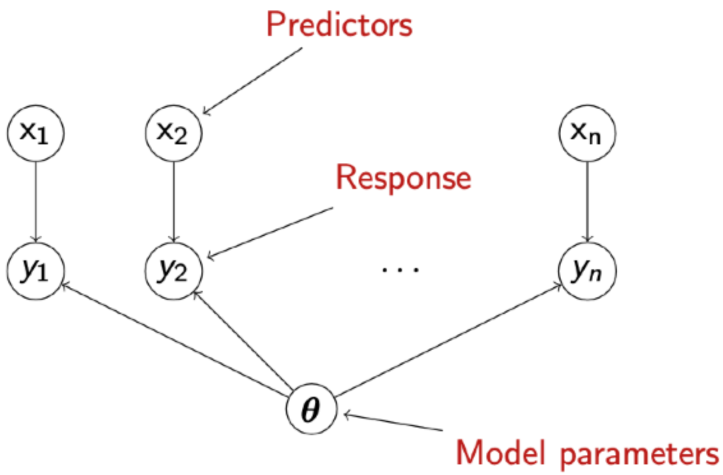
Bayes Net plate notation

Bayes Net plate notation

- Here is our simple (single-level, non-hierarchical) Bayes net for regression:

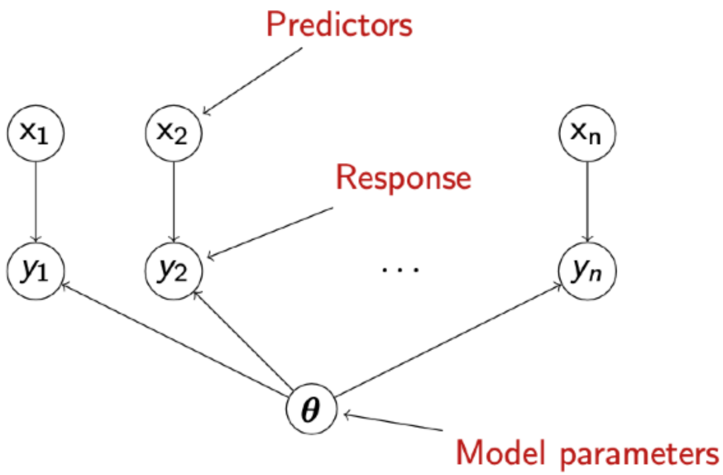
Bayes Net plate notation

- Here is our simple (single-level, non-hierarchical) Bayes net for regression:



Bayes Net plate notation

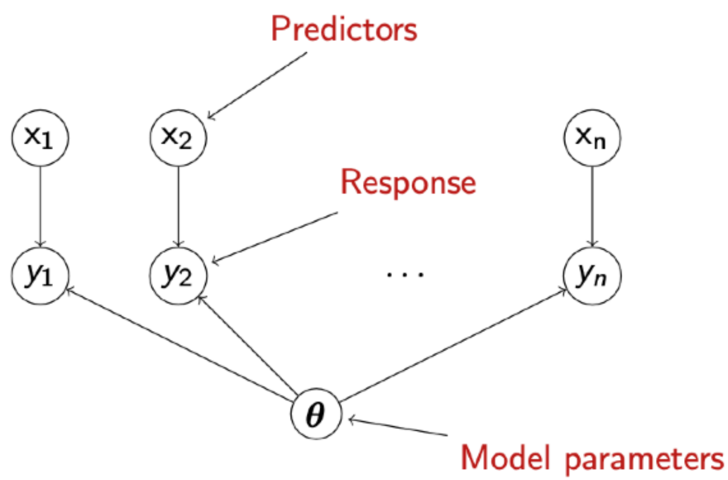
- Here is our simple (single-level, non-hierarchical) Bayes net for regression:



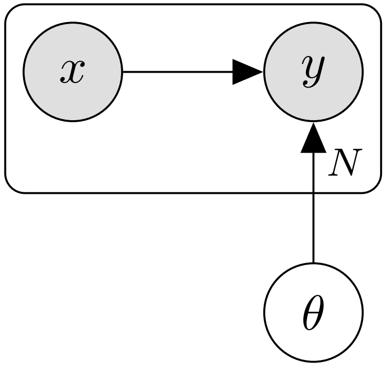
- Here is a version of it using **plate notation**:

Bayes Net plate notation

- Here is our simple (single-level, non-hierarchical) Bayes net for regression:

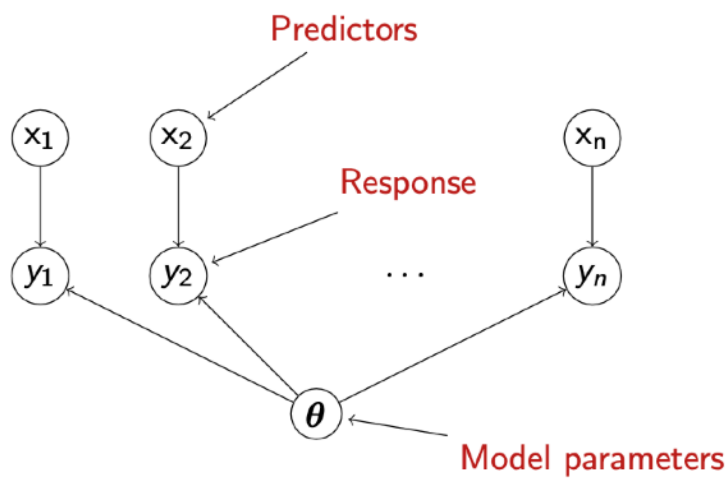


- Here is a version of it using **plate notation**:

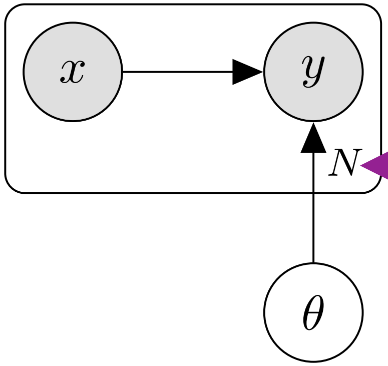


Bayes Net plate notation

- Here is our simple (single-level, non-hierarchical) Bayes net for regression:



- Here is a version of it using **plate notation**:



N instances, with θ parameterizing the distribution of y given x in the same way for each instance

Plate notation for hierarchical models

- Our multi-level/mixed-effects/ hierarchical model (with one level of hierarchy):

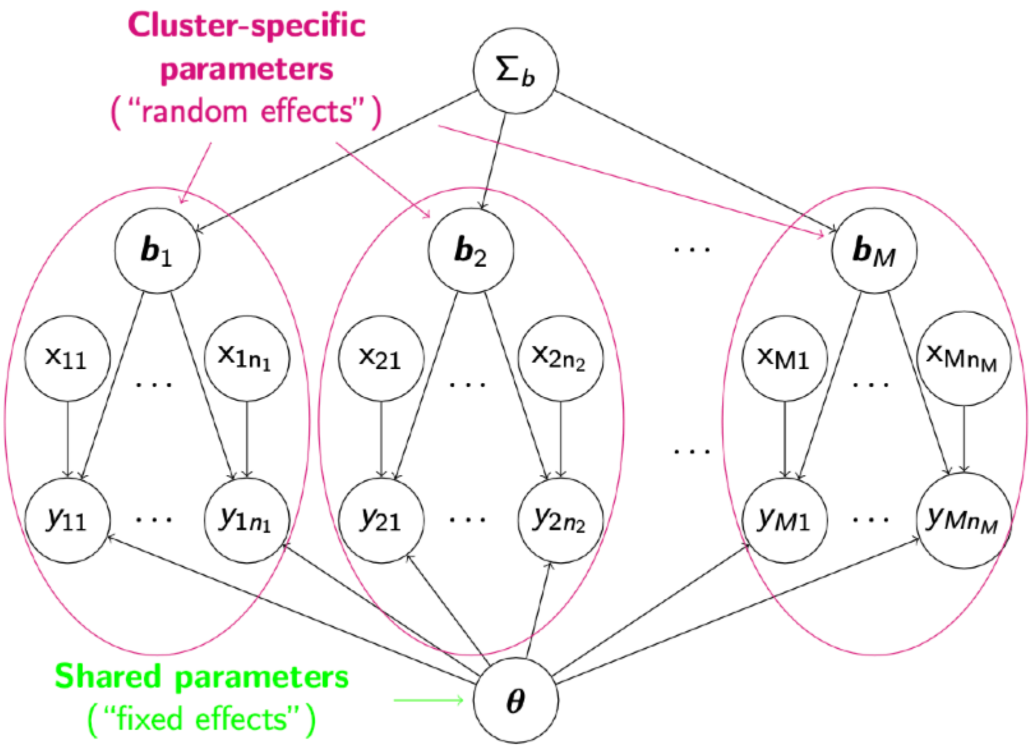
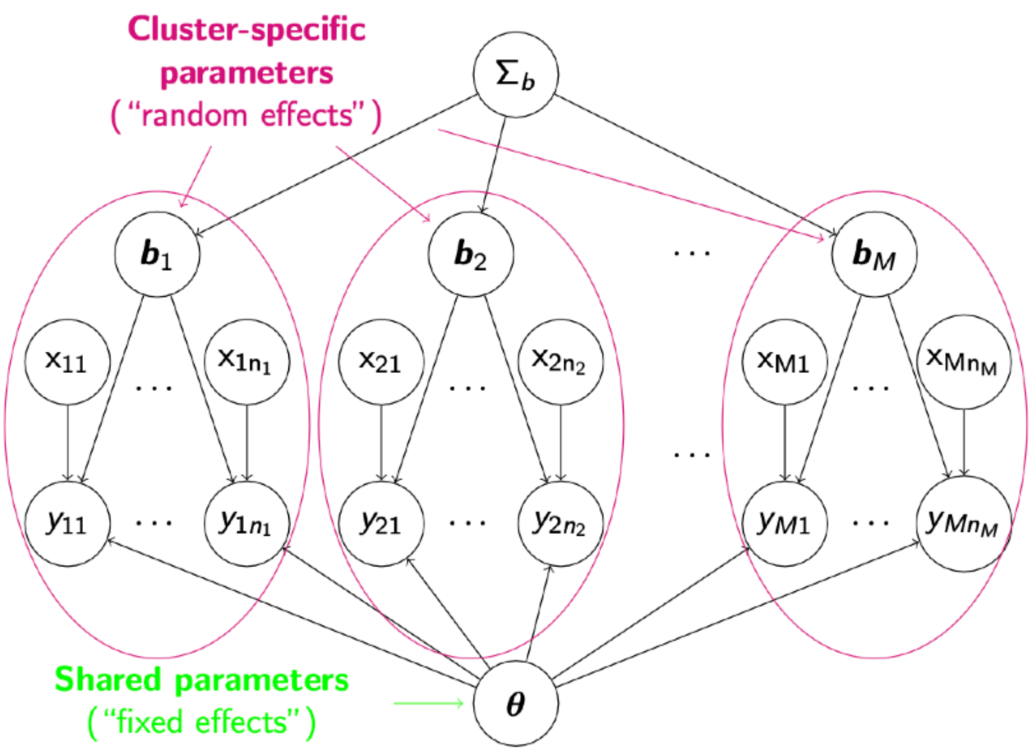
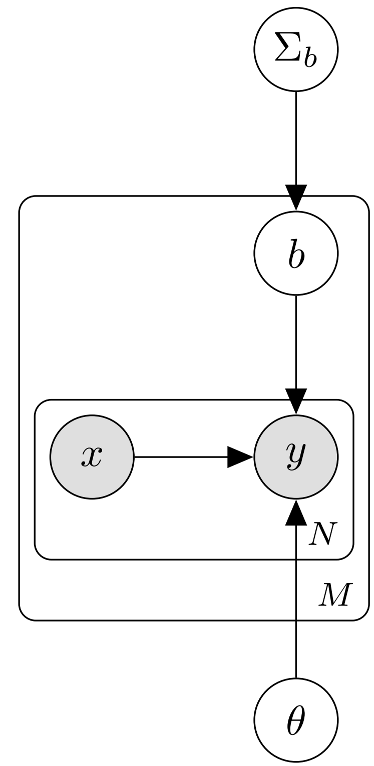


Plate notation for hierarchical models

- Our multi-level/mixed-effects/ hierarchical model (with one level of hierarchy):

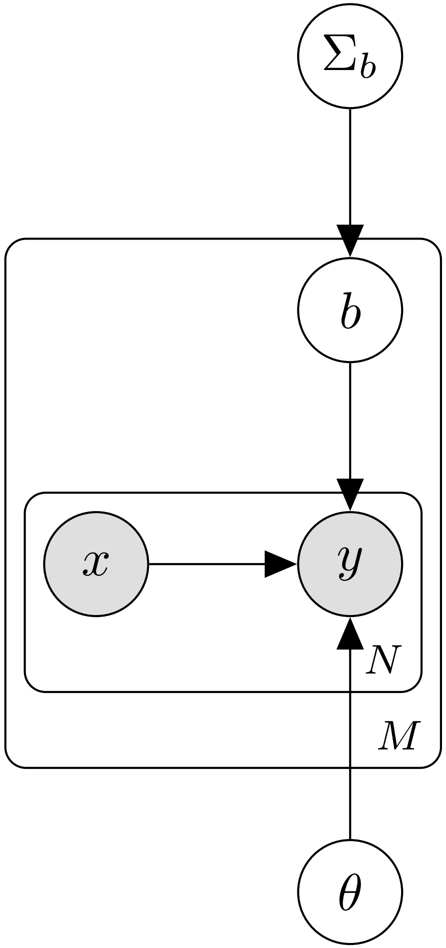


- The same model, using plate notation:



Compact mixed-effects model notation

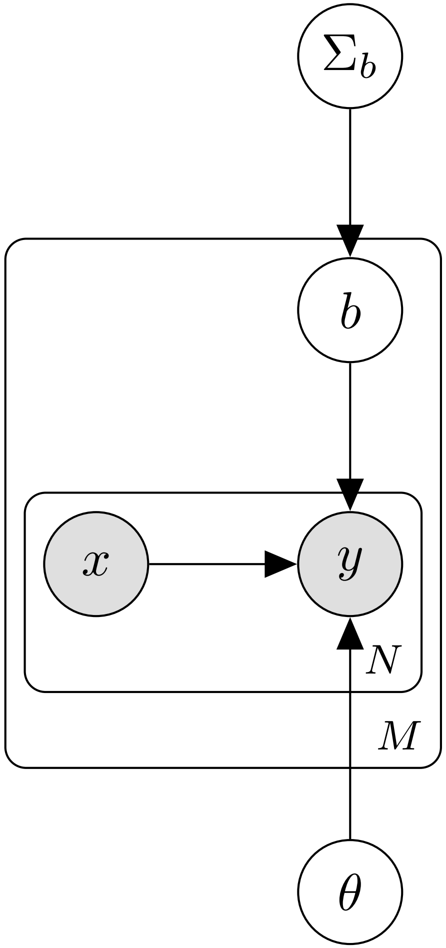
- Multi-level GLM assumptions:



$$\theta = \langle \beta, \sigma^2 \rangle$$

Compact mixed-effects model notation

- Multi-level GLM assumptions:

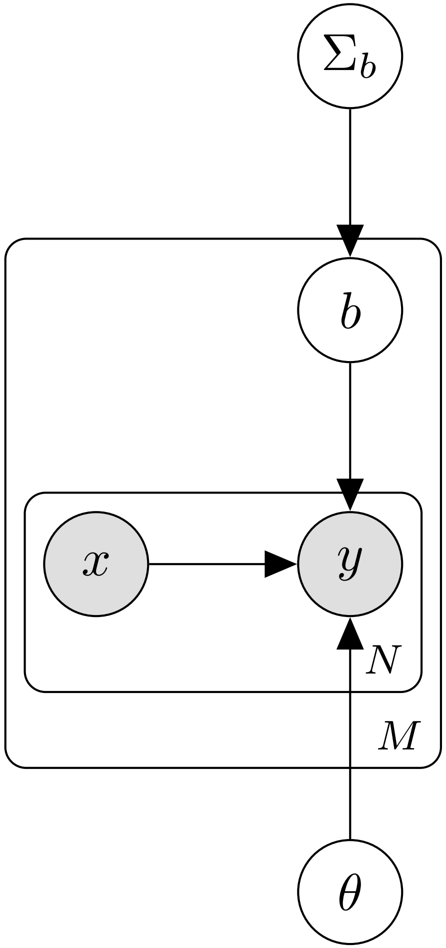


$$\theta = \langle \beta, \sigma^2 \rangle$$

$$b \sim N(0, \Sigma_b)$$

Compact mixed-effects model notation

- Multi-level GLM assumptions:



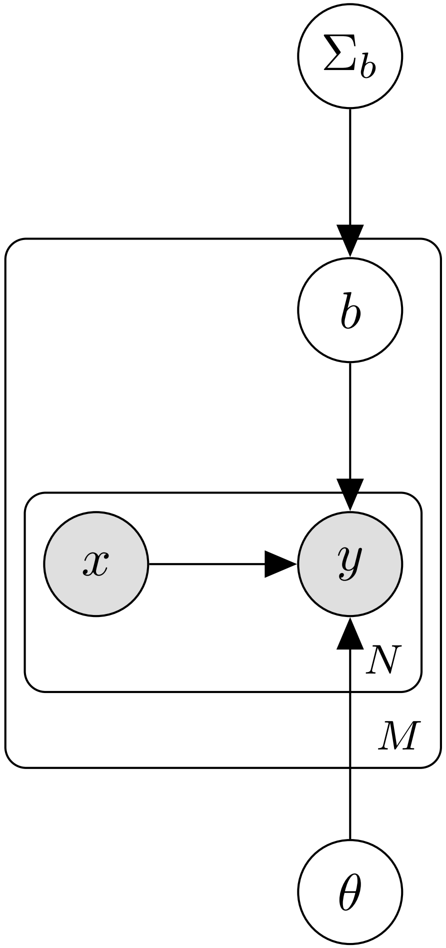
$$\theta = \langle \beta, \sigma^2 \rangle$$

$$b \sim N(0, \Sigma_b)$$

$$\eta = \left(\beta + b\right) x$$

Compact mixed-effects model notation

- Multi-level GLM assumptions:



$$\theta = \langle \beta, \sigma^2 \rangle$$

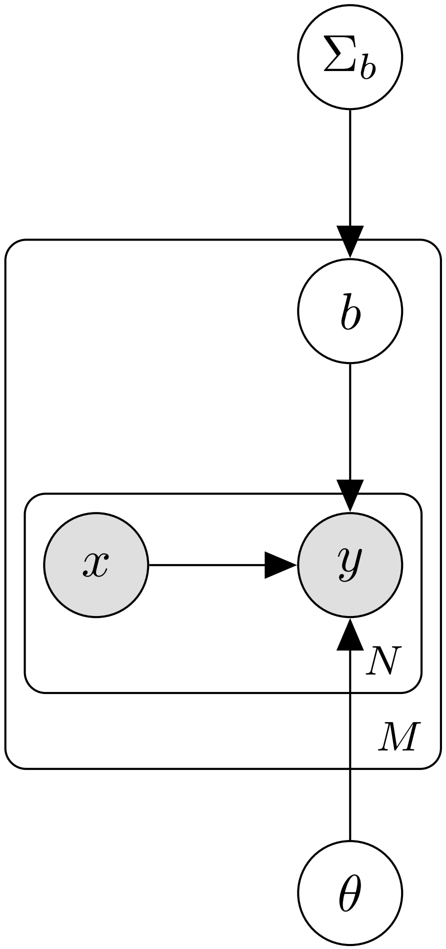
$$b \sim N(0, \Sigma_b)$$

$$\eta = \left(\beta + b\right) x$$

$$\widehat{y} \,|\, x = l^{-1}(\eta)$$

Compact mixed-effects model notation

- Multi-level GLM assumptions:



$$\theta = \langle \beta, \sigma^2 \rangle$$

$$b \sim N(0, \Sigma_b)$$

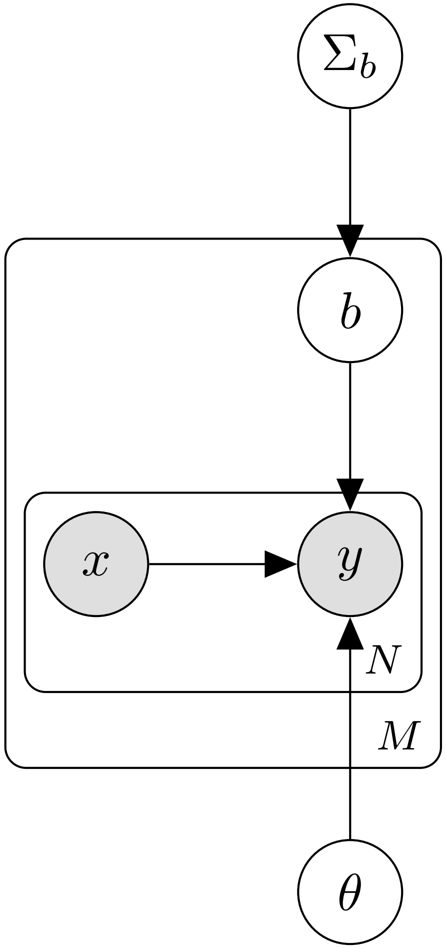
$$\eta = \left(\beta + b\right) x$$

$$\hat{y} \mid x = l^{-1}(\eta)$$

$P(Y \mid \hat{y})$ follows an *exponential* family

Compact mixed-effects model notation

- Multi-level GLM assumptions:



$$\theta = \langle \beta, \sigma^2 \rangle$$

$$b \sim N(0, \Sigma_b)$$

$$\eta = \left(\beta + b\right) x$$

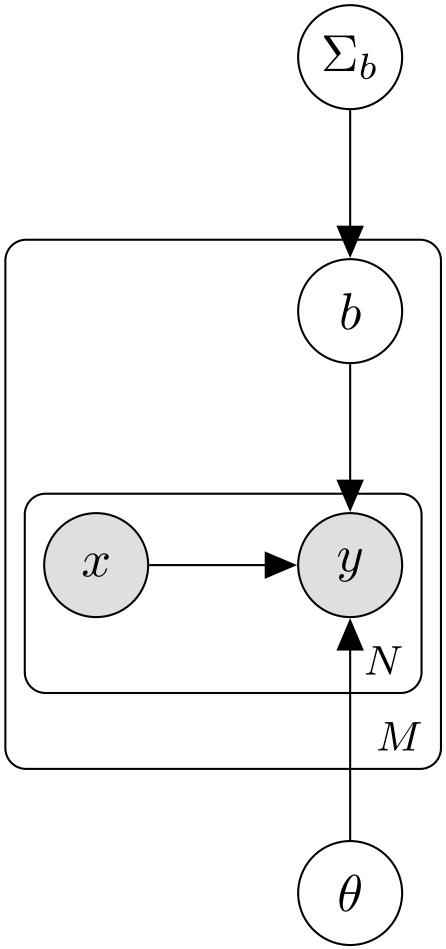
$$\hat{y} \mid x = l^{-1}(\eta)$$

$P(Y \mid \hat{y})$ follows an *exponential* family

And for ***multi-level linear regression,***

Compact mixed-effects model notation

- Multi-level GLM assumptions:



$$\theta = \langle \beta, \sigma^2 \rangle$$

$$b \sim N(0, \Sigma_b)$$

$$\eta = \left(\beta + b\right) x$$

$$\widehat{y} \mid x = l^{-1}(\eta)$$

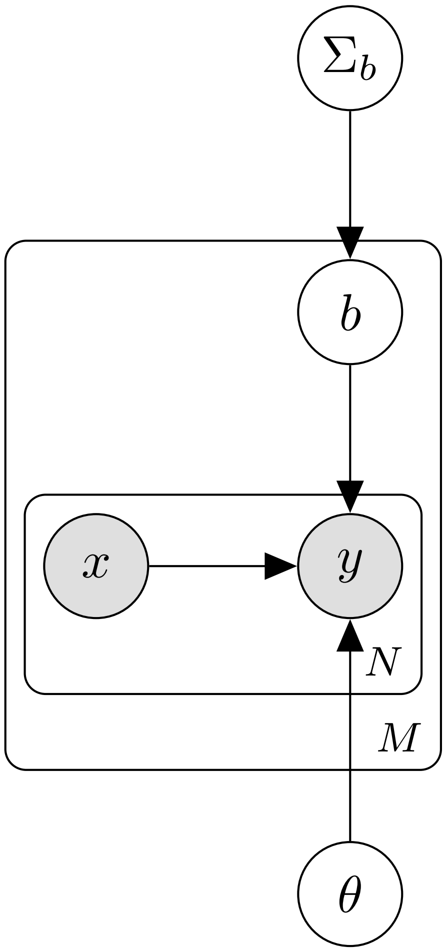
$P(Y \mid \widehat{y})$ follows an *exponential* family

And for ***multi-level linear regression,***

$$l(\eta) = \eta$$ (the link function is the **identity** function)

Compact mixed-effects model notation

- Multi-level GLM assumptions:



$$\theta = \langle \beta, \sigma^2 \rangle$$

$$b \sim N(0, \Sigma_b)$$

$$\eta = \left(\beta + b\right) x$$

$$\hat{y} \mid x = l^{-1}(\eta)$$

$P(Y \mid \hat{y})$ follows an *exponential* family

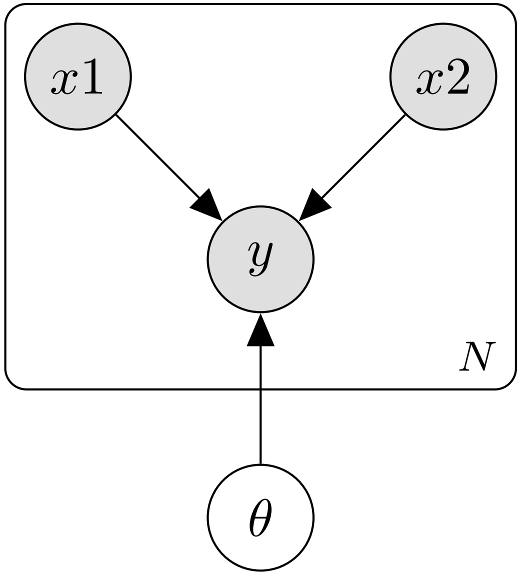
And for ***multi-level linear regression,***

$$l(\eta) = \eta$$ (the link function is the **identity** function)

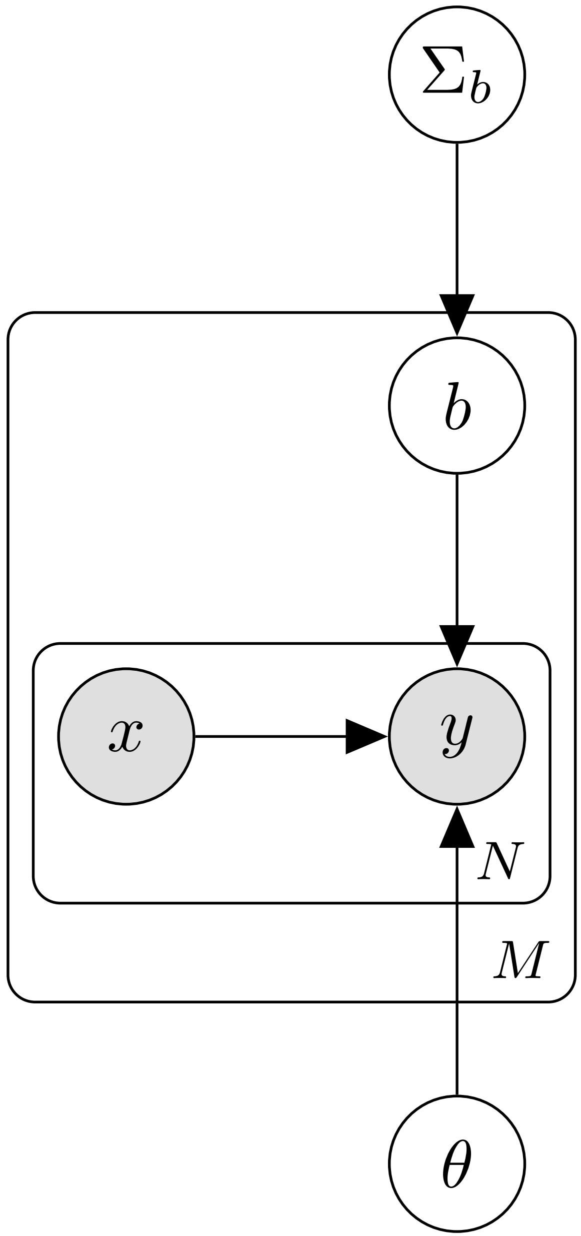
$$y \mid \hat{y} \sim N(\hat{y}, \sigma^2)$$

Credit-assignment problems and regression

- Example:
  - x_1: how long a student studies for an exam
  - x_2: a student's proclivity toward the subject matter
  - y: the student's score on the exam



Random effects as a "credit-assignment" confound



The rule for random effects for a predictor

The rule for random effects for a predictor

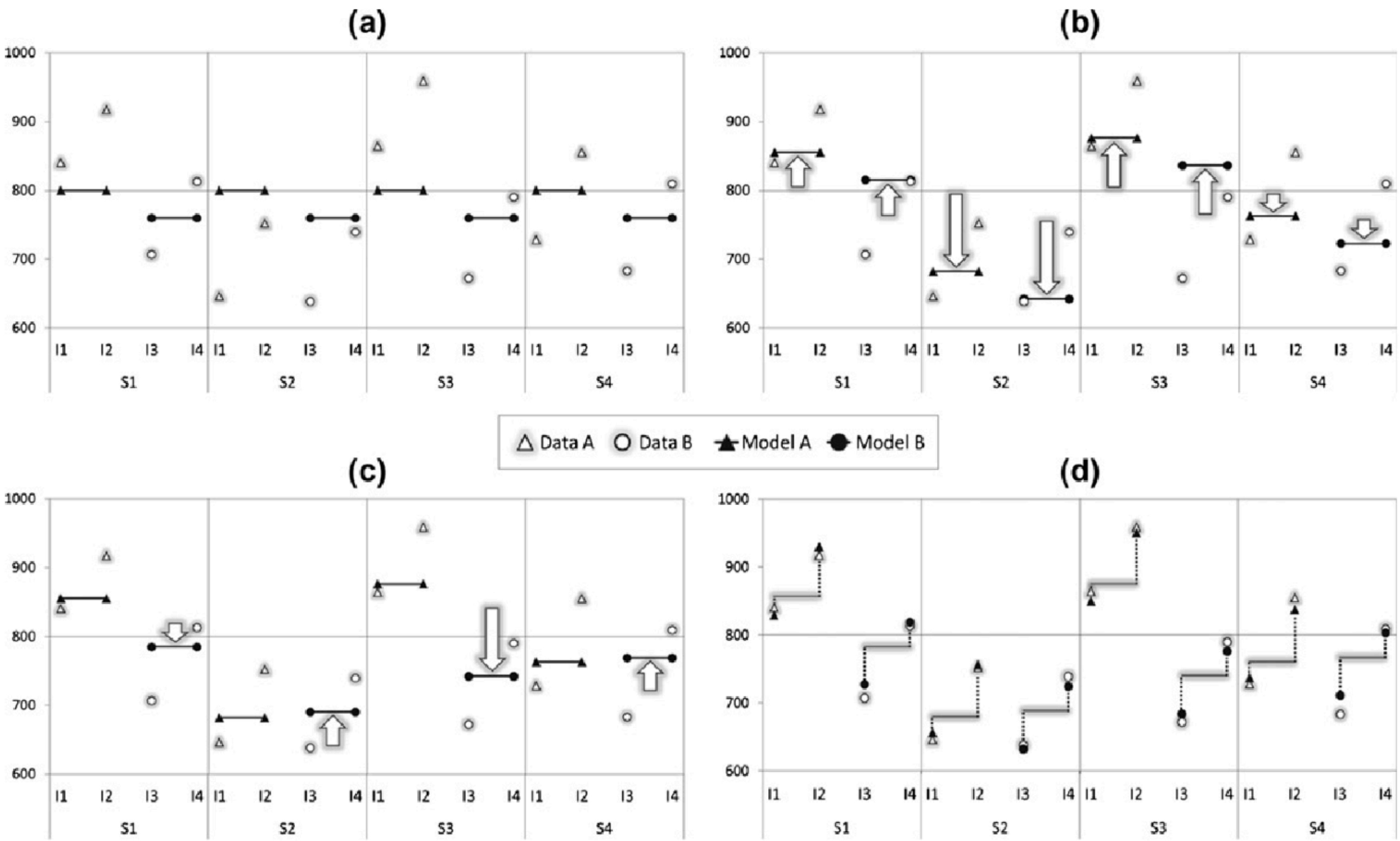
- If the average value of a predictor x varies **between** groups in a random-effects factor F, include a **random intercept** for F

The rule for random effects for a predictor

- If the average value of a predictor x varies **between** groups in a random-effects factor F, include a **random intercept** for F
- If the value of a predictor x varies **within** groups in a random-effects predictor F, include a **random slope of** x for F

Illustration of random intercepts & slopes

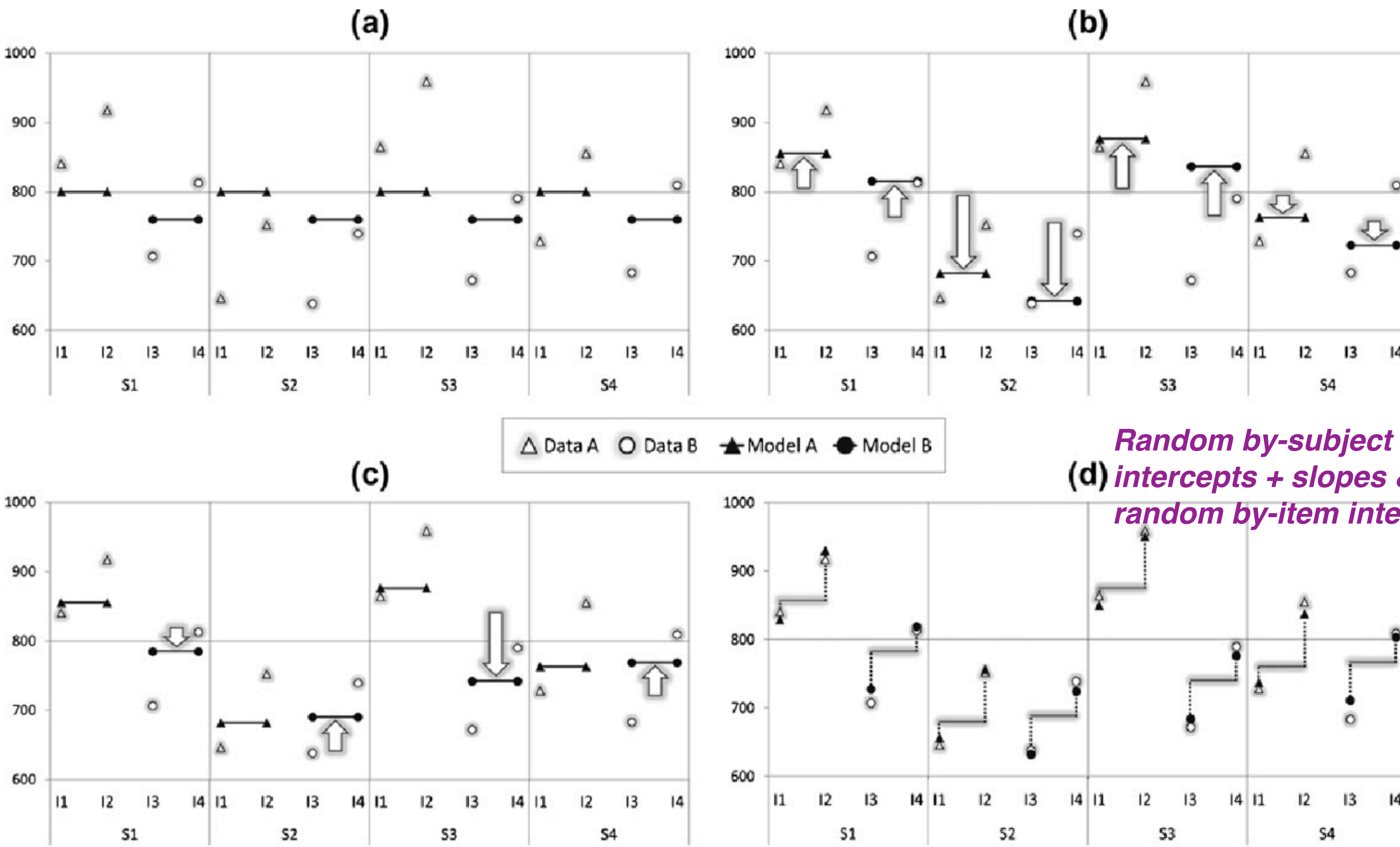
- Within-subject, between-items experiment design



(from Barr et al., 2013)

Illustration of random intercepts & slopes

- Within-subject, between-items experiment design

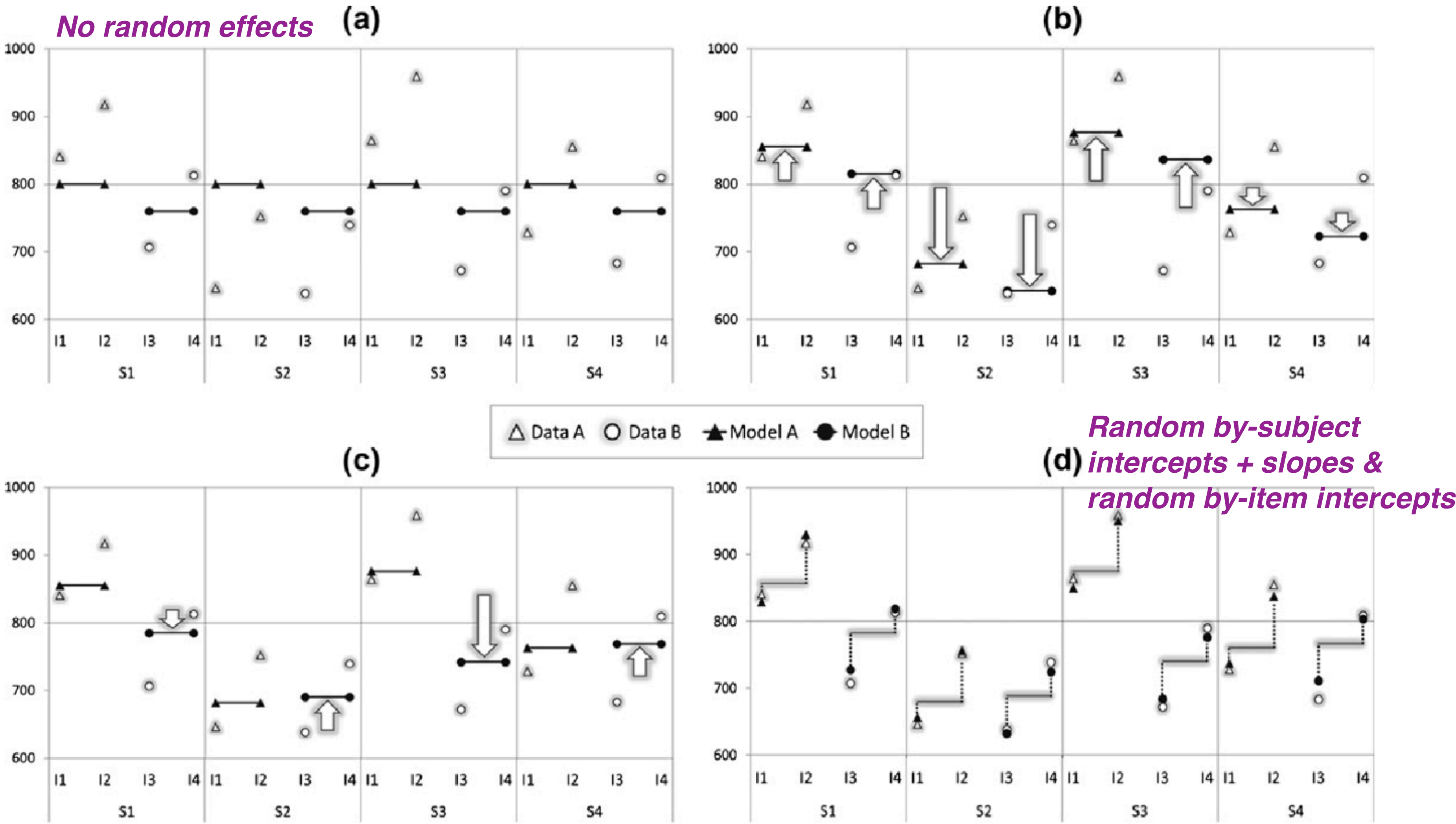


Random by-subject intercepts + slopes & random by-item intercepts

(from Barr et al., 2013)

Illustration of random intercepts & slopes

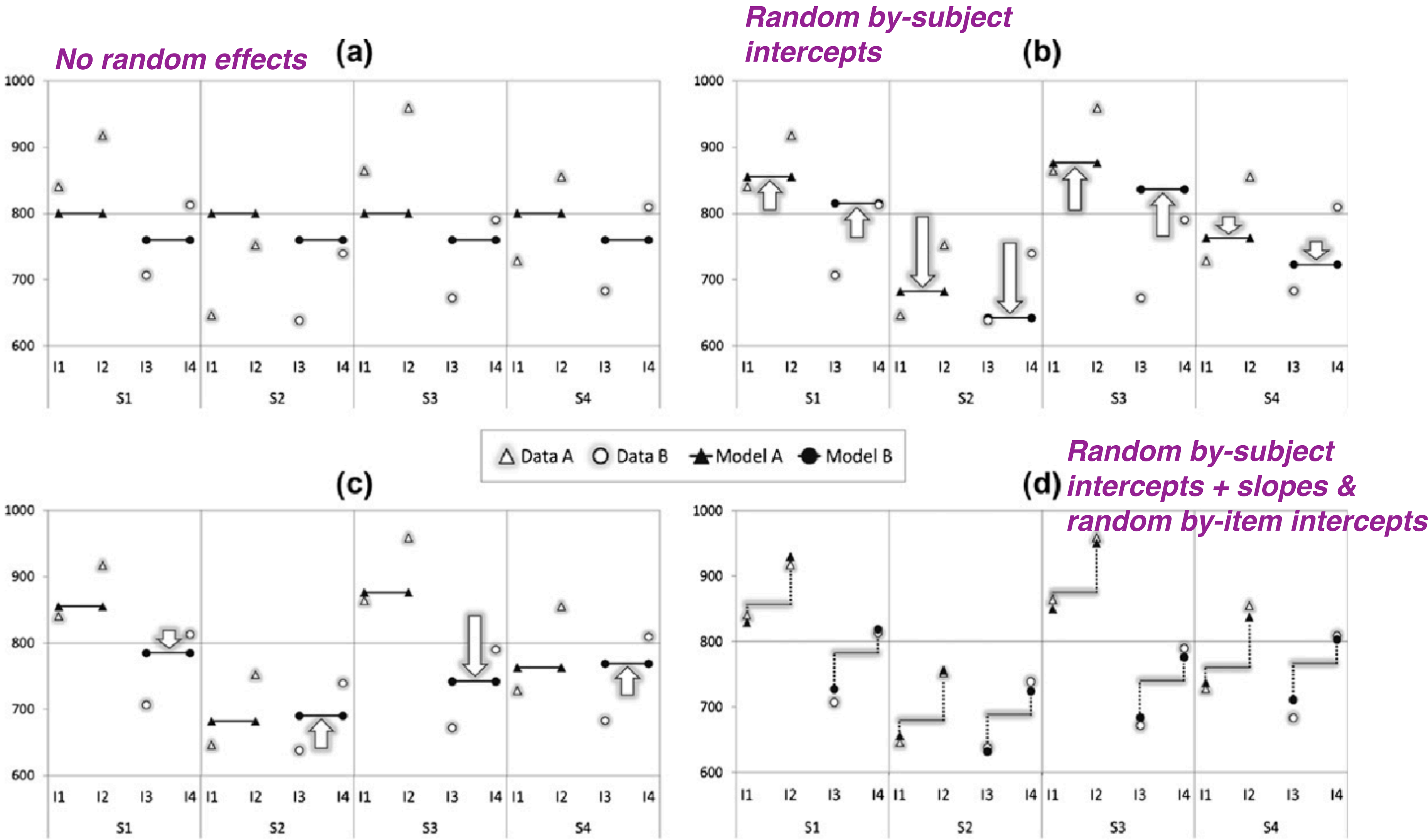
- Within-subject, between-items experiment design



(from Barr et al., 2013)

Illustration of random intercepts & slopes

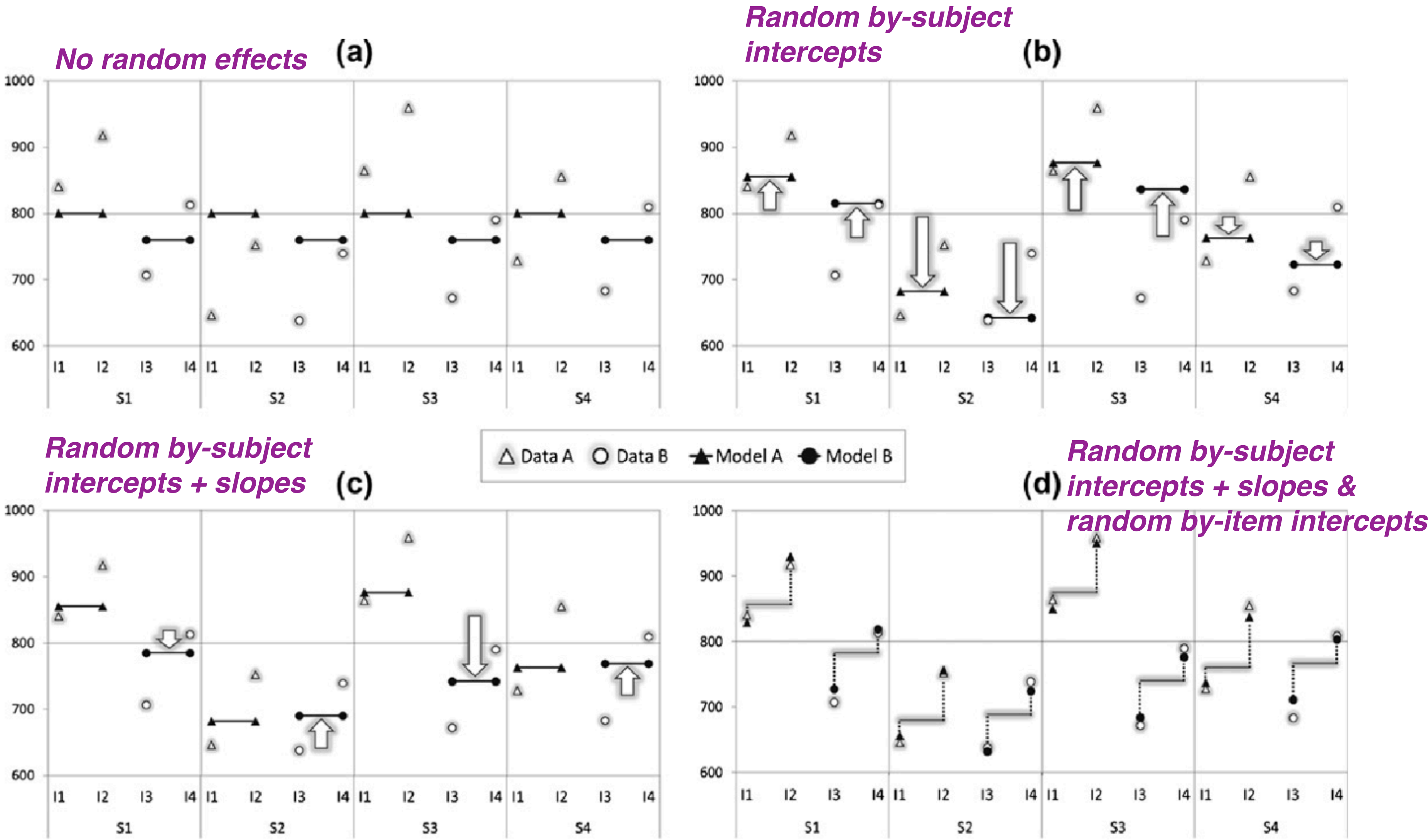
- Within-subject, between-items experiment design



(from Barr et al., 2013)

Illustration of random intercepts & slopes

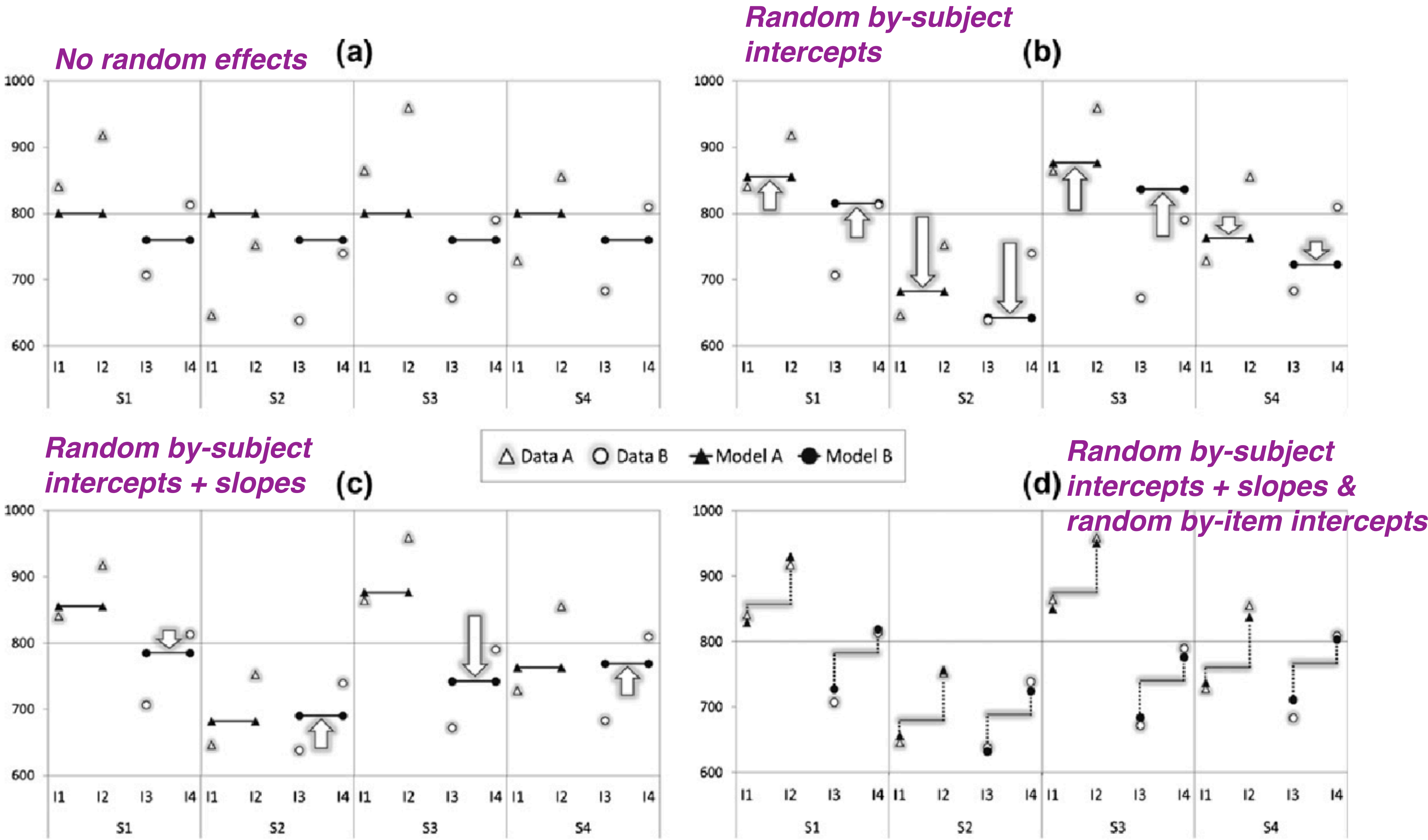
- Within-subject, between-items experiment design



(from Barr et al., 2013)

Illustration of random intercepts & slopes

- Within-subject, between-items experiment design



The MAXIMAL random effects structure for this design

(from Barr et al., 2013)

Barr et al. analysis of different models

Table 1
Summary of models considered and associated lmer syntax.

No.	Model	`lmer` model syntax
(1)	$Y_{si} = \beta_0 + \beta_1 X_i + e_{si}$	n/a (Not a mixed-effects model)
(2)	$Y_{si} = \beta_0 + S_{0s} + \beta_1 X_i + e_{si}$	`Y ~ X+(1|Subject)`
(3)	$Y_{si} = \beta_0 + S_{0s} + (\beta_1 + S_{1s})X_i + e_{si}$	`Y ~ X+(1 + X|Subject)`
(4)	$Y_{si} = \beta_0 + S_{0s} + I_{0i} + (\beta_1 + S_{1s})X_i + e_{si}$	`Y ~ X+(1 + X|Subject)+(1|Item)`
(5)	$Y_{si} = \beta_0 + S_{0s} + I_{0i} + \beta_1 X_i + e_{si}$	`Y ~ X+(1|Subject)+(1|Item)`
(6)[a]	As (4), but S_{0s}, S_{1s} independent	`Y ~ X+(1|Subject)+(0 + X|Subject)+(1|Item)`
(7)[a]	$Y_{si} = \beta_0 + I_{0i} + (\beta_1 + S_{1s})X_i + e_{si}$	`Y ~ X+(0 + X|Subject)+(1|Item)`

[a] Performance is sensitive to the coding scheme for variable X (see Online Appendix).

Performance for **between-items** designs

- Type 1 error, power, and corrected power

N_{items}	Type I		Power		Power′	
	12	24	12	24	12	24
Type I: Error at or near $\alpha = .05$						
min-F'	.044	.045	.210	.328	.328	.328
LMEM, maximal, χ^2_{LR}	.070	.058	.267	.364	.223	.342
LMEM, no random correlations, χ^2_{LR} [a]	.069	.057	.267	.363	.223	.343
LMEM, no within-unit intercepts, χ^2_{LR} [a]	.081	.065	.288	.380	.223	.342
LMEM, maximal, t	.086	.065	.300	.382	.222	.343
LMEM, no random correlations, t	.086	.064	.300	.382	.223	.343
LMEM, no within-unit intercepts, t [a]	.100	.073	.323	.401	.222	.342
$F_1 \times F_2$	.063	.077	.252	.403	.224	.337
Type I: Error far exceeding $\alpha = .05$						
LMEM, random intercepts only, χ^2_{LR}	.102	.111	.319	.449	.216	.314
LMEM, random intercepts only, t	.128	.124	.360	.472	.472	.314
LMEM, no random correlations, MCMC[a]	.172	.192	.426	.582		
LMEM, random intercepts only, MCMC	.173	.211	.428	.601		
F_1	.421	.339	.671	.706	.134	.212

[a] Performance is sensitive to coding of the predictor (see the Online Appendix); simulations use deviation coding.

Performance for **within-items** designs

N_{items}	Type I		Power		Power′	
	12	24	12	24	12	24
Type I: Error at or near α = .05						
min-F'	.027	.031	.327	.512	.327	.512
LMEM, maximal, χ^2_{LR}	.059	.056	.460	.610	.433	.591
LMEM, no random correlations, χ^2_{LR}[a]	.059	.056	.461	.610	.432	.593
LMEM, no within-unit intercepts, χ^2_{LR}[a]	.056	.055	.437	.596	.416	.579
LMEM, maximal, t	.072	.063	.496	.629	.434	.592
LMEM, no random correlations, t	.072	.062	.497	.629	.432	.593
LMEM, no within-unit intercepts, t[a]	.070	.064	.477	.620	.416	.580
$F_1 \times F_2$	.057	.072	.440	.643	.416	.578
Type I: Error far exceeding α = .05						
F_1	.176	.139	.640	.724	.345	.506
LMEM, no random correlations, MCMC[a]	.187	.198	.682	.812		
LMEM, random intercepts only, MCMC	.415	.483	.844	.933		
LMEM, random intercepts only, χ^2_{LR}	.440	.498	.853	.935	.379	.531
LMEM, random intercepts only, t	.441	.499	.854	.935	.379	.531

[a] Performance is sensitive to coding of the predictor (see the Online Appendix); simulations use deviation coding.

Anti-conservativity of model selection