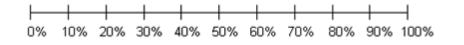


A HIERARCHICAL BAYESIAN IMPLEMENTATION OF PURELY BAYESIAN AND BAYESIAN MIXTURE MODELS OF CONDITIONAL REASONING

Henrik Singmann

A girl had sexual intercourse.

How likely is it that the girl is pregnant?



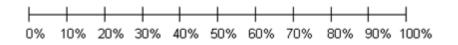
A girl had sexual intercourse.

How likely is it that the girl is pregnant?

0% 10% 20% 30% 40% 50% 60% 70% 80% 90% 100%

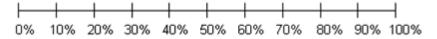
A girl is NOT pregnant.

How likely is it that the girl had NOT had sexual intercourse?



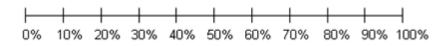


How likely is it that the girl is pregnant?



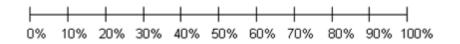
A girl is pregnant.

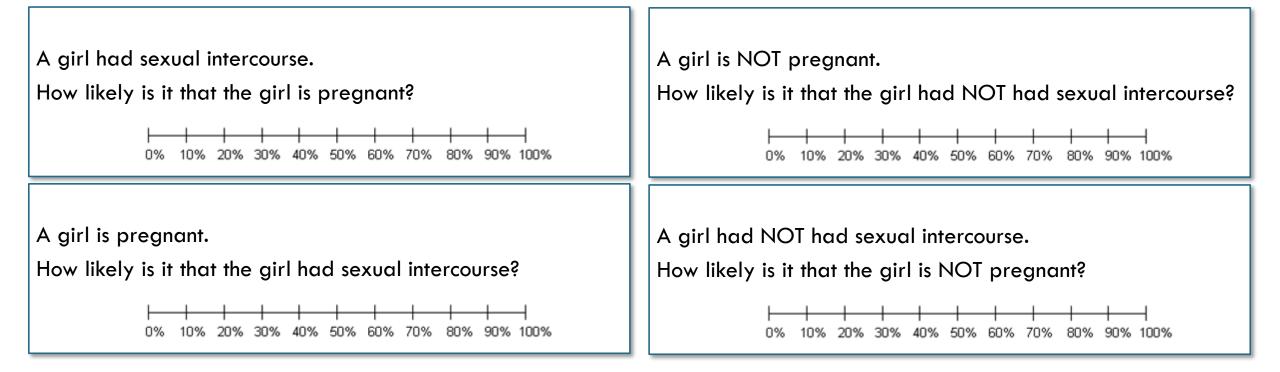
How likely is it that the girl had sexual intercourse?

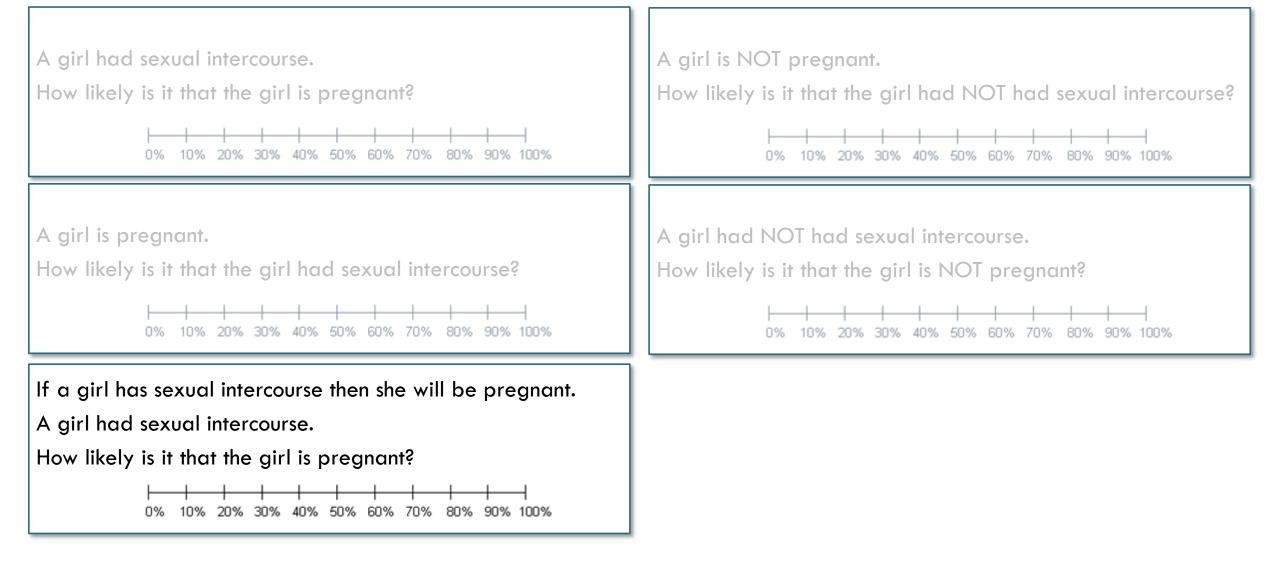


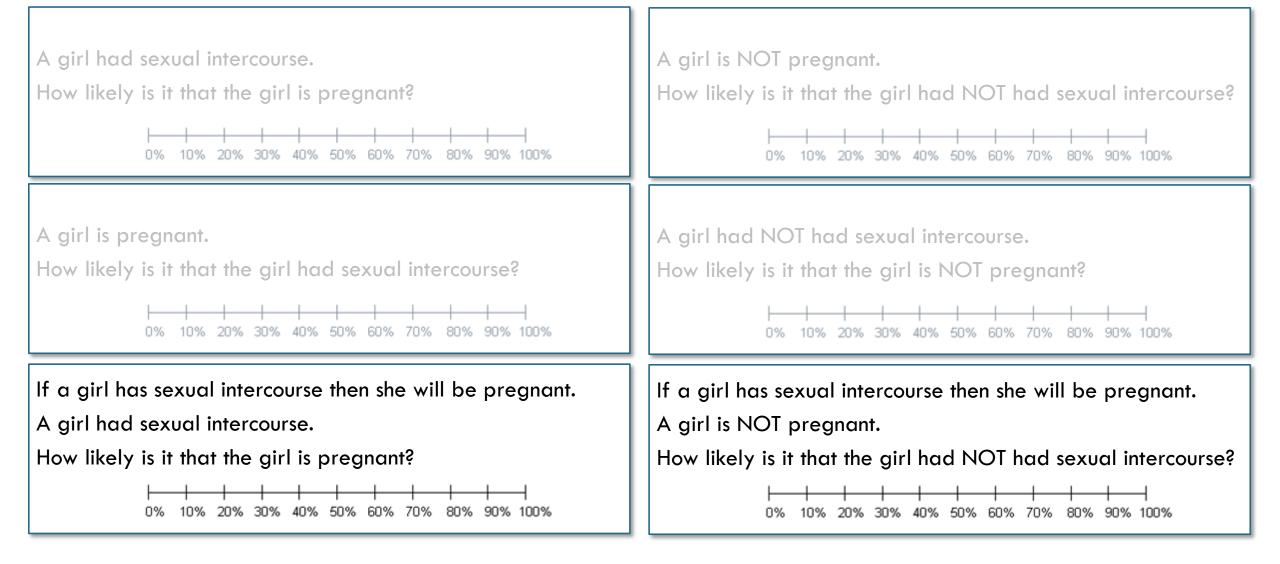
A girl is NOT pregnant.

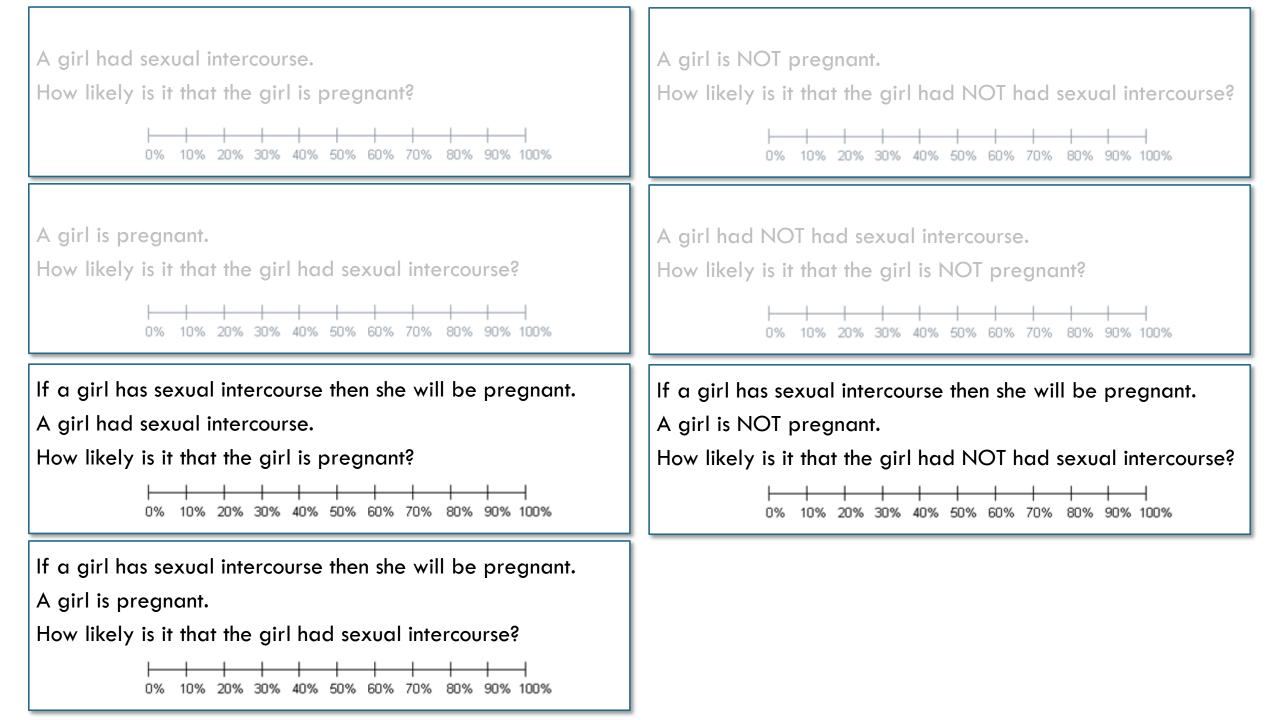
How likely is it that the girl had NOT had sexual intercourse?

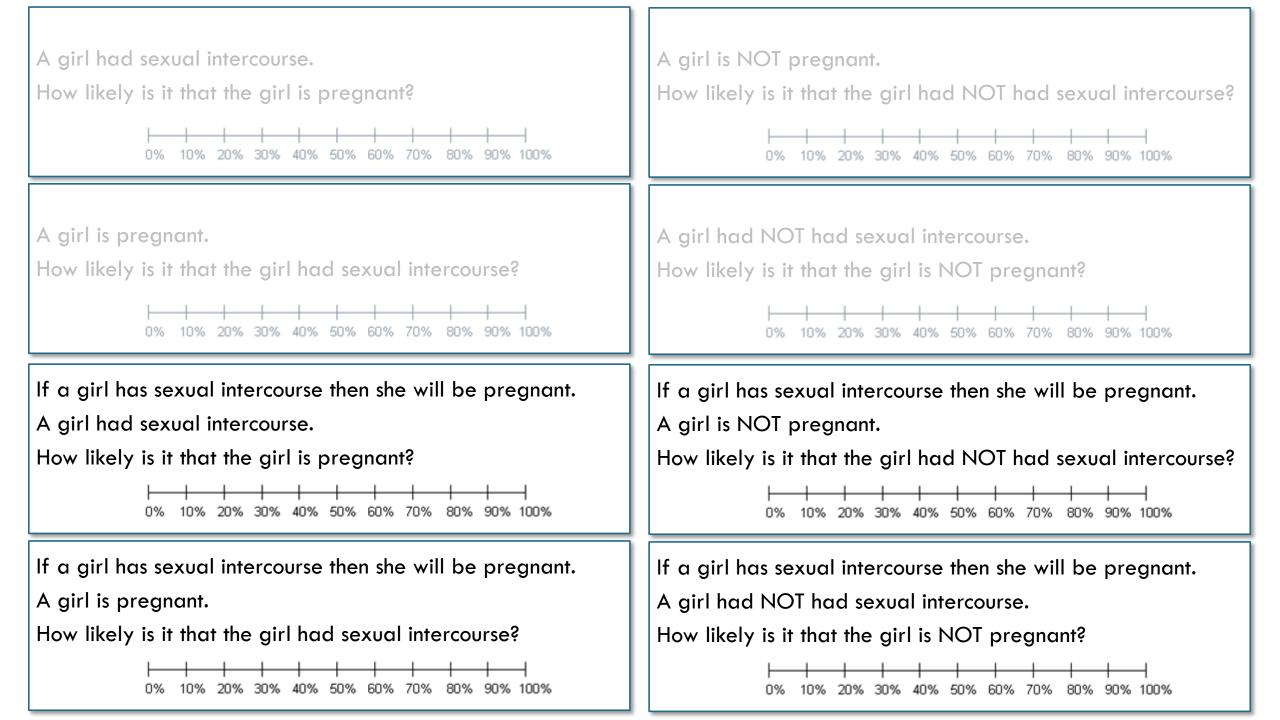


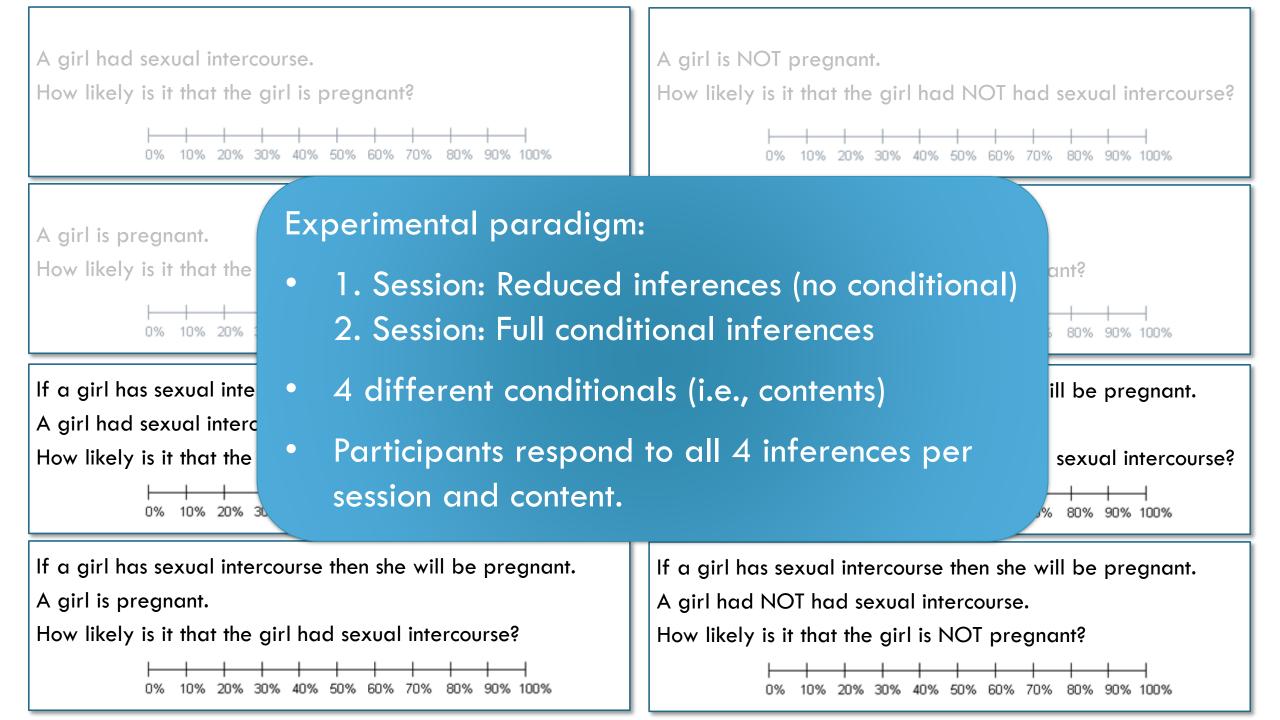












RESULTS

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

few disablers, many alternatives

Coke: If a person drinks a lot of coke then the person will gain weight.

many disablers, many alternatives

Girl: If a girl has sexual intercourse then she will be pregnant.

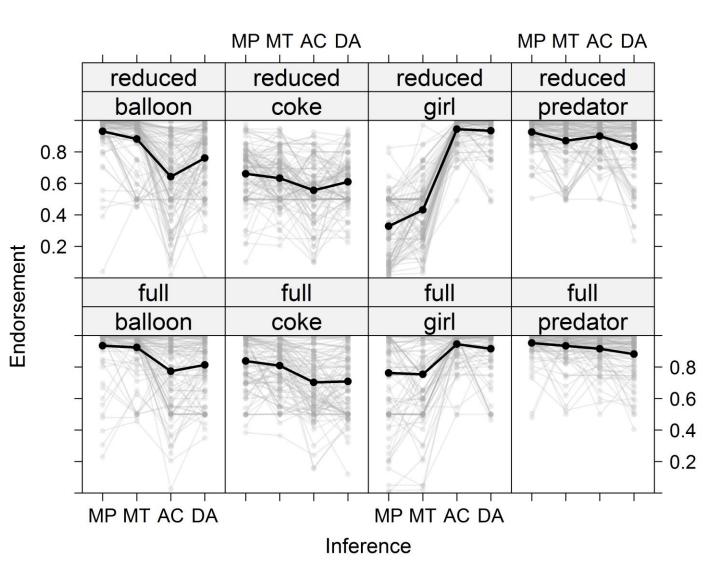
many disablers, few alternatives

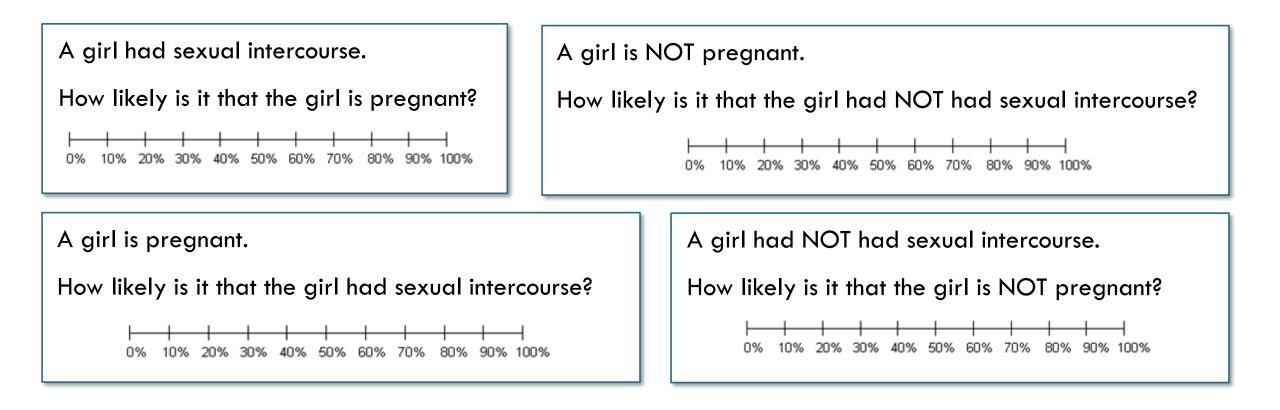
Predator: If a predator is hungry then it will search for prey.

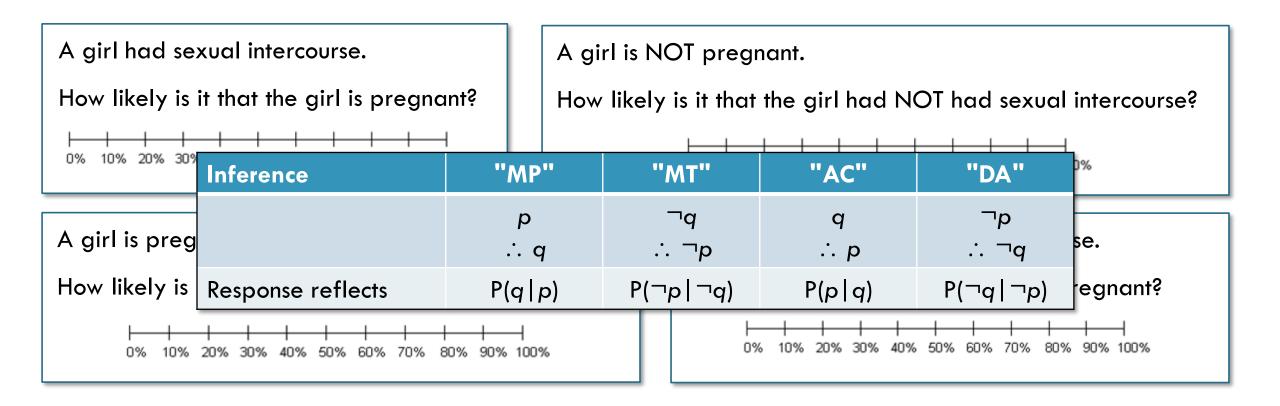
few disablers, few alternatives

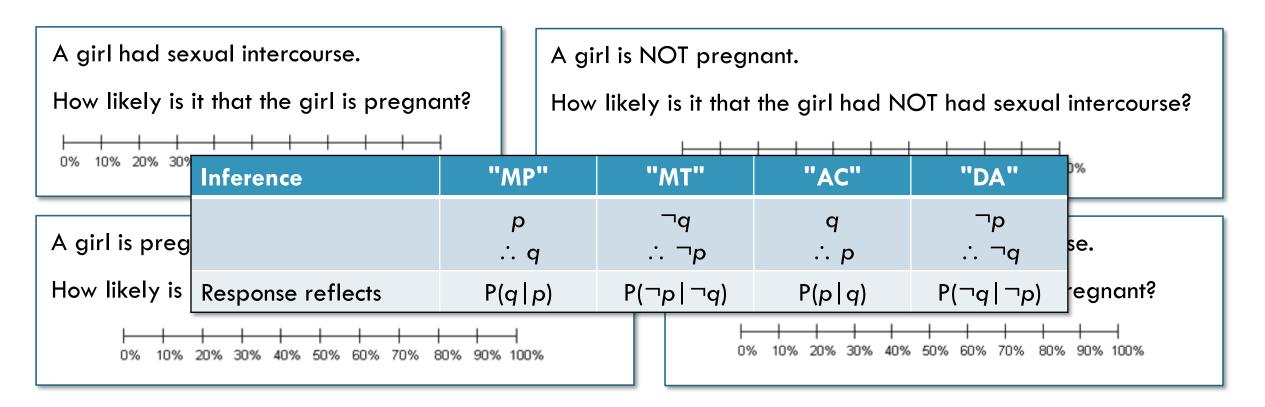
N = 101

Klauer, Beller, & Hütter (2010, Exp. 1) Singmann, Klauer, & Beller (2016, Exp. 1 & 3)









Joint probability distribution $oldsymbol{F}_{p,q}$						
	q	¬q				
р	P(p ∧ q)	P(p ∧ ¬q)				
¬р	$P(\neg p \land q)$	$P(\neg p \land \neg q)$				

3 free parameters

Provides conditional probabilities/predictions:

- $P(MP) = P(q | p) = P(p \land q) / P(p)$
- $P(MT) = P(\neg p | \neg q) = P(\neg p \land \neg q) / P(\neg q)$
- $P(AC) = P(p | q) = P(p \land q) / P(q)$
- $P(DA) = P(\neg q | \neg p) = P(\neg p \land \neg q) / P(\neg p)$

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

HIERARCHICAL MODELING

2 classical approaches for dealing with individual differences:

- complete pooling: ignores individual variability
- no pooling: ignores similarity across participants (e.g., Oaksford, Chater, & Larkin, 2000; Klauer, Beller, & Hütter, 2010; Singmann, Klauer, & Beller, 2016)

Partial pooling principled alternative:

- Individual level parameters are drawn from group-level distributions
- Provides higher precision for parameter estimates (even on the individual level)

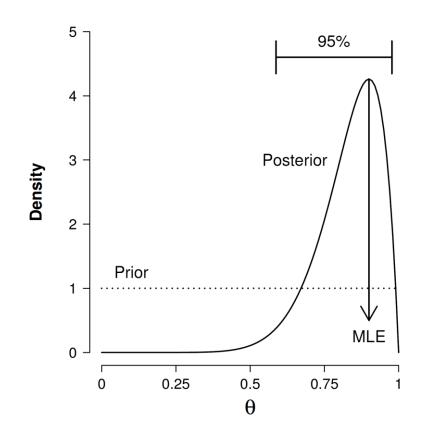


BAYESIAN STATISTICS

Requires likelihood (i.e., no least squares).

Information (uncertainty) regarding parameters expressed via (continuous) probability distributions.

- 1. Prior distributions capture ignorance before data is collected.
- 2. Prior distributions updated in light of data using Bayes' theorem.
- 3. Posterior distributions reflect new state of knowledge.



Ferrari & Cribari-Neto (2004) Simas, Barreto-Souza, & Rocha (2010)

BETA REGRESSION

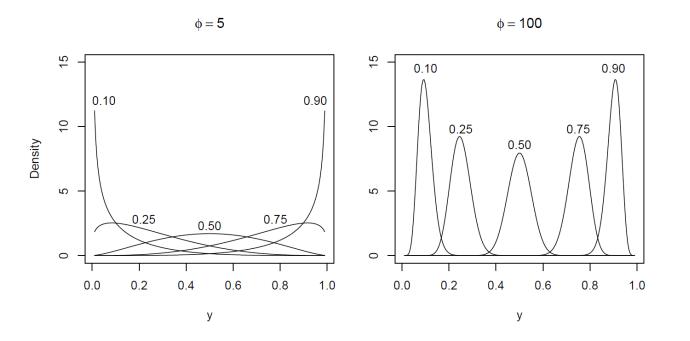
Allows to model data in unit interval (0, 1) using beta distribution.

Instead of shape parameters α and β , uses mean μ and precision ϕ :

•
$$\alpha = \mu \phi$$

$$\bullet \beta = (1 - \mu)\phi$$

Naturally addresses heteroscedasticity: More variation in mid ranges than at the upper and lower end.



HYPERDISTRIBUTION FOR PROBABILITY DISTRIBUTION

Predictions of Bayesian model result from probability distribution $F_{p,q}$.

Oaksford and Chater parameterize $F_{p,q}$ using three parameters:

•
$$a = P(p)$$

•
$$b = P(q)$$

•
$$e = P(not-q | p) = 1 - P(q | p)$$

Not all values of a, b, and e result in proper probability distribution:

• e is bound: $\left[\max\left(\frac{a-b}{a}, 0\right), \min\left(\frac{1-b}{a}, 1\right)\right]$

The joint distribution of a, b, and e cannot be a proper hyper/prior distribution for $F_{p,q}$.

Alternative provided by Dirichlet distribution, which usually has 2 parameters:

• $K \ge 2$, number of categories (integer)

• $\alpha_1, \ldots, \alpha_K$, concentration parameter

Support over K-dimensional vectors that sum to 1 (i.e., (K-1)-dimensional simplex).

Parameterization as in beta-regression possible (e.g., Kemp, Perfors, & Tenenbaum, 2007):

γ: mean of hyperparameter

• ψ : precision of hyperparameter



(simple model)

Data: $E_{\mathrm{red},ijk} \sim \mathrm{Beta}(\alpha_{\mathrm{red},ijk},\beta_{\mathrm{red},ijk})$

Group-level distribution:

 $F_{p,q,jk} \sim \text{Dirichlet}(\gamma_j \times \psi_j)$

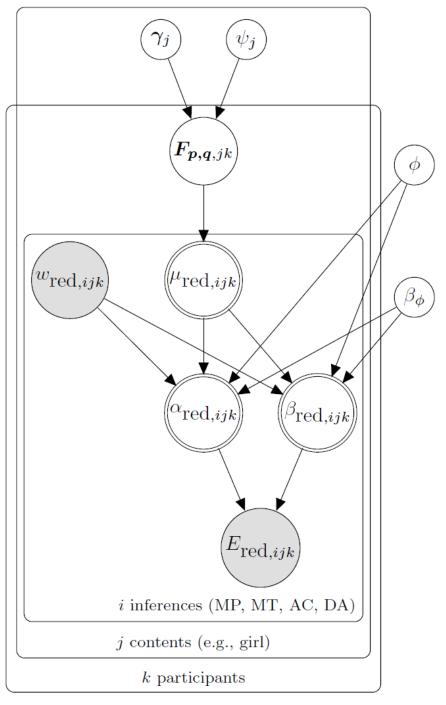
Priors:

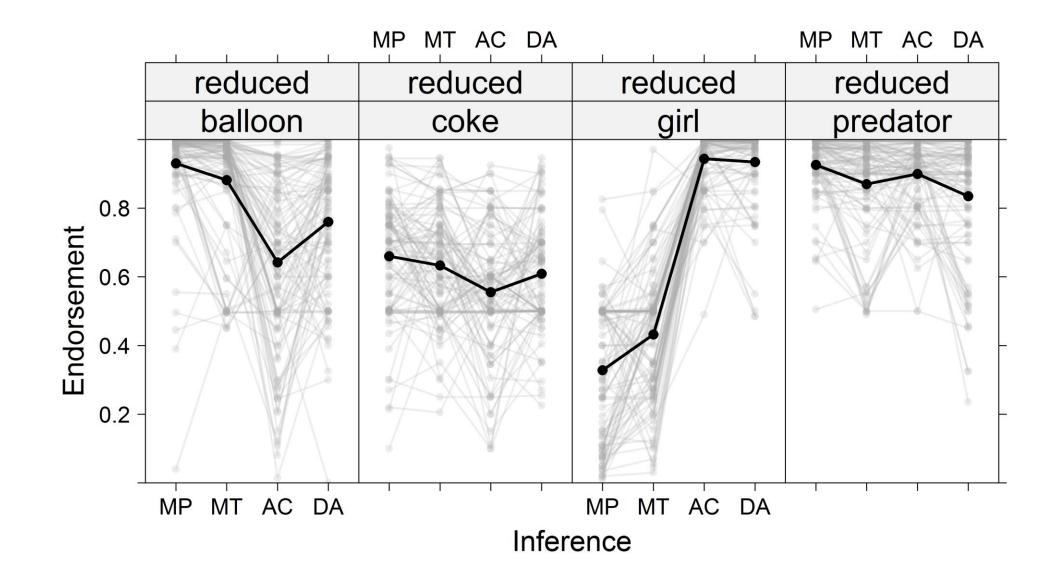
 $\gamma_j \sim \text{Dirichlet}(1)$ $\psi_j \sim \text{Cauchy}^+(1, 25)$ $\phi \sim \text{Cauchy}^+(2, 25)$ $\beta_\phi \sim \text{Cauchy}^+(0, 25)$

Beta regression:

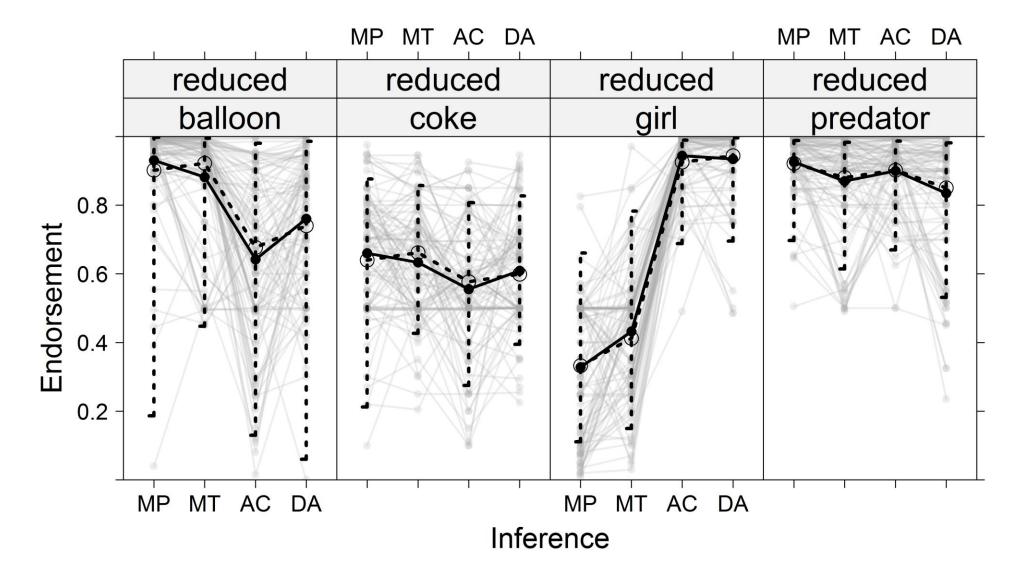
$$\alpha = \mu \times (\phi + \beta_{\phi} w)$$

$$\beta = (1 - \mu) \times (\phi + \beta_{\phi} w)$$

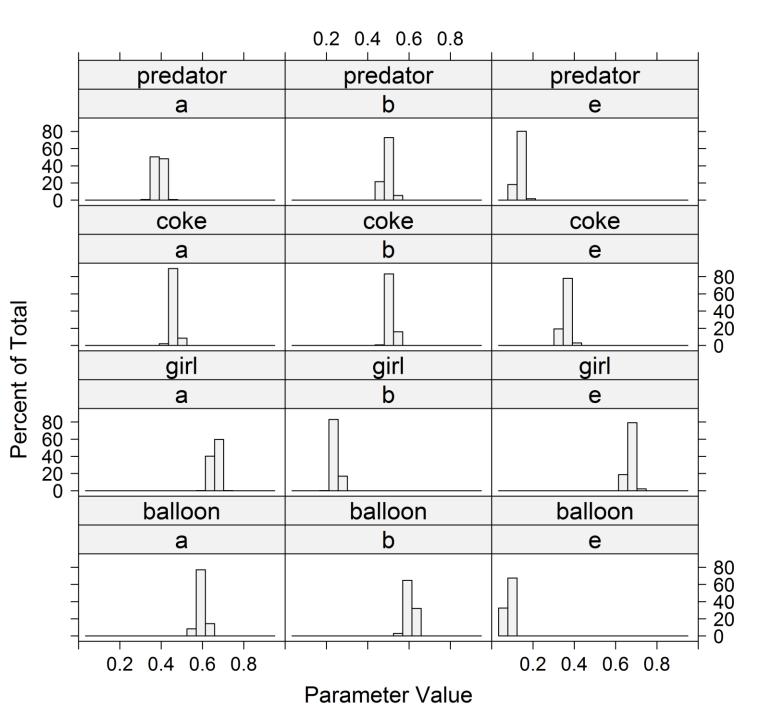




Simple model:



Black error bars: Range of individual level predictions from simple model



Balloon: If a balloon is pricked with a needle then it will pop. • few disablers, many alternatives

Coke: If a person drinks a lot of coke then the person will gain weight.

many disablers, many alternatives

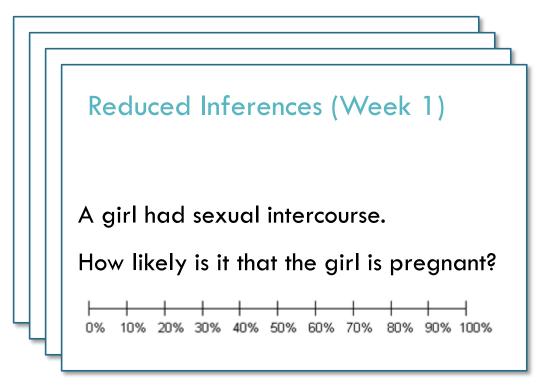
Girl: If a girl has sexual intercourse then she will be pregnant.

many disablers, few alternatives

Predator: If a predator is hungry then it will search for prey.

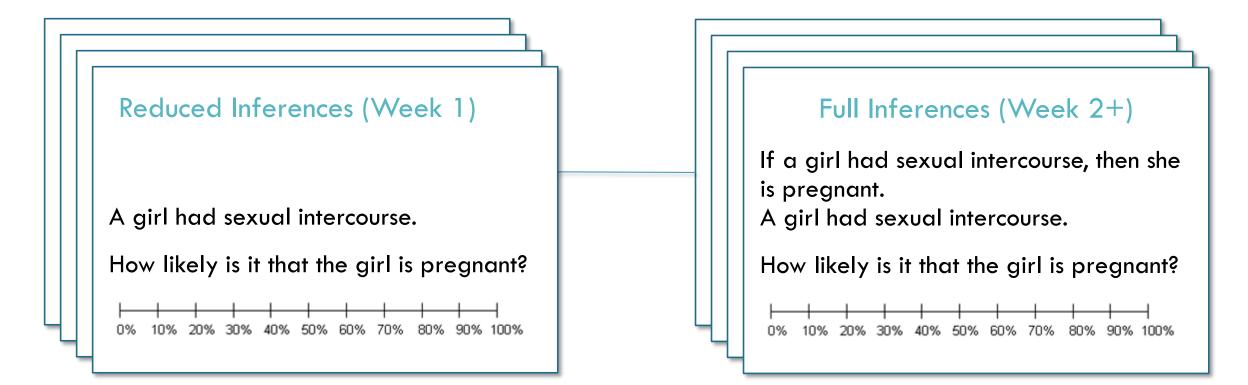
few disablers, few alternatives

EXPERIMENTAL PARADIGM



Klauer, Beller, & Hütter (2010) Singmann, Klauer, & Beller (2016)

EXPERIMENTAL PARADIGM



Klauer, Beller, & Hütter (2010) Singmann, Klauer, & Beller (2016)

EXPERIMENTAL PARADIGM

Inference	"MP"	"МТ"	"AC"	"DA"		<u>`</u>
	р ∴ q	ק ק⊏ ∴	q ∴ p	ק⊂ קר ∴	nces (We	eek 2+)
Response reflects	P(q p)	P(¬p ¬q) P(p q)		p) val intercou	urse, then she
A girl had sexual int				A girl had se	exual intercours	se.
How likely is it that t	Inference		MP	MT	AC	DA
0% 10% 20% 30% 40% 50%			$p \rightarrow q$ p $\therefore q$	$egin{array}{c} p ightarrow q \ eg q \ eg q \ eg . \ eg p \end{array}$	p → q q ∴ p	p ightarrow q $\neg p$ $\therefore \neg q$
	Response re	flects	P(q p)	P(¬p ¬q)	P(p q)	P(¬q ¬p)

Klauer, Beller, & Hütter (2010)

Singmann, Klauer, & Beller (2016)

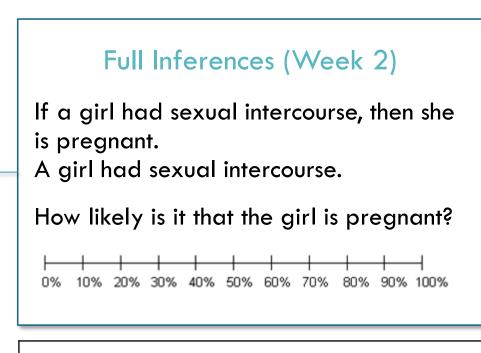


A girl had sexual intercourse.

How likely is it that the girl is pregnant?

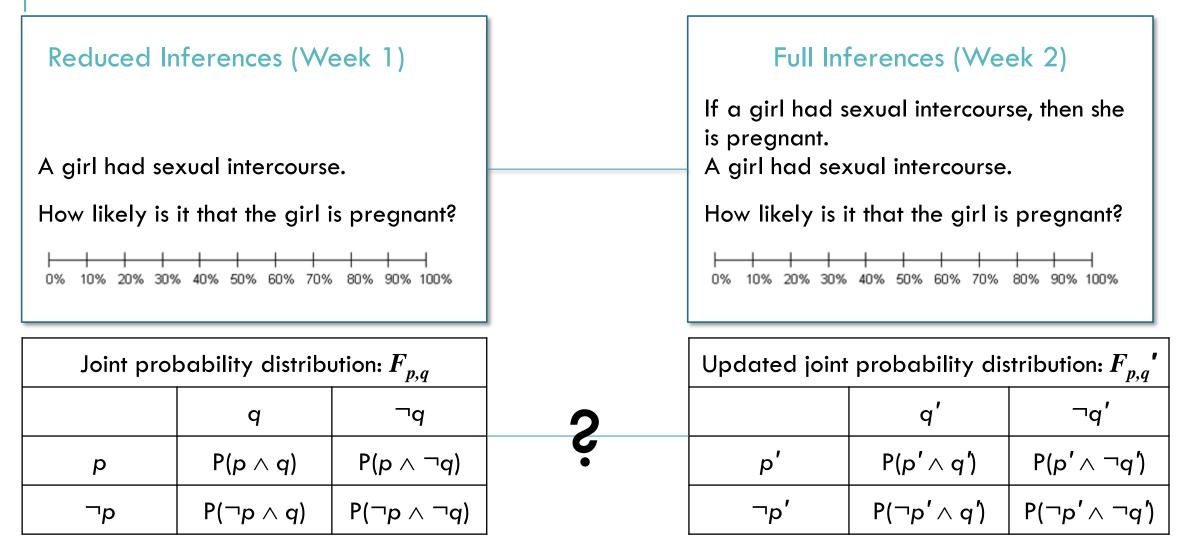
\vdash	+	+	+	+		+	-	+	+	-
0%	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%

Joint probability distribution: $F_{p,q}$							
	q ¬q						
р	$P(p\wedgeq)$	P(p ∧ ¬q)					
¬р	$P(\neg p \land q)$	P(¬p ∧ ¬q)					



Updated joint	int probability distribution: $m{F}_{p,q}$ '					
	q'	¬q'				
p'	P(p' ^ q')	P(p′∧ ¬q′)				
p′	P(¬p'∧q')	P(¬p'∧ ¬q')				





Reduced Inferences (Week 1)

Full Inferences (Week 2)

If a girl had sexual intercourse, then she

Role of conditional in Bayesian models:

- *PROB*: increases probability of conditional, P(q|p) (Oaksford et al., 2000): e' < e
- EX-PROB: increases probability of conditional P_{MP}(q|p) > P_{other}(q|p) (Oaksford & Chater, 2007)
- KL: increases $P(q \mid p)$ & Kullback-Leibler distance between $F_{p,q}$ and $F_{p,q}'$ is minimal (Hartmann & Rafiee Rad, 2012)

Consequence of updating: Effect is content specific.

q	P(p ∧ q)	P(p ∧ ¬q)	ē	p	P(p'∧q')	P(p'∧ ¬q')
٦p	P(¬p ∧ q)	Р(¬р∧¬q)		¬p'	P(¬p'∧q')	P(¬p'∧ ¬q')

Reduced Inferences (Week 1)

Full Inferences (Week 2)

If a girl had sexual intercourse, then she

Role of conditional in Bayesian models:

- PROB: increases probability of conditional, P(q | p) (Oaksford et al., 2000): e' < e
- EX-PROB: increases probability of conditional P_{MP}(q|p) > P_{other}(q|p) (Oaksford & Chater, 2007)
- KL: increases $P(q \mid p)$ & Kullback-Leibler distance between $F_{p,q}$ and $F_{p,q}$ ' is minimal (Hartmann & Rafiee Rad, 2012)

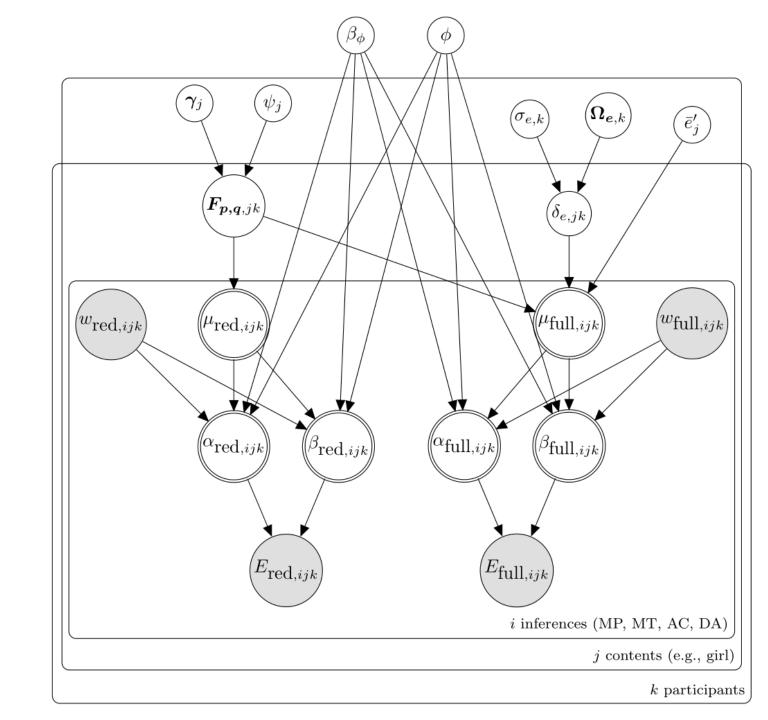
Consequence of updating: Effect is content specific.

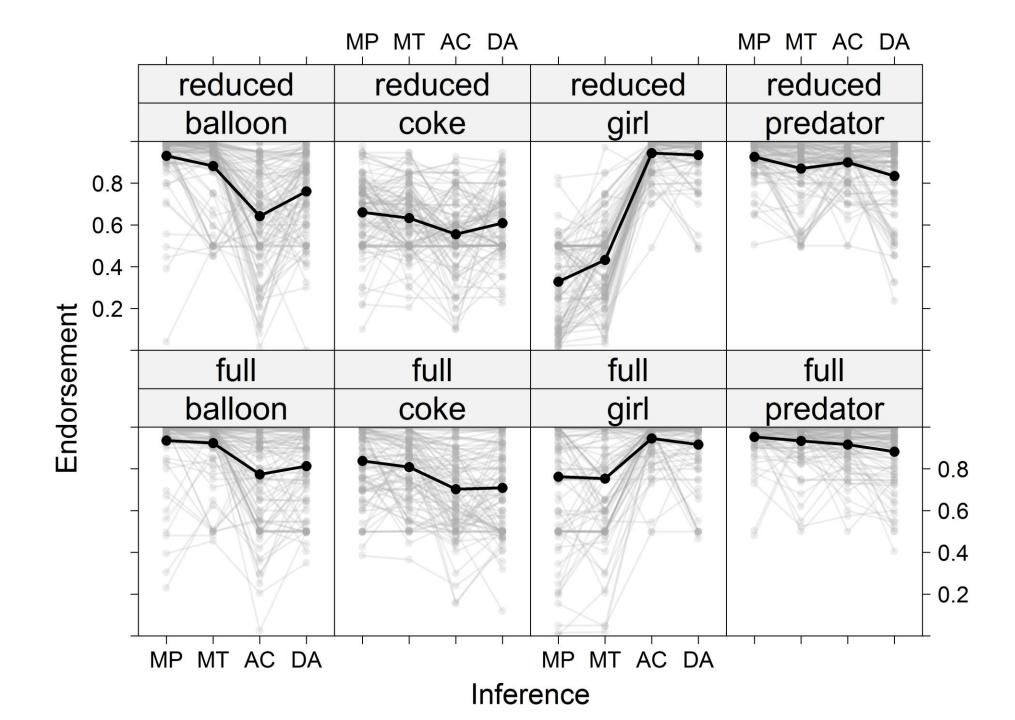
р	P(p ∧ q)	P(p ∧ ¬q)	ē	p	P(p'∧q')	P(p'∧ ¬q')
$\neg p$	P(¬p ∧ q)	Р(¬р∧¬q)		¬p'	P(¬p'∧q')	P(¬p′∧ ¬q′)

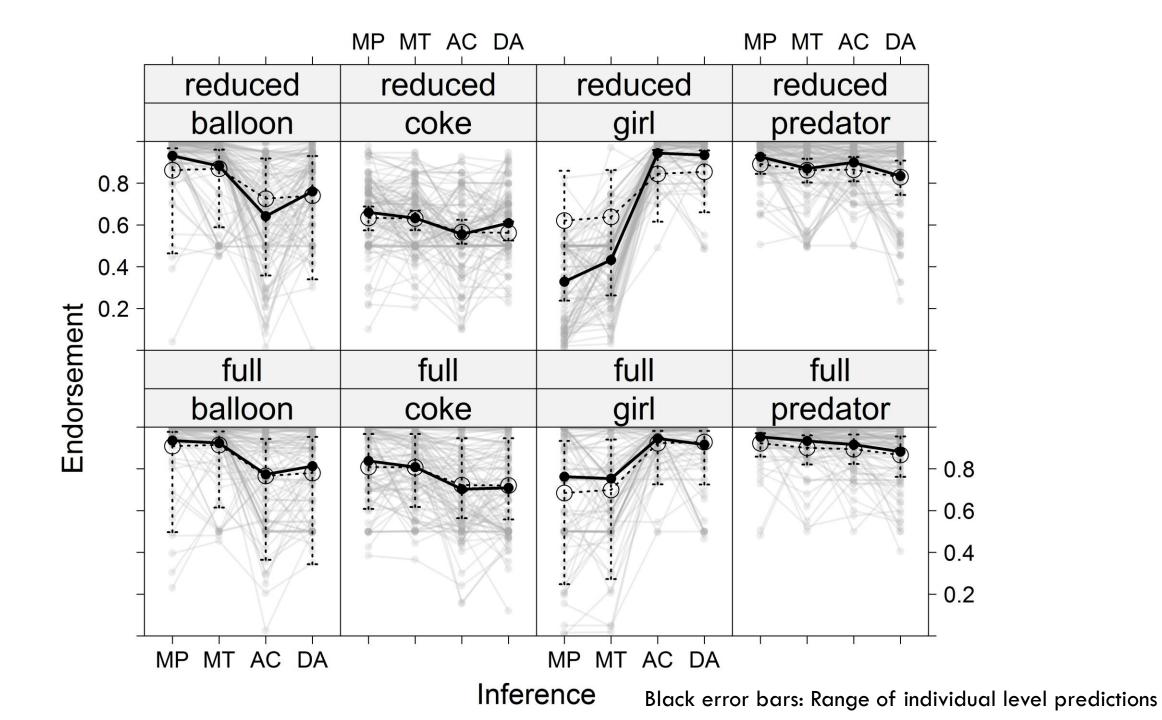
 $E_{\text{red},ijk} \sim \text{Beta}(\alpha_{\text{red},ijk},\beta_{\text{red},ijk})$ $E_{\text{full},ijk} \sim \text{Beta}(\alpha_{\text{full},ijk},\beta_{\text{full},ijk})$

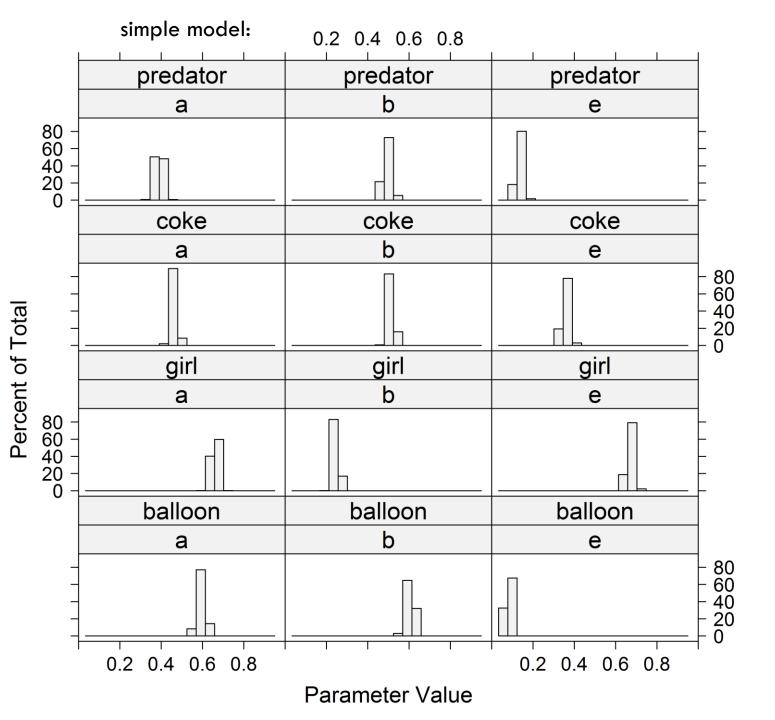
- $m{F}_{m{p},m{q},jk} \sim ext{Dirichlet}(m{\gamma}_j imes \psi_j) \ m{\delta}_e \sim ext{MvNormal}(m{0}, m{\Sigma}_{m{e}})$
 - $$\begin{split} \boldsymbol{\Sigma}_e &= \boldsymbol{\sigma}_e \boldsymbol{\Omega}_e \\ \boldsymbol{\Omega}_e &\sim \text{LKJ}(1) \\ \boldsymbol{\sigma}_{e,k} &\sim \text{Cauchy}^+(0,4) \\ \bar{e}'_j &\sim \text{Normal}(0,1) \end{split}$$

$$l_{ij} = \max\left(\frac{a_{ij} - b_{ij}}{a_{ij}}, 0\right)$$
$$e'_{ij} = e_{ij} - \left((e_{ij} - l_{ij}) \times \Phi(\bar{e}'_j + \delta_{e,jk})\right)$$









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

few disablers, many alternatives

Coke: If a person drinks a lot of coke then the person will gain weight.

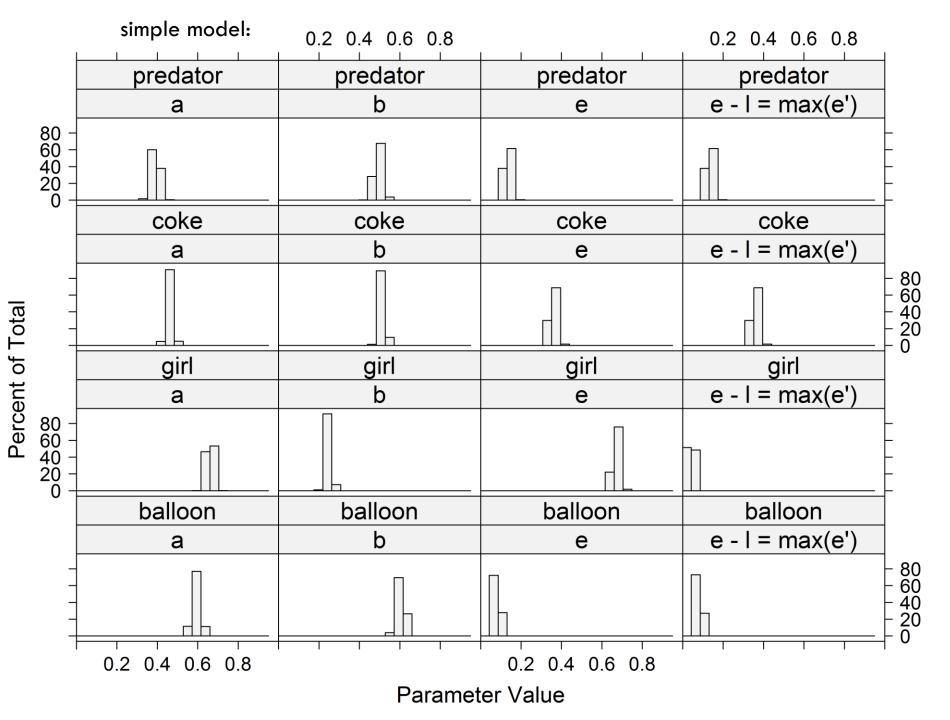
many disablers, many alternatives

Girl: If a girl has sexual intercourse then she will be pregnant.

many disablers, few alternatives

Predator: If a predator is hungry then it will search for prey.

few disablers, few alternatives



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

few disablers, many alternatives

Coke: If a person drinks a lot of coke then the person will gain weight.

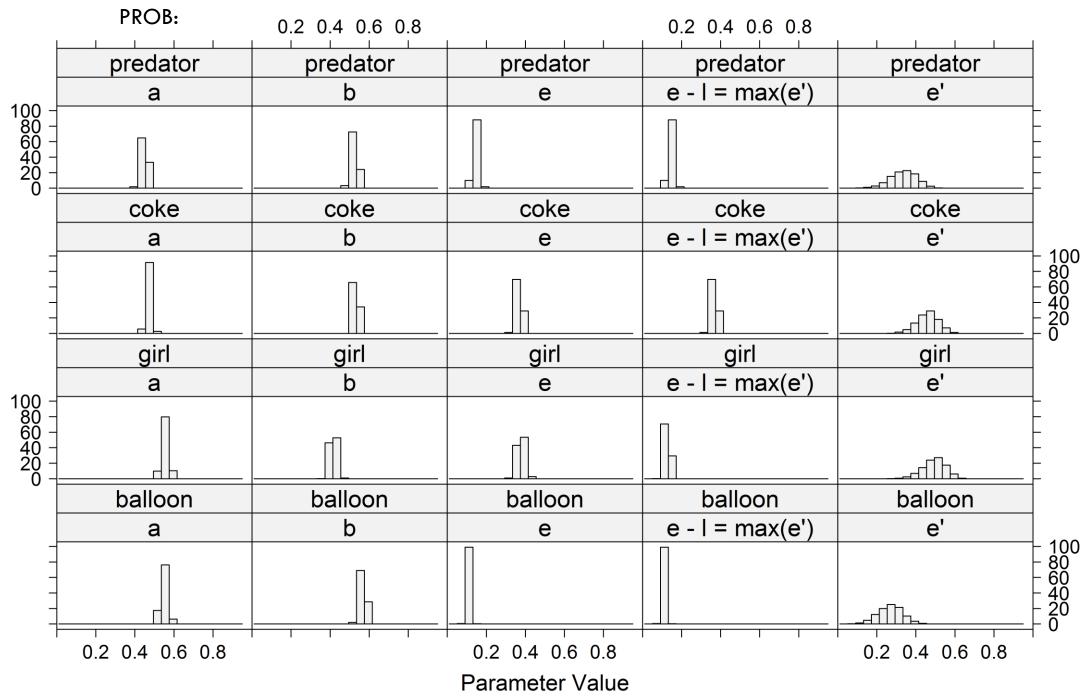
many disablers, many alternatives

Girl: If a girl has sexual intercourse then she will be pregnant.

many disablers, few alternatives

Predator: If a predator is hungry then it will search for prey.

few disablers, few alternatives



Percent of Total

Reduced Inferences (Week 1)

Full Inferences (Week 2)

If a girl had sexual intercourse, then she

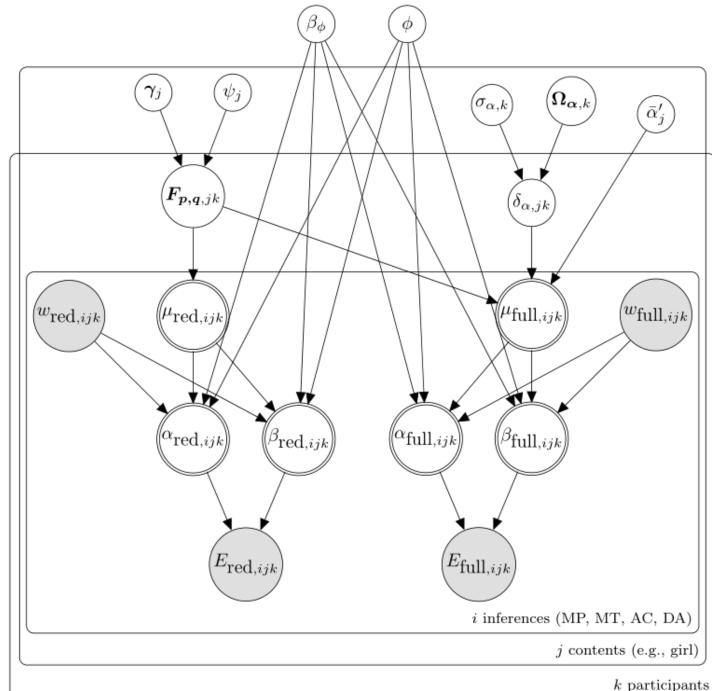
Role of conditional in Bayesian models:

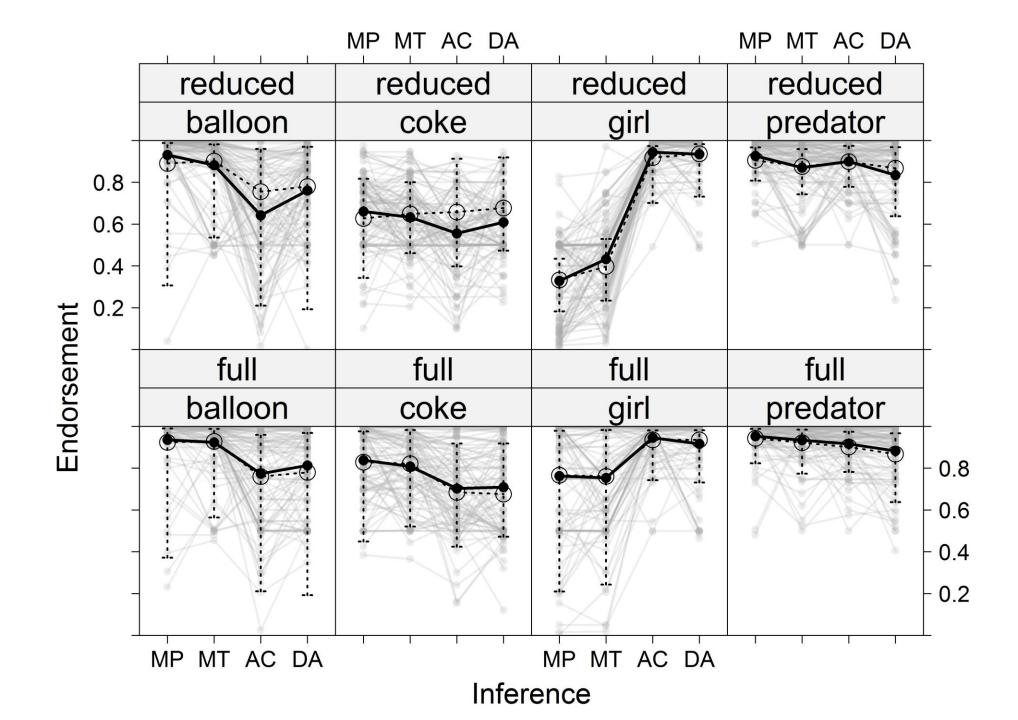
- PROB: increases probability of conditional, P(q|p) (Oaksford et al., 2000): e' < e
- EX-PROB: increases probability of conditional PMP(q|p) > Pother(q|p) (Oaksford & Chater, 2007)
- KL: increases $P(q \mid p)$ & Kullback-Leibler distance between $F_{p,q}$ and $F_{p,q}$ ' is minimal (Hartmann & Rafiee Rad, 2012)

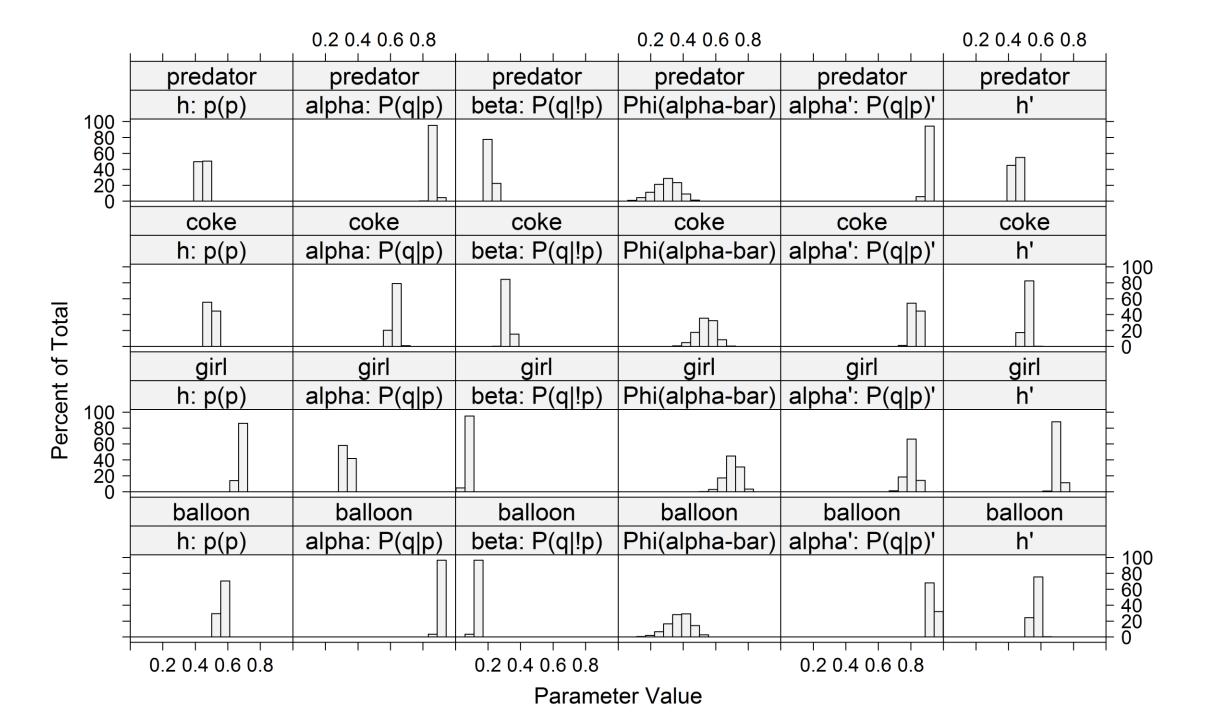
Consequence of updating: Effect is content specific.

p	P(p ∧ q)	P(p ∧ ¬q)	ē	p	P(p'∧q')	P(p'∧ ¬q')
٦p	P(¬p ∧ q)	P(¬p ∧ ¬q)		¬p'	P(¬p'∧q')	P(¬p'∧ ¬q')

Kullback-Leibler (KL) Modell Hartmann & Rafiee Rad (2012) Singmann, Klauer, & Beller (2016, Exp. 1 & 3) Parameterization of $F_{p,q}$: For $F_{p,q}$ ': • h = P(q)• $\alpha' > \alpha$ • $\alpha = P(q|p)$ Kullback-Leibler divergence between • $\beta = P(q|\neg p)$ $F_{p,q}$ and $F_{p,q}$ 'minimal. $E_{\mathrm{red},ijk} \sim \mathrm{Beta}(\alpha_{\mathrm{red},ijk},\beta_{\mathrm{red},ijk})$ $(w_{\mathrm{red},ijk})$ $E_{\text{full},ijk} \sim \text{Beta}(\alpha_{\text{full},ijk}, \beta_{\text{full},ijk})$ $F_{p,q,jk} \sim \text{Dirichlet}(\gamma_j \times \psi_j)$ $\boldsymbol{\delta}_{\alpha} \sim \operatorname{MvNormal}(\mathbf{0}, \boldsymbol{\Sigma}_{\boldsymbol{\alpha}})$ $\Sigma_{\alpha} = \sigma_{\alpha} \Omega_{\alpha}$ $\Omega_{\alpha} \sim \text{LKJ}(1)$ $\sigma_{\alpha,k} \sim \text{Cauchy}^+(0,4)$ $f_i \sim \text{Normal}(0,1)$ $\alpha_{ij}' = \alpha_{ij} + (1 - \alpha_{ij}) \times \Phi(\bar{\alpha}_j' + \delta_{\alpha,jk}))$



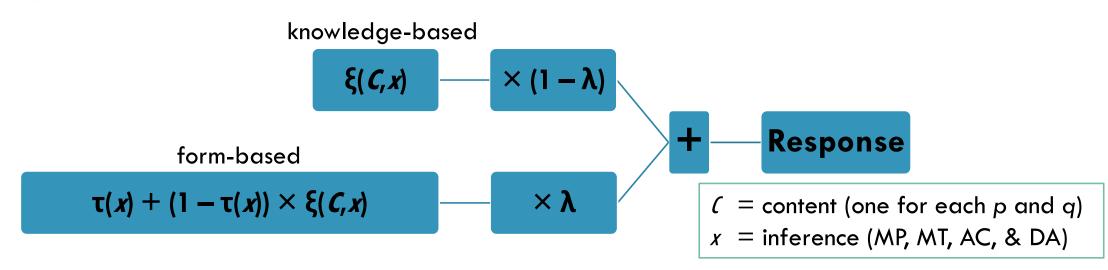


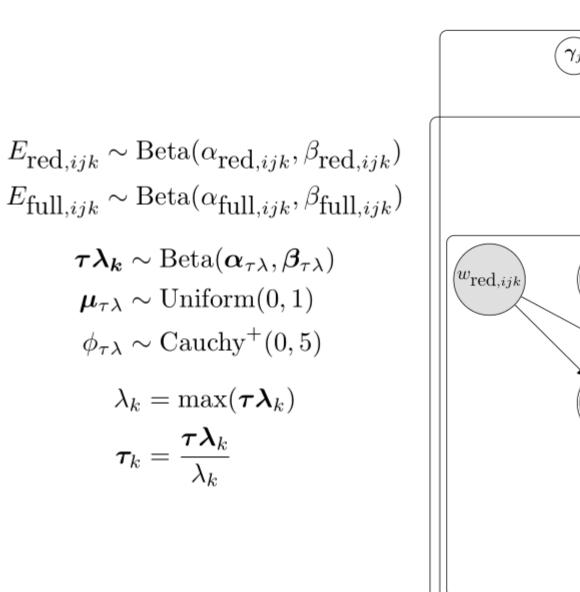


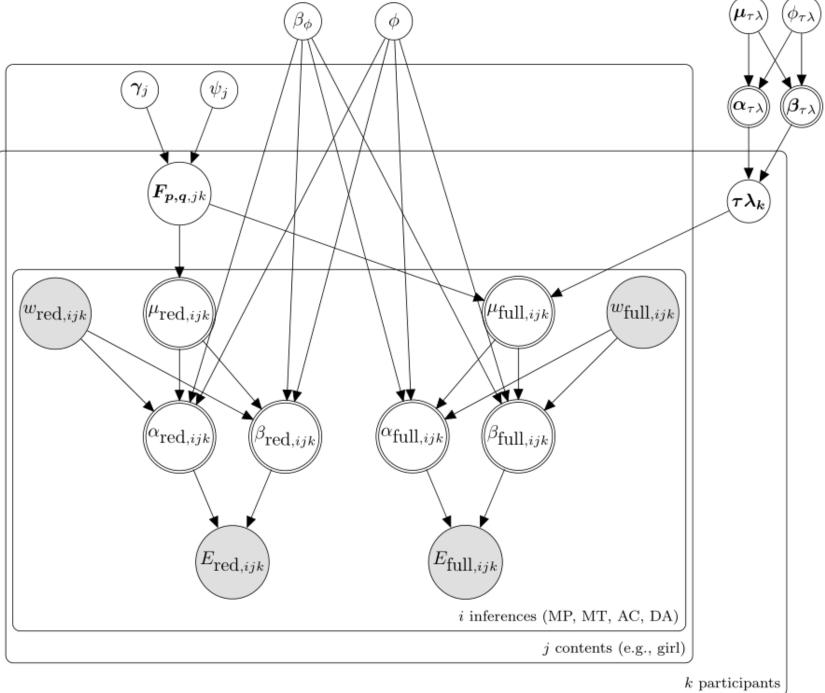
DUAL-SOURCE MODEL (DSM)

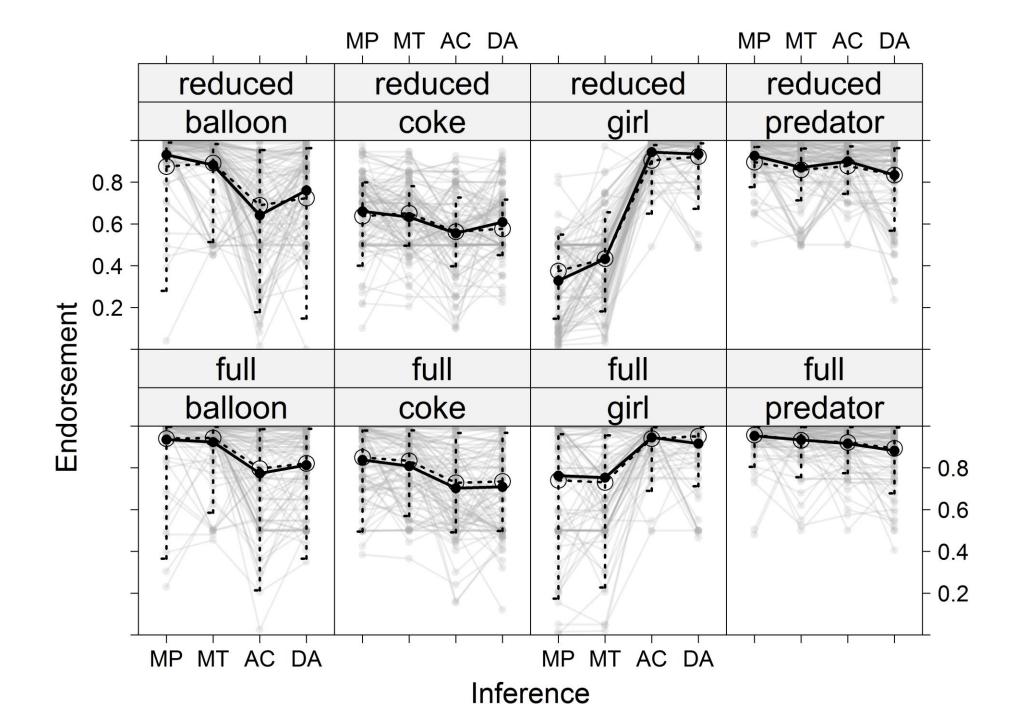
Klauer, Beller, & Hütter (2010, Exp. 1) Singmann, Klauer, & Beller (2016, Exp. 1 & 3)

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



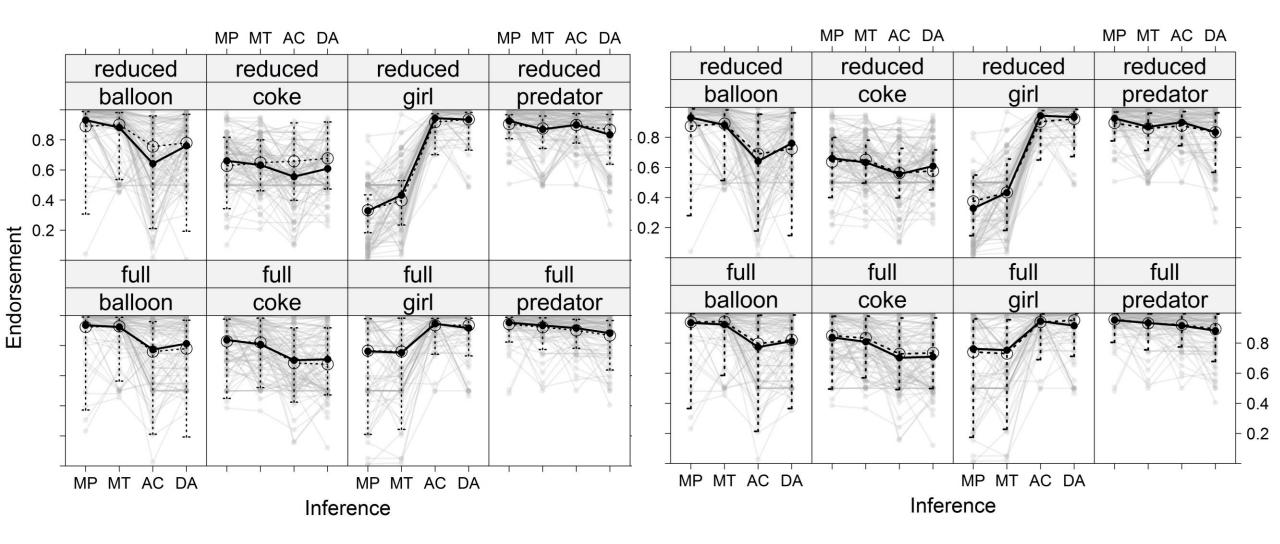






KL Model:

DSM:



SUMMARY: HIERARCHICAL BAYESIAN IMPLEMENTATION OF BAYESIAN MODELS OF REASONING

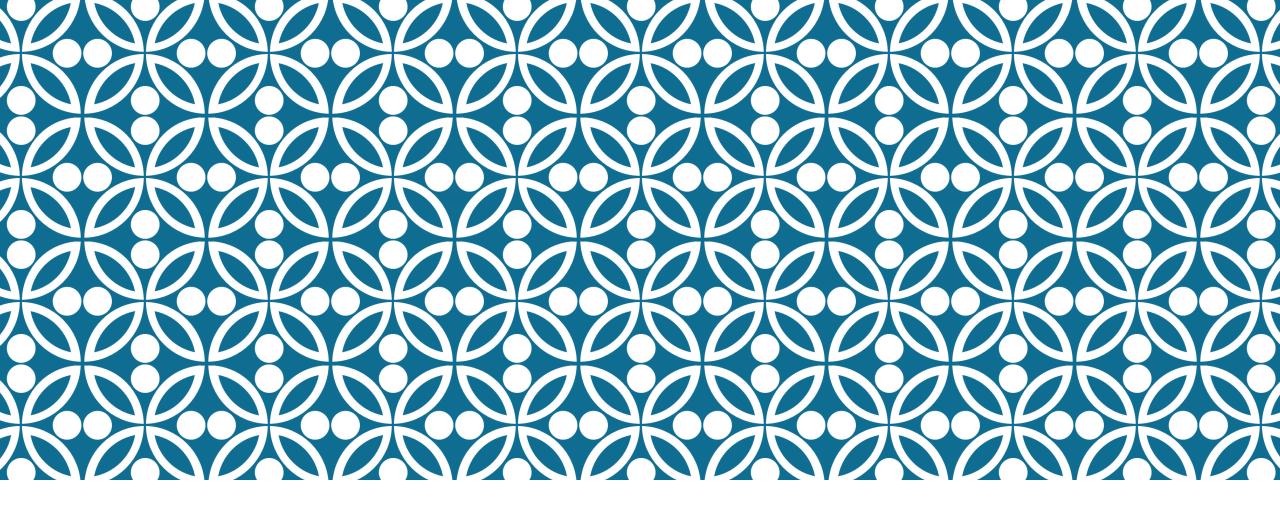
Bayesian statistics offer:

- Principled approach to model individual differences
- Allows investigation of individual level and group-level parameters
- Provides additional information (e.g., precision of probability distribution estimates, correltaion among individual parameters)

For inferences without conditional (i.e., purely knowledge) a simple Bayesian model provides good account.

Learning a conditional can be modeled with:

- Bayesian model that assumes unconstained updating of P(q|p) and KL minimization (Hartmann & Rafiee Rad, 2012).
- Dual-Source Model (Klauer et al., 2010; Singmann et al., 2016), which assumes individuals combine background knowledge with the subjective probability with which they see a specific inference as logically warranted.



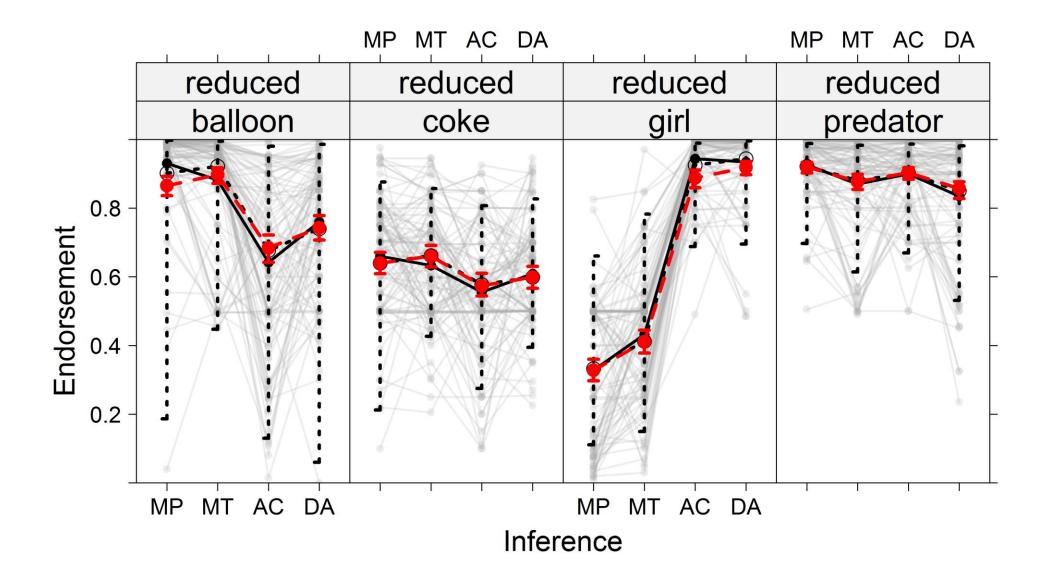
THAT WAS ALL

$F_{p,q}$: If a balloon is pricked with a needle then it will pop.			$F_{p,q}$: If a person drinks a lot of coke then the person will gain weight.		
<i>ψ_j</i> : 15 [27]	q	¬q	ψ _j : 58 [250]	q	$\neg q$
р	.36	.06	p	.29	.17
٦p	.16	.42	¬р	.23	.31

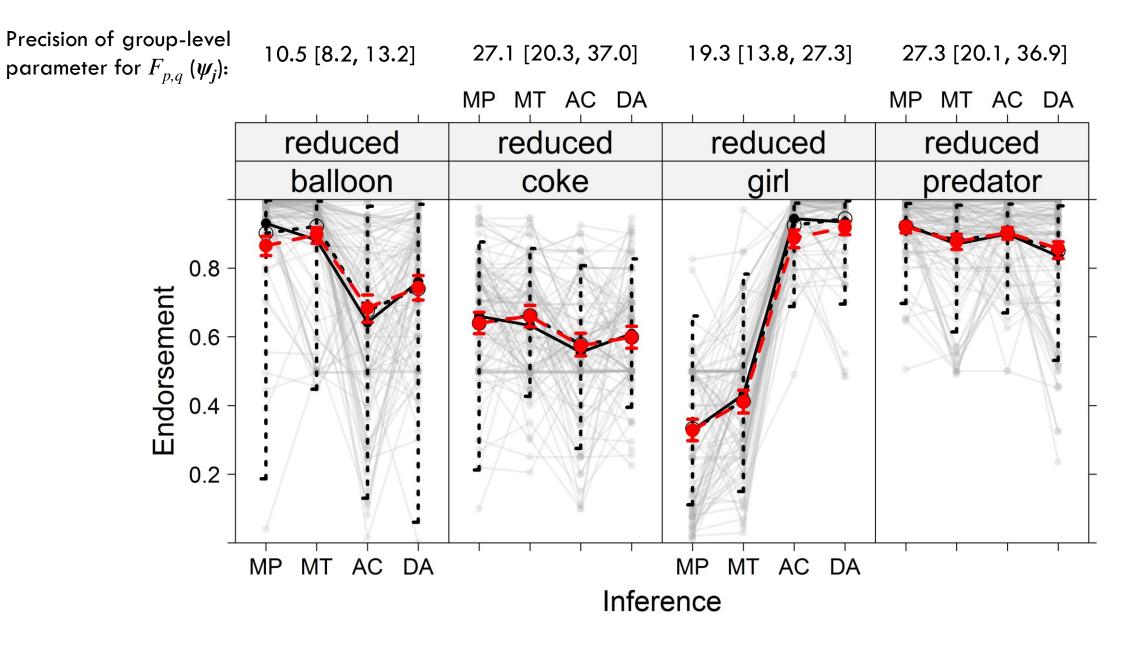
$F_{p,q}$: If a girl has sexual intercourse then she will be pregnant.				
ψ _j : 33 [21]	q	$\neg q$		
р	.24 [.35]	.41 [.21]		
٩٢	.03 [.07]	.31 [.37]		

$F_{p,q}$: If a predator is hungry then it will search for prey.				
ψ _j : 46 [130]	q	¬q		
q	.51	.06		
¬р	.07	.36		

$F_{p,q}$: If a balloon is pricked with a needle then it will pop.				$F_{p,q}$: If a person drinks a lot of coke then the person will gain weight.		
ψ _j : 15 [27]	q	٦q		<i>ψ_j</i> : 58 [250]	q	Γq
q	.36	.06		p	.29	.17
¬р	.16	.42		¬р	.23	.31
Precision of group-level 10.5 [8.2, 13.2] parameter for $F_{p,q}(\psi_j)$, initial model: 19.3 [13.8, 27.3]			27.1 [20.3, 37.0] 27.3 [20.1, 36.9]			
$F_{p,q}$: If a girl has sexual intercourse then she will be pregnant.				$F_{p,q}$: If a predator is hungry then it will search for prey.		
ψ _j : 33 [21]	q	٦q		<i>ψ_j</i> : 46 [130]	q	$\neg q$
q	.24 [.35]	.41 [.21]		q	.51	.06
٦р	.03 [.07]	.31 [.37]		¬р	.07	.36



Black error bars: Range of individual level predictions from simple model



Black error bars: Range of individual level predictions from simple model