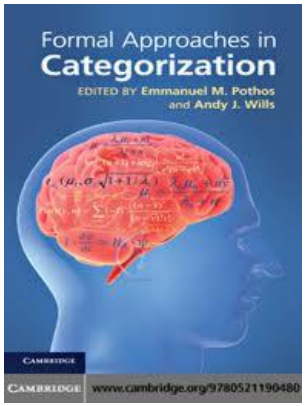
# On the adequacy of current empirical evaluations of formal models of categorization

Andy J. Wills - *University of Exeter, UK*

Emmanuel Pothos – *University of Swansea, UK*




Wills, A.J. and Pothos, E.M. (2012). *Psychological Bulletin, 138,* 102-125.

Many thanks to: Brad Love, Todd Maddox, Rob Nosofsky

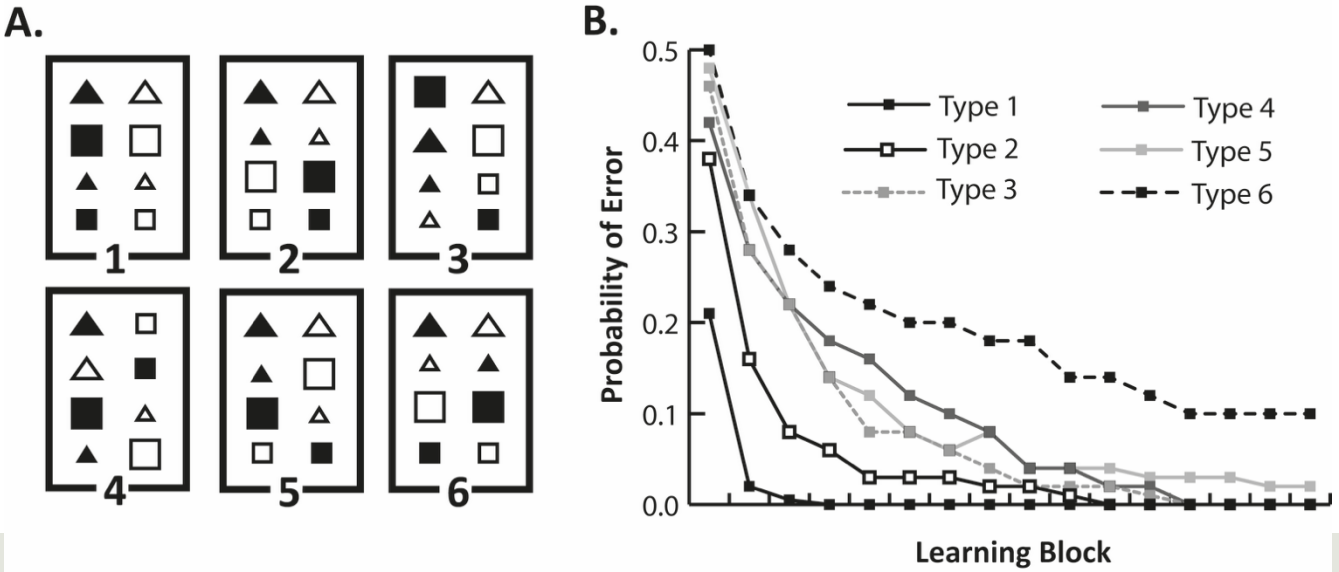
*15 words / second !*

# Categorization

- Dividing the world into groups of things.
- Building block of cognition.
  - *Obama* but not *President*
- Confluence of key processes
  - Memory, decision-making and contingency learning.

# Formal models

- Unambiguously specify transformations from one or more independent variables to one or more dependent variables.
- The model's outputs can be determined from its inputs by a computer program without further human intervention.



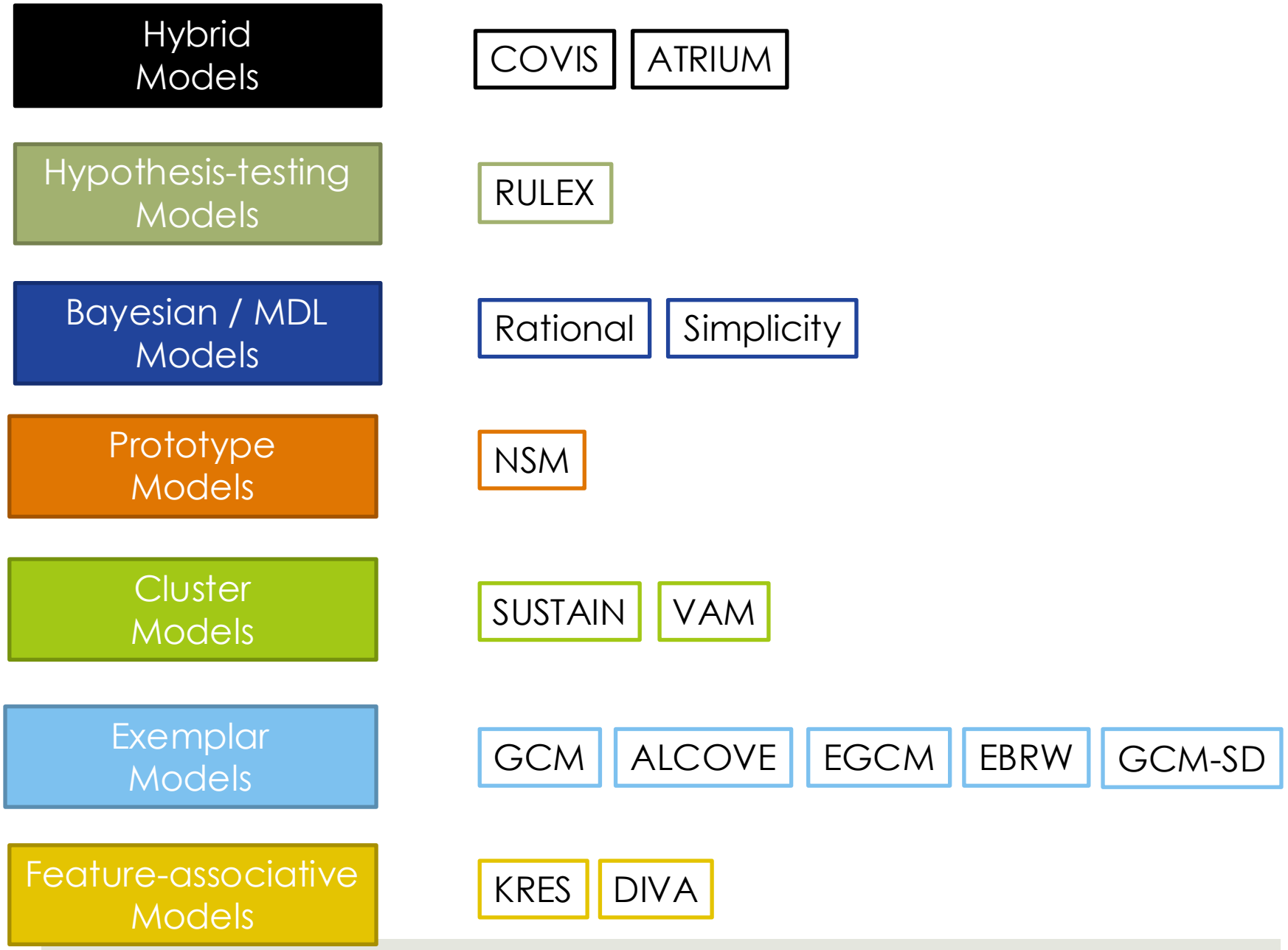
# Reasons not to model

- Compared to informal theories, they are:
  - More time consuming
  - Arguably more likely to lead to neglect of phenomena outside the model's scope.

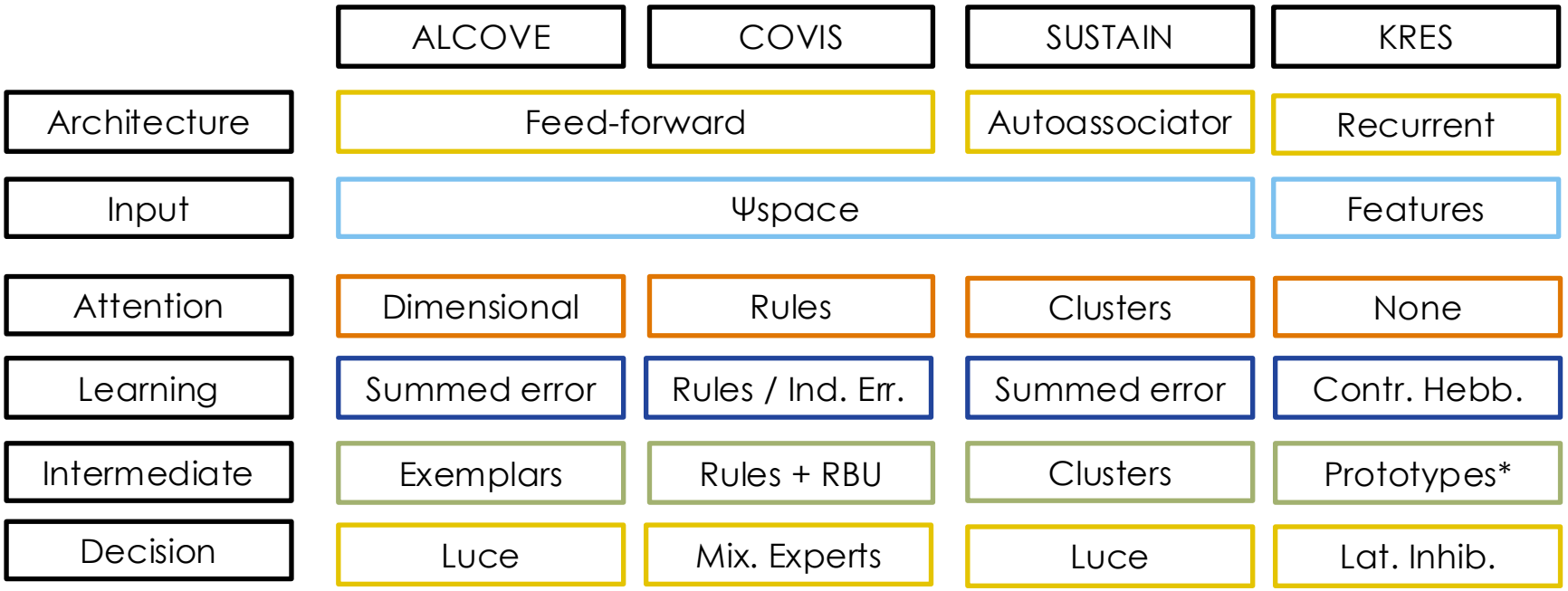
Murphy (2011)

# Reasons *to* model

- Recognition of problem complexity.
- Deeper insight through data reduction.
- Ambiguity reduction…
  - The ability of model X to encompass set of empirical results Y should be unambiguously determinable.
- …and therefore facilitation of theory comparison.
- Behavior prediction and reproduction.
- The state of the art in categorization models…


Hybrid Models
COVIS
ATRIUM
Hypothesis-testing Models
RULEX
Bayesian / MDL Models
Rational
Simplicity
Prototype Models
NSM
Cluster Models
SUSTAIN
VAM
Exemplar Models
GCM
ALCOVE
EGCM
EBRW
GCM-SD
Feature-associative Models
KRES
DIVA

# Formal models of categorization



# Summary

## Models of categorization

- Many models.
- The models are divergent.
- The models are domain-general and hence non-complementary.

## Advantages of formal models

- Recognition of problem complexity. ✓
- Deeper insight. ✗
- Ambiguity reduction ✗
- Facilitation of theory comparison.
- Behavior prediction and reproduction. ✗

# Comparison of formal models: Progress

- Formal model comparison in categorization has been happening for forty years.
  - Reed (1972) … Little et al. (2011)
- There has been evolution within model classes
  - Exemplar models: Nilsson (1965) ➢ Medin & Schaffer (1978) ➢ Nosofsky (1984) ➢ Kruschke (1992)
- …but this has not been accompanied by a reduction in the number of non-complementary model *classes*.

# Comparison of formal models: Progress

- It's a difficult problem, and progress is inevitably slow.
- Narrowness of model comparison
  - Smith & Minda (2000) – 2 models, 1 DV, 1 data set (30 variants thereof).
  - Numerous other comparisons limited to 1-2 experiments
    - Nosofsky, Kruschke & McKinley, 1992
    - Stanton, Nosofsky & Zaki, 2002
    - Little et al., 2011

# Choice point?

1. Assess **relative adequacy** of models from different classes across a broad range of phenomena

OR

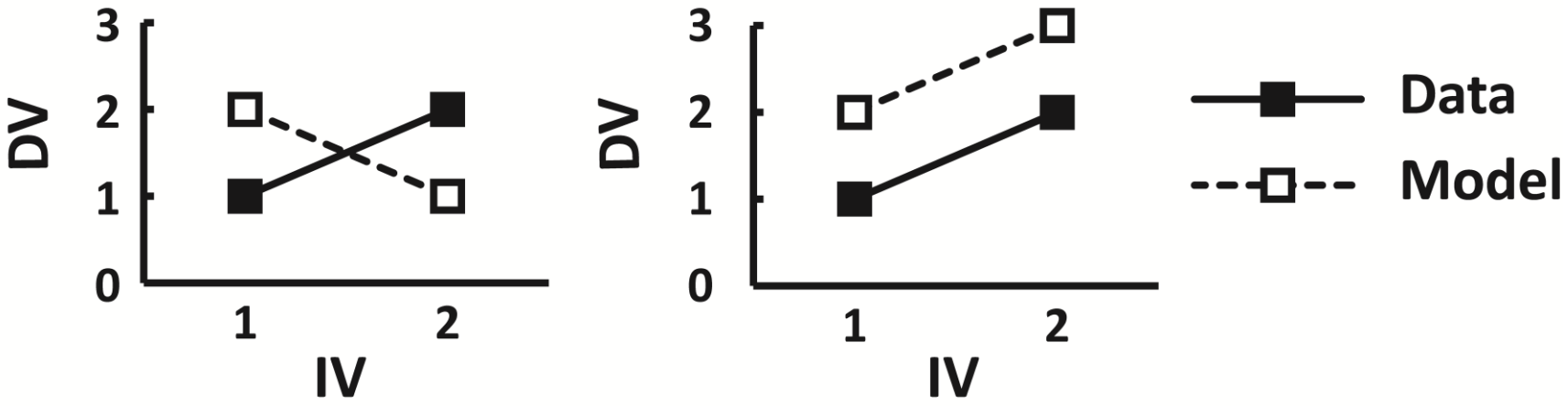
2. Negate most of the reasons for favoring formal models in the first place.

# Defining relative adequacy

The number and proportion of irreversible, ordinal, penetrable successes.

# number and proportion of irreversible, ordinal, penetrable successes

- Model adequacy often assessed by SSE
- SSE does not distinguish between quantitative and qualitative adequacy
- *Over-fitting:* The model with the lower SSE can sometimes be the less adequate model (Olsson, Wennerholm and Lyxzen, 2004)



# number and proportion of irreversible, ordinal, penetrable successes

- Irreversible success – a success which cannot be turned into a failure (or, more appositely, a failure into a success) without re-evaluating the full range of phenomena within the model's domain.
- Avoidance of arbitrarily variable parameters.
  - Parameters are to be expected.
  - These parameters may vary between experiments.
  - But, if so, this variation should be lawful / independently measurable.
  - Arbitrarily variable = set on a by-experiment (or by-condition!) basis solely on the criterion of maximizing fit of that data set.

# number and proportion of irreversible, ordinal, penetrable successes

- *Uncontroversial* - A model that captures more of what we know is better, other things being equal, than a model that captures less.
- But, current practice does not follow this.
  - Introduction of GCM – Fit against one experiment.
  - Introduction of SUSTAIN – Fit against seven experiments.
  - Number of categorization papers published in JEP:LMC in 2009… **11**
- There have been fits of both models to further data sets…
- …but model parameters were optimised on these later data sets independently of the fit to earlier data sets.
- This does not increase number of successes, at least as defined here.
- *Solution* – Irreversible success -> avoidance of arbitrarily variable parameters.

# number and proportion of irreversible, ordinal, penetrable successes

- A Utopian ideal? Some proposals for making this practical…
- Explicitly define a model's **domain** (explanatory scope).
  - List the DVs that the model is intended to address.
  - Define the model's domain with respect to major classes of IVs.
    - For example, "the model does not account for changes in stimulus presentation time, and is applicable to self-paced situations".
  - This is seldom done, currently.

# number and proportion of irreversible, ordinal, penetrable successes

- Prioritize DVs
  - Primary DVs for a model of categorization are those that directly index the classification decisions that are made.
  - For example: classification probability (primary) vs. eye-gaze duration (secondary)
- Sharing the load
  - Establishing key phenomena.
  - Comparing model adequacy against a broad range of key phenomena.

# number and proportion of irreversible, ordinal, penetrable successes

- Will we ever be able to use fixed parameters over multiple studies?
  - It's an end goal – we will not get there in one step.
  - There has already been some success – Love et al. (2004)
  - Variation between studies is fine, if its lawful / independently measurable.
    - For example, *r* in GCM is determined by whether stimuli are integral or separable.
    - ALCOVE sets dimension weights through minimization of error.

# number and proportion of irreversible, ordinal, penetrable successes

- Where two well-defined models are comparable in number of successes, we should favor the model with the **smaller** domain.
  - Greater ambiguity reduction.
  - Better behavior prediction.

# number and proportion of irreversible, ordinal, penetrable successes

- Other things being equal, more penetrable is better.
  - **Effort** - Simple, incorrect, models are frequently preferred if the incorrect model is "good enough"
  - **Natural language approximation**
    - Allows approximate understanding with lower effort.
    - Clarifies which properties are central.
    - Connects with informal theorizing.

# Conclusion

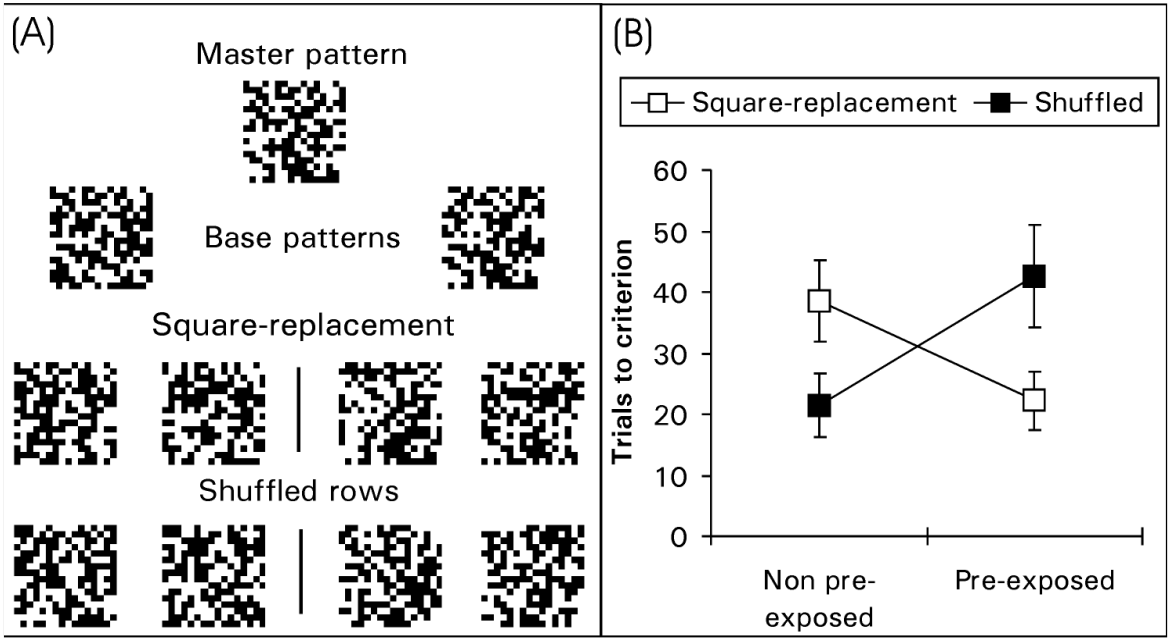
- Categorization research has seen a proliferation of models.
- These models cannot be considered to be complementary.
- They must therefore be considered to be competing accounts.
- In principle, formal specification permits assessment of relative adequacy.

# Conclusion

- In practice, comparisons are narrow with parameters optimized independently for each experiment.
- Progress through consensus on the criteria by which relative model adequacy should be assessed.
- We propose comparing **well-defined models** on the basis of the **number** and **proportion** of **irreversible**, **penetrable**, **ordinal successes** that can be attributed to them across a **broad range** of **key phenomena**.

# Some examples…

# Example: Wills et al. (2004)



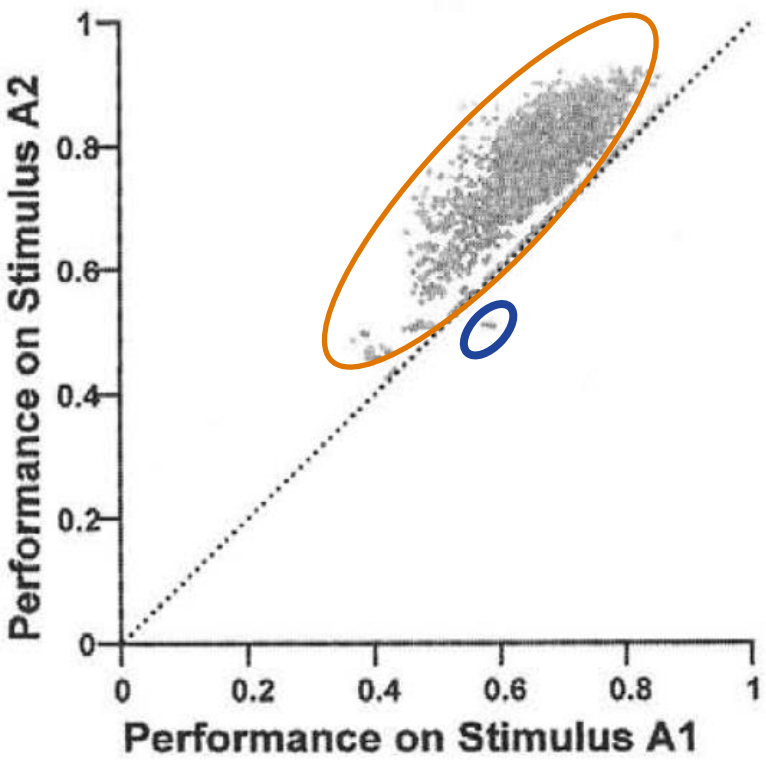

- An arbitrarily variable parameters approach: GCM, optimize *c* for each condition.
- An approach avoiding arbitrarily variable parameters: GCM, do similarity scaling from data collected after exposure.

# Example: Medin & Schaffer (1978)

| | C | F | S | N |
|---|---|---|---|---|
| A1 | 1 | 1 | 1 | 0 |
| A2 | 1 | 0 | 1 | 0 |
| A3 | 1 | 0 | 1 | 1 |
| A4 | 1 | 1 | 0 | 1 |
| A5 | 0 | 1 | 1 | 1 |

| | C | F | S | N |
|---|---|---|---|---|
| B1 | 1 | 1 | 0 | 0 |
| B2 | 0 | 1 | 1 | 0 |
| B3 | 0 | 0 | 0 | 1 |
| B4 | 0 | 0 | 0 | 0 |

- Performance on A2 better than on A1, despite A1 being more typical of category A. Supports exemplar models.
- GCM can accommodate this result across a broad range of parameters,
- but also its opposite..
- The GCM parameters that lead to an A1 > A2 prediction also lead to other, falsifiable predictions (Nosofsky, personal communication.
- What GCM predicts about M&S78 cannot be resolved by considering the result in isolation. It could be resolved through the evaluation of GCM against a broad range of phenomena




Smith & Minda (2000), Figure 8