Learning in Dynamic Conditional Inference

Mike Oaksford & Simon Hall, *Birkbeck College, University of London*

Introduction

- **The New Paradigm**
- **Conditional Inference (Oaksford & Chater, 2007)**
  - Conditionalisation
  - Invariance Violations
  - Learning and Abstract Data
- **Problems (Oaksford & Chater, 2013)**
  - Counterfactual Probabilities
  - Parameter Setting
  - Model Refit and Two Clusters
  - Predictions
- **Experiments: Revising down Pr(q|p)**
  - Experiment 1: Causal Conditionals
  - Experiment 2 & 3: Abstract Conditionals and Contrast Classes
  - Experiment 4: Inference plus Pr(q|p) Judgement task
- **Conclusions**

1996: XXVI International Congress of Psychology, Montreal

542: Pragmatic constraints on human reasoning
Convenor: K.J. Holyoak *(SYM 12:30–15:00)*

542.2: Reasoning and relevance
Politzer, G. *Centre National de la Recherche Scientifique, Paris, France*

542.3: Probabilistic pragmatics and human reasoning
Oaksford, M. *University of Warwick, Coventry, UK*
Pragmatics can be defined as the study of the relation between language and background knowledge (Levinson, 1983). Such knowledge profoundly affects human verbal reasoning in a variety of tasks. Recently it has been suggested that the effects of background knowledge can be modelled probabilistically (Oaksford & Chater, 1994). This probabilistic pragmatics of reasoning is briefly illustrated in three areas of human reasoning-Wason's selection task, conditional inference, and syllogistic reasoning. It is concluded that most human reasoning may reflect an adaptation to gathering useful probabilistic information about the world rather than the operation of an underlying logical competence.

Dynamic Inference & the New Paradigm

- **New Paradigm**
  - Pr(*if p then q*) = Pr($q|p$) (Probability conditional: $p \rightarrow q$)
  - Asserting *if p then q* means Pr($q|p$) is high
- **The Ramsey Test**
  - Subjective probability: Degrees of belief
  - Suppose *p*, by adding to one's stock of beliefs, revise to accommodate, and read off Pr(*q*)
- **Conditional Inference as Bayesian Conditionalization**
  - Oaksford et al (2000); Oaksford & Chater (2007, 2013)

Bayesian Conditionalisation

- If p then q $Pr_0(q|p) = .9$ (1)
- p $Pr_1(p) = 1$ (2)
- q $Pr_1(q) = .9$

- P_0 = old distribution
- P_1 = new distribution

Pr_0	q	$\neg q$	
p	.45	.05	.5
$\neg p$	.25	.25	.5
	.7	.3	1

Pr_1	q	$\neg q$	
p	.9	.1	1
$\neg p$	0	0	0
	.9	.1	1

MP as a Rule for Updating Our Degrees of Belief by Conditionalization

	If p then q	$P_0(q\|p)$	$P_0(white\|swan)$
	p	$P_1(p) = 1$	$P_1(Tweety\ is\ a\ swan) = 1$
∴	*q*	$P_1(q) = P_0(q\|p)$	$P_1(Tweety\ is\ white) = P_0(white\|swan)$
AC	*If p then q*	$P_0(q\|p)$	
	q	$P_1(q) = 1$	
∴	*p*	$P_1(p) = P_0(p\|q) = P_0(q\|p)P_0(p)/P_0(q)$	(Bayes' Theorem)
MT	*If p then q*	$P_0(q\|p)$	
	¬q	$P_1(\neg q) = 1$	
∴	*¬p*	$P_1(\neg p) = P_0(\neg p\|\neg q) = (1\text{-}P_0(q)\text{-}P_0(p)(1\text{-}P_0(q\|p)))/(1\text{-}P_0(q))$	
DA	*If p then q*	$P_0(q\|p)$	
	¬p	$P_1(\neg p) = 1$	
∴	*¬q*	$P_1(\neg q) = P_0(\neg q\|\neg p) = (1\text{-}P_0(q)\text{-}P_0(p)(1\text{-}P_0(q\|p)))/(1\text{-}P_0(p))$	

(Oaksford, Chater, & Larkin, 2000; Oaksford & Chater, 2007)

Constraints on Conditionalisation

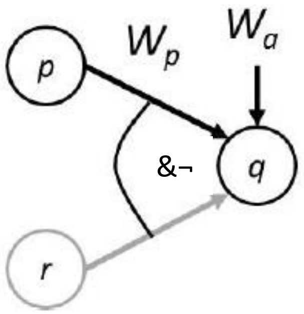
- **Invariance (Pearl, 1989; Jeffrey, 2004)**
- **Rigidity (Sobel, 2004)**
- **Dynamic Uncertainty Sum Rule (Adams, 1998)**
  - **$\mathrm{Pr}_0(x) = \mathrm{Pr}_0(x|I)$ Probability of a premise *x* should not change (much) given new information *I*. Premises are *independent* of new information**
- **MP: $\mathrm{Pr}_0(p \rightarrow q) = \mathrm{Pr}_0(p \rightarrow q|p) = \mathrm{Pr}_1(p \rightarrow q)$**
  - **$\mathrm{Pr}_0(q|p) = \mathrm{Pr}_0(q|p, p) = \mathrm{Pr}_1(q|p)$**
  - **Information in the categorical premise does not alter the conditional probability**

Invariance Violations

- MT (Sobel, 2004; Oaksford & Chater, 2007)
  - If turn key, car starts
  - Car did not start
  - ∴ Key not turned???
- Far more likely some other cause (defeater)
- $\mathrm{Pr}_0(q|p, \neg q) = \mathrm{Pr}_0(q|p)$?
- But even if no b-gnd K, assertion of categorical premise only informational if key was turned
- Suggests $p, \neg q$ counterexample, i.e., $\mathrm{Pr}_0(p \rightarrow q) \neq \mathrm{Pr}_1(p \rightarrow q)$

Invariance Violations: Some Responses

- $\mathrm{Pr}_0(q|p) \neq \mathrm{Pr}_1(q|p)$
  - What are the rational constraints on $\mathrm{Pr}_1(q|p)$?
- Explain Away (World Knowledge)
  - r = Petrol tank empty
  - $\mathrm{Pr}_0(q|p) = \mathrm{Pr}_1(q|p)$

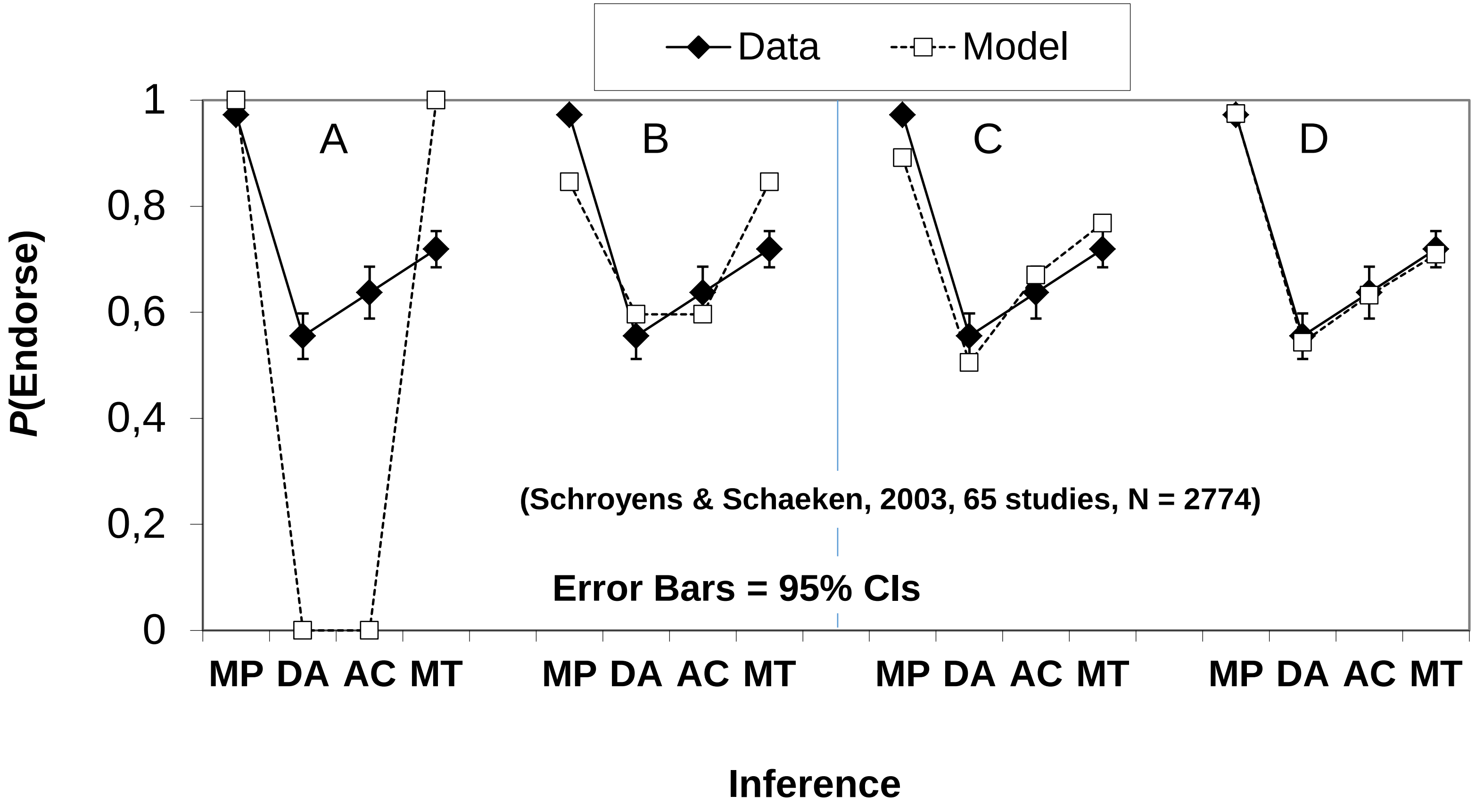


- Counterexample
  - Learn $\mathrm{Pr}_0(q|p)$ lower than you thought
  - $\mathrm{Pr}_0(q|p) > \mathrm{Pr}_1(q|p)$
  - Bayesian Updating on a single counterexample

Invariance Violations

- Does it matter?
  - Are examples of violations rare aberrations or the norm
  - Zhao & Osherson (2012) discuss cases when it happens and when it does not
- Empirical Consequences
  - Can better explain existing data
  - Predicts that information about categorical premises might alter Pr(q|p), i.e., the probability of the conditional premise

Conditional Inference Data (Abstract)



MP If p, then q
p
∴ q

MT If p, then q
$\neg q$
∴ $\neg p$

AC If p, then q
q
∴ p

DA If p, then q
$\neg p$
∴ $\neg q$

Learning (Oaksford & Chater, 2007)

D

Pr_0	q	$\neg q$	
p	.45	.05	.5
$\neg p$	.25	.25	.5
	.7	.3	1

I

Pr_0	q	$\neg q$	
p	.35	.15	.5
$\neg p$	.35	.15	.5
	.7	.3	1

- **With no b-gnd K, adjust $Pr_0(p \rightarrow q)$ assuming a single $p, \neg q$ counterexample**
  - **Probability of endorsing MP = $Pr_0(q|p)$**
  - **Foil = Independence Model**
  - **Marginals constant between D and I**
  - **Prior Pr(D) = $Pr_0(q|p)$**
  - **Prior Pr(I) = 1 - Pr(D)**
  - **$Pr(D|p,\neg q) = Pr(D|\neg p,q)$ = revised $Pr_0(q|p)$**
  - **Revised $Pr_0(q|p)$ used to calculate MT, DA and AC**
- **Produces original fits in Panel D**

Problems (Oaksford & Chater, 2013)

The Queen's Anniversary Prizes For Higher and Further Education 2006

- **MT**
  - If counterexample, then key was turned. Shouldn't Ps assign probability 0 to "the key was not turned"?

- **Parameter Setting**
  - In Oaksford & Chater (2007) $Pr_0(p)$ set to .5, violates rarity (Oaksford & Chater, 1994)
  - Is there a more rational basis for parameter setting?
  - Here we argue that this may be a consequence of the assumption that $Pr(D|p,\neg q) = Pr(D|\neg p,q)$ = revised $Pr_0(q|p)$

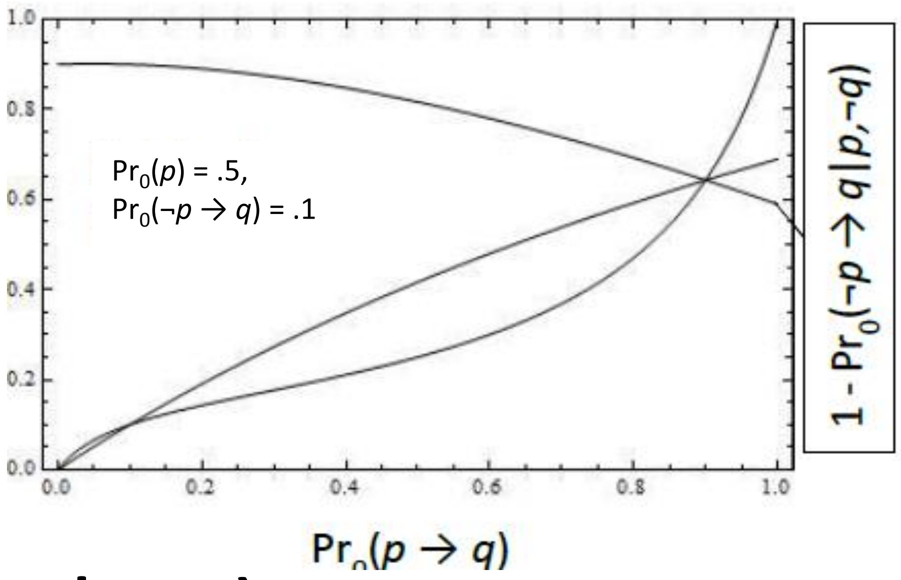
- **Re-Modelling**
  - Re-model the data, with $Pr_0(p)$ and $Pr_0(\neg p|q)$ free to vary

MT and Counterfactual Probabilities

- Tom: *If its sunny tomorrow, I'll play tennis*
  - You turn up to play tennis (because you believe Pr(q|p) is high and its sunny) but Tom not there?
  - Counterexample, so drive Pr(q|p) to 0? Infer its not sunny?
- Legitimate excuse (defeater) or he is unreliable?
- Learn from the counterexample
  - *The probability you would have assigned to Tom turning up given what you now know*
- Counterfactual Reasoning
  - *if I knew then what I know now I would have…*reasoning is ubiquitous
- No change in the learning process for these counterfactual probabilities

Probabilistic Bi-Conditional

x = Prior ($\mathrm{Pr}_0(p \rightarrow q)$)
y = Posterior ($\mathrm{Pr}_0(p \rightarrow q \mid p, \neg q)$, $\mathrm{Pr}_0(p \rightarrow q \mid \neg p, q)$)



- **For $\mathrm{Pr}_0(p \rightarrow q \mid p, \neg q) = \mathrm{Pr}_0(p \rightarrow q \mid \neg p, q)$**
  - $Pr_0(q \mid p) = 0$
  - $Pr_0(q \mid p) = Pr_0(q \mid \neg p)$
  - $\mathbf{Pr_0(q \mid p) = 1 - ((1/\mathrm{Pr}_0(p)) - 1) Pr_0(q \mid \neg p)}$
- **So if $\mathrm{Pr}_0(p) = .5$, $Pr_0(q \mid p) = Pr_0(\neg q \mid \neg p)$**
  - p is as necessary as it is sufficient for q
  - This probabilistic biconditional is a very simplifying assumption

Re-Fitting the Model

- **Re-fitted with:**
  - $\Pr_0(p)$ and $\Pr_0(q|\neg p)$ free to vary
  - $\Pr_0(q|p) = \Pr(\mathrm{MP})$
  - Revised $\Pr_0(q|p) = \Pr_0(p \rightarrow q|p,\neg q)$
- **Fit was the same as O&C (2007)**
  - So cluster analysis on $\Pr_0(q|p)$, $\Pr_0(p)$ and $\Pr_0(q|\neg p)$

- **Cluster 1 (N = 29)**
  - $\Pr_0(p) = .48\ (.09) \approx .5$
  - $\Pr_0(q|p) = .98\ (.03)$
  - $\Pr_0(q|\neg p) = .04\ (.07)$
  - $\Pr_0(q|p) + \Pr_0(q|\neg p) \approx 1$
  - Initially *p* is as necessary as it is sufficient for *q*.

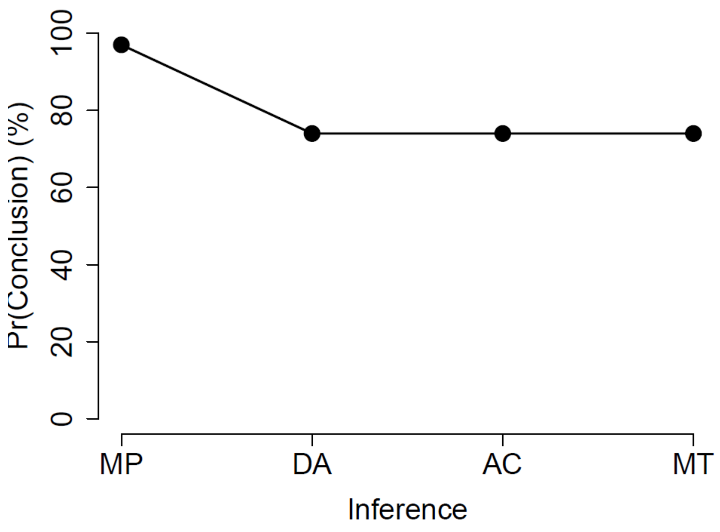
- **Cluster 2 (N = 26)**
  - $\Pr_0(p) = .40\ (.08) << .5$
  - $\Pr_0(q|p) = .96\ (.04)$
  - $\Pr_0(q|\neg p) = .46\ (.17) \approx .5$
  - $\Pr_0(\neg q|\neg p) \approx \Pr_0(q|\neg p)$
  - Initially $\neg p$ is uninformative about *q*.
  - $\Pr_0(p)$ much closer to rarity

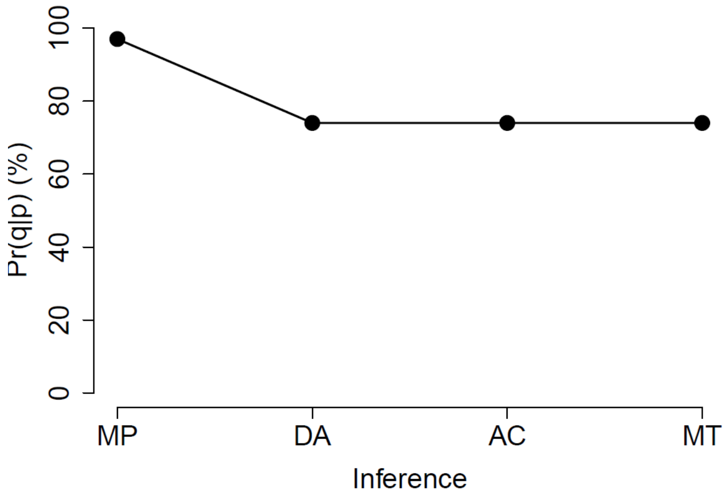
Revised $\Pr_0(q|p) = .74$
(For DA, AC, MT)

Predictions: Cluster 1

The Queen's Anniversary Prizes
For Higher and Further Education
2006

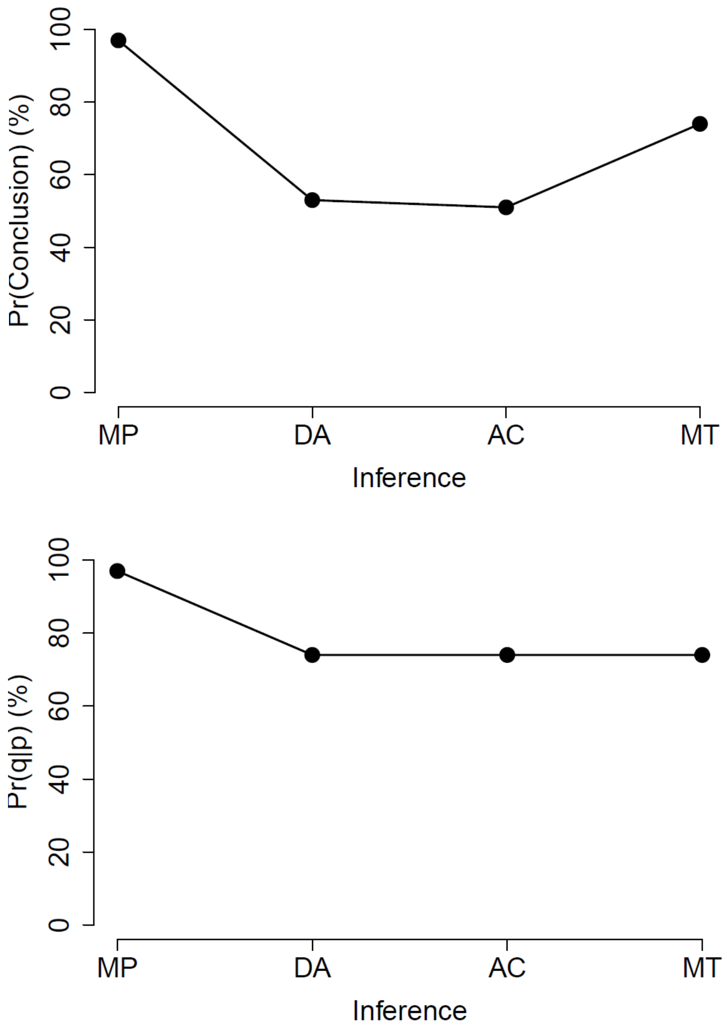
Pr_0	q	$\neg q$	
p	.37	.13	.5
$\neg p$	.13	.37	.5
	.5	.5	1





Predictions: Cluster 2

Pr_0	*q*	¬*q*	
p	.3	.1	.4
¬*p*	.28	.32	.6
	.58	.42	1



Experiment 1

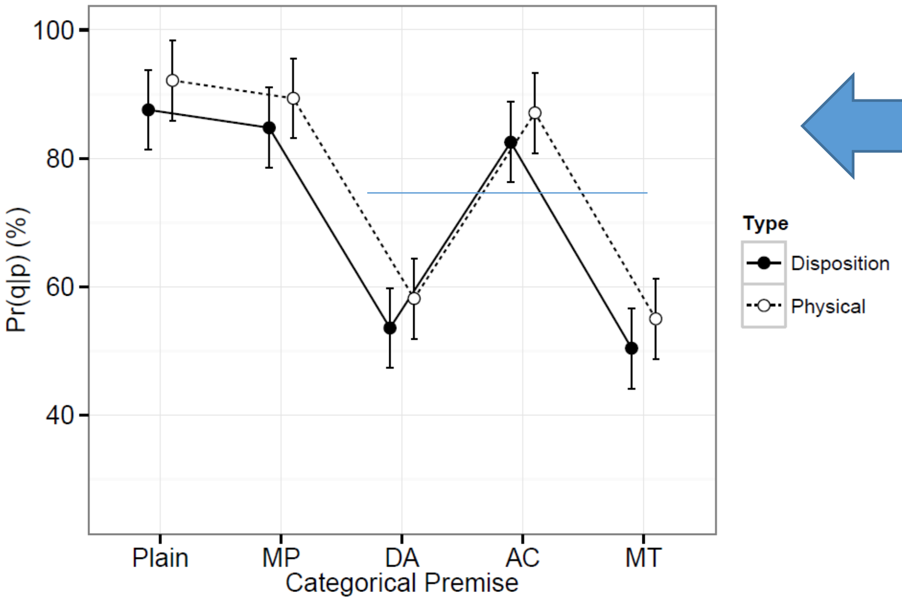
The Queen's Anniversary Prizes
For Higher and Further Education
2006

- **Causal Materials**
  - Learning strategy based on analogy with causal case
- **Materials**
  - Physical causes (6) and dispositions (7), random presentation.
- **Task**
  - Given conditional and categorical premise judge Pr(q|p)
  - Included "Plain condition" just conditional
  - **N = 27** (Plain and MP, DA, AC, MT for each conditional)

Plain If his Renault car breaks down, then Fred is late.
How likely do you think it is that Fred is late if his Renault car breaks down?

MT If his Renault car breaks down, then Fred is late.
Fred isn't late.
How likely do you now think it is that Fred is late if his Renault car breaks down?

Experiment 1 Results



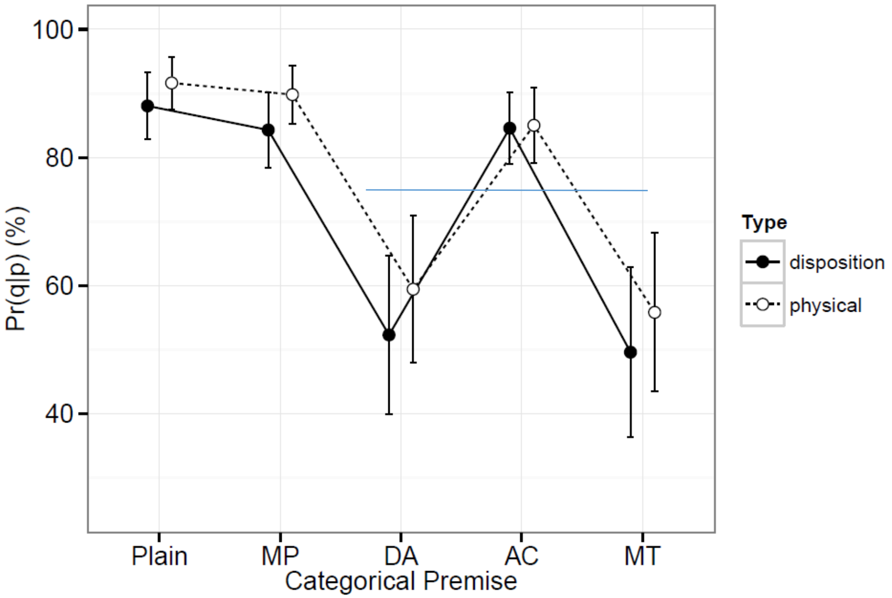
Linear Mixed Models Analysis
(lmer, lsmeans)

Plain = MP = AC > DA = MT

All Diffs *p < .0001*

Collapsed over items
(ANOVA)

Implemented in Qualtrics
Run on CrowdFlower



Experiment 1: Conclusions & Questions

- **Conclusions**
  - Ps do revise Pr(q|p) down given cat premise (DA and MT)
  - Not just because cat premise is present (Plain = MP)
    - Not just fulfilling perceived task demands
- **Questions?**
  - Not for AC, so only for negated premises
    - Because obvious counterexamples OR just negation?
    - Perceived predictive (MP)/diagnostic (AC) function?
  - If abstract by analogy with causal case
    - Does this result replicate for abstract material?
  - Can DA and MT be presented without negations?

Experiment 2

- **Abstract Materials**
  - Cards with coloured shapes on them
  - Use 6 rules with either negations or contrast class member
  - Cover Story:

Your friend has discovered a box of about 1000 cards discarded outside a psychology lab. The cards are blank on one side, and has a coloured shape on the other side. Each shape is of one colour. Your friend has been trying to work out whether there are some rules about possible combinations of colour and shape which the psychologists used to prepare the cards. He has examined 100 cards from the box.

In the following questions your friend first tells you one of the rules he believes applies. You will then be asked to judge how likely each rule is to be true. He then takes a further sample of 25 cards from the remaining 900 cards in the box. After examining these 25 cards he chooses 4 to tell you something about one at a time. This will be either its shape or its colour. In light of the information about each of the 4 cards, you will be asked to judge afresh how likely the rule he has just proposed is to be true.

Before starting the experiment proper you will be given an example of the judgements you need to make, after which you will start the experiment itself.

Experiment 2

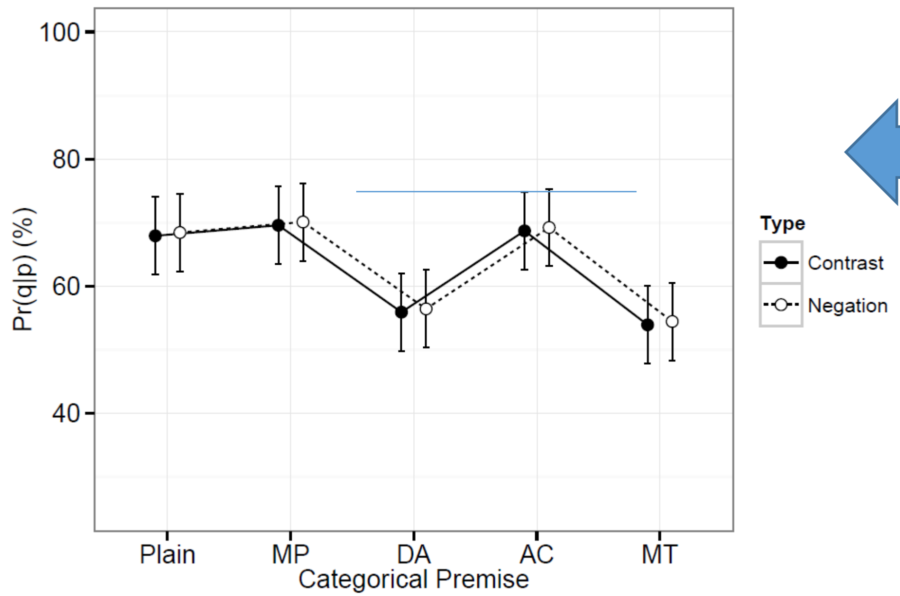
- **Contrast Classes (Oaksford & Stenning, 1992)**
  - 6 rules: 3 *if shape, colour*; 3 *if colour, shape* (6 shapes, 6 colours)

Negation DA
Your friend tells you that, "If the shape is a ring, then it is orange.“
For one of the 4 cards he selected from the new sample of 25 he tells you that the **shape on the card is not a ring**.
How likely do you think it is that the shape is orange if it is a ring?

Contrast DA
Your friend tells you that, "If the shape is a ring, then it is orange.“
For one of the 4 cards he selected from the new sample of 25 he tells you that the **shape on the card is a square**. (square ∈ Shapes\ring)
How likely do you think it is that the shape is orange if it is a ring?

- **Design**
  - Blocked on Contrast/Negation (CN or NC order)
  - Random presentation within blocks
  - **N = 47**

Experiment 2 Results



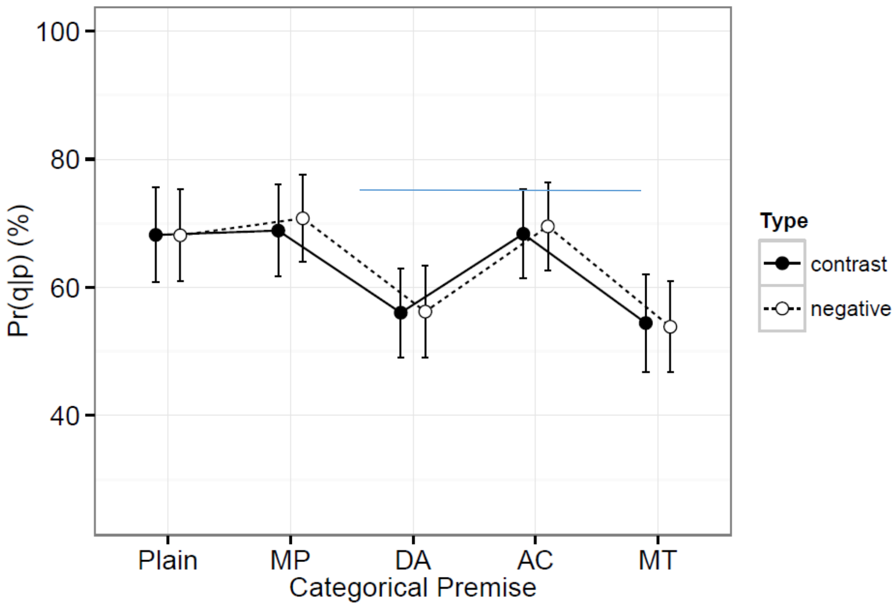
Linear Mixed Models Analysis
(lmer, lsmeans)

Plain = MP = AC > DA = MT

All Diffs *p < .001*

Collapsed over items
(ANOVA)

Implemented in Qualtrics
Run on CrowdFlower



Experiment 2: Conclusions & Questions

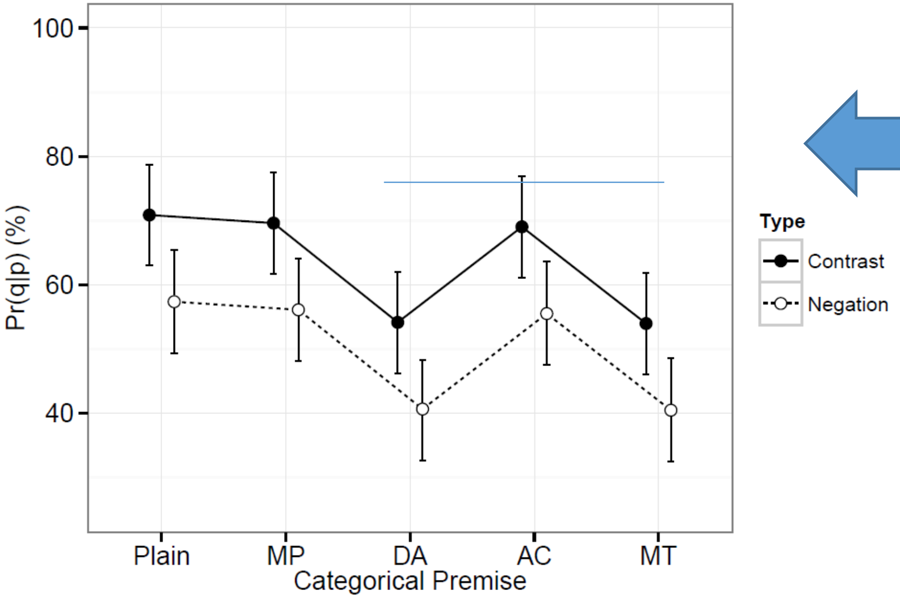
- **Conclusions**
  - Ps do revise Pr(q|p) down given cat premise (DA and MT)
    - Happens without explicit negations and with abstract material
  - Not just because cat premise is present (Plain = MP)
    - Not just fulfilling perceived task demands
  - Same pattern as causal case
    - Rehder (2014) also showed similar behaviour with causal and blank material in discounting/augmentation task
- **Questions?**
  - Again not for AC
    - Because obvious counterexamples (not just negation)?
    - Perceived predictive (MP)/diagnostic (AC) function?
  - Pr(q|p) much lower than causal for Plain, MP, and AC
  - Pr(q|p) much lower than model fits for DA and MT

Experiment 3

- **Blocking**
  - Not recorded in Experiment 2
  - Are results different for Block 1 vs. Block 2?
  - Replicate Experiment 2 making sure block recorded
- **Task**
  - **N = 51** (Plain and MP, DA, AC, MT for each conditional)

Experiment 3 Results: Block 1

The Queen's Anniversary Prizes For Higher and Further Education 2006



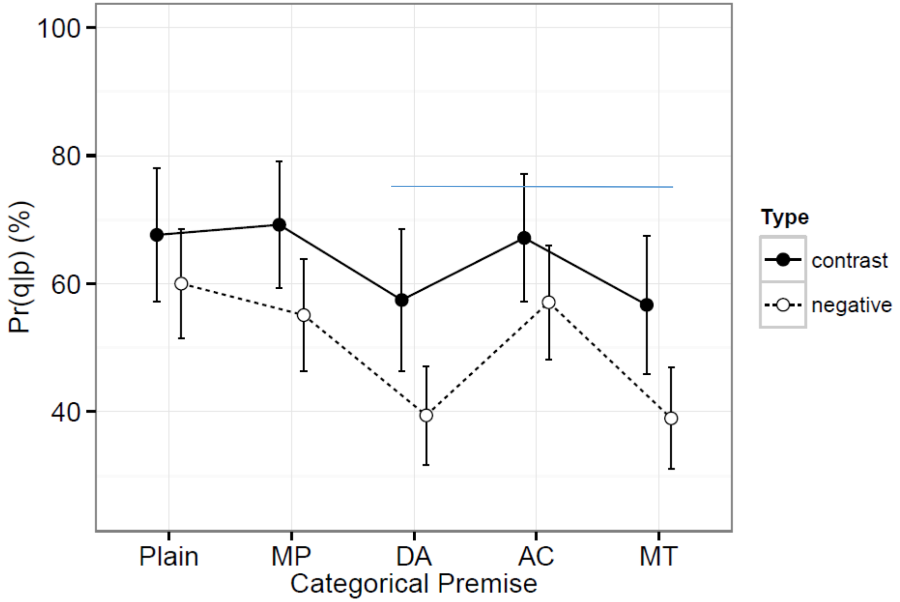
Linear Mixed Models Analysis
(lmer, lsmeans)

Plain = MP = AC > DA = MT

All Diffs *p* < *.001*

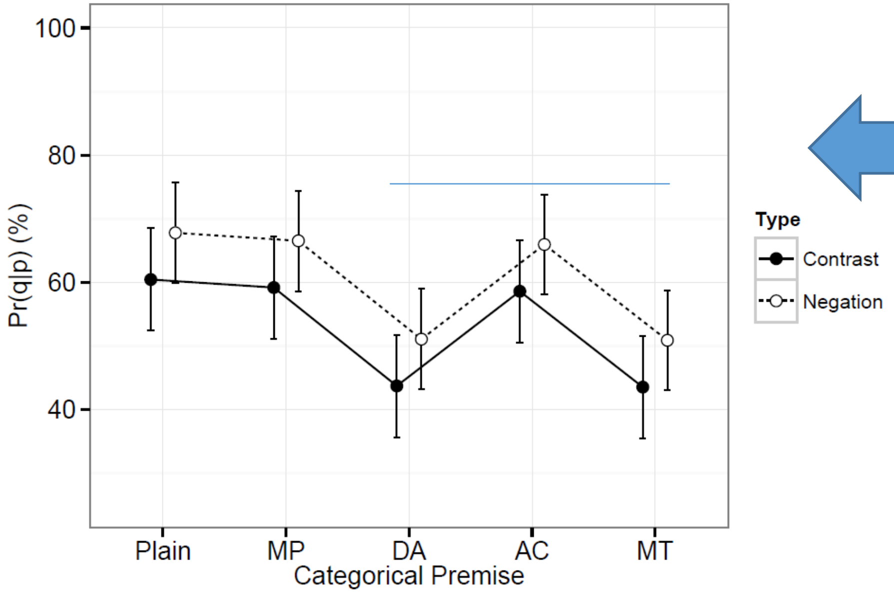
Collapsed over items
(ANOVA)

Implemented in Qualtrics
Run on CrowdFlower



Experiment 3 Results: Block 2

The Queen's Anniversary Prizes
For Higher and Further Education
2006



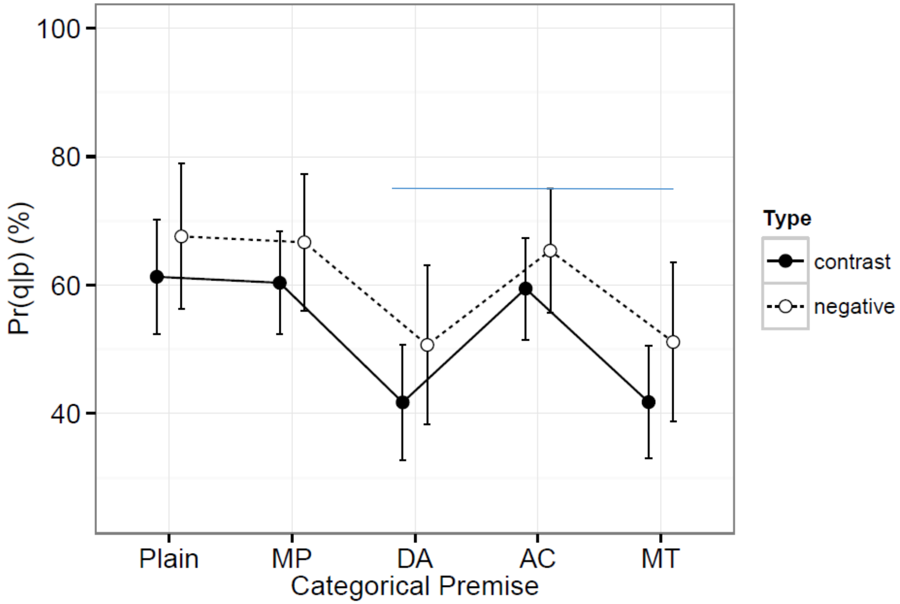
Linear Mixed Models Analysis
(lmer, lsmeans)

Plain = MP = AC > DA = MT

All Diffs *p* < *.001*

Collapsed over items
(ANOVA)

Implemented in Qualtrics
Run on CrowdFlower



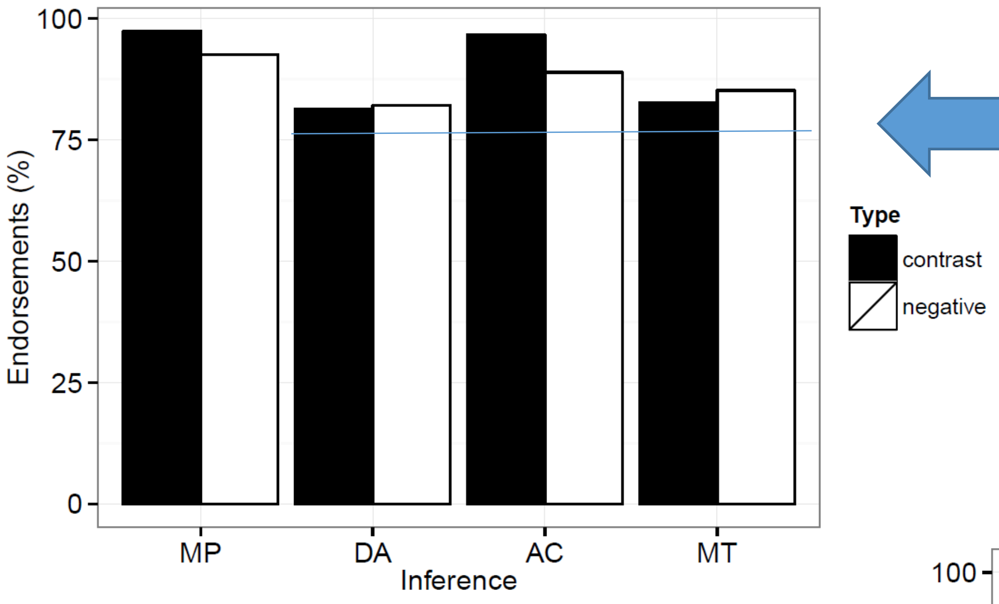
Experiment 3: Conclusions & Questions

- **Conclusions**
  - Replicated Experiment 2 in both blocks
  - Block effects
    - Individual differences: CN > NC?
- **Questions?**
  - Pr(q|p) much lower than causal for Plain, MP, and AC
  - Pr(q|p) much lower than model fits for DA and MT
  - Participant are more probabilistically coherent in the context of inferences (Evans, Thompson, & Over, 2015)
    - Perhaps can only expect revisions of Pr(q|p) to match model fits if Ps do inference tasks first?

Experiment 4

- **Inference plus Pr(q|p) Judgement task**
  - Same materials as Experiments 2 and 3
- **Task Order, Within N-C blocks two ordered sub-blocks:**
  - 1. **Plain** All Plain Pr(q|p) judgements in a single block, then
  - 2. **Inference**
    - Perform the inference, MP, DA, AC, MT given the conditional and categorical premises (Binary Selection Task)
    - After each inference asked for new Pr(q|p) judgement
- **Design**
  - Blocked on Contrast/Negation (CN or NC order)
  - Random presentation within sub-blocks
  - **N = 52**

Experiment 4 Results: Block 1



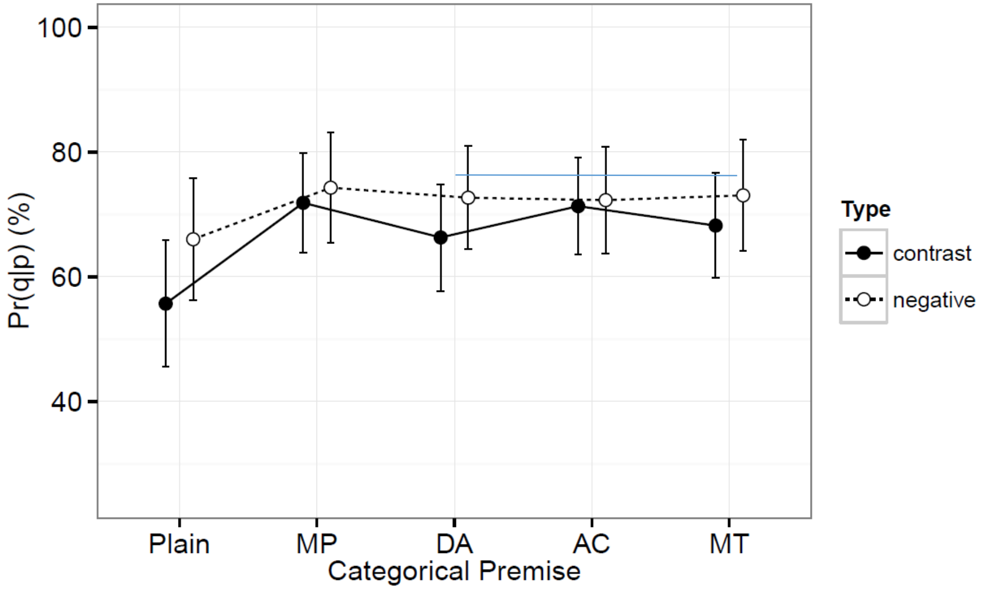
Collapsed over items (**ANOVA**) Same result using **clmm**

MP = AC > DA = MT

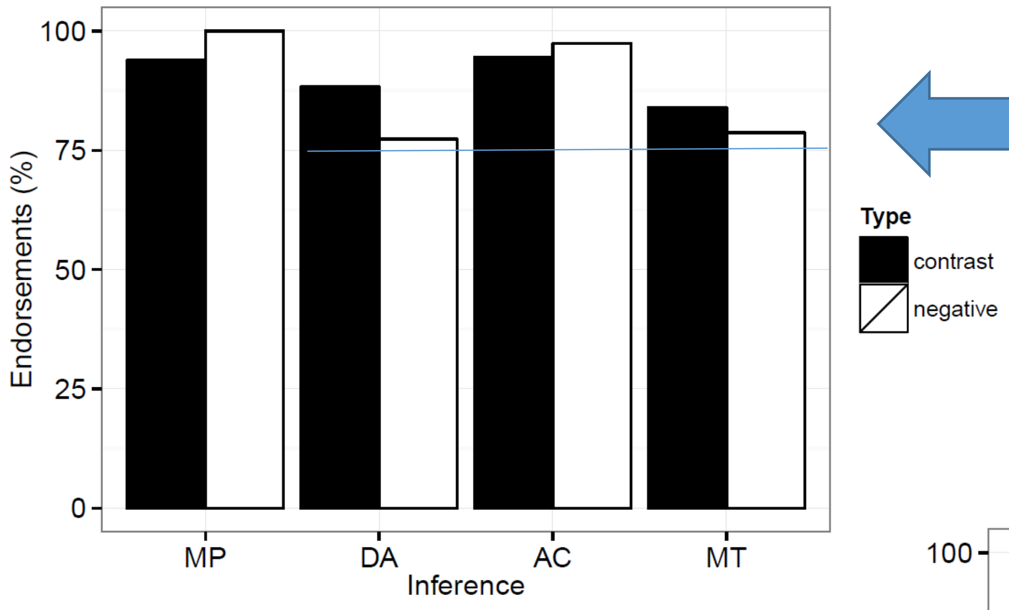
All Diffs *p < .001*

Collapsed over items (**ANOVA**)

Implemented in Qualtrics
Run on ProlificAcademic



Experiment 4 Results: Block 2



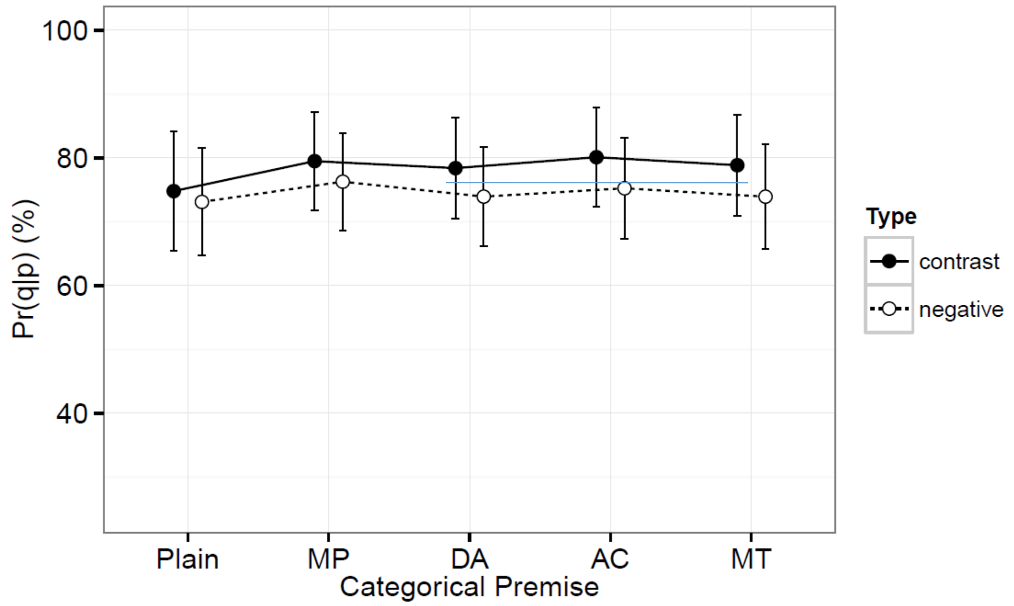
Collapsed over items **(ANOVA)** Same result using **clmm**

MP = AC > DA = MT

All Diffs *p < .001*

Collapsed over items **(ANOVA)**

Implemented in Qualtrics
Run on ProlificAcademic



Experiment 4: Conclusions & Questions

- Conclusions
  - Pr(q|p) same as model fits for DA, AC, and MT (Cluster 1)
  - Endorsements for DA and MT also same as model fits
  - Block effect for Inference and Pr(q|p)
    - Inference: Individual differences CN > NC
    - Pr(q|p): Individual differences NC > CN
- Questions?
  - Pr(q|p) for Plain & MP $\ngtr$ DA, AC & MT
  - Endorsements for AC > DA and MT
  - Pr(q|p) $\approx$ .74 cannot explain pattern of endorsements (other than for DA, MT)

Conclusions

- Positives and Negatives
  - Participants do revise Pr(q|p) down given categorical premise (Expts, 1, 2, 3) BUT only for DA, MT, not AC (not cat prem or "¬" per se)
  - For abstract material, after an inference task, they are at the same levels as model fits (Expt 4) BUT Plain and MP not higher
- Possibilities (worth pursuing)
  - Conditional needs to be ASSERTED
    - In Abstract task, person just "believes" these are the rules (convinced, strongly believes)
    - In Abstract task, sample size = 100 (N = 1000) needs to be varied
    - Use argumentative context: A asserts p → q, B selects p card…
  - Inference Task
    - Binary Selection format: MP/DA = q ¬q
    - Standard format: All infs = p ¬p q ¬q None of these
  - Don't block Plain
  - Inform Ss of Range of colours and shapes (fix marginals)
  - Inference plus Pr(q|p) with Causal material

Conclusions continued

- Summary
  - Tantalising but not conclusive evidence of dynamic update of conditional probabilities given categorical premises
  - May explain much of the data from earlier period of conditional inference research using abstract material
  - Moreover could do so by analogy to the later period of research primarily on causal conditional reasoning and probabilities (New Paradigm)
- Future
  - Look at learning conditional information in more complex inferences and how probabilities are updated (minimising mutual information)
  - How are representations updated to include defeaters
    - Are we measuring the right probabilities? $Pr(q|p, \neg defeaters)$ v $Pr(q|p, defeaters)$
    - Doesn't seem to make sense for abstract case BUT WCST (people may always generate more complex rules)