Mechanisms of rational and irrational generalization in adult humans

Andy J. Wills
University of Exeter
willslab.co.uk

Research colleagues / publications

Primary sources:

Wills, A.J. & Lavric, A. (in preparation). Attention and the inverse base-rate effect: Evidence from event-related potentials.

Wills, A.J., Barrasin, T. J. and McLaren, I. P. L. (accepted). Working Memory Capacity and Generalization in Predictive Learning. In L. Carlson, C. Hölscher, & T. Shipley (Eds.), *Proceedings of the 33rd Annual Conference of the Cognitive Science Society* (pp. ...). Austin, TX: Cognitive Science Society.

Wills, A.J., Graham, S., Koh, Z., McLaren, I.P.L. and Rolland, M.D. (in press). Effects of Concurrent Load on Feature- and Rule-based Generalization in Human Contingency Learning. *Journal of Experimental Psychology: Animal Behavior Processes.*

Graham, S., Jie. H.L., Minn, C.H., McLaren, I.P.L. and Wills, A.J. (2011). Simultaneous backward conditioned inhibition and mediated conditioning. *Journal of Experimental Psychology: Animal Behavior Processes.* Advance online publication. doi: 10.1037/a0021828

Wills, A.J., Lavric, A, Croft, G. and Hodgson, T.L (2007). Predictive learning, prediction errors and attention: Evidence from event-related potentials and eye-tracking. *Journal of Cognitive Neuroscience. 19*, 843-854.

Review articles:

Kovacs, T. and Wills, A.J. (accepted). Generalization vs. discrimination learning. *Encyclopaedia of the Sciences of Learning.*

Wills, A.J. (2009). Prediction Errors and Attention in the Presence and Absence of Feedback. *Current Directions in Psychological Science*, *18*, 95-100.

Generalization

- From a known set of category members, how does one decide how to respond to novel items?
- Key question in cognitive psychology / cognitive neuroscience and (I hope!) in other studies of Intelligent Systems.
- Overview of some work on rational and irrational generalization, through the lens of:
  - Stimulus competition and fast attention
  - Absent but expected stimuli
  - Rule- vs. similarity-based generalization.
- Interested to hear what (if any!) useful connections this might make to what you do / are interested in.

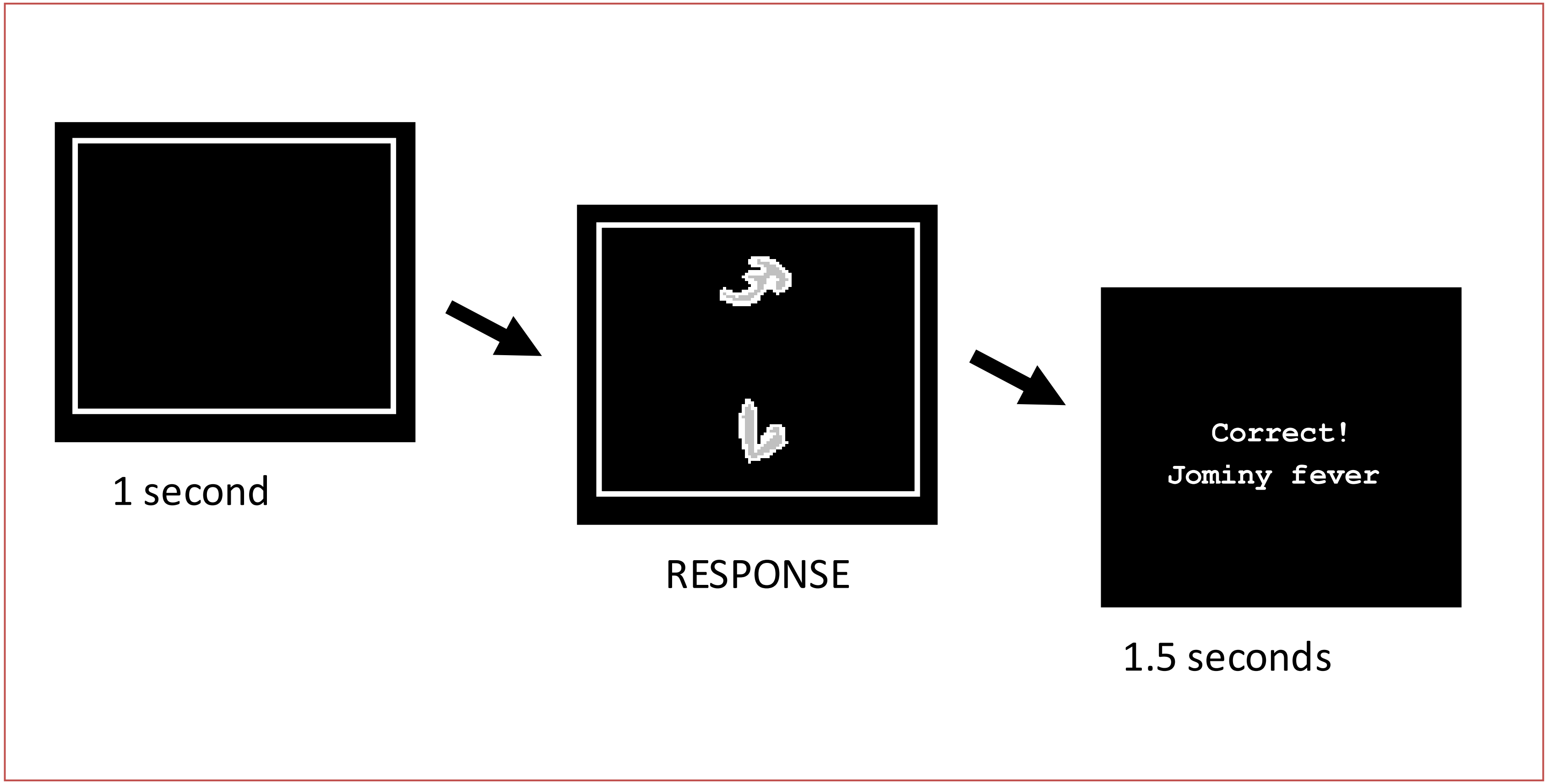
1. Stimulus competition and fast attention

Cue competition

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

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

Procedure



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

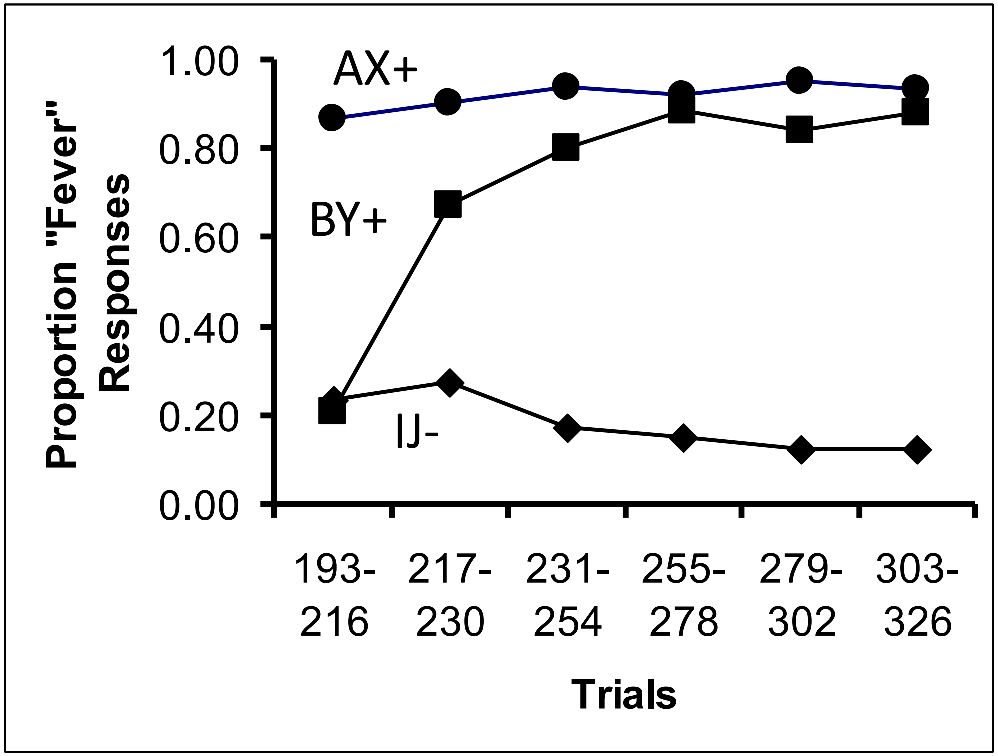
Behavioural results

Phase 1

A+ 0.90

B- 0.03

I- 0.03



Phase 3

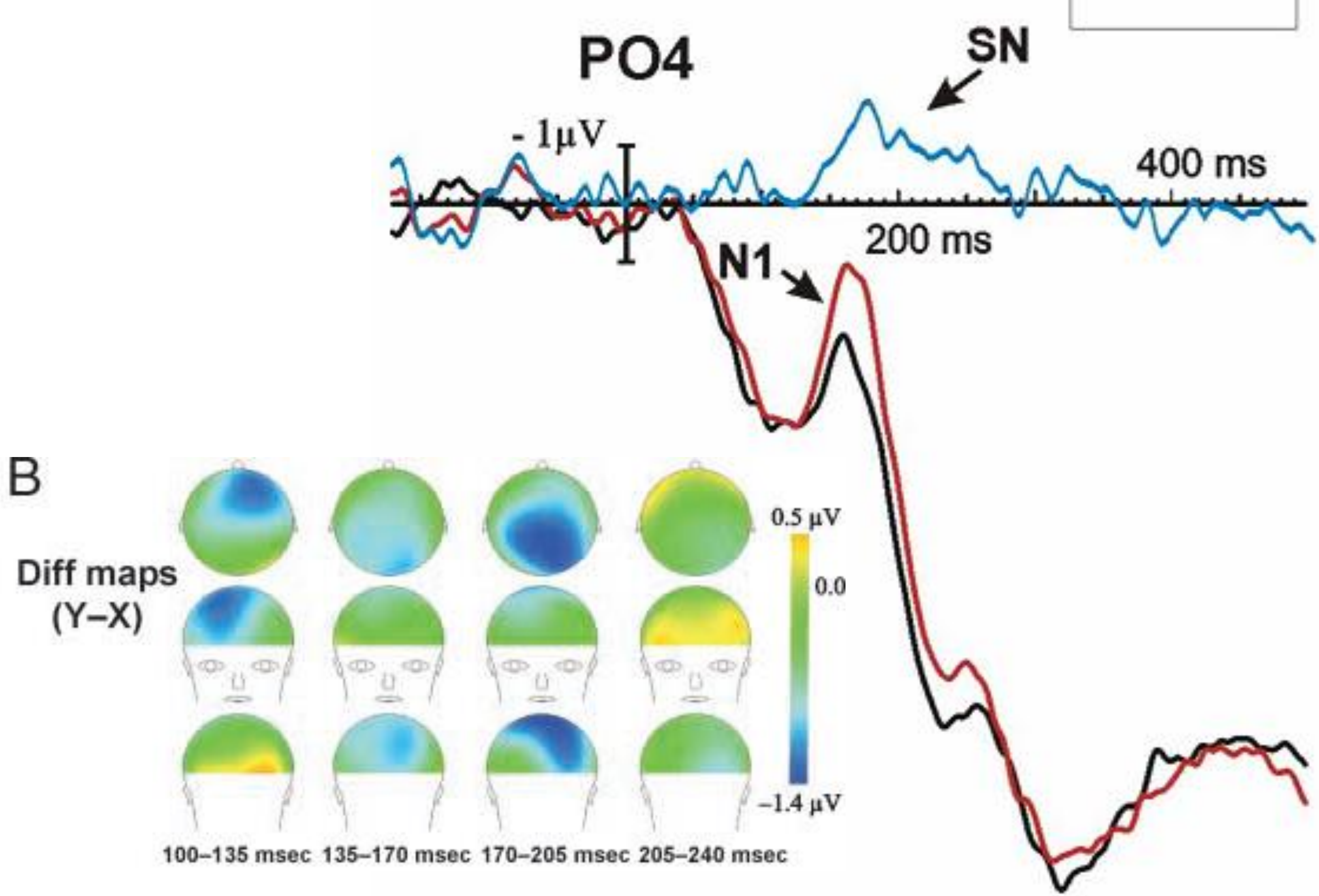
X – 0.45

(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?

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.

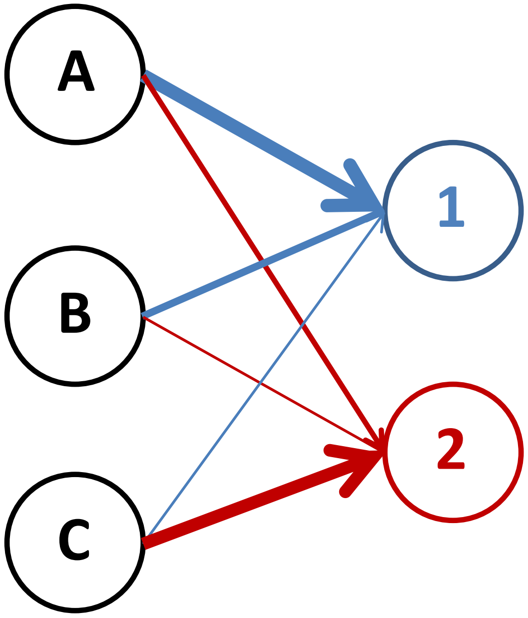
2 x DB → 1   1 x EC → 2   BC?

(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

2 x AB → Disease 1

1 x AC → Disease 2

2 x FD → Disease 1

1 x GE → Disease 2

Phase 2

As phase 1, plus:

2 x [B?, C?, D?, E?]

1 x [A?, BC?, DE?]

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	

- 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)

Results

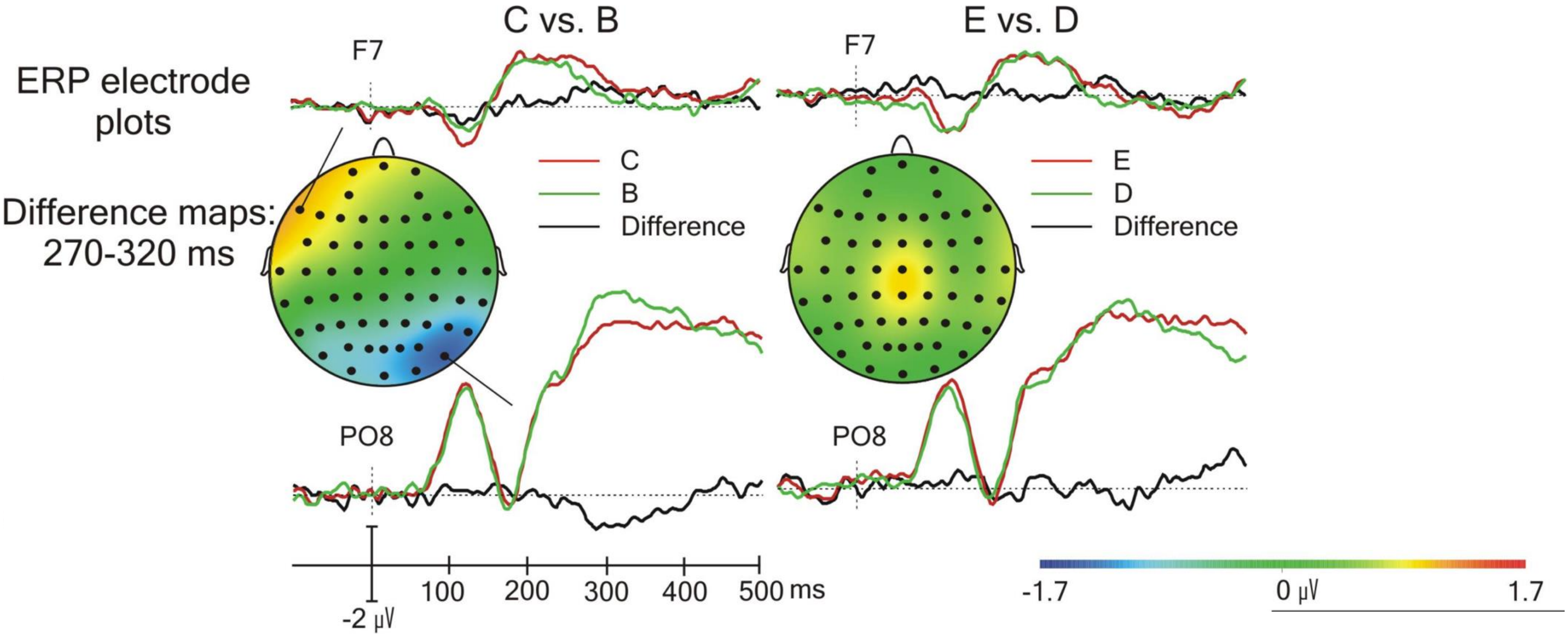
Phase 1	*Phase 2*
2 x AB → 1	As phase 1, plus:
1 x AC → 2	2 x [B?, C?, D?, E?]
2 x FD → 1	1 x [A?, BC?, DE?]
1 x GE → 2	

	A→1	B→1	C→2	D→1	E→2	BC→1	DE→1
Mean	0.69	0.88	0.67	0.87	0.56	0.36	0.95
SE	(0.04)	(0.03)	(0.06)	(0.04)	(0.06)	(0.06)	(0.02)
RT	835	732	763	711	755	917	785
SE	(54.9)	(37.3)	(37.8)	(35.0)	(35.4)	(37.3)	(41.0)

- 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.

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	



2. Absent but expected stimuli

Retrospective Revaluation

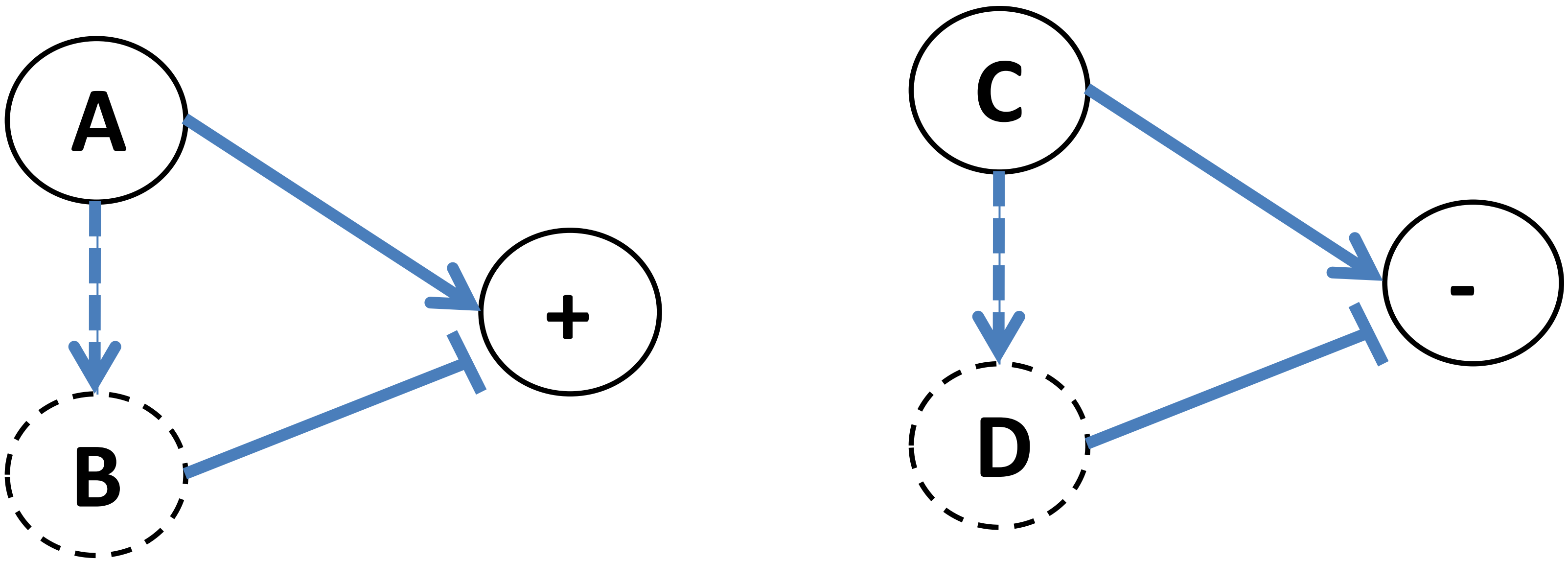
Phase 1	Phase 2	Test
A+	AB+	C?
C-	CD+	D?

Reverse the order…

Phase 1	Phase 2	Test
AB+	A+	C?
CD+	C-	D?

- Deduction
- Error-driven learning

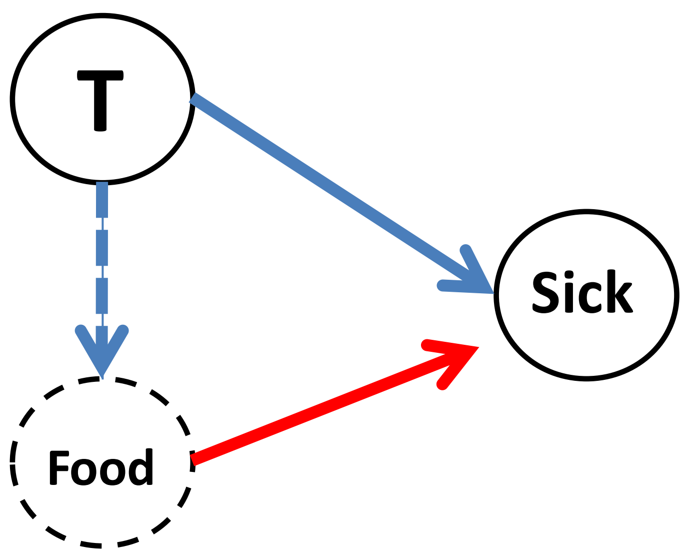
Modified SOP



Mediated conditioning (Holland, 1981)

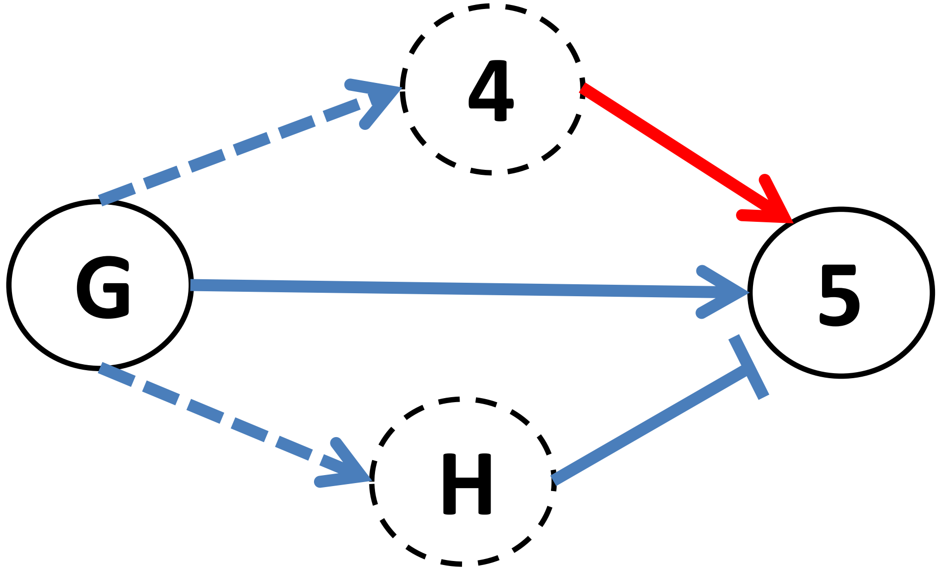
Tone -> Food Tone -> LiCl Food -> ?

Opposite to RR. Perhaps absent-but-expected outcome->outcome acts differently to absent-but-expected cue -> outcome?

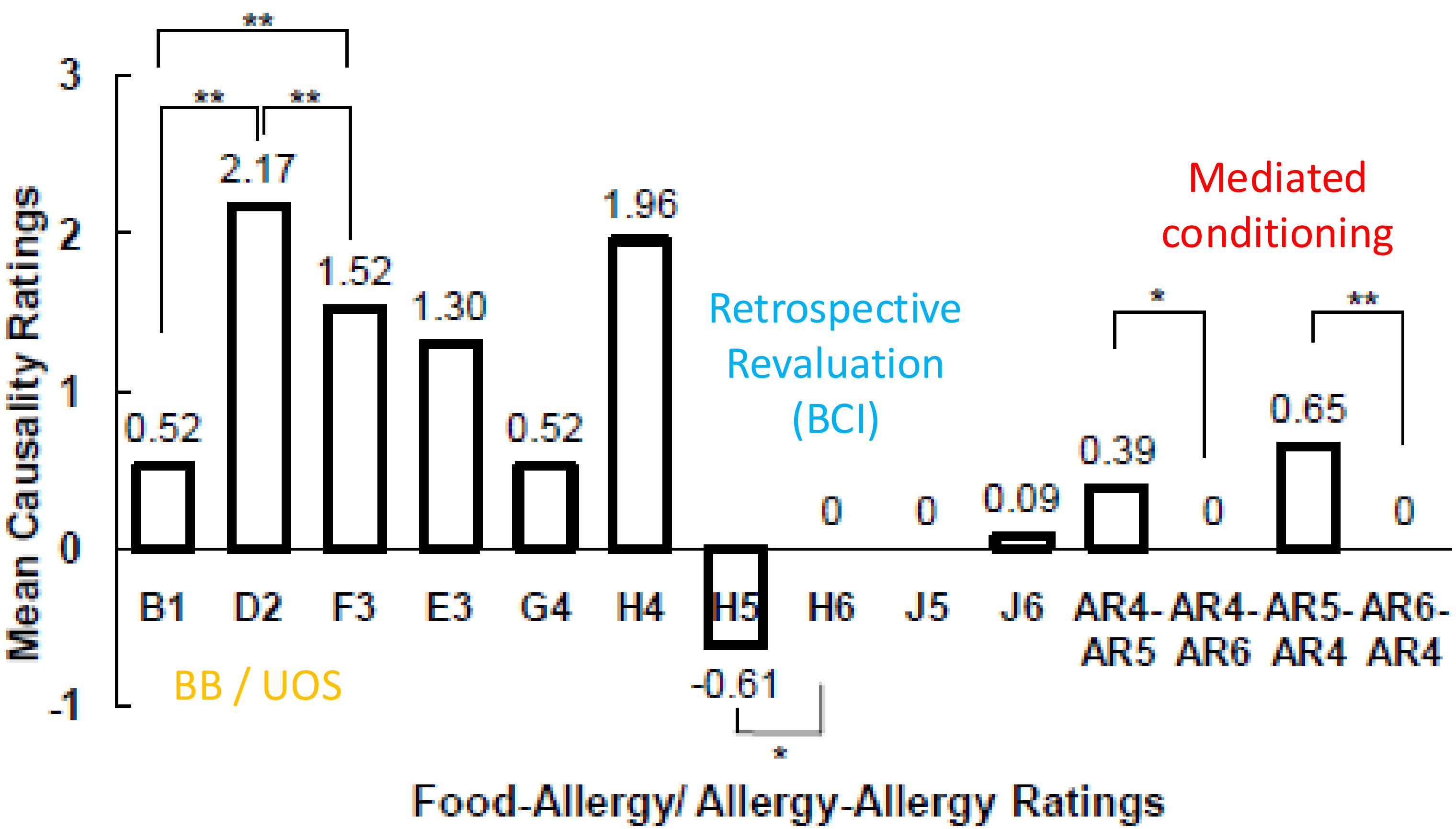


Simultaneous RR and MC

Condition	*Phase 1*	*Phase 2*
Backward Blocking	AB+1	A+1
Unovershadowing	CD+2	C-
RR Control	EF+3	E°
RR & MC	GH+4	G+5
MC Control	IJ°	I+6



Simultaneous RR and MC



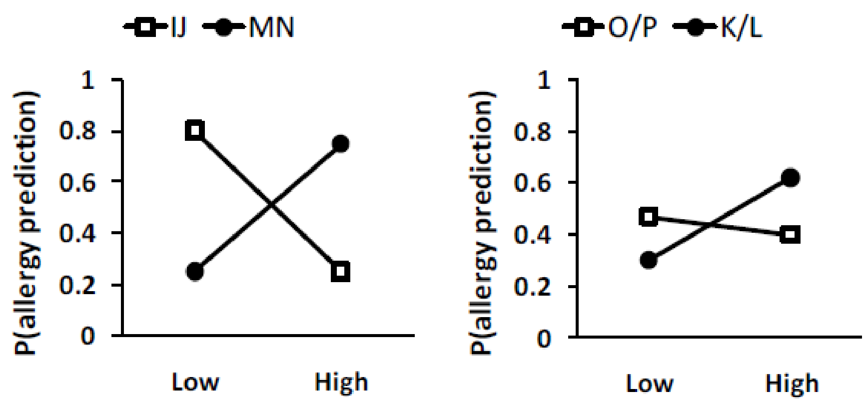
3. Rules and similarity

Shanks & Darby (1998)

A.

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?

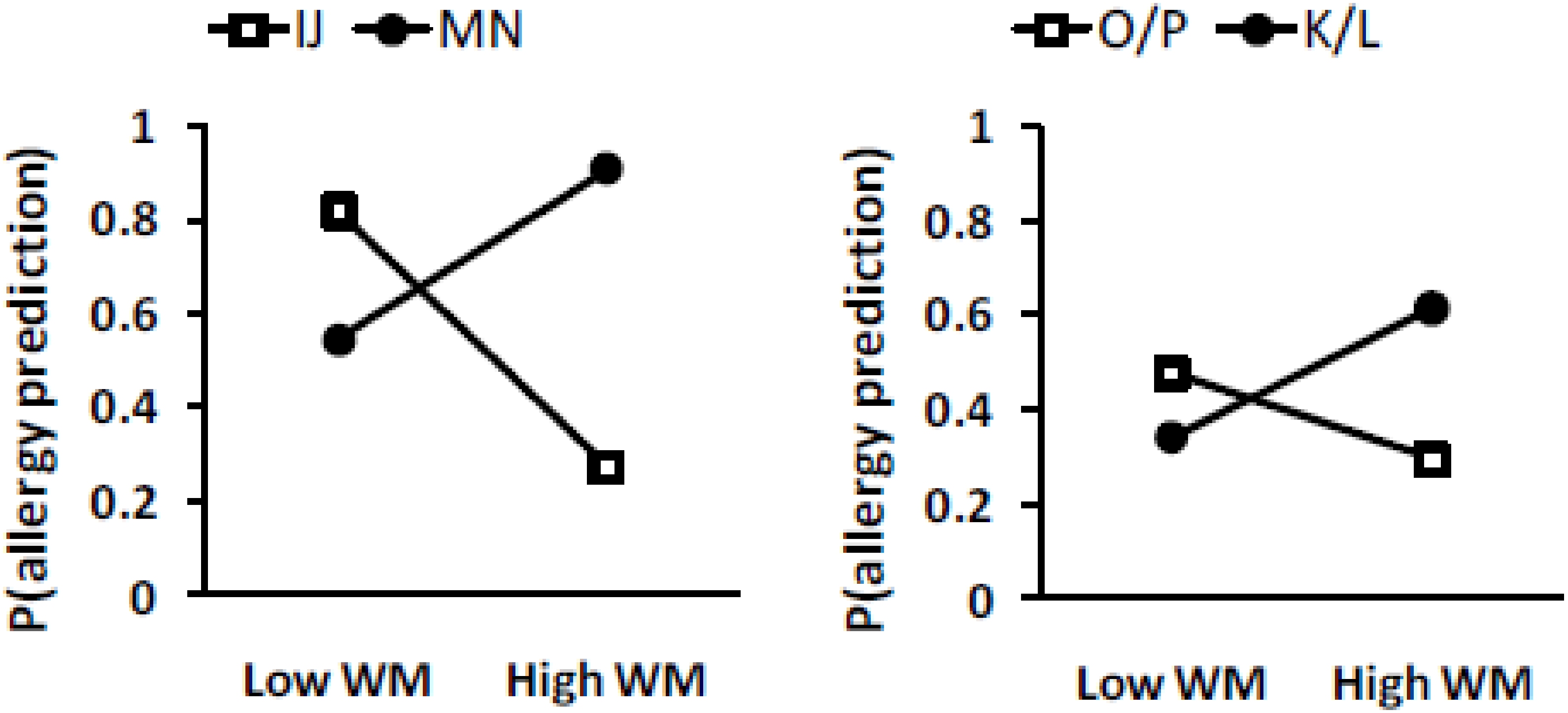
B.



Predictions

- High WMC more likely to show “rule-based” generalization than low WMC
- Rule-based generalization more likely under full attention than under concurrent load.

Shanks-Darby and WMC



Function of learning level?

Concurrent WM load procedure

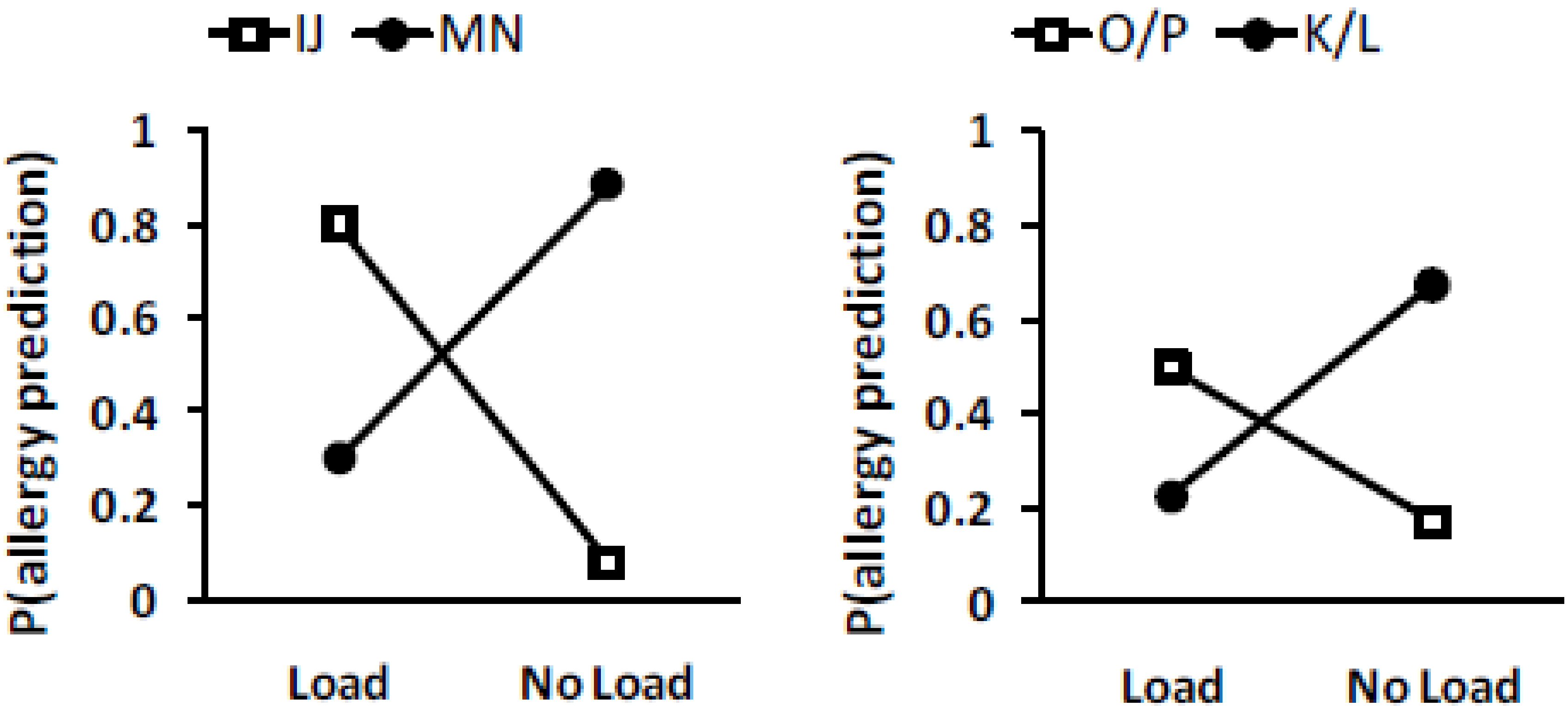
6..2..9..1..4..3

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

Correct.
Mr. X had a reaction

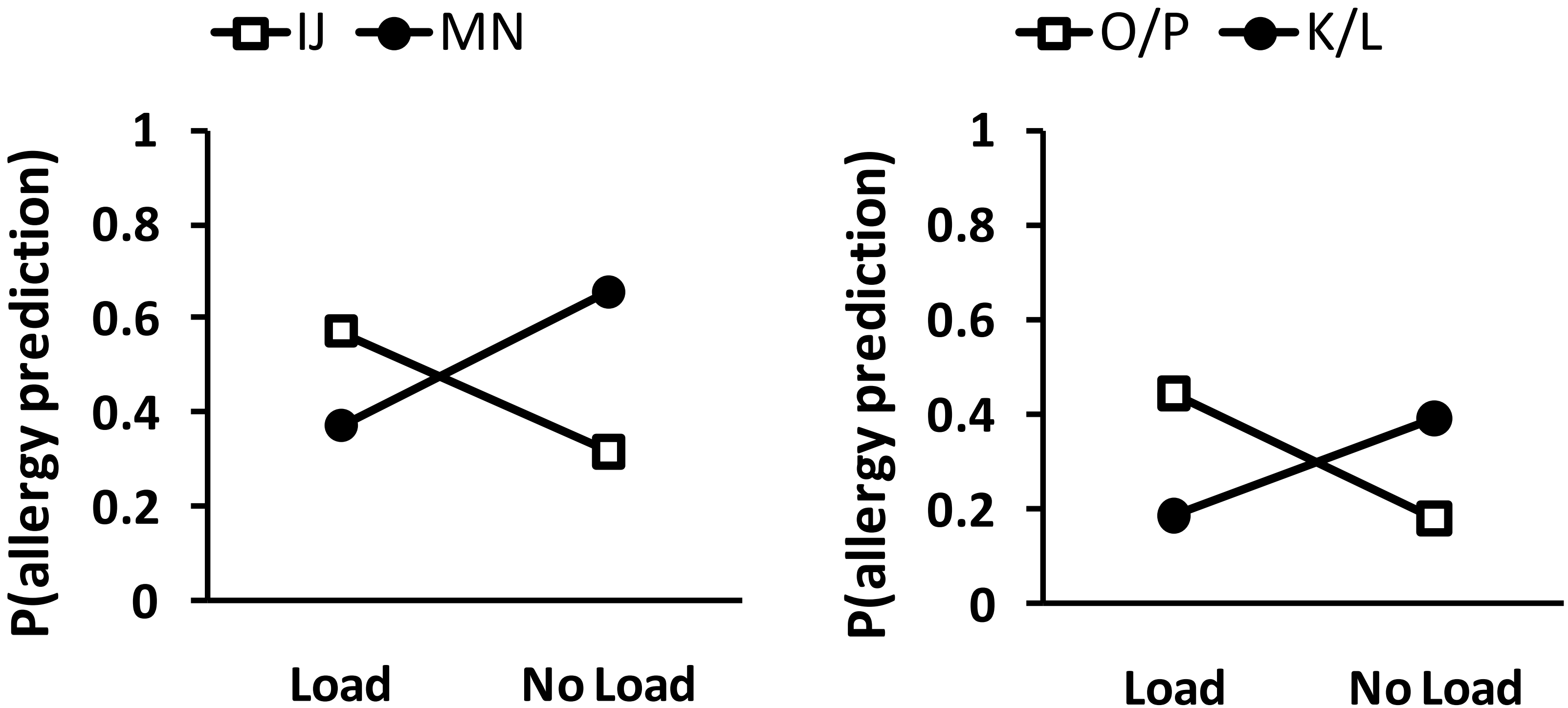
9

Shanks-Darby and Concurrent Load



Load during training and test

Concurrent load experiment 2



Load during training vs. No load during training (collapsed across test load)

4. Summary

Concluding on Generalization

- Surprise-driven attention
  - Sometimes gives a shortcut to the right answer (forward cue competition).
  - Sometimes leads to really daft inferences (inverse base-rate effect).
- Absent-but-expected cues
  - Some generalizations seem rational (retrospective revaluation)
  - Others seem frankly odd (mediated conditioning)
  - Both RR and MC can be simultaneously demonstrated.
  - Something to do with same/different categories?
- Rules vs. similarity
  - Both seem reasonable
  - But can lead to opposite behaviour (Shanks-Darby procedure)
  - Which behaviour you get depends on:
    - Cognitive load at time of learning
    - Individual differences in working memory capacity
- Which of these behaviours do artificial systems display? Which would you want them to?