UNIVERSITY OF EXETER

Deliberation and Categorization

Andy J. Wills

Collaborators

Tim Hodgson

Exeter → Lincoln

Aureliu Lavric

Exeter

Chris Longmore

Exeter → Plymouth

Ian McLaren

Exeter

Fraser Milton

Exeter

Terms

- Categorization
  - Dividing the world into groups of things
  - Building block of cognition
  - Contingency learning is a simple form of category learning.
- Deliberation
  - Deliberative cognition is effortful, voluntary, demanding of mental resources, typically extended in time.
  - Non-deliberative cognition is largely effortless, automatic, requiring of minimal mental resources, typically fast.

Plan

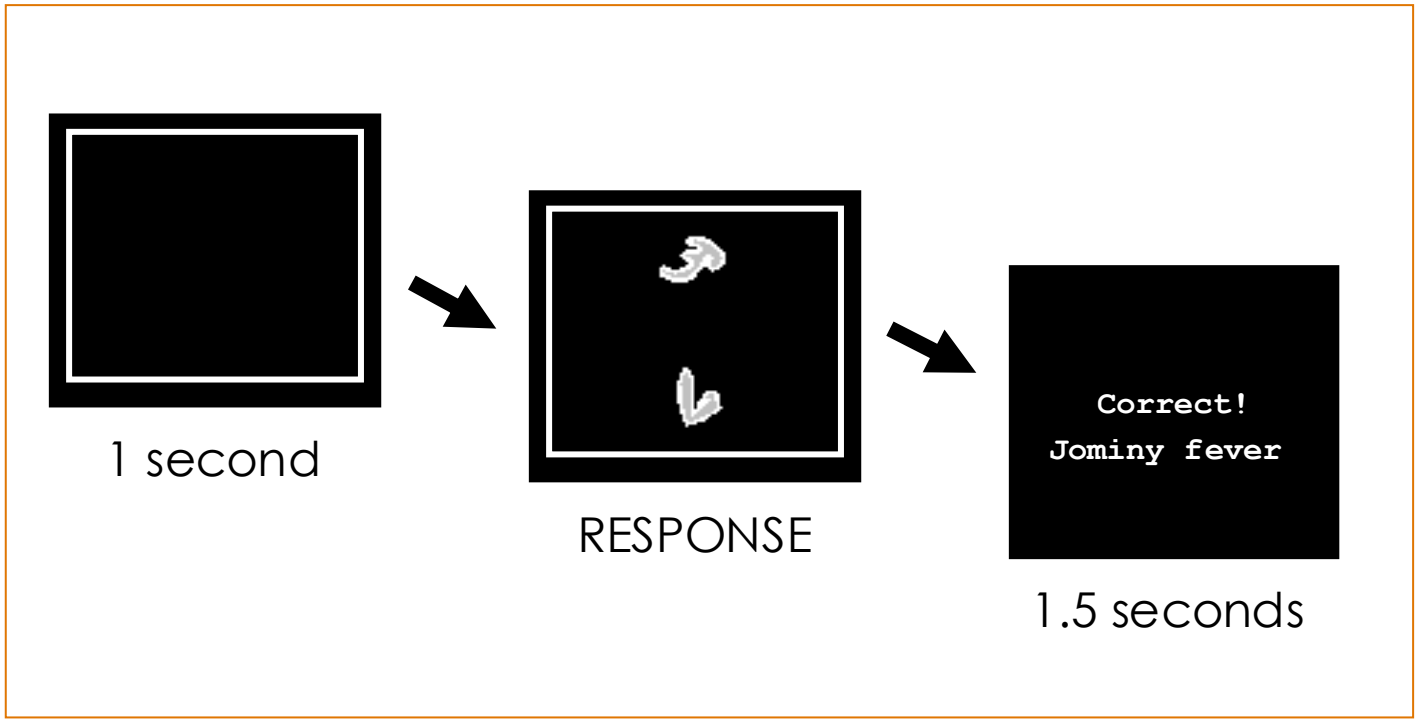
1. **Fast attentional processes in contingency learning**
   a) **EEG study of Blocking**
   b) **EEG study of the Inverse Base Rate Effect**
2. Deliberative processes in contingency/category learning
   a) Critique of an existing procedure
   b) Evidence from the Shanks-Darby procedure
3. Family resemblance categories as a "fallback" mode of cognition.
   a) Critique of three existing procedures
   b) Evidence from the Regehr-Brooks procedure.

Cue competition

Phase 1	Phase 2	Test
A+	AX+	X?
B-	BY+	Y?
I-	IJ-	

- Deduction
- Error-driven learning
- Error-driven attention

Wills, Lavric, Croft & Hodgson (2007)



- 2 second time-out (0.3% trials terminated)

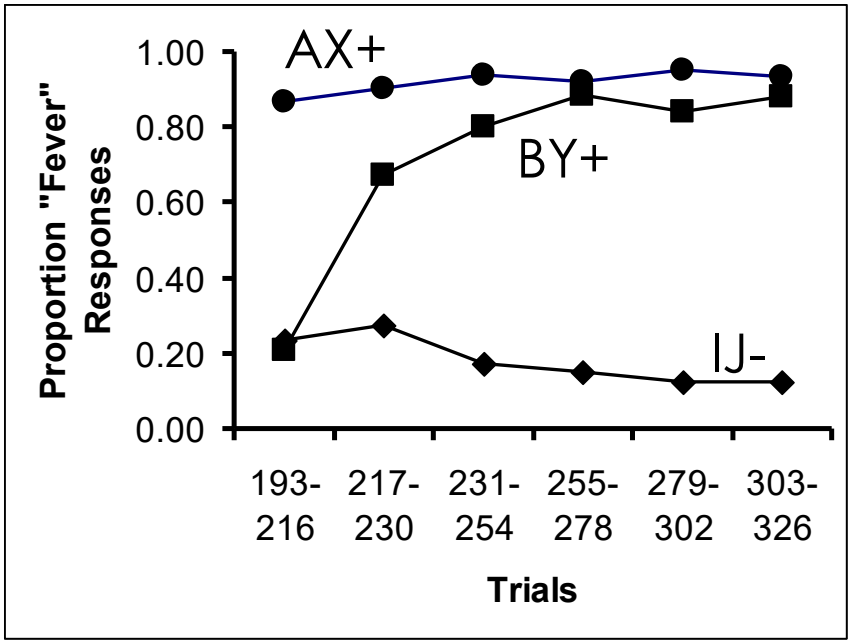
Behavioural results

Phase 1

A+ 0.90

B- 0.03

I- 0.03

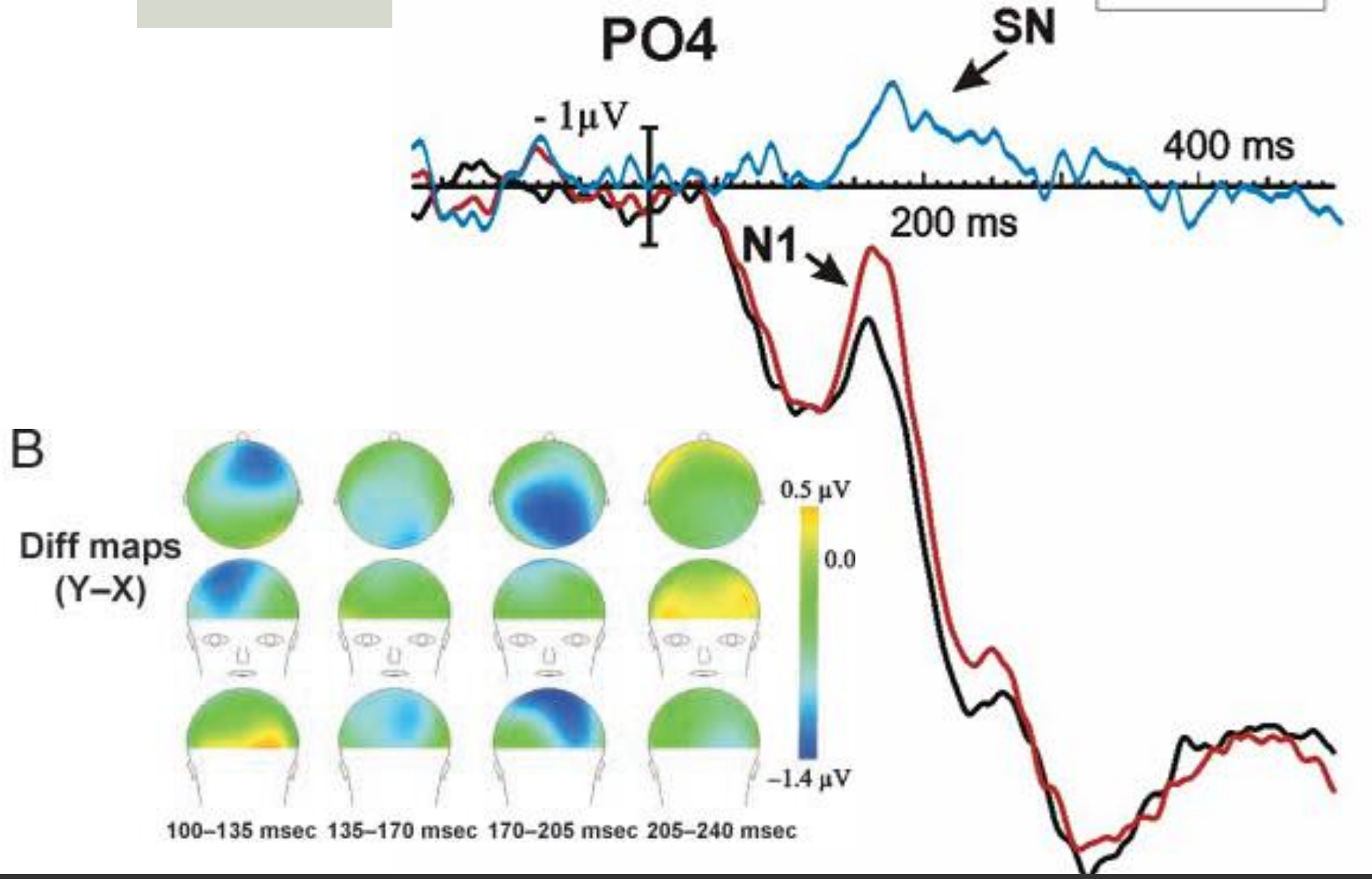


Phase 3

X – 0.45 (807ms)

Y – 0.72 (767ms)

- Other phase 3 trial types: A:0.96; AX: 0.98; BY: 0.91; B: 0.18



Attentional difference as cause or consequence?

- Wills & Lavric (in preparation)

Inverse base-rate effect

Dizziness and Skin Rash → Jominy fever (common)

Dizziness and Back Pain → Phipp's syndrome (rare)

Skin Rash and Back Pain

Is the patient more likely to have:

Jominy fever or Phipp's syndrome ?

Inverse base-rate effect

Dizziness and Skin Rash → Jominy fever (common)

Dizziness and Back Pain → Phipp's syndrome (rare)

Skin Rash and Back Pain

Is the patient more likely to have:

Jominy fever or Phipp's syndrome

- In the experimental context, skin rash perfectly predicts Jominy fever, and back pain perfectly predicts Phipp's syndrome. Jominy fever is more common, so the rational answer is "Jominy".
- Across a number of experiments (Medin & Edelson, 1998; Kruschke, 1996; Juslin et al., 2001; Kruschke, 2001) the rare disease (Phipps) is chosen.

Why?

Eliminative inference explanation

2 x AB → 1 | 1 x AC → 2 | BC?

(a) Eliminative inference: "When faced with a novel situation, produce a novel response".

(b) As AC → 2 is rarer than AB → 1, participants are more likely to forget C →2 than B → 1.

(c) If C →2 is forgotten, then the familiar response for BC is 1 (from B →1). Hence under eliminative inference, they respond "2".

Juslin et al. (2001).

Prediction of Eliminative inference

The inverse base-rate effect should not be dependent on the presence of a common symptom.

2 x DB → 1 1 x EC → 2 BC?

(a) Eliminative inference: "When faced with a novel situation, produce a novel response".

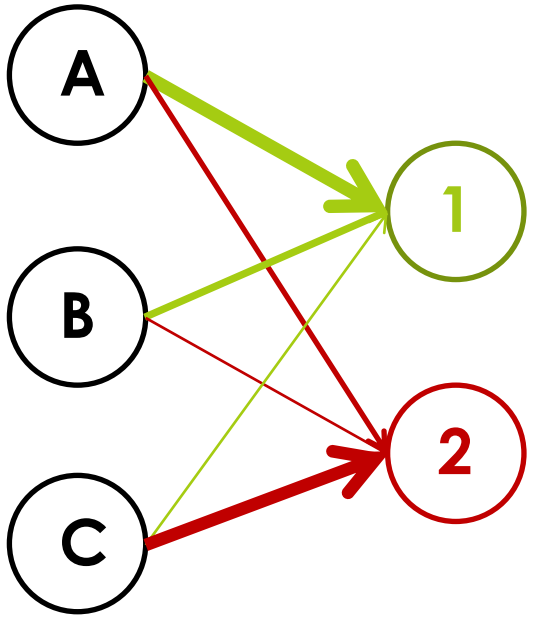
(b) As EC → 2 is rarer than DB → 1, participants are more likely to forget C →2 than B → 1.

(c) If C →2 is forgotten, then the familiar response for BC is 1 (from B →1). Hence under eliminative inference, they respond "2

Kruschke (2001)

Error-correcting learning

- Predicted by a simple error-correcting learning algorithm (delta rule / Rescorla-Wagner / temporal difference models).
  - A has greater associative strength to 1 than to 2.
  - Cues compete to predict outcomes.
  - So C →2 gains more associative strength than B →1.
- Prediction of this account:
  - C → 2 is greater than B → 1



2 x AB → 1 1 x AC → 2 BC?

Error-correcting attention

2 x AB → 1 | 1 x AC → 2 | BC?

- AB → 1 is learned first (because it is more common).
- On seeing AC, participant tends to predict 1 because they have learned A → 1.
- In order to reduce future error, attention to C (the perfect predictor of 2) is increased.

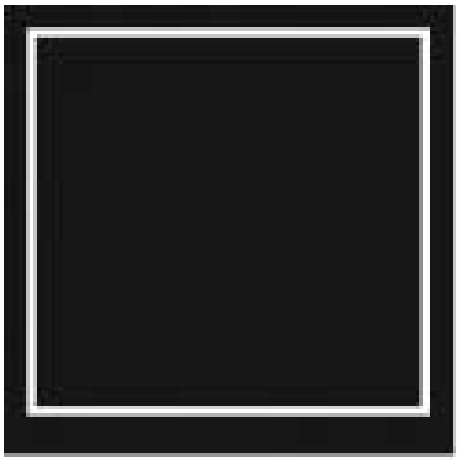
Kruschke (1996)

Design

Phase 1	*Phase 2*
2 x AB → Disease 1	As phase 1, plus:
1 x AC → Disease 2	2 x [B?, C?, D?, E?]
2 x FD → Disease 1	1 x [A?, BC?, DE?]
1 x GE → Disease 2	

Procedure

Phase 1	*Phase 2*
2 x AB → 1	As phase 1, plus:
1 x AC → 2	2 x [B?, C?, D?, E?]
2 x FD → 1	1 x [A?, BC?, DE?]
1 x GE → 2	



1 sec

RESPONSE

1.5 sec

- Cell bodies in blood samples.
- Each letter in the design instantiated in three "cell bodies".
- Cell bodies randomly allocated to letters for each participant.
- 20 blocks in phase 1 (18 trials per block)
- 8 blocks in phase 2 (51 trials per block)

EEG

- 58 scalp electrodes. 500 Hz sample rate.
- Low-pass filtered (40Hz)
- Segmented by stimulus onset (-100ms to +500ms)
- Assess B, C, D, and E during phase 2.
- Attentional selection by stimulus features is commonly associated with a posterior selection negativity and also sometimes an anterior selection positivity (Hillyard & Anllo-Vento, 1998).
- Selection negativity previously seen in a forward cue competition design (Wills, Lavric, Croft & Hodgson, 2007)

Phase 1	*Phase 2*
2 x AB → 1	As phase 1, plus:
1 x AC → 2	2 x [B?, C?, D?, E?]
2 x FD → 1	1 x [A?, BC?, DE?]
1 x GE → 2	

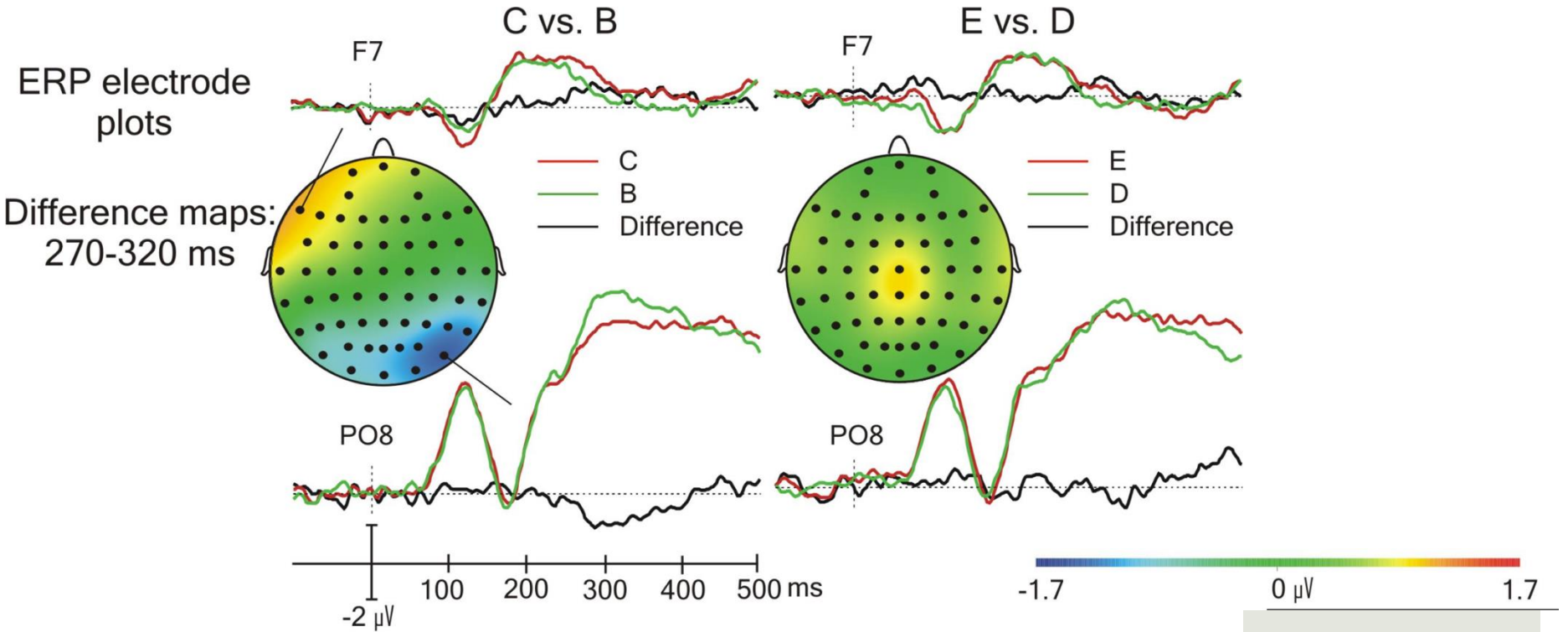
Results

Phase 1	*Phase 2*
2 x AB → 1	As phase 1, plus:
1 x AC → 2	2 x [B?, C?, D?, E?]
2 x FD → 1	1 x [A?, BC?, DE?]
1 x GE → 2	

	A→1	B→1	C→2	D→1	E→2	BC→1	DE→1
Prob	0.69	0.88	0.67	0.87	0.56	0.36	0.95
RT	835	732	763	711	755	917	785

- Inverse base-rate effect (BC1 < 0.5).
- Eliminative inference account ruled out (DE1 > 0.5).
- Associative strength account ruled out (B1 > C2).

• TANOVA and permutation-based correction used to identify significant time windows in the scalp distribution.
• 270-320ms revealed in C – B comparison. None revealed in E – D.
• (C-B) vs. (E-D) comparison is significant in this time window.
• Posterior selection negativity to C (compared to B)
• Anterior selection positivity to C (compared to B)
• Absence of effect in E – D comparison rules out explanation in terms of differential frequency.



Part 1: Summary

- In the inverse base-rate effect phenomenon, participants make the apparently irrational decision that when two perfect predictors of two different outcomes co-occur, the *less* frequent outcome is more likely.
- At least in the current study, this is not due to a process of "eliminative inference" (no effect in absence of common element – see also Kruschke, 2001).
- Nor is it due to error-correcting learning (as B > C behaviourally).
- The remaining class of explanation – error-correcting attention – is supported by ERPs showing a selection negativity to C.
  - Commonly seen correlate of selective attention to features (Hillyard & Anllo-Vento, 1998)
  - Selection negativity previously reported in an EEG study of forward cue competition (Wills, Lavric, Croft & Hodgson, 2007).
  - However, in Wills et al. 2007, it was not possible to rule out an associative strength mediated effect.
  - In the current study, it was.

Plan

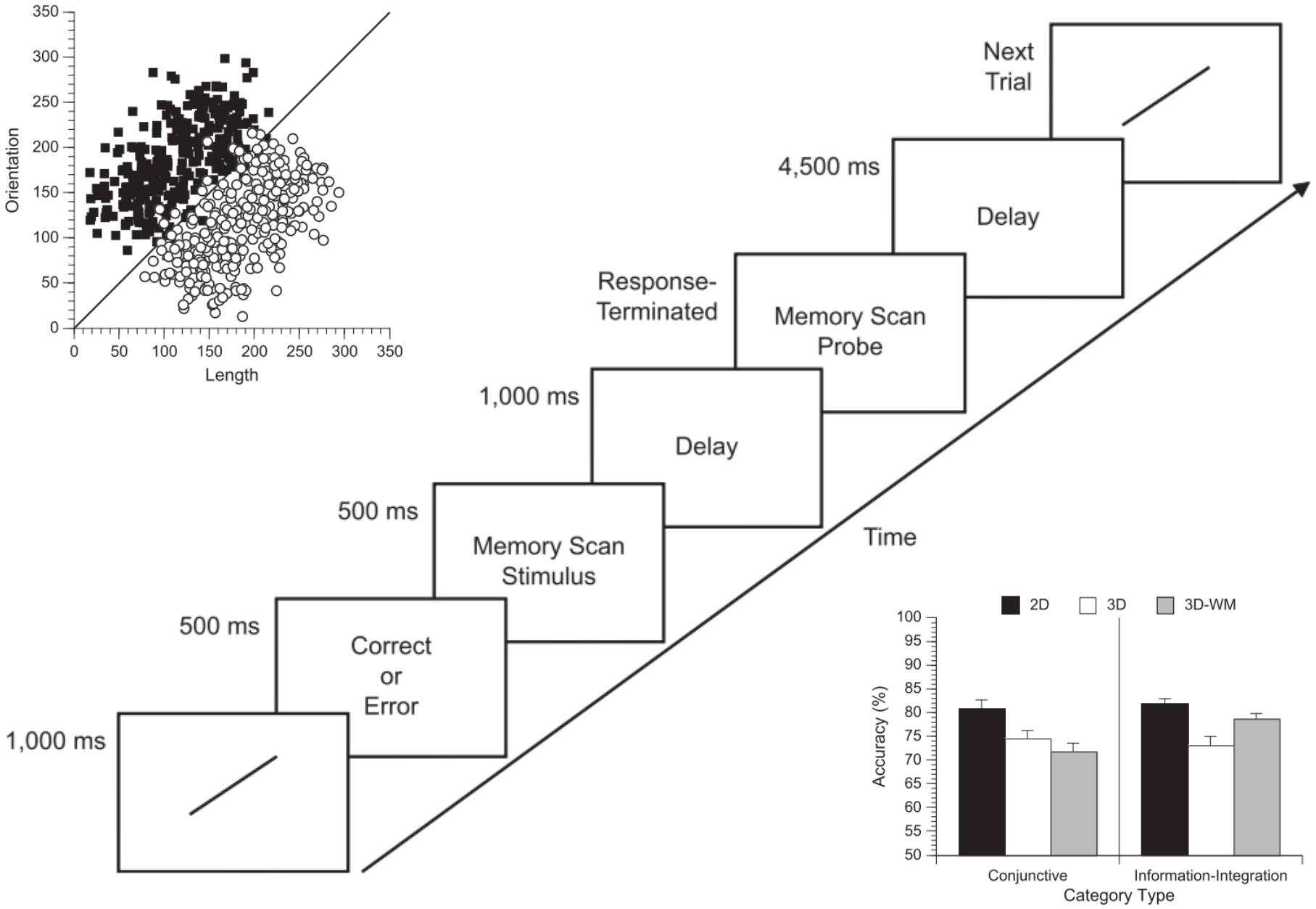
1. Fast attentional processes in contingency learning
   a) EEG study of Blocking
   b) EEG study of the Inverse Base Rate Effect

2. **Deliberative processes in contingency/category learning**
   a) Critique of an existing procedure
   b) Evidence from the Shanks-Darby procedure

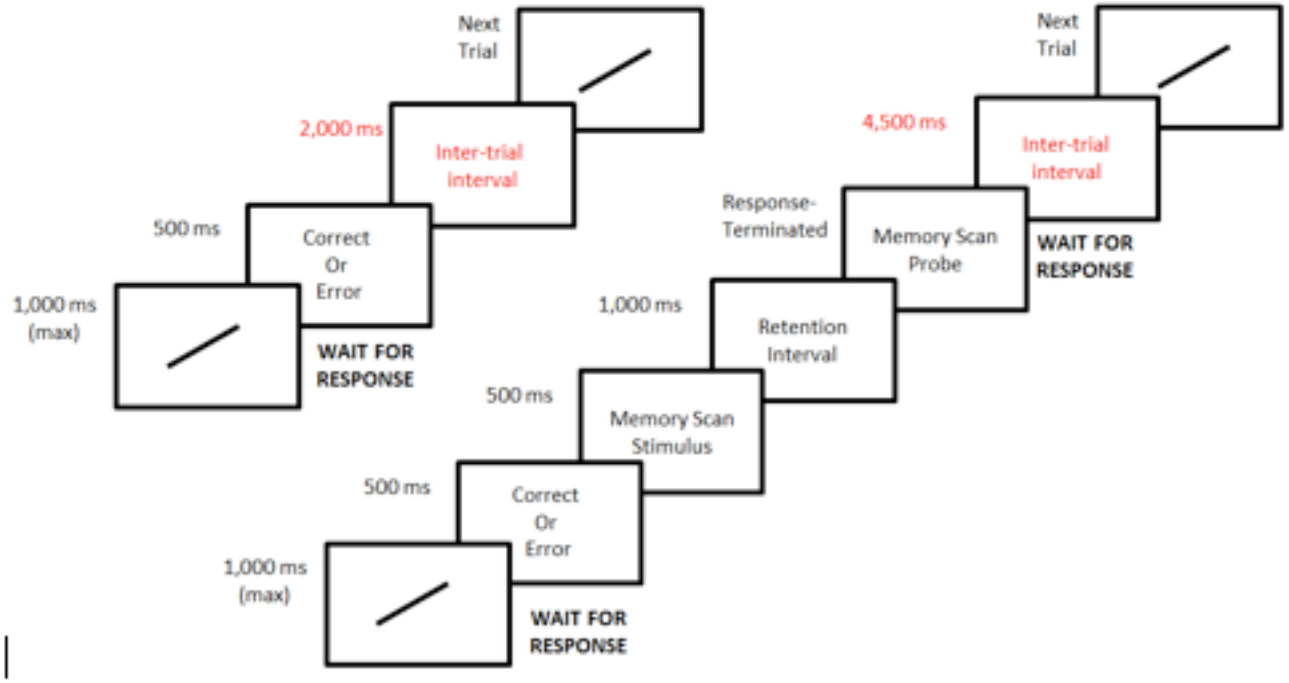
3. Family resemblance categories as a "fallback" mode of cognition.
   a) Critique of three existing procedures
   b) Evidence from the Regehr-Brooks procedure.

“Removing the frontal lobes”

- Filoteo, J. V., Lauritzen, S., & Maddox, W. T. (2010). Removing the frontal lobes: the effects of engaging executive functions on perceptual category learning. *Psychological science, 21,* 415-23.

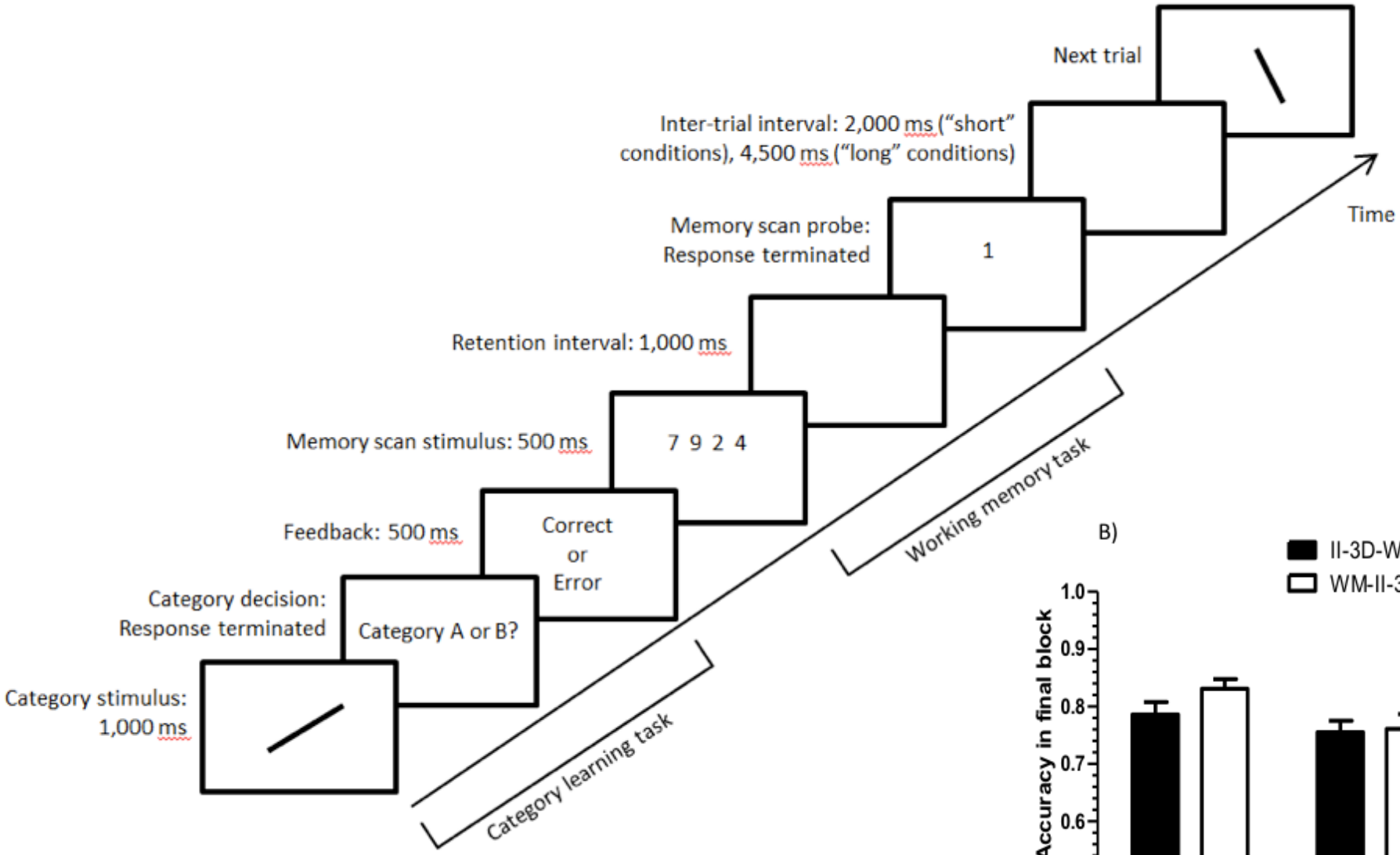

Orientation
Length
0
50
100
150
200
250
300
350
1,000 ms
Correct or Error
500 ms
Memory Scan Stimulus
500 ms
Delay
1,000 ms
Response-Terminated
Memory Scan Probe
4,500 ms
Delay
Next Trial
Time
2D
3D
3D-WM
Accuracy (%)
Conjunctive
Information-Integration
Category Type

A confound



- Newell, Moore, Wills & Milton (submitted)

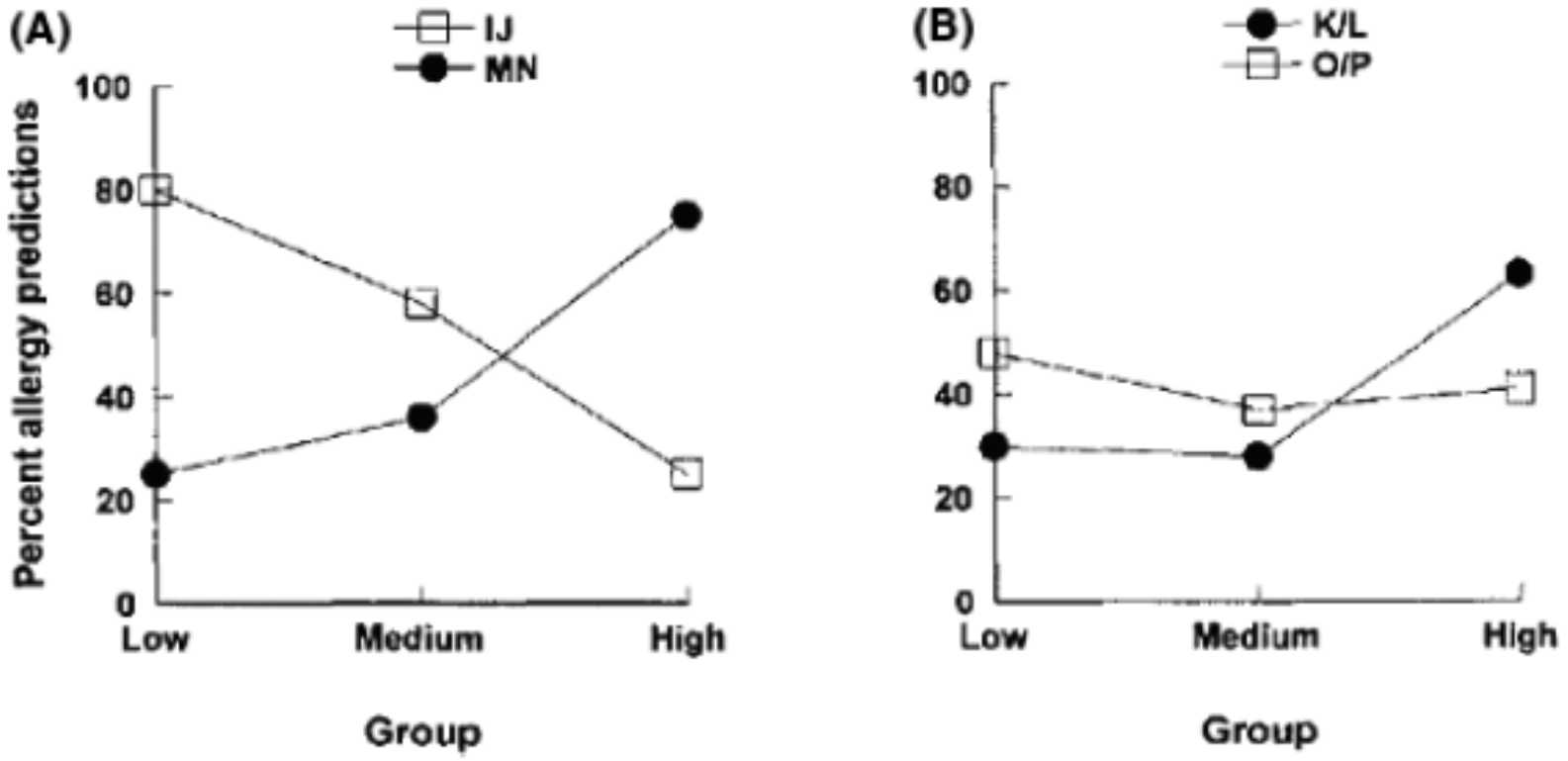
A confound corrected



Shanks & Darby (1998)

Training			**Test**		
A+	B+	AB-	A?	B?	AB?
C-	D-	CD+	C?	D?	CD?
E+	F+	EF-	E?	F?	EF?
G-	H-	GH+	G?	H?	GH?
I+	**J+**		I?	J?	**IJ?**
		KL-	**K?**	**L?**	KL?
M-	**N-**		M?	N?	**MN?**
		OP+	**O?**	**P?**	OP?

Shanks & Darby (1998)



- Percentage allergy responses at test

Concurrent load task

6..2..9..1..4..3

Mr. X consumes
a meal containing
oranges
C = Reaction M = no reaction

Correct.
Mr. X had a reaction

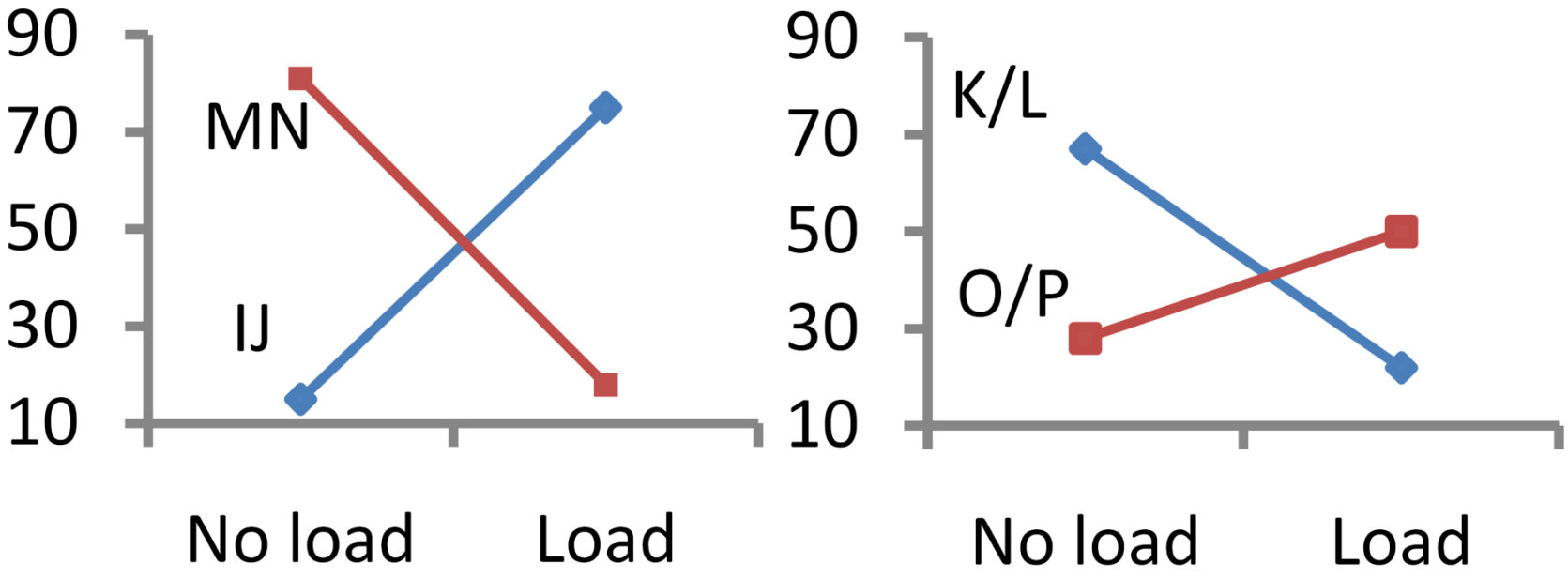
9

The correct response to “9” on this trial is “1”.

Some details

- 32 participants.
- Trained to criterion (86%).
- 7 participants failed to meet criterion within 16 blocks and were excluded.
- Concurrent load participants took longer (14.2 vs. 9.2 blocks).

Results



- Percentage allergy responses at test

Part 2: Summary

- Concurrent load leads to generalization on the basis of similarity.
- Full attention leads to generalization on the basis of an opposites rule.
- Some possible explanations:
  - Competition between qualitatively different learning systems.
  - Reduction in level of deliberative processing.

Plan

1. Fast attentional processes in contingency learning
   a) EEG study of Blocking
   b) EEG study of the Inverse Base Rate Effect
2. Deliberative processes in contingency/category learning
   a) Critique of an existing procedure
   b) Evidence from the Shanks-Darby procedure
3. **Family resemblance categories as a "fallback" mode of cognition.**
   a) Critique of three existing procedures
   b) Evidence from the Regehr-Brooks procedure.

Brooks (1978)

- *Analytic categorization*
  - Participant separates aspects of the stimulus and evaluates their ability to predict category membership.
  - Typically results in a subset of those aspects being used.
- *Nonanalytic categorization*
  - Participant evaluates category membership on the basis of overall similarity to known examples.
  - So, all stimulus attributes have some control over responding.
  - More likely to occur where cognitive resources are limited.
- A striking prediction:
  - A categorization process that employs all the information in the stimulus (overall similarity) is less effortful than a categorization process that employs a subset of that information (analytic, or "rule-based", categorization).

Less is more

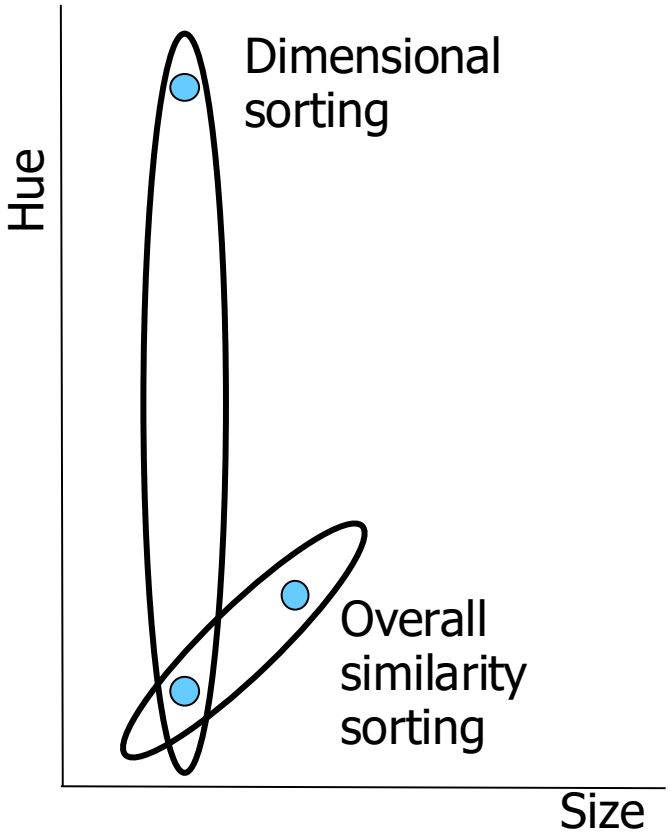
"…it is easier for people to base similarity and categorization judgments on more, rather than fewer, properties"

Goldstone & Barsalou (1998)

"…family resemblance categorization emerges as a primitive, fallback strategy when more sophisticated cognitive strategies are lacking for any reason"

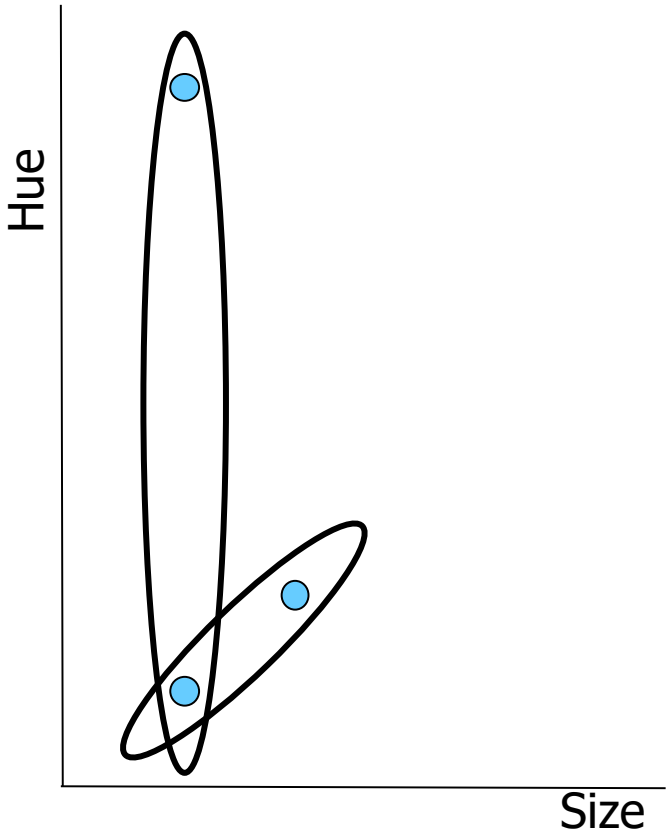
Couchman et al. (2010)

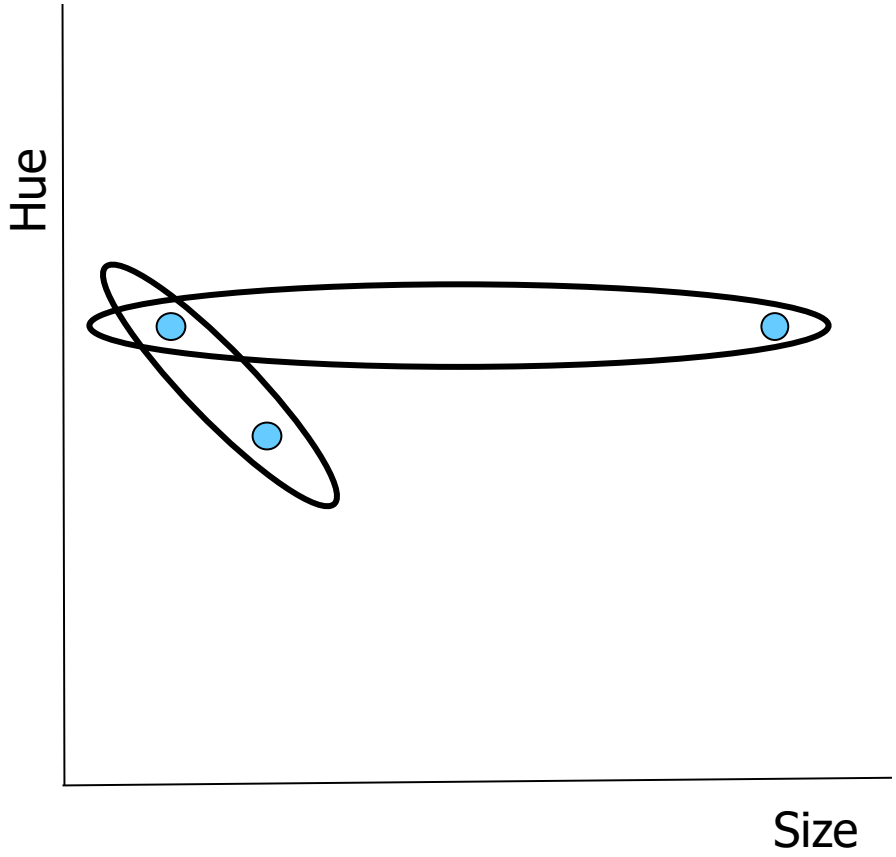
Triad procedure



- Ward (1983)
- Smith and Kemler Nelson (1984)
- Overall similarity classification is enhanced by:
  - Time pressure
  - Cognitive load
  - Impulsivity

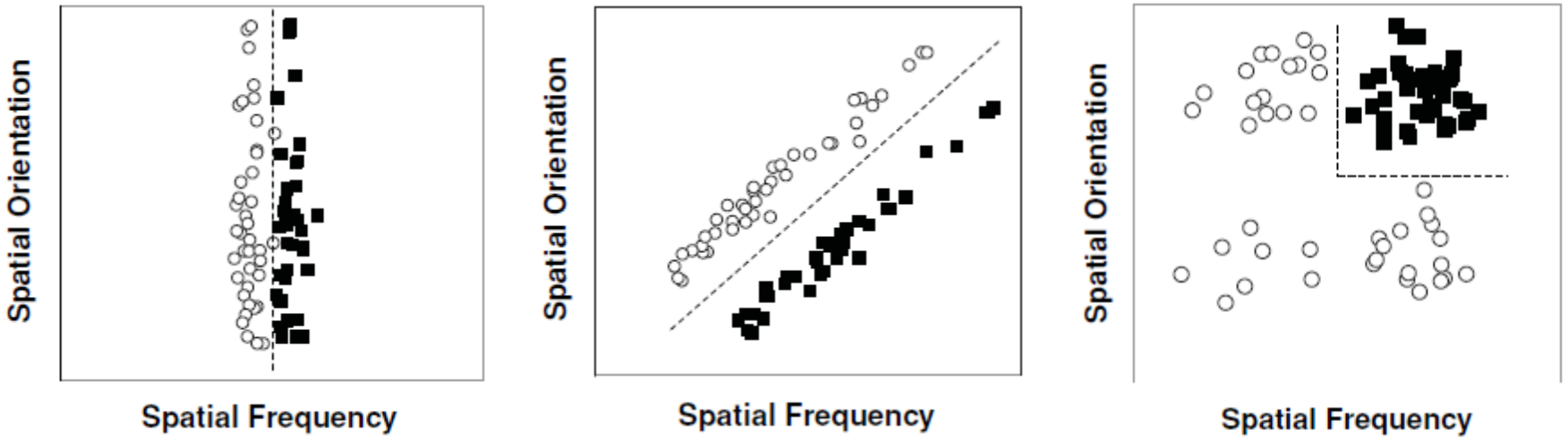
Triad procedure – issues of interpretation





Thompson (1994); Raijmakers et al.,(2004)

Ashby / Maddox Procedure

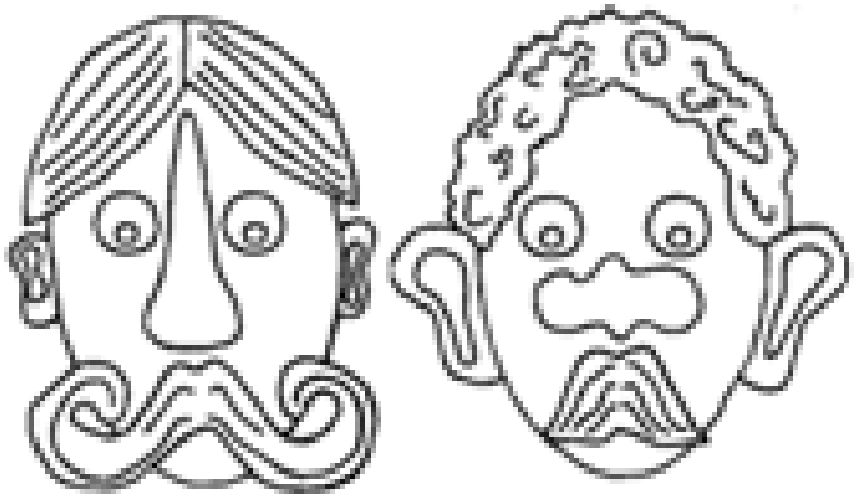


- Zeithamova and Maddox (2006) – Concurrent load affects A more than B.
  - Although see Newell, Dunn & Kalish (2010)
  - Also, more about verbalizability?

Criterial Attribute Procedure

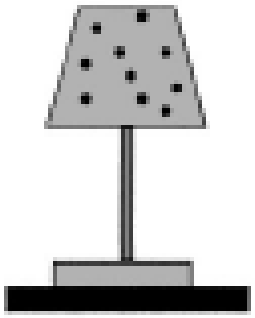
Table 1.

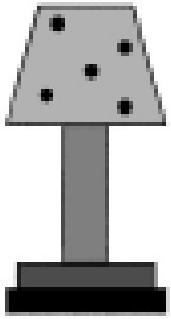
Category A				Category B			
D1	D2	D3	D4	D1	D2	D3	D4
0	0	0	0	1	1	1	1
0	1	0	0	1	0	1	1
0	0	1	0	1	1	0	1
0	0	0	1	1	1	1	0
Test				0 1 1 1			
				1 0 0 0			

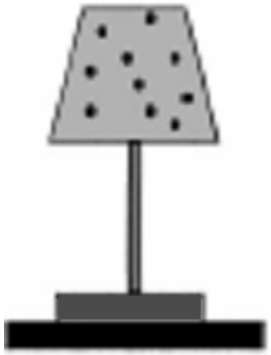


- Incidental training increases OS classification at test, compared to intentional training (Kemler Nelson, 1984).
- Concurrent load during training increases OS classification at test, compared to full attention training (Smith & Shapiro, 1989).
- Test phase conducted under intentional, full-attention conditions in all cases.
- Incidental / divided attention training interferes with participants discovering the criterial attribute during training?

Match-to-standards procedure



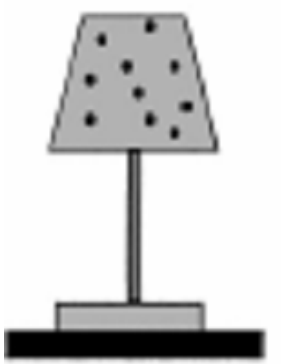




- Regehr & Brooks (1995); Milton & Wills (2004).
- Classification (sorting) without feedback.
- Classification of a set of 10 items (the 2 standards, plus the 8 one-away stimuli).
- 90+% of participants produce one of two kinds of sort:

OS and UD sorting

1 1 1 1



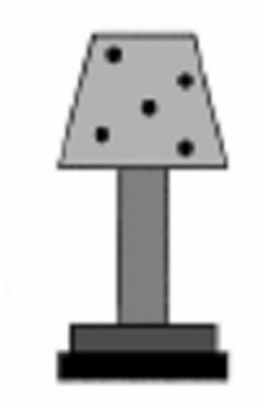
Overall similarity

1 1 1 1
0 1 1 1
1 0 1 1
1 1 0 1
1 1 1 0

Unidimensional

1 1 1 1
1 0 0 0
1 0 1 1
1 1 0 1
1 1 1 0

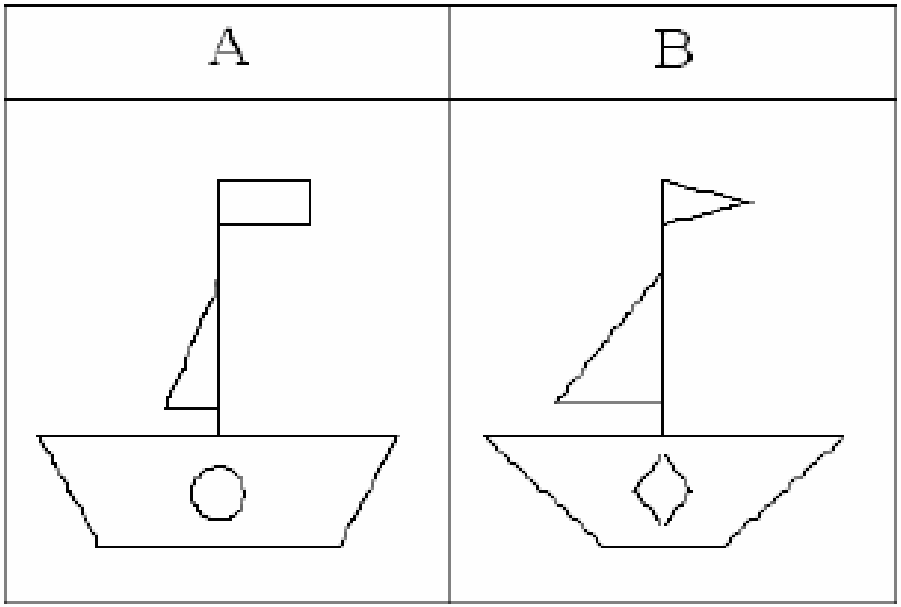
0 0 0 0

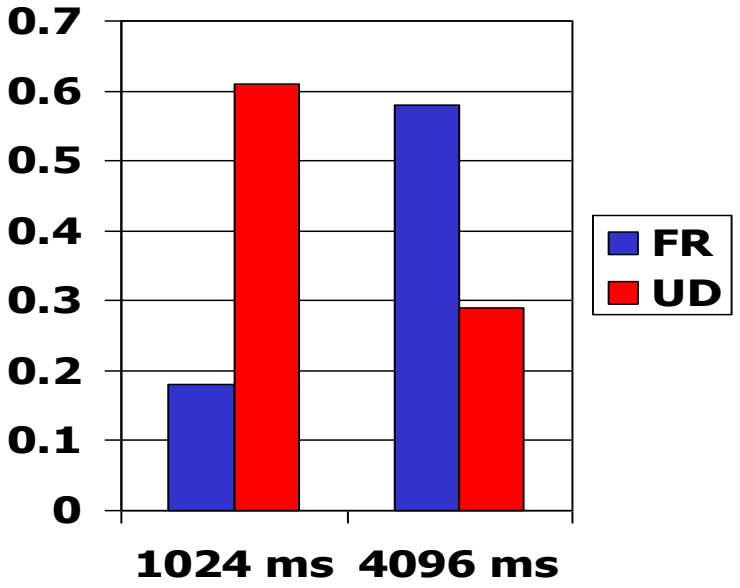


0 0 0 0
1 0 0 0
0 1 0 0
0 0 1 0
0 0 0 1

0 0 0 0
0 1 1 1
0 1 0 0
0 0 1 0
0 0 0 1

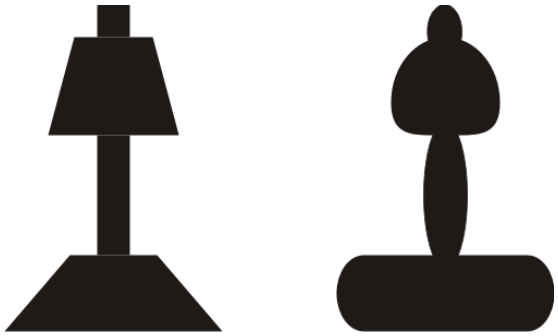
Presentation time



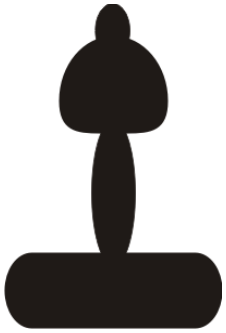


- 12 blocks, DV = probability of sort type.
- Milton, Longmore & Wills (2008)

Concurrent load



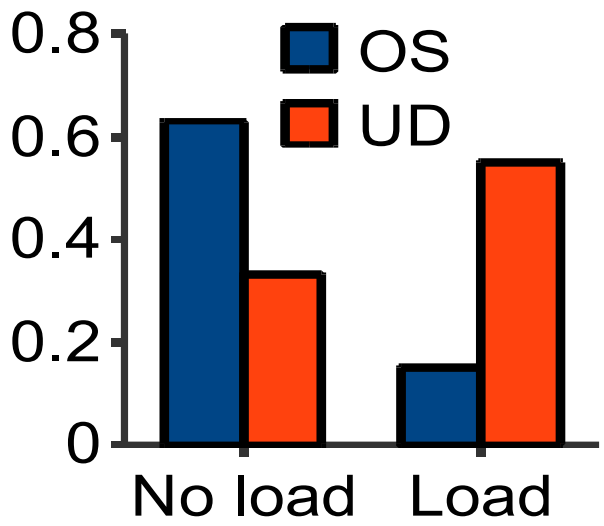
"...11 ...48 ... 9 ... 87 ... 45 ... 78 ... 23 ... 91 .. 43 ... 82 ..."



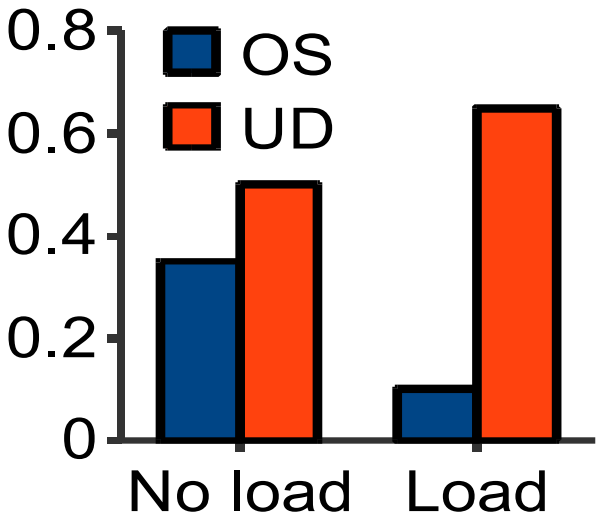
A or B?

- Wills, Milton, Longmore, Hester and Robinson (submitted)

Concurrent load



1500ms

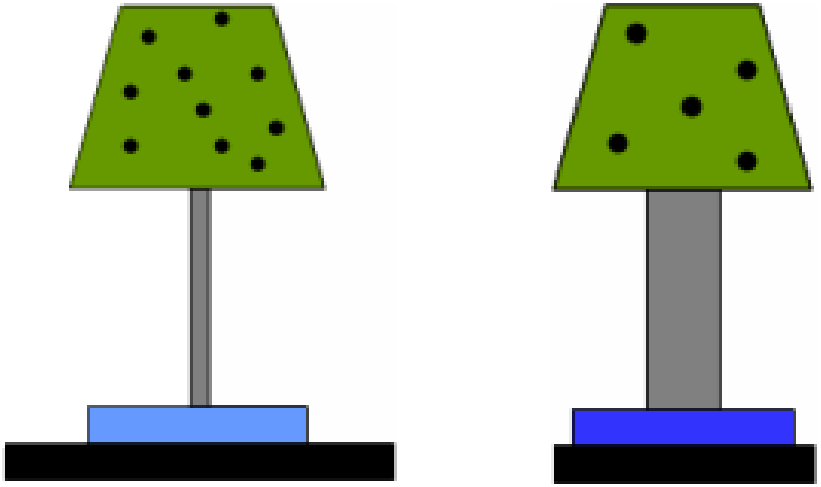


300ms

- Also works with a simultaneous procedure.

Working memory capacity

- Single block of match-to-standards classification.
- Operation span (OSPAN)
- OS sorters' mean span: 3.7
- UD sorters' mean span: 2.3

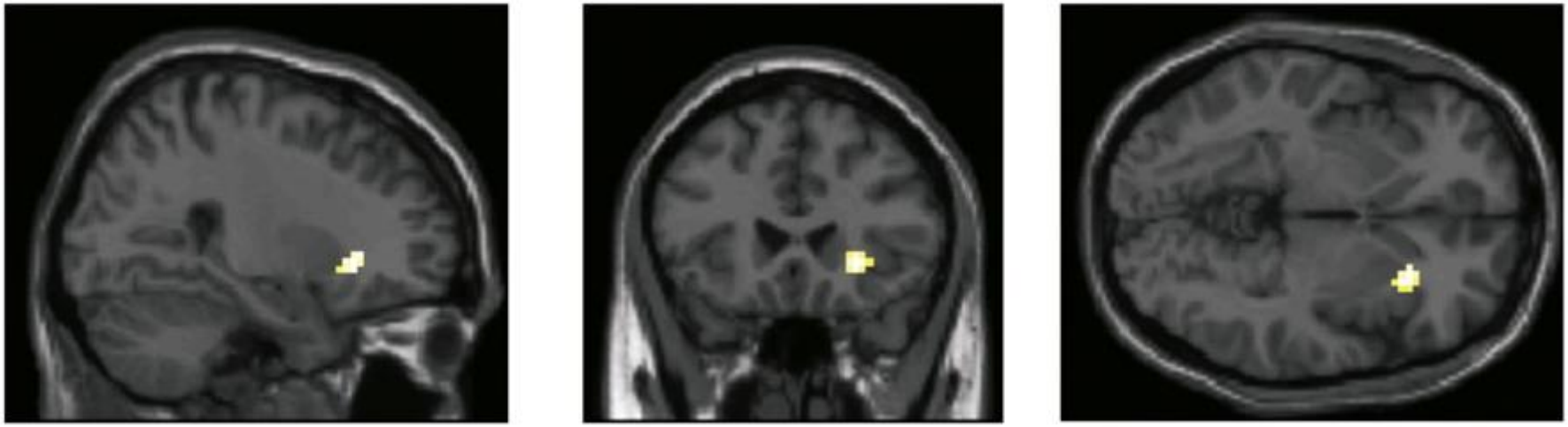


Brain imaging

"Generally, it seemed that the authors did not meet the burden of defending the idea that FR responding is an effortful, explicit, analytic thing… They are arguing counter to the excellent literature on the cognitive neuroscience of category learning. There, it seems clear that prefrontal areas do unidimensional, rule-based things. It seems clear that the basal ganglia or the tail of the caudate do multidimensional, information-integration things".

- Anonymous JEP:LMC reviewer

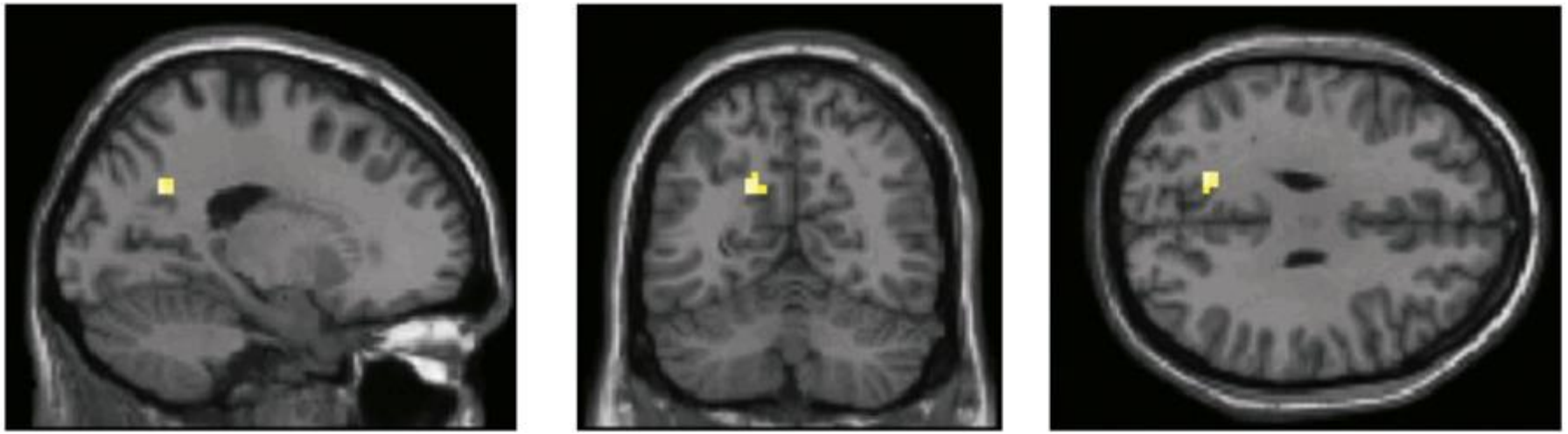
OS - UD



Includes: ventrolateral frontal cortex

Milton, Wills & Hodgson (2009)

UD - OS



Includes: superior parietal lobe

Milton, Wills & Hodgson (2009)

Comparative Cognition

- Wills, A.J., et al. (2009). A Comparative Analysis of the Categorization of Multidimensional Stimuli. *Journal of Comparative Psychology*, 123, 4, 391-405.

Part 3: Summary

- Overall similarity classification involves the consideration of multiple (experimenter-defined) stimulus dimensions; single-dimension classification just one.
- Time pressure reduces the prevalence of overall similarity classification (Milton, Longmore & Wills, 2008).
- Concurrent load reduces the prevalence of overall similarity classification, as does limited working memory capacity.
- Overall similarity classification leads to higher frontal brain activity than single-dimension classification.
- More is more.

Conclusions: Three aspects of deliberation in categorization

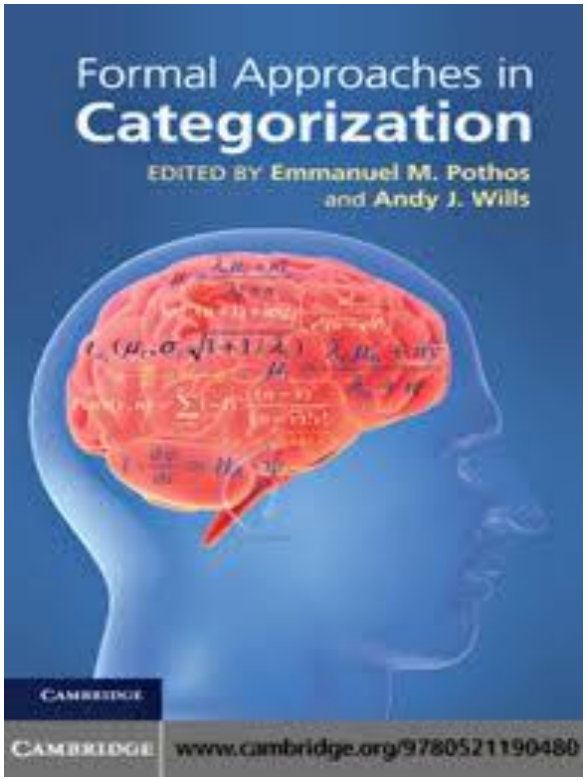
1. **Whilst not discounting the role of deliberative processes in contingency learning, it also involves fast, non-deliberative, attentional processes that can lead to irrational behaviour** (EEG studies).

2. **Although much of the existing evidence is flawed, new studies by our lab suggest the presence of ordinally different generalization behaviour depending on the availability of cognitive resources** (Shanks-Darby procedure).

3. **Opposite to the commonly-held belief, overall similarity classification is more effortful than single-dimension classification** (MTS procedure).

Another talk:
Formal models of categorization

- Wills, A.J. and Pothos, E.M. (2012). On the adequacy of current empirical evaluations of formal models of categorization. *Psychological Bulletin, 138,* 102-125.



Perceptual learning

- Welham, A.K. and Wills, A.J. (2011). Unitization, similarity and overt attention in categorization and exposure. *Memory and Cognition, 39,* 1518-1533

Artificial Grammar Learning

- Beesley, T., Wills, A.J. & Le Pelley, M.E. (2010). Syntactic transfer in artificial grammar learning. *Psychonomic Bulletin & Review, 17,* 122-128.