance of features for categorization?**



M.Lee, and E.Wagenmakers (2014). "**Bayesian cognitive modeling: A practical course**". Cambridge university press.

Generalized Context Model (GCM)

- Exemplar model of categorization
- Category representation is just a list of category members (exemplars)
- Assume simple case (2 features per exemplar, 2 categories):
- We need a distance function (w is the attentional weight):

$$d_{ij} = w|p_{i1} - p_{j1}| + (1-w)|p_{i2} - p_{j2}|$$

- ...and a similarity function (c scales the drop-off in similarity with increasing distance):

$$s_{ij} = \exp(-c \cdot d_{ij})$$

R.Nosofsky (1986). "**Attention, similarity, and the identification–categorization relationship.**" In: Journal of experimental psychology: General 115.1, p. 39.

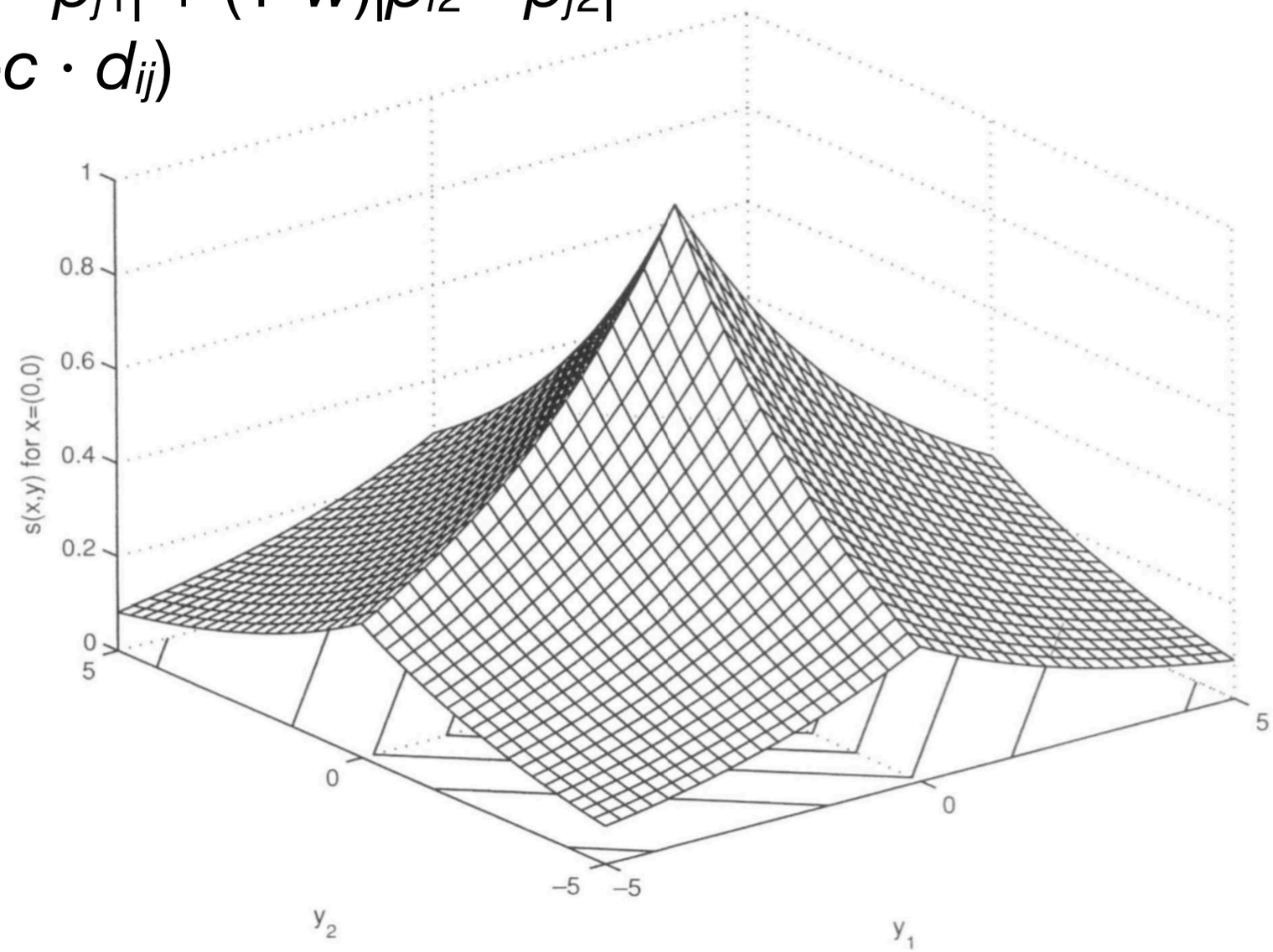
Similarity function

$d_{ij} = w|p_{i1} - p_{j1}| + (1-w)|p_{i2} - p_{j2}|$

$s_{ij} = \exp(-c \cdot d_{ij})$

$c = 0.5$

$w = 0.5$



J.Kruschke (2008). “**Models of categorization**”. In:
The Cambridge handbook of computational psychology, pp. 267–301.

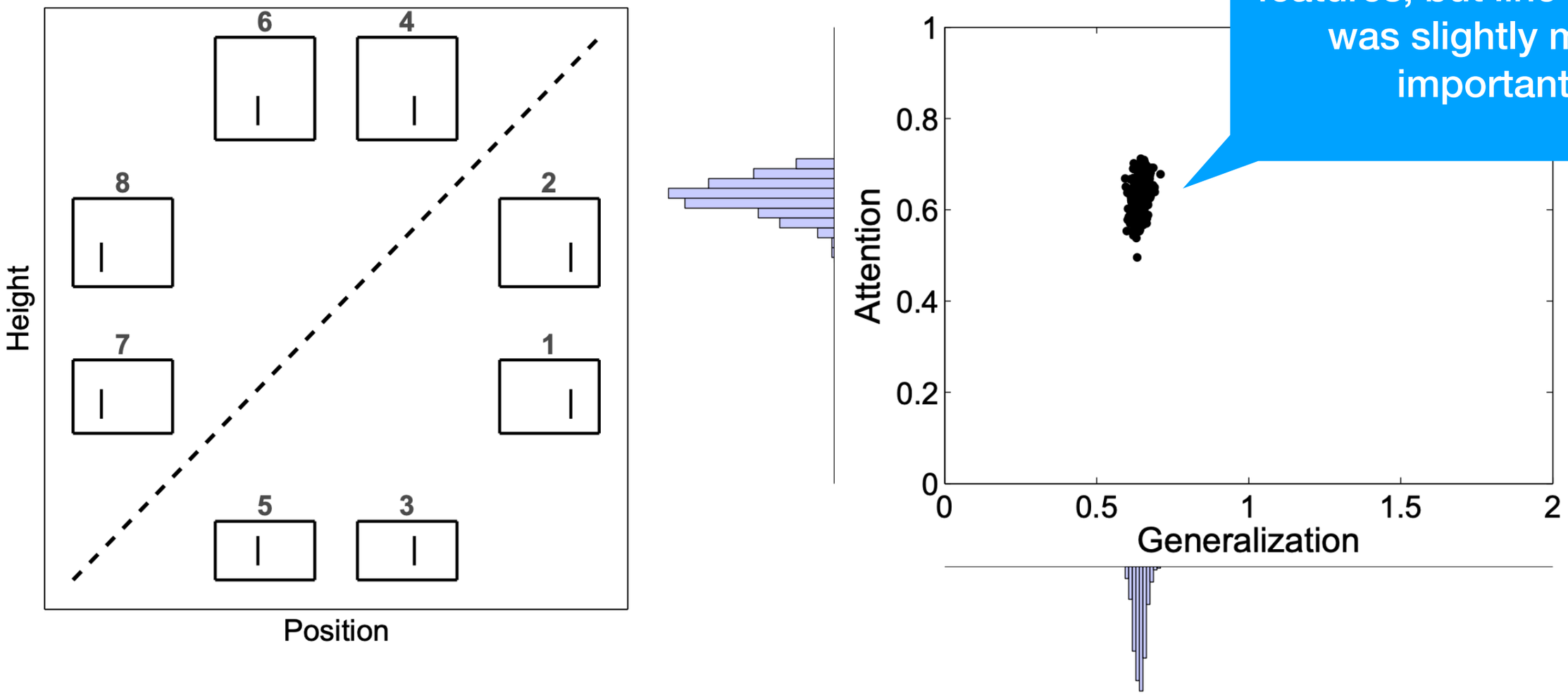
Generalized Context Model (GCM)

- Probability of classifying exemplar *i* into category *A* (as opposed to category B) is:

$$P(R_i = A|i) = \frac{\sum_{j \in A} s_{ij}}{\sum_{j \in A} s_{ij} + \sum_{j \in B} s_{ij}}$$

R.Nosofsky (1986). "**Attention, similarity, and the identification–categorization relationship.**"
In: Journal of experimental psychology: General 115.1, p. 39.

Categorization Example

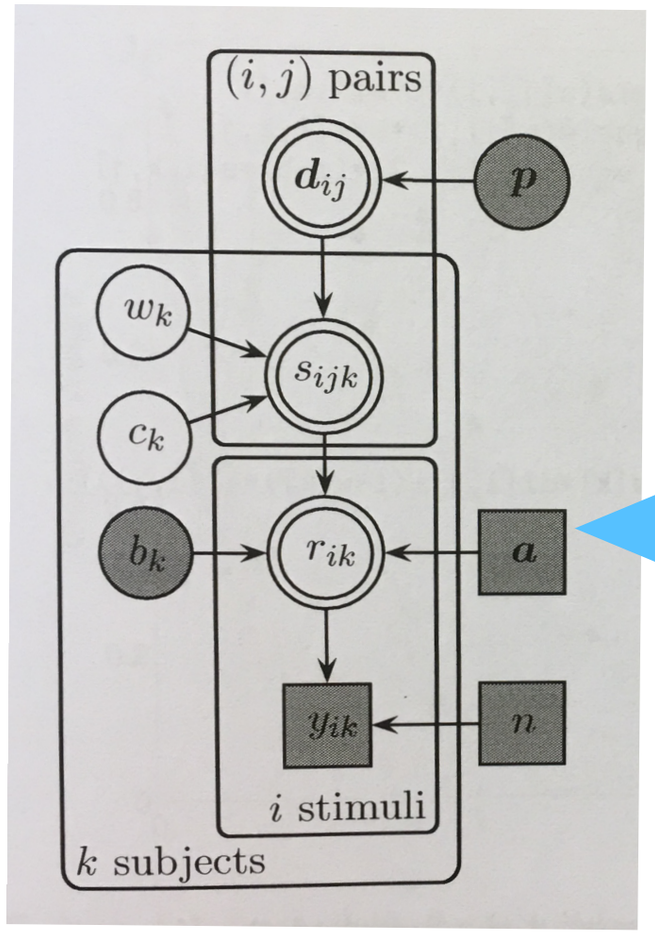


Participants used both features, but line position was slightly more important

M.Lee, and E.Wagenmakers (2014). "**Bayesian cognitive modeling: A practical course**". Cambridge university press.

Generalized Context Model (GCM)

- Bayesian GCM with repeated measures:

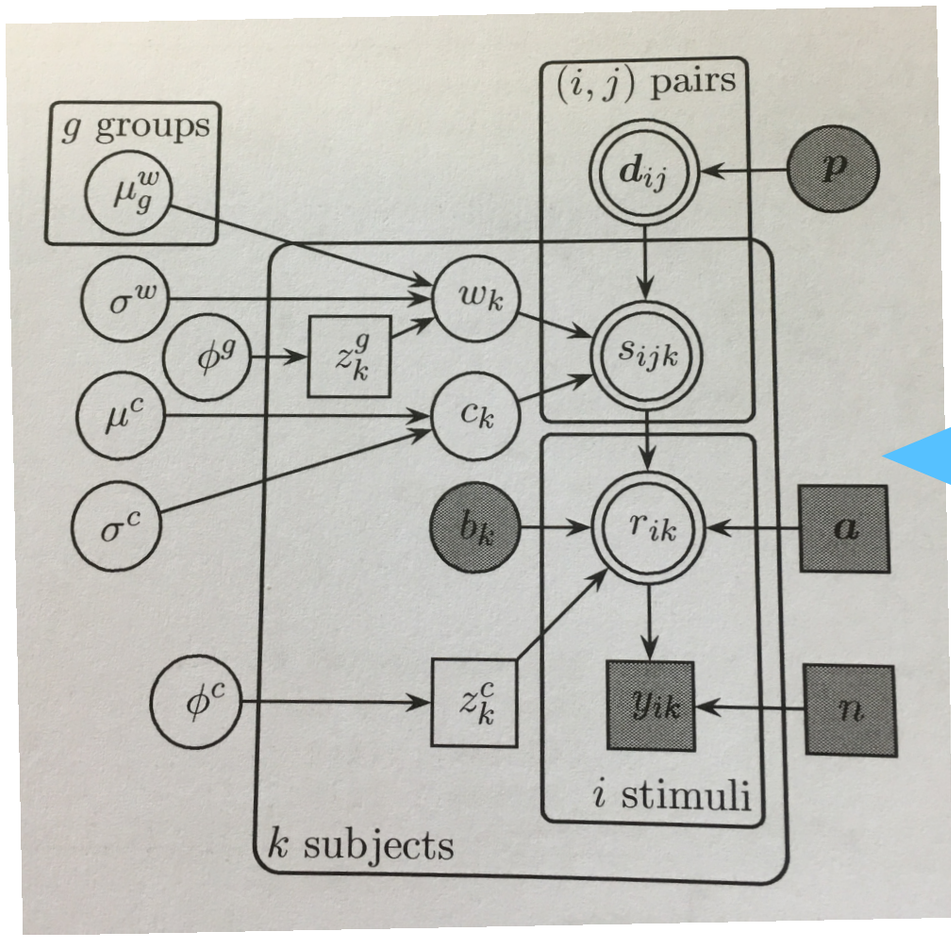


Essay idea 1: for relevance feedback, how does the attentional weight distribution (w_k) vary between:
- systems?
- user groups?
- search tasks?

M.Lee, and E.Wagenmakers (2014). "**Bayesian cognitive modeling: A practical course**". Cambridge university press.

Generalized Context Model (GCM)

- Bayesian GCM with repeated measures and a latent-mixture including contamination:



Essay idea 2: for relevance feedback, how does the proportion of <u>contamination</u> users change between:
- **systems?**
- **user groups?**
- **search tasks?**

M.Lee, and E.Wagenmakers (2014). "**Bayesian cognitive modeling: A practical course**". Cambridge university press.

Generalized Context Model (GCM)

- Probability of classifying exemplar *i* into category *A* is (alternate version):

$$P(R_i = A|i) = \frac{\left(\sum_{j \in A} s_{ij}\right)^{\gamma}}{\left(\sum_{j \in A} s_{ij}\right)^{\gamma} + \left(\sum_{j \in B} s_{ij}\right)^{\gamma}}$$

- $\gamma = 1$: same response as original GCM
- $\gamma < 1$: responses are increasingly random
- $\gamma > 1$: responses are increasingly deterministic

Essay idea 3: for relevance feedback, how does γ differ between lookup and exploratory search?

S.McKinley and R.Nosofsky (1995). "**Investigations of exemplar and decision bound models in large, ill-defined category structures.**" In: Journal of Experimental Psychology: Human Perception and Performance, 21.1, p. 128.

A.Medlar and D.Glowacka (2018). "**How Consistent is Relevance Feedback in Exploratory Search?**" In: Proc. of the 27th ACM CIKM. pp. 1615–1618.

Essay suggestions

- **IIR interface components:**
    - ranking
    - relevance feedback
    - implicit relevance feedback
- **Search tasks:**
    - Lookup/exploratory search
- **Cognitive models:**
    - **Decision making**
        - Categorization processes (prototype vs exemplar)
        - Signal detection theory (discriminability vs bias)
        - Diffusion decision processes (berry-picking)
    - **Other**
        - Working memory, perceptual speed, risk taking, implicit learning, etc.

I will add starting points to many of these topics on the course webpage!

Next deadline...

- **Changes will appear on the course webpage**
- ~~04.09.19 Lecture 1: Introduction to IR and IIR~~
- ~~11.09.19 Lecture 2: Cognitive modelling~~
- **25.09.19 Deadline for topic selection (title + 3 papers min.)**
- 09.10.19 Presentation of chosen topic (5 mins, 5 slides)
- 30.10.19 Feedback session
- 20.11.19 Final presentations (20 mins, 20 slides) - if necessary
- 27.11.19 Final presentations (20 mins, 20 slides)
- 11.12.19 Deadline for final paper submission