Comparing Models of Probabilistic Conditional Reasoning: Evidence for an Influence of Form

Albert-Ludwigs-Universität Freiburg

Henrik Singmann

Karl Christoph Klauer

Sieghard Beller

NEW • FRAMEWORKS • OF RATIONALITY

DFG • SCHWERPUNKTPROGRAMM • 1516

UNI FREIBURG

4 Conditional Inferences

Modus Ponens (MP):

If *p*, then *q*.
p

Therefore, *q*

Affirmation of the consequent (AC):

If *p*, then *q*.
q

Therefore, *p*

Modus Tollens (MT):

If *p*, then *q*.
Not *q*

Therefore, not *p*

Denial of the antecedent (DA):

If *p*, then *q*.
Not *p*

Therefore, not *q*

4 Conditional Inferences

Modus Ponens (MP):
If *p*, then *q*.
p
Therefore, *q*

Modus Tollens (MT):
If *p*, then *q*.
Not *q*
Therefore, not *p*

valid in standard logic (i.e., truth of premises entails truth of conclusion)

Affirmation of the consequent (AC):
If *p*, then *q*.
q
Therefore, *p*

Denial of the antecedent (DA):
If *p*, then *q*.
Not *p*
Therefore, not *q*

NOT valid in standard logic (i.e., truth of premises does NOT entail truth of conclusion)

Research Question:

What is the **effect of the presence of the conditional** in everyday probabilistic conditional reasoning?

Example Item: Knowledge Phase

Observation: A balloon is pricked with a needle.

How likely is it that it will pop?

Example Item: Knowledge Phase

Observation: A balloon is pricked with a needle.

How likely is it that it will pop?

Example Item: Rule Phase

Rule: If a balloon is pricked with a needle, then it will pop.

Observation: A balloon is pricked with a needle.

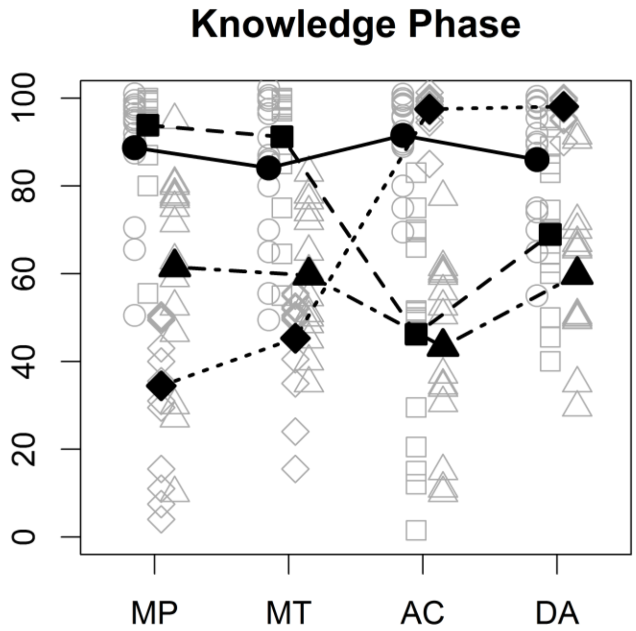
How likely is it that it will pop?

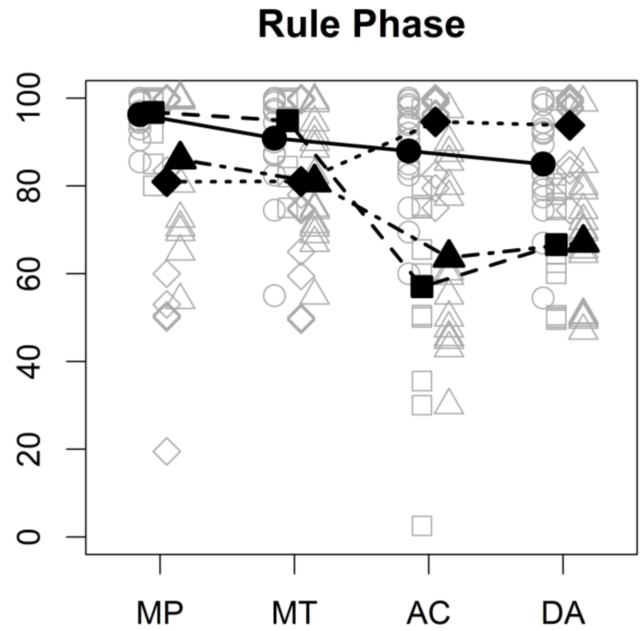
Example Item: Rule Phase

Rule: If a balloon is pricked with a needle, then it will pop.

Observation: A balloon is pricked with a needle.

How likely is it that it will pop?



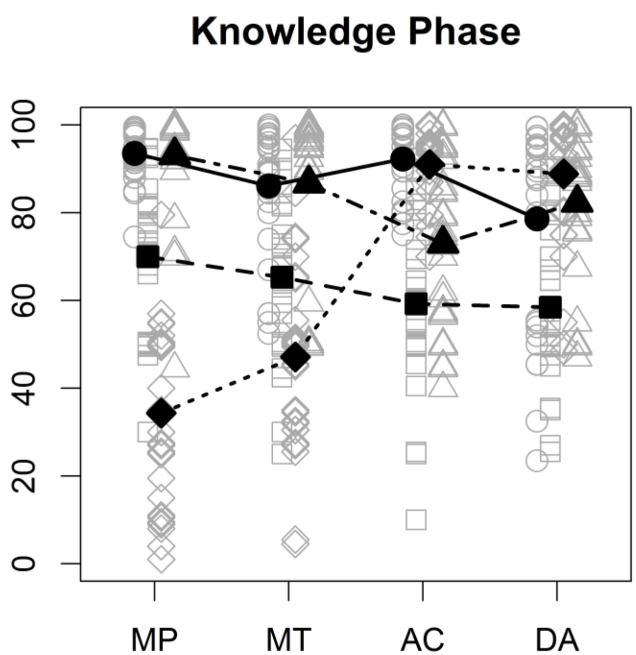


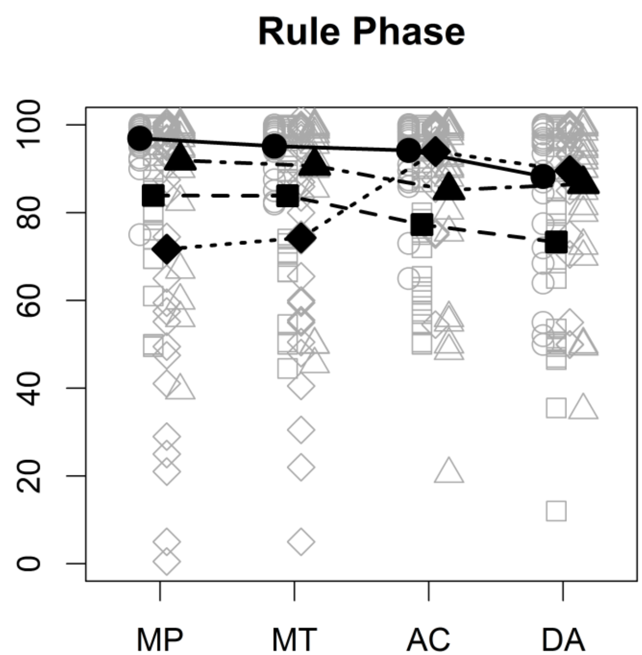
Klauer, Beller, & Hütter (2010): Experiment 1 (n = 15)

↑ conditional absent ↓

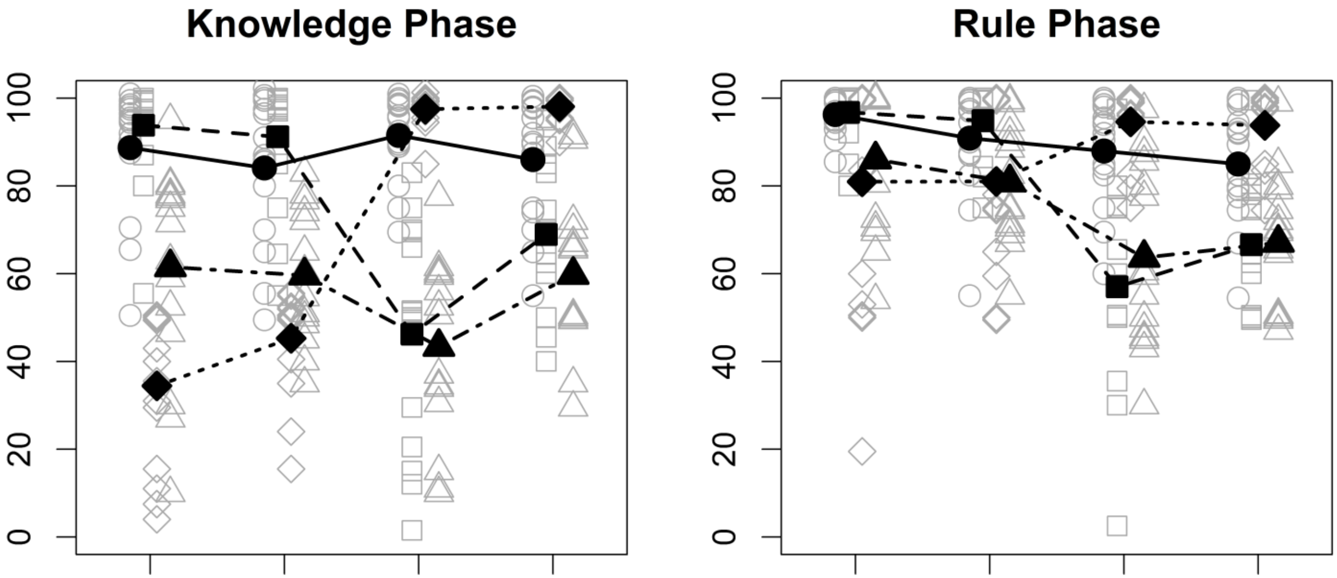
↑ conditional present ↓

different lines represent different conditionals (i.e., different contents/items)



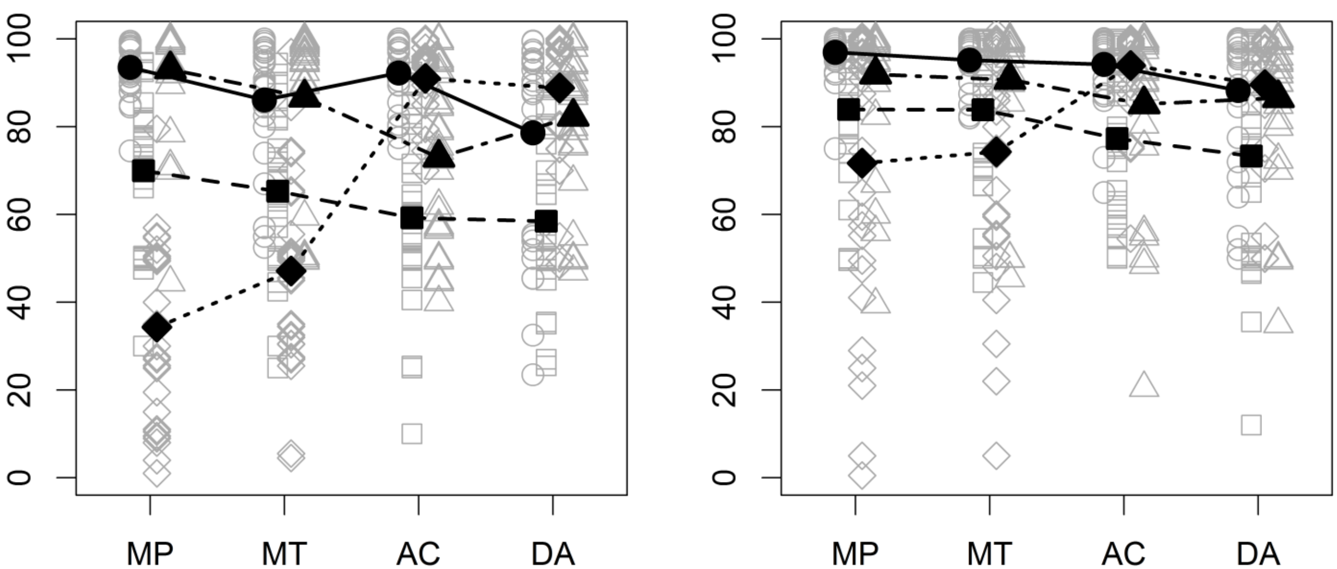


new data (n = 29)

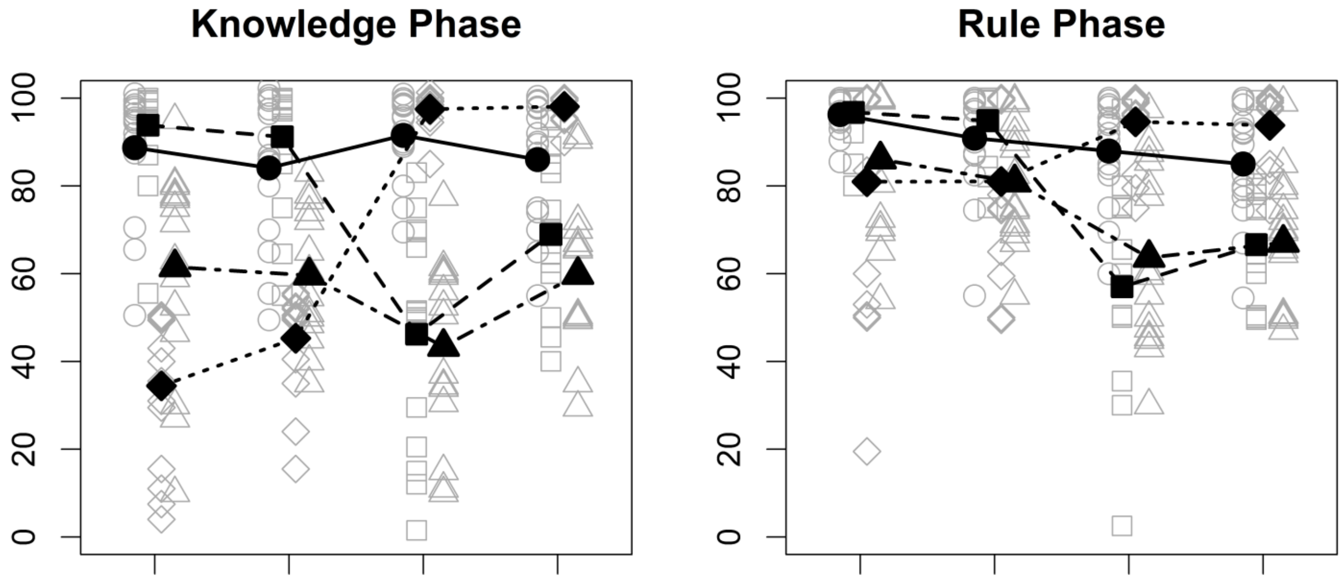


Klauer, Beller, & Hütter (2010): Experiment 1 (n = 15)

The presence of the conditional increases participants' estimates of the probability of the conclusion. Especially so for the formally valid inferences MP and MT.

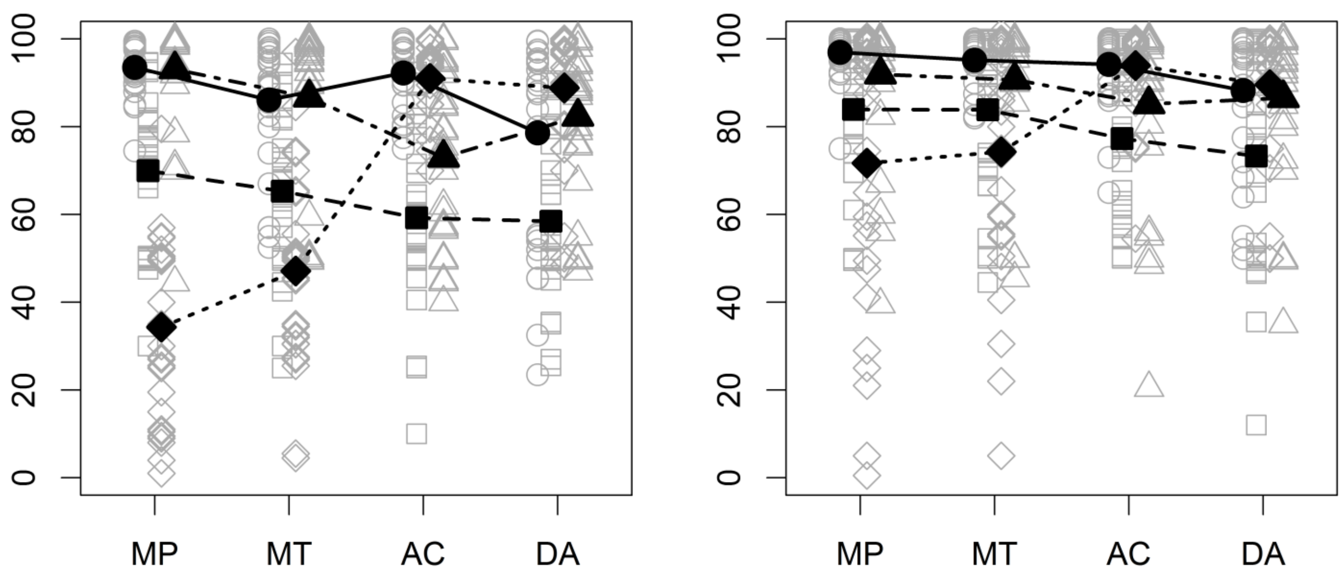


new data (n = 29)



Klauer, Beller, & Hütter (2010): Experiment 1 (n = 15)

How can we explain this effect of the presence of the conditional? Our data challenge pure probabilistic approaches that solely rely on background knowledge.



new data (n = 29)

Example Item: Knowledge Phase

Observation: A balloon is pricked with a needle.

How likely is it that it will pop?

The Knowledge Phase

- Participants are asked to estimate a conclusion given a minor premise only. E.g., Given *p*, how likely is *q*?

- The response should reflect the conditional probability of the conclusion given minor premise, e.g., $P(q|p)$

"Inference"	"MP"	"MT"	"AC"	"DA"
	p $\therefore q$	$\neg q$ $\therefore \neg p$	q $\therefore p$	$\neg p$ $\therefore \neg q$
Response reflects	$P(q\|p)$	$P(\neg p\|\neg q)$	$P(p\|q)$	$P(\neg q\|\neg p)$

Formalizing the Knowledge Phase

- We have the joint probability distribution over $P(p)$, $P(q)$, and their negations in the knowledge phase per content:

	q	¬q
p	$\mathrm{P}(p \wedge q)$	$\mathrm{P}(p \wedge \neg q)$
$\neg p$	$\mathrm{P}(\neg p \wedge q)$	$\mathrm{P}(\neg p \wedge \neg q)$

- From this we can obtain the conditional probabilities, e.g.: $\mathrm{P}(\mathrm{MP}) = \mathrm{P}(q|p) = \mathrm{P}(p \wedge q) / \mathrm{P}(q)$
- We need at least three independent parameters (e.g., $\mathrm{P}(p)$, $\mathrm{P}(q)$, and $\mathrm{P}(\neg q|p)$, Oaksford, Chater, & Larkin, 2000) to describe the joint probability distribution.

How do we explain the effect of the conditional?

- When the **conditional is absent**, participants use their **background knowledge** to estimate the **conditional probability of the conclusion given minor premise**. E.g., Given *p*, how likely is *q*? P(*q*|*p*)

- The **presence of the conditional** adds a different type of information: **form-based evidence** (i.e., the **subjective probability to which an inference is seen as logically warranted** by the form of the inference). E.g., How likely is the conclusion given that the inference is MP?

- The dual-source model (Klauer, Beller, & Hütter, 2010) posits that people integrate these two types of information in the conditional inference task.

How do we explain the effect of the conditional?

- When the **conditional is absent**, participants use their **background knowledge** to estimate the **conditional**

Our intuition:
The conditional provides form-based information which is integrated with background knowledge on the subject matter to come to a blended reasoning conclusion.

- The dual-source model (Klauer, Beller, & Hütter, 2010) posits that people integrate these two types of information in the conditional inference task.

Formalizing the Dual-Source Model

Par.	Interpretation	Influencing Factors
λ	Relative weight given to form-based versus knowledge-based evidence	E.g., speaker expertise, instructional emphasis on rule
τ	Degree to which an inference is seen as logically warranted	E.g., inference (MP, MT, AC, DA), connective (e.g., "if -then" vs. "or")
ξ	Knowledge-based response proposal	E.g., contents of the premises/ salience of counterexamples

- Observable response on one inference
  With conditional = $\lambda\{\tau(x) \times 1 + (1 - \tau(x)) \times \xi(C,x)\} + (1 - \lambda)\xi(C,x)$
  Without conditional = $\xi(C,x)$

- $\tau(x)$ = form-based evidence, subjective probability for accepting inference x (i.e., MP, MT, AC, DA) based on the logical form.
- $\xi(C,x)$ = knowledge-based evidence, subjective probability for accepting inference *x* for content *C* based on the background knowledge.

How else could we explain it?

- An alternative to our idea can be construed from parts of the literature on reasoning in the new paradigm (e.g., Oaksford & Chater, 2007; Pfeifer & Kleiter, 2010) or the literature on causal Bayes nets (e.g., Sloman, 2005; Fernbach & Erb, in press).

- According to this view, **reasoning is basically probability estimation from background knowledge**.

- Consequently, the **presence of the conditional** simply **changes the knowledge base** from which people reason (i.e., the form affects each inference idiosyncratically given the specific content)

Formalizations of "mere background knowlegde"

We consider **two models** based on the work of **Oaksford, Chater, and Larkin (2000)**:

- In their model the joint probability distribution is defined by P(p), P(q), and P(¬q|p) (the exceptions parameter), from which all four conditional probabilities can be calculated.
(This is also the parametrization of the knowledge in the dual-source model)

- In the **original model** (Oaksford et al., 2000), **the presence of the conditional decreases possible exceptions**, P'(¬q|p) < P(¬q|p), so only affecting one of the underlying parameters.
= one free parameter per conditional.

- In the **extended version** of their model (Oaksford & Chater, 2007; Oaksford & Chater, in press), **the presence of the conditional decreases possible exceptions differently for MP versus MT, AC, and DA** (less exceptions for MP).
= two free parameters per conditional.

An alternative formalization

- According to **Hartmann and Rafiee Rad** (2012) when reasoning from conditionals and learning the premises, participants update their joint probability distribution by minimizing the **Kullback-Leibler divergence** with respect to the new information.
- We therefore describe the joint probability distribution by a = P(p), α = P($q|p$), and β = P($\neg q|\neg p$).
- When **learning the conditional**, participants **update the conditional probability $\alpha' > \alpha$**, and the other parameters (a' and β') update **by minimizing the Kullback-Leibler distance** to the original distribution.
- With only one additional free parameter (α') we therefore also update the other parameters.

Empirical Question:

Which of the three models provides the best account to the empricial data?

The Data Sets

- We only considerd data sets (or parts thereof) that implement the paradigm sketched in the beginning:
  - A knowledge phase without conditional.
  - A rule phase with conditional being present.
  - No additional manipulations.

- We analyzed 6 datasets (*N* = 148):
  - Klauer et al. (2010, Exps. 1 & 3, *n* = 15 & 18) and 2 new data sets (*n* = 26 & 29) exactly implemented the paradigm. (For Klauer et al.'s experiments the third sessions were omitted)
  - For Klauer et al. (2010, Exp 4, n = 13) we omitted the trials which did not follow the paradigm sketched above.
  - One new data set (n = 47) in which we manipulated speaker expertise (i.e., two types of conditionals, experts and novices; Stevenson & Over, 2001)

General Methodology

- 4 different conditionals (i.e., contents)
- Participants responded to all four inferences per content.
- Participants always give two estimates per inference: to the original and to the converse inference, which are pooled for analysis.
- Two measurement points (at least one week in between):
  1. without conditional
  2. with conditional

→ 32 data points per participant.

Exceptions:

- Klauer et al. (2010, Exp. 3): 5 conditionals = 40 data points per participant
- New data on speaker expertise: 6 conditionals (3 experts, 3 novice) = 48 data points

The Candidate Models

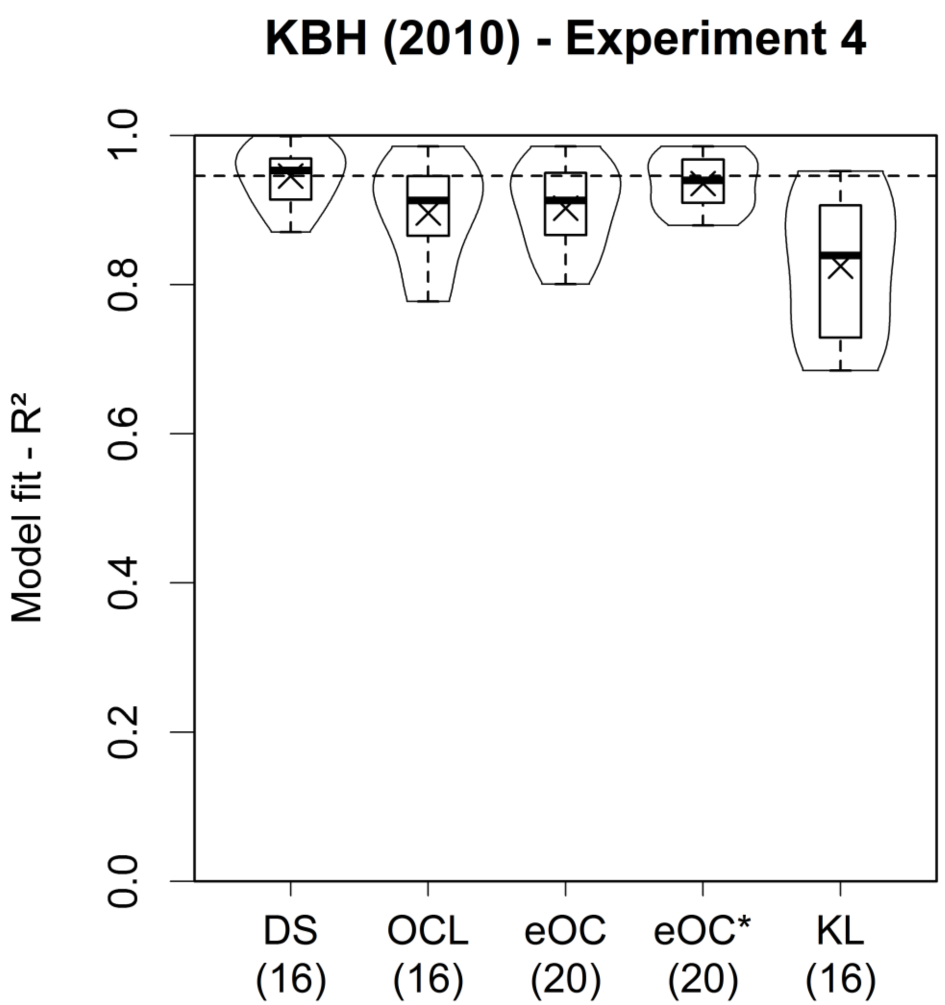
- All Models have three parameters per conditional for the knowledge phase (DS and O&C models share the same parametrization) = 12 parameters for the standard paradigm with 4 conditionals.
- The **dual-source model (DS)** adds one set of four form-based parameters (4) for the knowledge phase independent of the number of conditionals = **16 parameters**
- The **original Oaksford et al. (2000) model (OCL)** adds one parameter per conditional = **16 parameters**.
- The **extended Oaksford & Chater (2007) model (eOC & eOC*)** adds two parameters per conditional = **20 parameters**. (eOC* does not restrict a, b & e'_{MP} to form a probability distribution)
- The model based on minimizing the **Kullback-Leibler divergence (KL)** adds one parameter per conditional = **16 parameters**.
- We fitted all five models to each individual data set by minimizing the sum of the squared deviance from data and predictions.

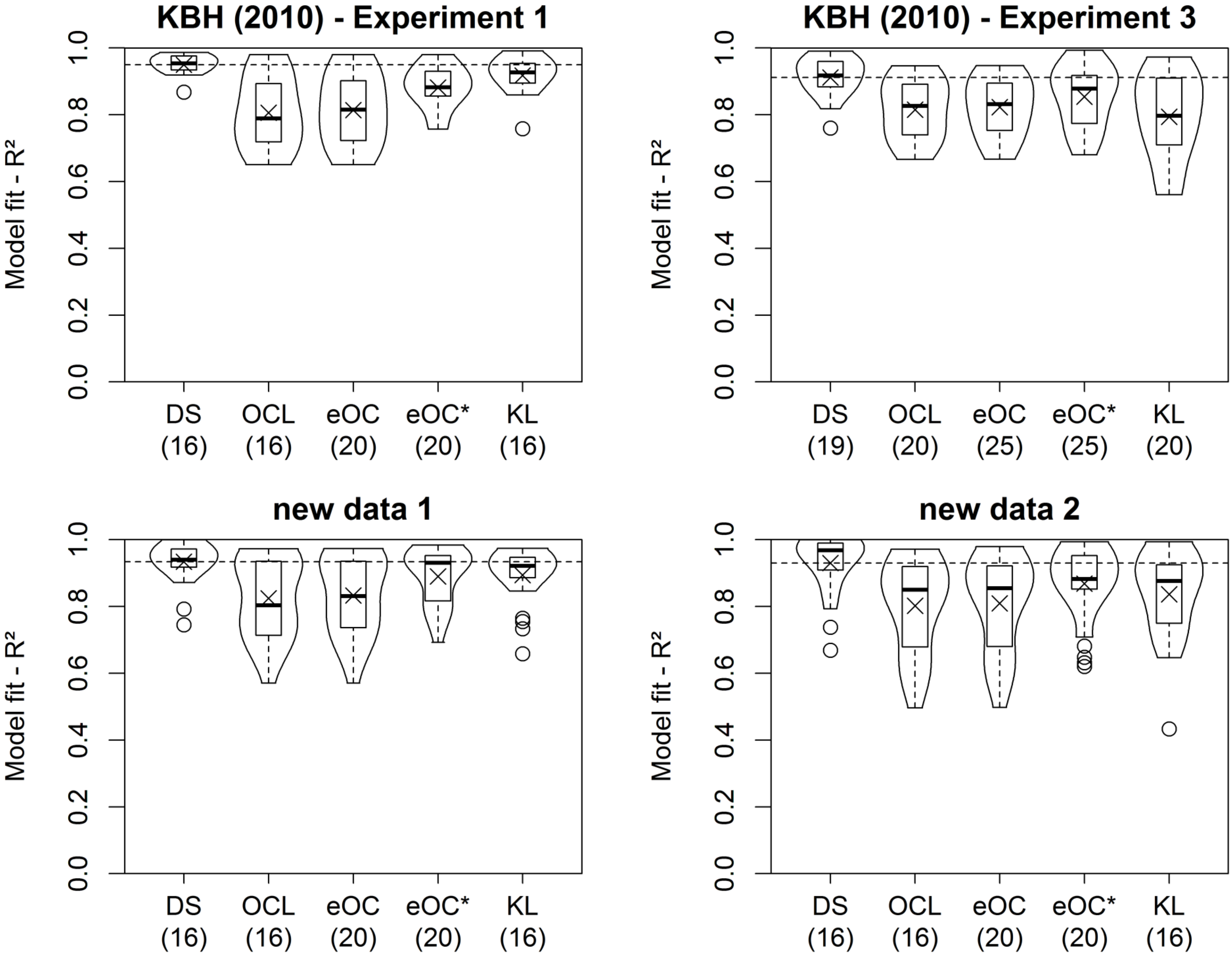
Results I

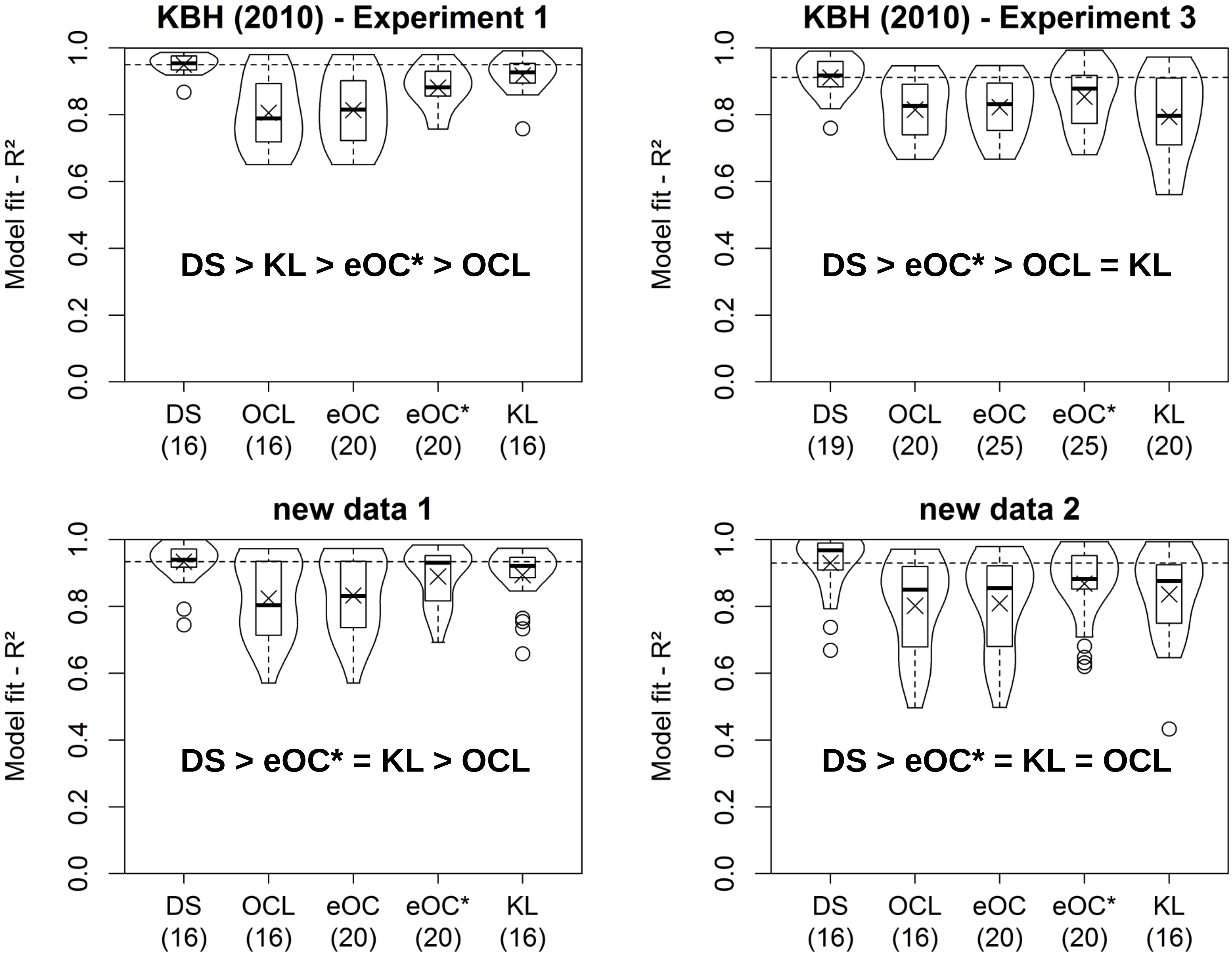
Klauer et al. (2010), Exp. 4

- 5 measurement points, but we only use those data that is comparable to the other conditions (which are spread across the 5 measurement points)
- The dashed line represents the mean fit value from DS.
- Numbers behind model names are the numbers of parameters for each model.

- Ordering:
  DS = eOC* > OCL > KL



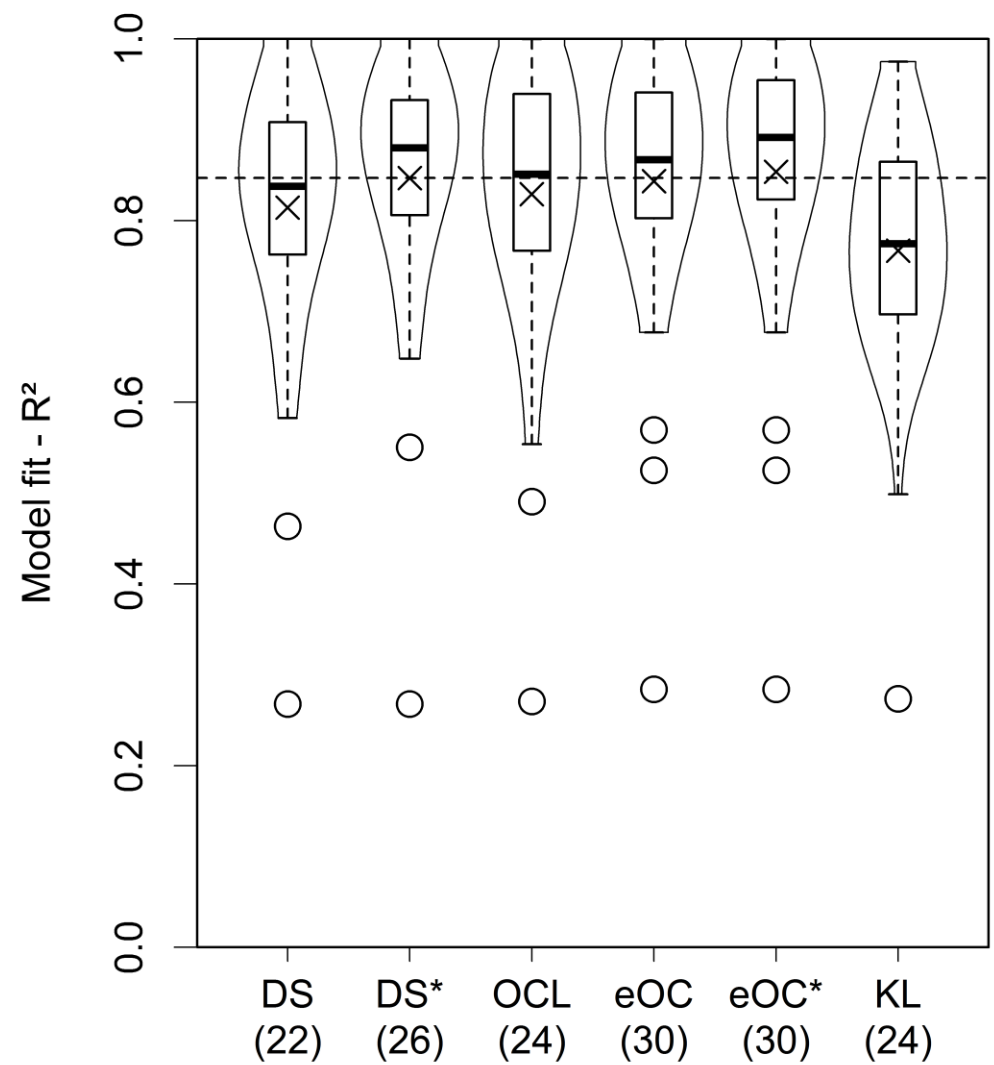

KBH (2010) - Experiment 1
KBH (2010) - Experiment 3
new data 1
new data 2
Model fit - R²
0.0
0.2
0.4
0.6
0.8
1.0
DS (16)
OCL (16)
eOC (20)
eOC* (20)
KL (16)
DS (19)
OCL (20)
eOC (25)
eOC* (25)
KL (20)



KBH (2010) - Experiment 1
Model fit - R²
0.0
0.2
0.4
0.6
0.8
1.0
DS > KL > eOC* > OCL
DS (16)
OCL (16)
eOC (20)
eOC* (20)
KL (16)
KBH (2010) - Experiment 3
Model fit - R²
DS > eOC* > OCL = KL
DS (19)
OCL (20)
eOC (25)
eOC* (25)
KL (20)
new data 1
Model fit - R²
DS > eOC* = KL > OCL
DS (16)
OCL (16)
eOC (20)
eOC* (20)
KL (16)
new data 2
Model fit - R²
DS > eOC* = KL = OCL
DS (16)
OCL (16)
eOC (20)
eOC* (20)
KL (16)

Results Speaker Expertise

- 48 data points.
- In addition to the standard DS, we fitted a version (DS*) in which we allowed for separate sets of form parameters for the expert and novice, respectively.
- Ordering:
  eOC* = DS* ≥ OCL = DS > KL



Summary Results

- **The dual-source model provides the overall best account** for 6 datasets. For the two cases where it shares the first place (with the extended model by Oaksford & Chater, 2007), it has less parameters.
- The dual-source model provides the **best account for 83 of 148 data sets (56%)**.

- The **extended Oaksford & Chater model** (48 best accounts) performs better than the **Kullback-Leibler model** (17 best account) **but has more parameters**.

- This is strong evidence for our interpretation of **the effect of the conditional**:
  It seems to **provide formal information that can not easily be accounted for by changes in participants' knowledge base only**.

Beyond Model Fit

- The parameters of the dual-source model offer an insight into different underlying cognitive processes:
  - A dissociation of different suppression effects: disablers decrease the weighting parameters, whereas alternatives decrease the background-knowledge based evidence for AC and DA (and both affect the form-based component)
  - Speaker expertise maps parsimoniously on the weighting parameters and does not affect the form-based parameters.

- The dual-source model can be easily extended to other types of connectives such as e.g., or, as it is not strictly tied to conditional reasoning.