

Heuristics as Bayesian Inference

Paula Parpart

UCL Experimental Psychology &
DTC for Financial Computing & Analytics,
UCL Computer Science

Collaborators

Prof Brad Love – UCL

Prof Matt Jones – University of Colorado

Dr Takao Noguchi – UCL

UK PhD Centre in
Financial Computing &
Analytics



Overview

- 1. What are Heuristics**
- 2. Fast and Frugal vs. Heuristics-and-Biases Approach**
- 3. Less-is-more phenomena**
- 4. Bias-variance & overfitting**
- 5. Heuristics as Bayesian Inference: Model I**
- 6. Heuristics as Bayesian Inference: Model II**
- 7. Discussion: Q & A**
- 8. Implications**

Heuristics

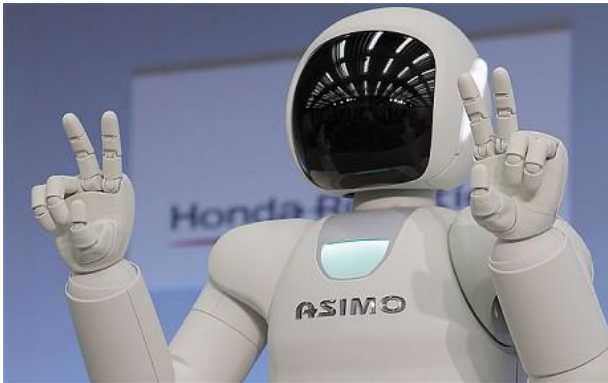


Less-Can-Be-More: Managers' One-Good-Reason Decisions



Why are Heuristics so important in AI and Computer Science?

- They can solve NP-complete (computationally intractable) problems when classic methods (probability theory) fail to find an exact solution



What strategy would you use?

1/N Rule: Allocate resources equally to each of N alternatives. (Benartzi & Thaler 2001)



Optimizing portfolio models such as the Nobel Prize-winning “Markowitz’s mean-variance portfolio” (DeMiguel et al. 2009)

Early decision theories

- Many economic theories portrayed decision agents as idealised, perfectly rational humans
 - rational choice theory (Scott, 2000; Friedman, 1953)
 - expected utility theory (von Neumann & Morgenstern, 1944; 1947; 1953)
- Statistical optimal models are regarded as “rational” because they are grounded in the **laws of logic** and the **axioms of probability theory**.
- *Homo economicus* always acts rationally with complete knowledge, out of self-interest and with the desire for wealth



Early decision theories

- Many economic theories portrayed decision agents as idealised, perfectly rational humans
 - rational choice theory (Scott, 2000; Friedman, 1953)
 - expected utility theory (von Neumann & Morgenstern, 1944; 1947; 1953)

→ Highly unrealistic image of humans: People usually do **not** have complete, perfect knowledge at hand, **nor** unlimited time, **nor** unlimited memory capacities.

- *Homo economicus* always acts rationally with complete knowledge, out of self-interest and with the desire for wealth



Psychological models of decision making

- **Herbert Simon (1990)**: people are bounded in their rationality. Therefore people usually satisfice rather than maximize.
- **Kahneman and Tversky (1974)**: people use heuristics and often *deviate* from rational norms, i.e., they display cognitive biases:
 - conjunction fallacy (representativeness heuristic)
 - availability bias (availability heuristic)
 - anchoring bias (anchoring heuristic)
 - , ...



Heuristics – General Definition

A heuristic is a strategy that ignores part of the information, with the goal of making decisions more quickly, frugally, and/or accurately than more complex methods.











(Gigerenzer & Gaissmaier, 2012)

Overview

- 1. Fast and frugal Heuristics:
How do they work?**

Take-The-Best Heuristic











What team will win the game?

Cues	v			Coding
(1) League position	.90			+1
(2) Last game result	.81			0
(3) Home vs. away	.73			-1
(4) No. of goals	.54			-1

Mechanism:

1. Search through cues in order of their (absolute) validity.
2. Stop on finding the first cue that discriminates between the teams.
3. The team with the higher value on that discriminating cue is predicted to win, i.e., have a higher criterion value.











Tallying Heuristic

Cues	v			Coding
(1) League position	.90			+1
(2) Last game result	.81			0
(3) Home vs. away	.73			-1
(4) No. of goals	.54			-1

- Mechanism:
1. Count the positive and negative evidence in favour of either team
 2. Decision rule: Decide for the alternative that is favoured by more cues
 3. Ignore all cue validity magnitudes, and only rely on cue directionalities (+ and -).

Linear Regression

$$Y_i = \beta_1 * LeaguePos + \beta_2 * LastgameResult + \beta_3 * HomeAway + \beta_4 * NoGoals$$

Cues	v			Coding
(1) League position	.90			+1
(2) Last game result	.81			0
(3) Home vs. away	.73			-1
(4) No. of goals	.54			-1

Mechanism:

- Considers all the cues
- Selectively weights each cue
- Takes into account covariance among cues

$$\hat{\beta} = (X^T X)^{-1} X^T Y$$











Heuristics use cue validities (v) as weights.

$$v = \frac{R}{R + W}$$

R = number of correct predictions,

W = number of incorrect predictions, and consequently $0 \leq v \leq 1$

A

Cues	v			Coding
(1) League position	.90			+1
(2) Last game result	.81			0
(3) Home vs. away	.73			-1
(4) No. of goals	.54			-1

B



What are main differences between heuristics and from full-information models (e.g., full regression)?

Heuristics use cue validities (v) as weights:

$$v = \frac{R}{R + W}$$

R = number of correct predictions, $0 \leq v \leq 1$

W = number of incorrect predictions, and consequently

- Does not take account co-variance among cues

Linear regression weights:

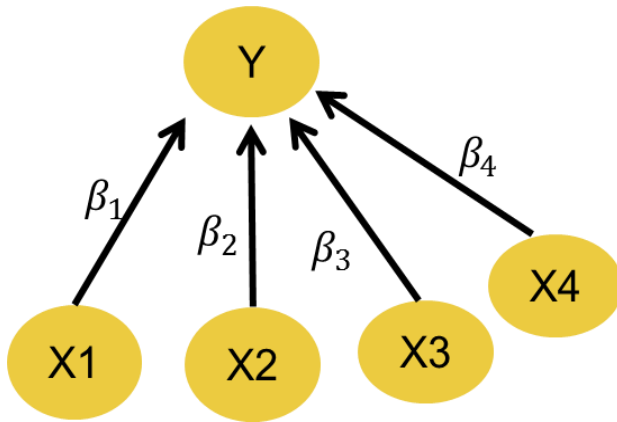
$$Y_i = \beta_1 * LeaguePos + \beta_2 * LastgameResult + \beta_3 * HomeAway + \beta_4 * NoGoals$$

- Considers all the cues
- Selectively weights each cue
- Takes into account co-variance among cues

$$\hat{\beta} = (X^T X)^{-1} X^T Y$$

Full-information models

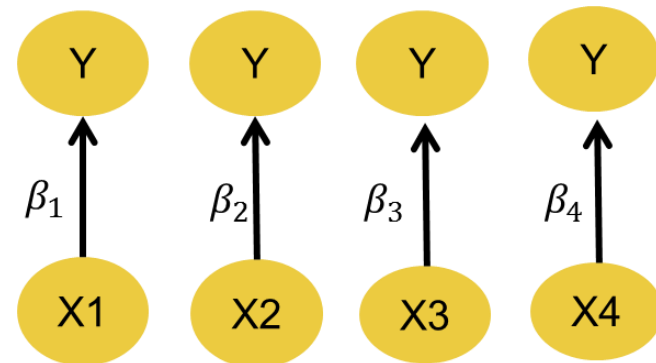
OLS regression weights:



β_1	β_2	β_3	β_4
-0.29	1.16	-0.11	0.08

Heuristics

Individual regression coefficients/ Cue validities



β_1	β_2	β_3	β_4
0.00	1.00	0.25	0.71

Important: Cue validities are a linear transformation of single predictor regression coefficients (right figure). They ignore any dependencies among cues.

Prominent notions of heuristics



Daniel Kahneman
& Amos Tversky
(1974, 1981, 2003)



Gerd
Gigerenzer &
the ABC
research
group (1999)

Heuristics and biases

Heuristics: suboptimal, a source for biases and irrational behaviour, assuming an accuracy-effort-tradeoff.

Rationality: still using laws of logic, axioms of probability theory, optimization

Heuristics are **biased** approximations to rational inference.



Kahneman & Tversky (1974)

Fast and frugal heuristics

Heuristics: not biased, but adaptive, exploit structure in environment, lead to good accuracy levels, no accuracy-effort-tradeoff.

Rationality: No more logic & probability theory. Instead ->
Ecological Rationality

Heuristics are smart, **adaptive** strategies to act in an uncertain world.



Gigerenzer & the abc research group (1999)

Probabilistic Approach

- Bayesian models have taken over cognitive science (reasoning, judgment, learning and decision making)
- Very useful and widely applicable
- Often only on computational level though (Marr, 1982)
- Need to be integrated with mechanistic approaches?



Ecological Approach

- Heuristics are not compatible with probabilistic inference.
- Should not even compare human behaviour against the norms of probability theory
- Heuristics are psychologically plausible process models accounting for cognitive constraints.



➤ **Heuristics and biases** and **fast and frugal heuristics** program differ in many ways

➤ But both agree that **Heuristics \neq Bayesian**

➤ Most studies focused on demonstrating over and over again that people behave according to heuristics, or, in a nearly optimal Bayesian fashion. → collecting existence proofs is not very useful

Parpart et al., (submitted)

- I. We show that heuristics are compatible with Bayesian Inference.
- II. We contribute the novel idea that heuristics can be thought of **as embodying a strong Bayesian prior.**
- III. We thereby attempt to reconcile irrational and adaptive approaches of heuristics.

Heuristics are often contrasted with *full-information models*

- *Full-information models*: make full and proper use of available information. Such as:
 - "Rational" Multiple Regression:
 - Uses all cues and optimally weighs and integrates them
- Probabilistic Models of Cognition
 - Bayesian inference models
 - Optimal inference as benchmark to compare human behaviour against (Oaksford & Chater, 2007)

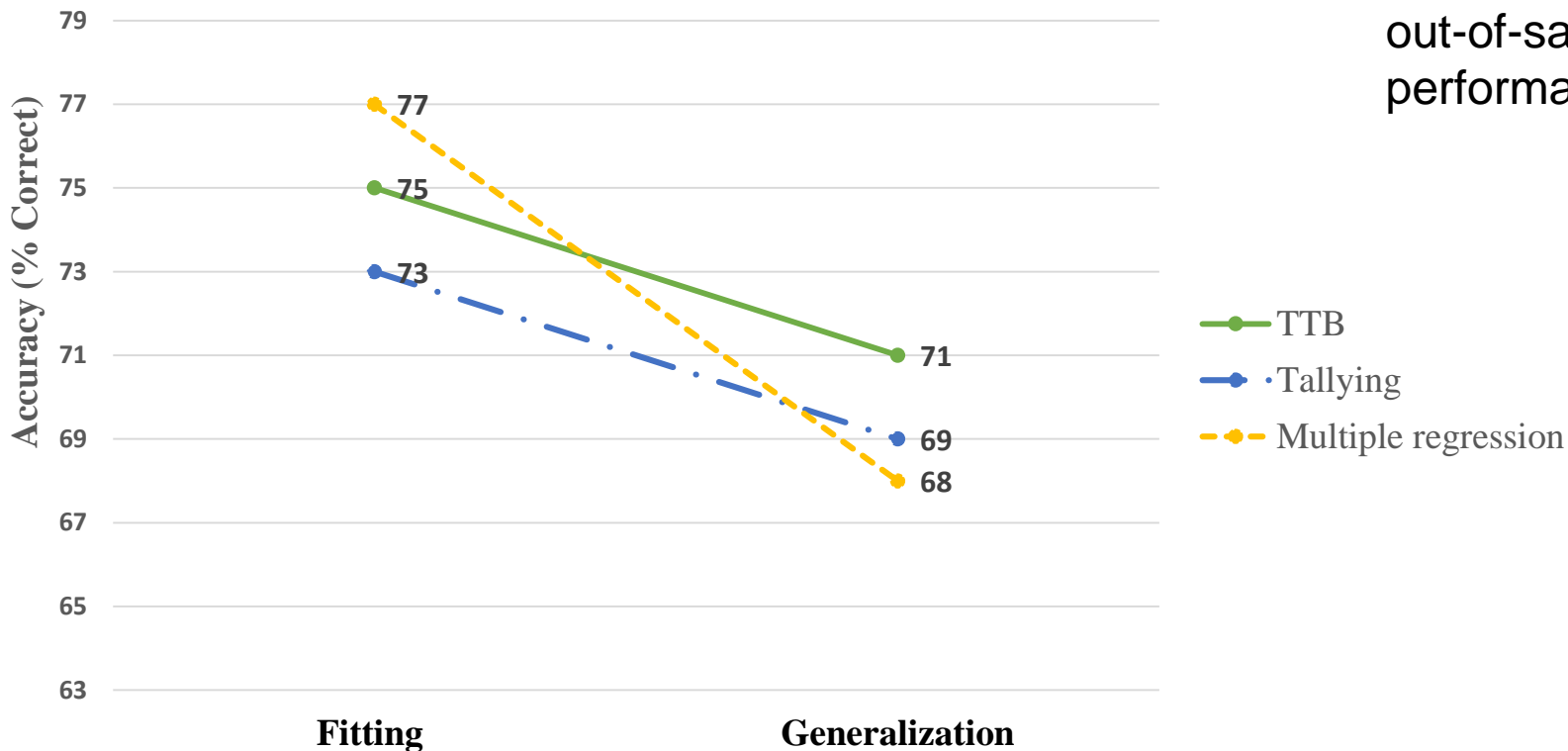
Less-is-more: Heuristics can outperform full-information models

- Czerlinski et al. (1999) showed that heuristics can sometimes outperform “rational” linear multiple regression
- Heuristics can outperform three-layer feed-forward connectionist neural network trained using the back propagation algorithm, two exemplar-based models, and a decision tree- induction algorithm (Chater et al., 2003; Brighton, 2006).

→ Such results can appear paradoxical because heuristics neglect relevant information, while the full-information methods make full use of the data.

Heuristics vs. “rational” accounts

Generalization performance across 20 data sets
 Training size = 50% of each dataset



- Generalization performance measures the out-of-sample performance.

Original data sets (Czerlinski et al., 1999): City size task, professors salaries, High school dropout rates, Homelessness House price, Mortality, Land rent, Car accidents, Fuel consumption, Obesity at age 18, Body Fat, Fish fertility, Mammals’ sleep, Cow manure, Biodiversity, Rainfall from cloud, Oxidant in L.A., Ozone in San Francisco

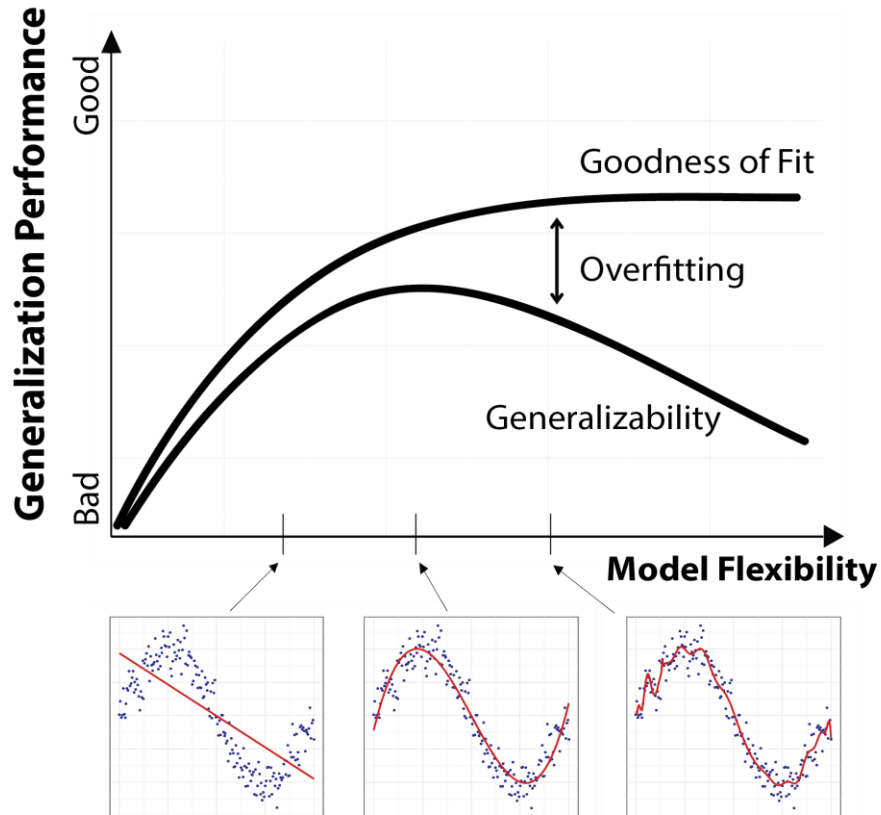
How can *Less-is-more* findings be explained?

1. **Bias-variance** (Brighton & Gigerenzer, 2009)
2. **Later: Our Bayesian integration model will provide a new explanation.**

In this view, an explanation for the success of heuristics is that their **relative simplicity and inflexibility amounts to a strong inductive bias, akin to a Bayesian prior**, that makes the model best-suited to certain learning and decision problems.

1. Bias-variance

A



(Adopted from Pitt & Myung, 2002)

bias-variance tradeoff:

$$\text{Prediction error} = (\text{bias})^2 + \text{variance} + \text{noise}$$

Bias-variance

- A model's bias and the input data are responsible for what a model learns from the training data.
- In addition to differing in **bias**, models can also differ in how sensitive they are to the *variability* in the training sample, i.e., this is reflected in the **variance** of the model's parameters after training.

$$\text{Prediction error} = (\text{bias})^2 + \text{variance} + \text{noise}$$

- Both the inductive bias and the parameters' variance determine how well a model classifies novel test cases – this is crucial, as the utility of any model is measured by its generalization performance (Kohavi, 1995)

Overfitting

- Higher flexibility (higher variance) can in fact hurt a models' performance as it means the model is overly affected by the idiosyncrasies of the training sample.



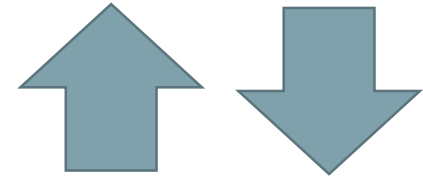
Tour de France

This phenomenon, commonly referred to as *overfitting*, is characterized by high performance on experienced cases from the training sample but poor performance on novel test items.

Overfitted models have high goodness-of-fit but low generalization performance (Pitt & Myung, 2002)

Overfitting

- Bias and variance trade off with one another



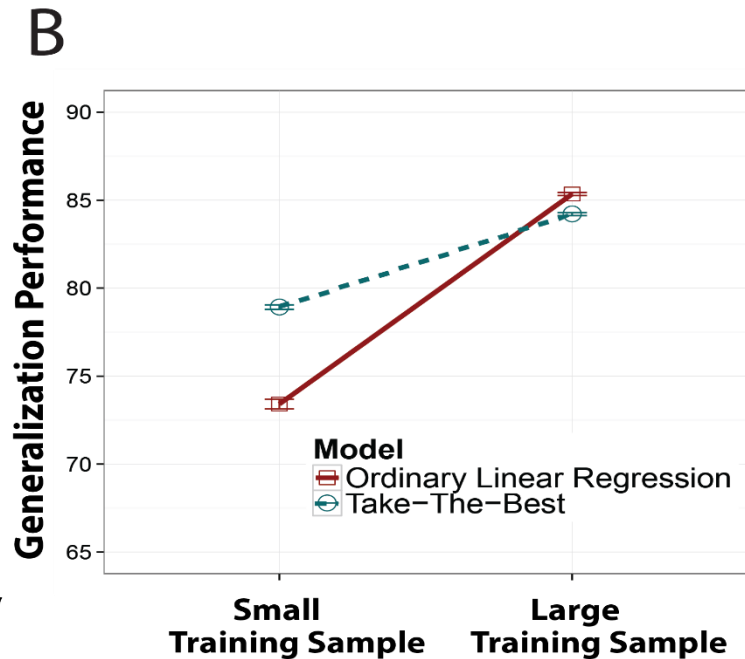
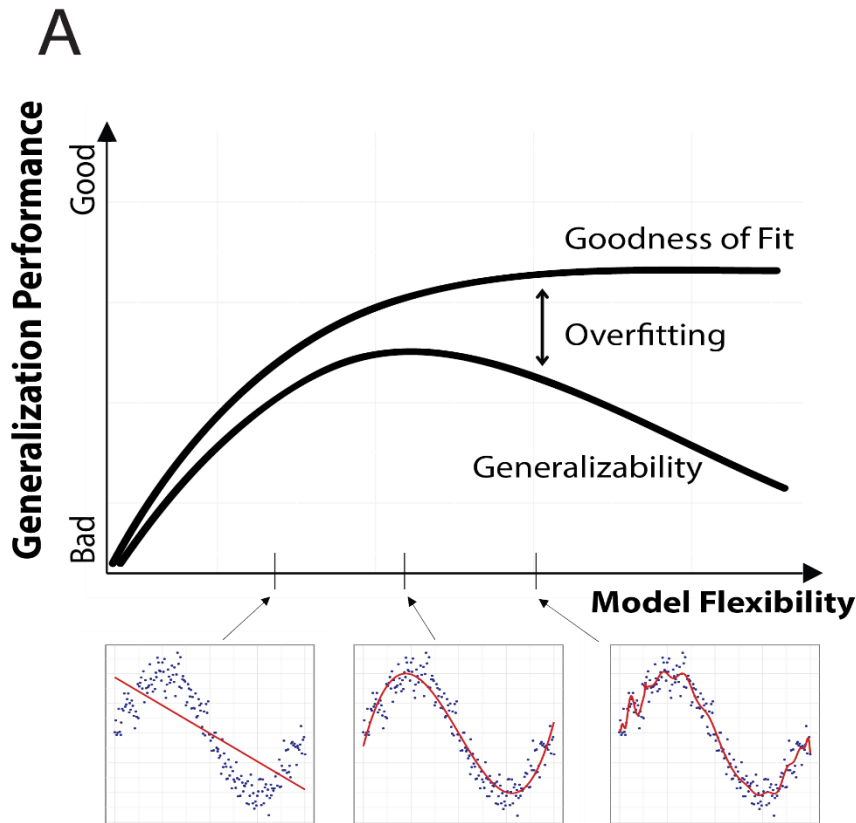
$$\text{Prediction error} = (\text{bias})^2 + \text{variance} + \text{noise}$$



“less well trained cyclist
can paradoxically do
better”

→ implies that simpler (i.e., more biased) models, such as heuristics can outperform more flexible (i.e., less biased) models

Why can simple heuristics sometimes outperform more complex algorithms?



B) House data set by Czerlinski et al., (1999)

- As the size of the training sample increases, more complex models (OLS) should fare better. → we find that the advantage for the heuristic disappears when training sample size is increased (Figure B)

Novel Bayesian Approach: 3 goals

- **We need to move beyond demonstrations like these, and get a deeper, formal understanding that is general and powerful.**
 - **Why** can heuristics sometimes perform better than full-information models?
 - Create a formal link between OLS and heuristics.
 - Show that intermediate models may perform best.

Heuristics as Bayesian Inference: Model I

Model I: Bayesian Model for Tallying

Tallying as a limiting case of regularized regression

- The 1st **Bayesian model** we developed is conceptually related to *ridge regression*, a successful regularized regression approach in machine learning.
- Ridge regression extends ordinary linear regression by incorporating a penalty term that adjusts model complexity to improve weight estimates and avoid overfitting

Regularized regression: Ridge regression (L2)

$$\hat{W}_{\text{ridge}} = \arg \min_w \left\{ \underbrace{\|y - Xw\|^2}_{\text{Goodness-of-Fit}} + \underbrace{\theta \|w\|^2}_{\text{Penalty Term}} \right\}$$

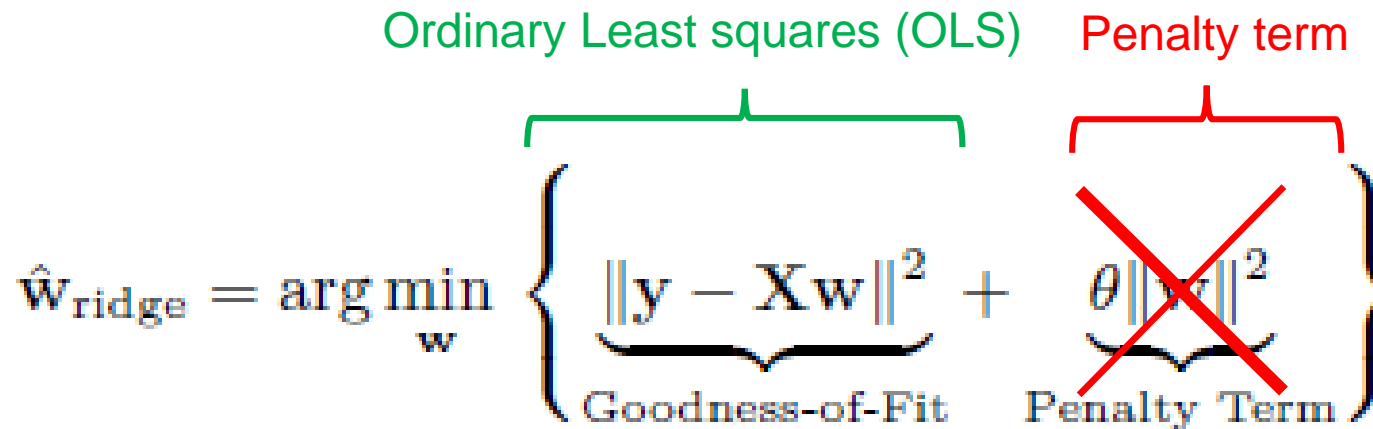
Ordinary Least squares (OLS) Penalty term

- Linear regression coefficients usually suffer from high variance as the model gets more complex (overfitting)
- Ridge regression's penalty term reduces model complexity as the penalty parameter θ increases, as **more bias is introduced in the model**, reducing variance
- Less overfitting.

Regularized regression: Ridge regression (L2)

$$\hat{W}_{\text{ridge}} = \arg \min_w \left\{ \underbrace{\|y - Xw\|^2}_{\text{Goodness-of-Fit}} + \underbrace{\theta \|w\|^2}_{\text{Penalty Term}} \right\}$$

Ordinary Least squares (OLS) Penalty term



Special case 1: When $\theta = 0$,

- ridge regression is concerned only with goodness of fit (i.e., minimizing squared error on the training set).
- ridge regression is equivalent to OLS ($\hat{w}^{\text{ridge}} \rightarrow \hat{w}^{\text{OLS}}$)

Regularized regression: Ridge regression (L2)

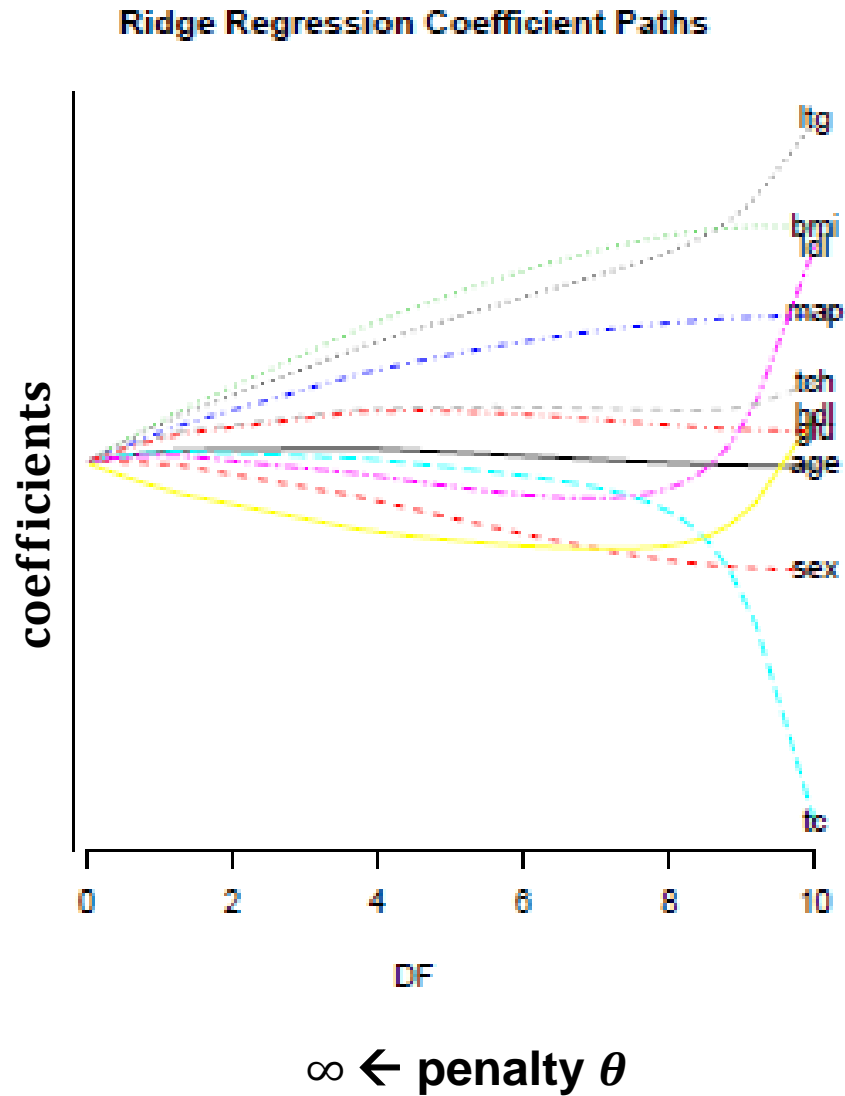
$$\hat{W}_{\text{ridge}} = \arg \min_w \left\{ \underbrace{\|y - Xw\|^2}_{\text{Goodness-of-Fit}} + \underbrace{\theta \|w\|^2}_{\text{Penalty Term}} \right\}$$

Ordinary Least squares (OLS) Penalty term

Special case 2: When $\theta \rightarrow \infty$,

- the pressure to shrink the weights increases, reducing them to zero as $\theta \rightarrow \infty$. $\hat{w}^{\text{ridge}} \rightarrow 0$
- Larger values of θ lead to stronger inductive bias, which can reduce overfitting by reducing sensitivity to noise in the training sample

Regularized regression (L2): Ridge regression



Regularized regression: Ridge regression (L2)

$$\hat{W}_{\text{ridge}} = \arg \min_w \left\{ \underbrace{\|y - Xw\|^2}_{\text{Goodness-of-Fit}} + \underbrace{\theta \|w\|^2}_{\text{Penalty Term}} \right\}$$

Ordinary Least squares (OLS) Penalty term

The optimal setting of θ will always depend on the environment from which the weights, cues, and outcomes were sampled.

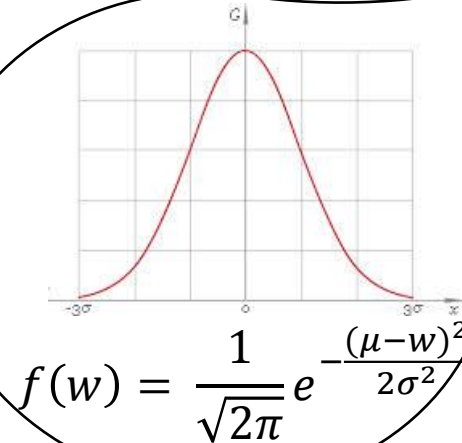
Penalty term is like a Gaussian Bayesian prior

$$\hat{w}_{\text{ridge}} = \arg \min_w \left\{ \underbrace{\|y - Xw\|^2}_{\text{Goodness-of-Fit}} + \underbrace{\theta \|w\|^2}_{\text{Penalty Term}} \right\}$$

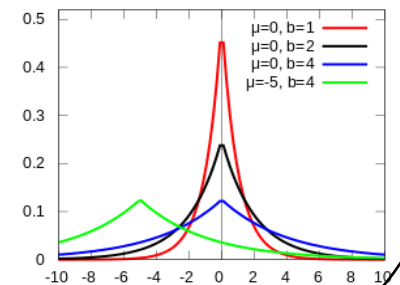
Penalty term

**Gaussian
(Normal) prior on
the weights**

where $\theta = \frac{\sigma^2}{\eta^2}$



Other priors



Bayesian Interpretation

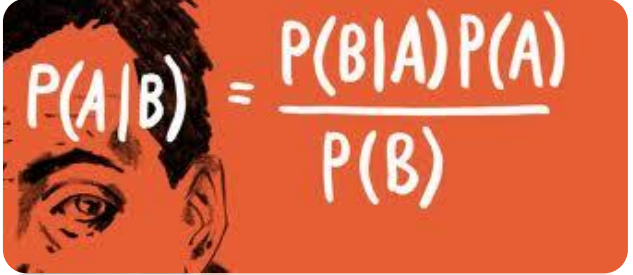
$$\hat{\mathbf{W}}_{\text{ridge}} = \arg \min_{\mathbf{w}} \left\{ \underbrace{\|\mathbf{y} - \mathbf{X}\mathbf{w}\|^2}_{\text{Goodness-of-Fit}} + \underbrace{\theta \|\mathbf{w}\|^2}_{\text{Penalty Term}} \right\}$$

Gaussian prior on the weights with $\theta = \frac{\sigma^2}{\eta^2}$

- In the Bayesian interpretation, we don't call it "penalty parameter", but "strength of the prior".
- Strength of the prior is reflected by $\frac{1}{\eta^2}$ growing stronger as $\eta^2 \rightarrow 0$.











Bayesian Framework for Tallying

- This Gaussian prior distribution is combined with current observations (i.e., the training sample) to form a posterior distribution (also Gaussian) over the weights.


$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

Tallying as a limiting case of regularized regression

- Our Bayesian derivation of the tallying heuristic extends ridge regression by assuming the directionalities of the cues (i.e., the signs of the true weights) are known in advance.

A		v		B	
Cues					
(1) League position	.90			+	
(2) Last game result	.81			+	
(3) Home vs. away	.73			+	
(4) No. of goals	.54			+	

- This is concordant with how the tallying heuristic was originally proposed in the literature (Dawes, 1979)

Tallying as a limiting case of regularized regression

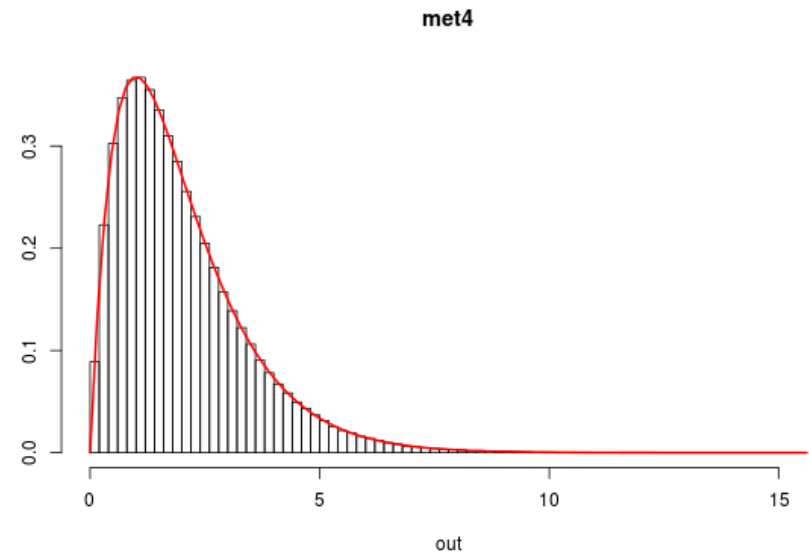
- Thus we define the prior for each weight as half-Gaussian, truncated at zero, and we refer to this Bayesian model as the *half-ridge* model.

- Prior is defined by

$$\mathbf{w} \sim \mathcal{N}(\mathbf{0}, \Sigma) \mid \mathbf{w} \geq \mathbf{0}$$

$$\Sigma = \eta^2 \mathbf{I},$$

→ Prior weights are all positive.



Tallying as a limiting case of regularized regression

- Posterior is then also truncated at zero.
- Important question is what happens to this posterior as the prior becomes arbitrarily strong, i.e., $\frac{1}{\eta^2} \rightarrow \infty$.

$$\frac{\mathbf{w}}{\eta} \xrightarrow{d} \mathcal{N}(\mathbf{0}, I)_{|\mathbf{w} \in \mathcal{O}} \text{ as } \frac{1}{\eta^2} \rightarrow \infty.$$

Just as in ridge regression, strengthening the prior in the half-ridge model shrinks the weights toward zero.

- However, the ratios of the weights—that is, the relative inferred strengths of the cues— **all converge to unity**.

Tallying as a limiting case of regularized regression

- Posterior is then also truncated at zero.
- Important question is what happens to this posterior as the prior becomes arbitrarily strong, i.e., $\frac{1}{\eta^2} \rightarrow \infty$ or $\eta \rightarrow 0$.

$$\lim_{\eta \rightarrow 0} \mathbf{E} \left[\frac{w_i}{\eta} \mid \mathbf{X}, \mathbf{y} \right] = \pm \sqrt{\frac{2}{\pi}},$$

Weights all have same expectation in the limit.

However, the ratios of the weights—that is, the relative inferred strengths of the cues—all converge to unity.

Tallying as a limiting case of regularized regression

- That means, the optimal decision-making strategy under the Bayesian half-ridge model converges to a simple summation of the predictors—that is, **a tallying heuristic**.

$$\lim_{\eta \rightarrow 0} \mathbf{E} \left[\frac{w_i}{\eta} \mid \mathbf{X}, \mathbf{y} \right] = \pm \sqrt{\frac{2}{\pi}},$$

Weights all have same expectation
in the limit = equal -weight

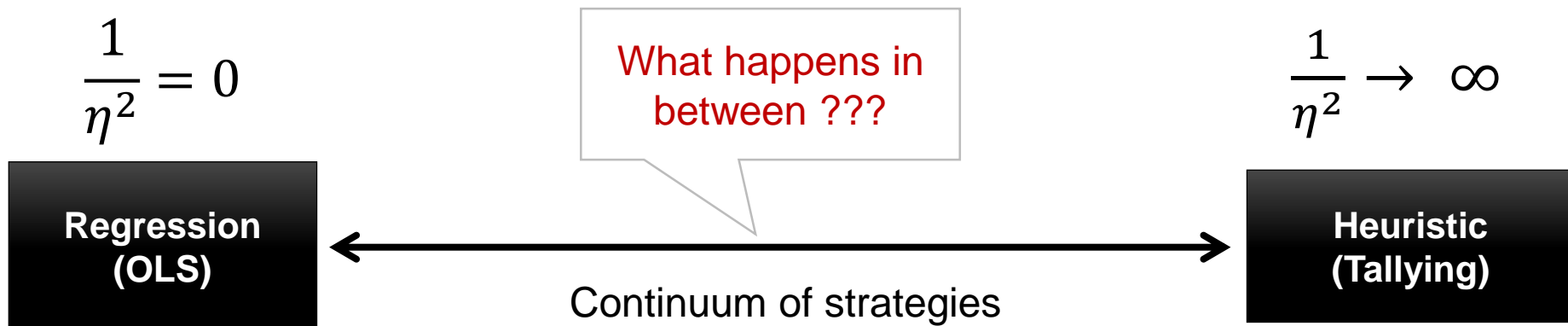
What have we just done?

- Based only on assumptions about the distribution of weights in the environment, we established a Bayesian model that converges to the tallying heuristic with a very strong prior (when $\frac{1}{\eta^2} \rightarrow \infty$).

→ Tallying heuristic represents an extreme case of a Bayesian inference model. 😊

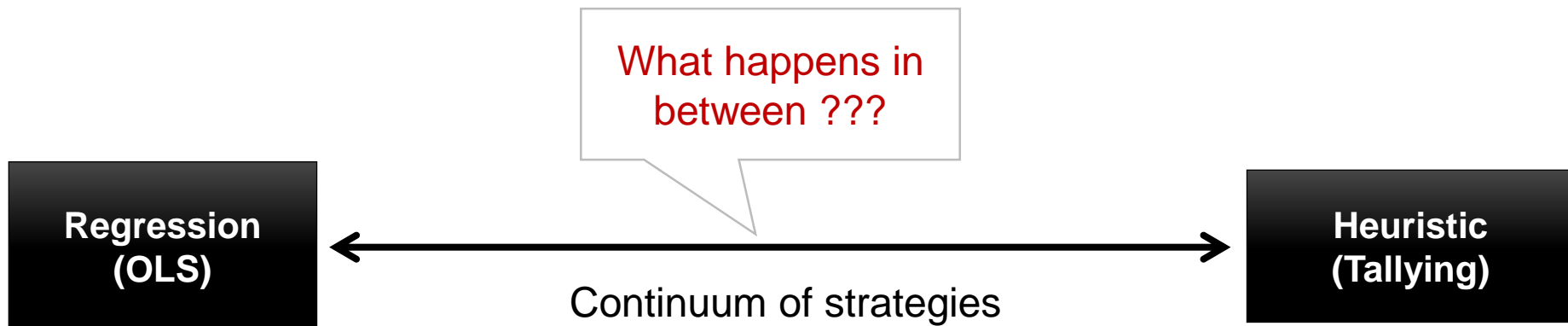
What have we just done?

- What happens at the other end of the Bayesian half-ridge model, i.e., when prior strength is zero ($\frac{1}{\eta^2} = 0$)?
- The model converges to a full regression model.



Hypotheses

- For many environments, the best-performing model should lie somewhere between these two extremes.

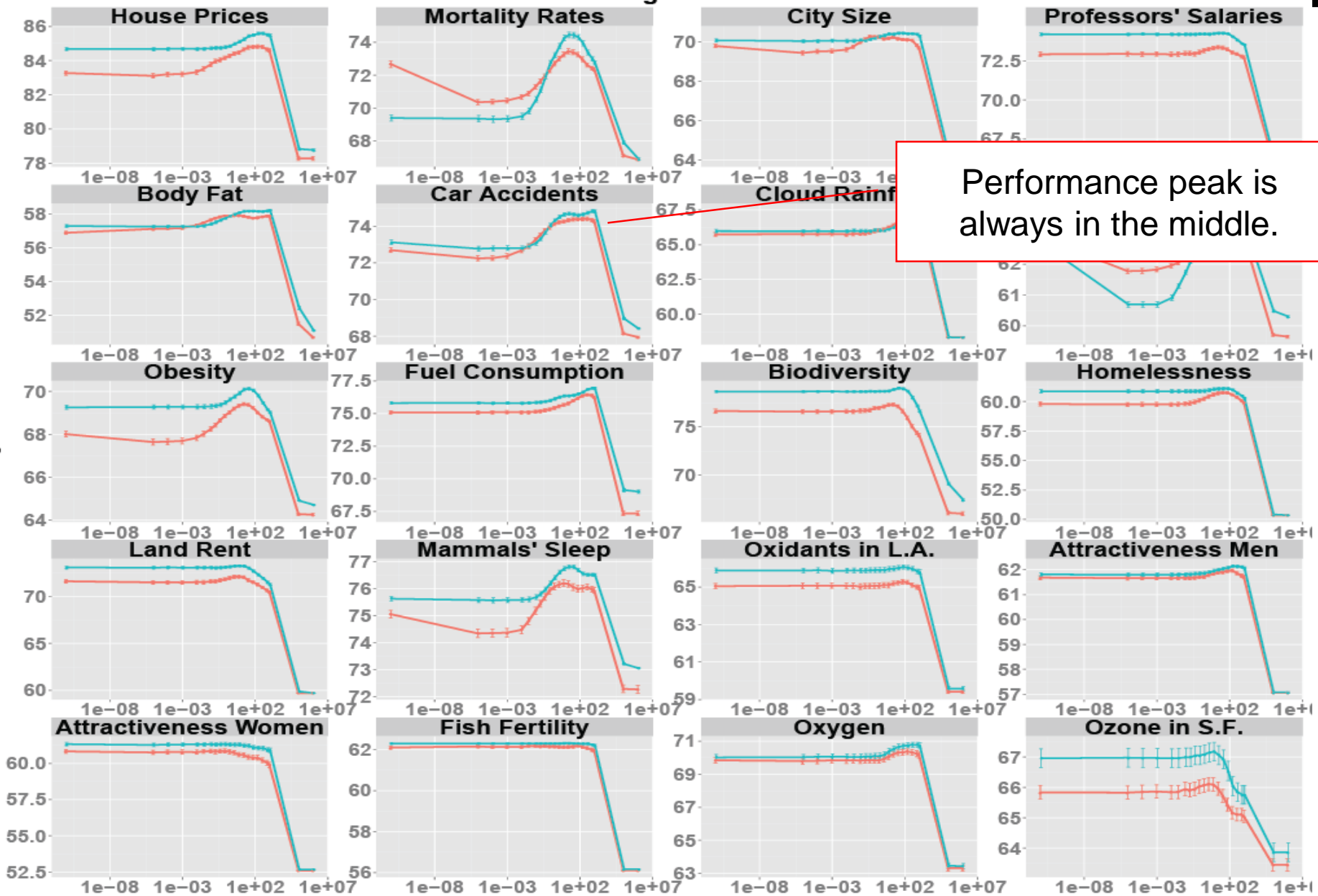


Results: Performance of the Bayesian half-ridge model compared to heuristics and linear regression

We test this on the famous twenty ABC datasets (Gigerenzer et al., 1999)

TrainingSize = 10-20

Generalization performance (%)



strength of the prior ($\frac{1}{\eta^2}$)

Interim Summary: Bayesian Model for Tallying

1. Intermediate models performed best in all cases.
2. This suggests that ignoring information was never the best solution → **Less was not more.**
3. Contrary to the less-is-more claim, **the best performing Bayesian model used all the information in the training data, but down-weights it appropriately.**

Overview

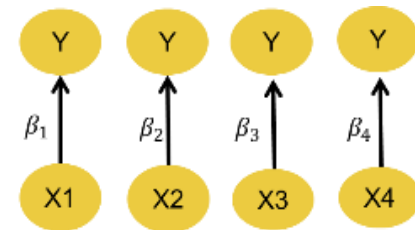
- 1. What are Heuristics**
- 2. Fast and Frugal vs. Heuristics-and-Biases Approach**
- 3. Less-is-more phenomena**
- 4. Bias-variance & overfitting**
- 5. Heuristics as Bayesian Inference: Model I**
- 6. Heuristics as Bayesian Inference: Model II**
- 7. Discussion: Q & A**
- 8. Implications**

What about the Take-The-Best heuristic?

- 2nd Bayesian model that provides a unification of TTB, tallying, and linear regression.
- Unlike linear regression, both TTB and tallying rely on isolated cue-outcome relationships (i.e., cue validity) that disregard covariance information among cues.
- We use this insight to construct our second Bayesian model.

The Role of Covariance

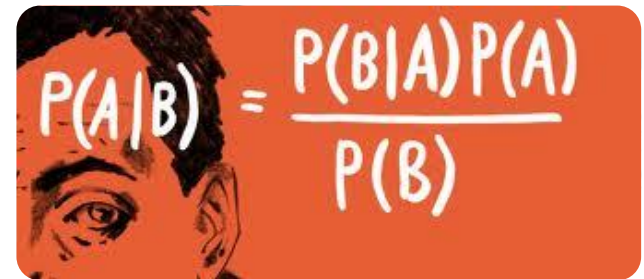
- **Heuristics** do not assess covariance.



-
- **Complex models**: OLS estimates covariance from the data in the learning phase, and this can hurt at generalization (overfitting).

Bayesian Framework

- Prior = reflecting the amount of **covariance** in the environment
- Likelihood = a latent state variable model that enables us to smoothly move between linear regression and the heuristics (tallying and TTB heuristic).

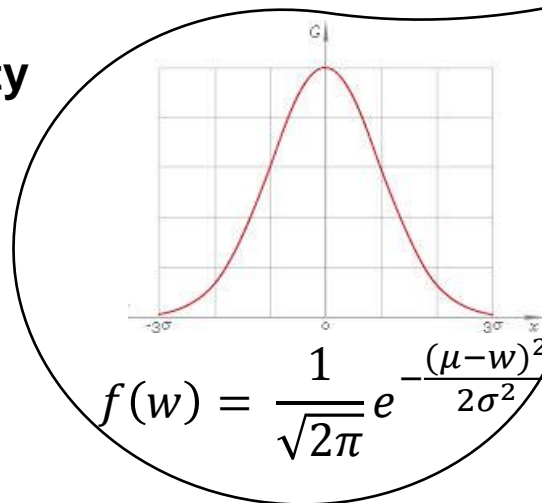

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

Regularization with covariance prior

$$\hat{w}_{\text{ridge}} = \arg \min_w \left\{ \underbrace{\|y - Xw\|^2}_{\text{Goodness-of-Fit}} + \underbrace{\theta \|w\|^2}_{\text{Penalty Term}} \right\}$$

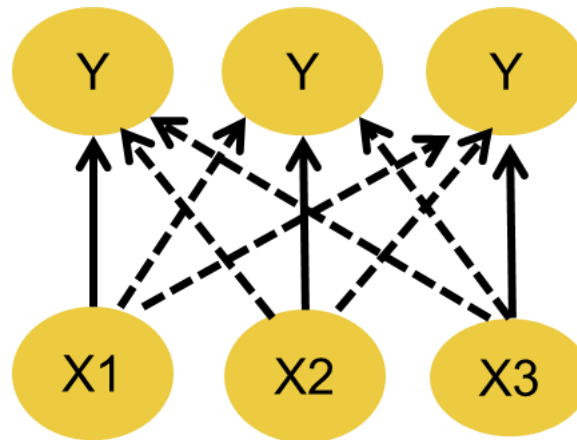
Penalty term

Ridge penalty
= Gaussian
prior on the
weights



Prior
reflecting
covariance
information
in
environment!

Covariance Orthogonalizing Regularization (COR)



Multivariate (= multiple DV's) Regression

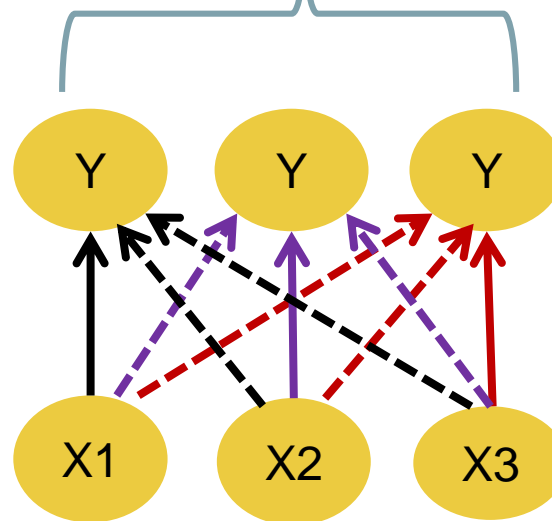
Our latent state variable model

$$Y = w_{11} \cdot X_1 + w_{21} \cdot X_2 + w_{31} \cdot X_3$$

$$Y = w_{12} \cdot X_1 + w_{22} \cdot X_2 + w_{32} \cdot X_3$$

$$Y = w_{13} \cdot X_1 + w_{23} \cdot X_2 + w_{33} \cdot X_3$$

-> like doing linear regression 3 times!



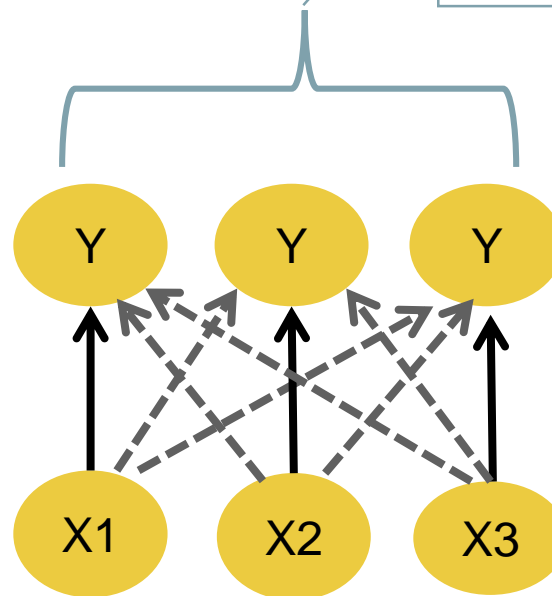
Multivariate (= multiple DV's) Regression

Our latent state variable model

$$Y = w_{11}X_1 + w_{21}X_2 + w_{31}X_3$$

$$Y = w_{12}X_1 + w_{22}X_2 + w_{32}X_3$$

$$Y = w_{13}X_1 + w_{23}X_2 + w_{33}X_3$$



---> = Cross-connections

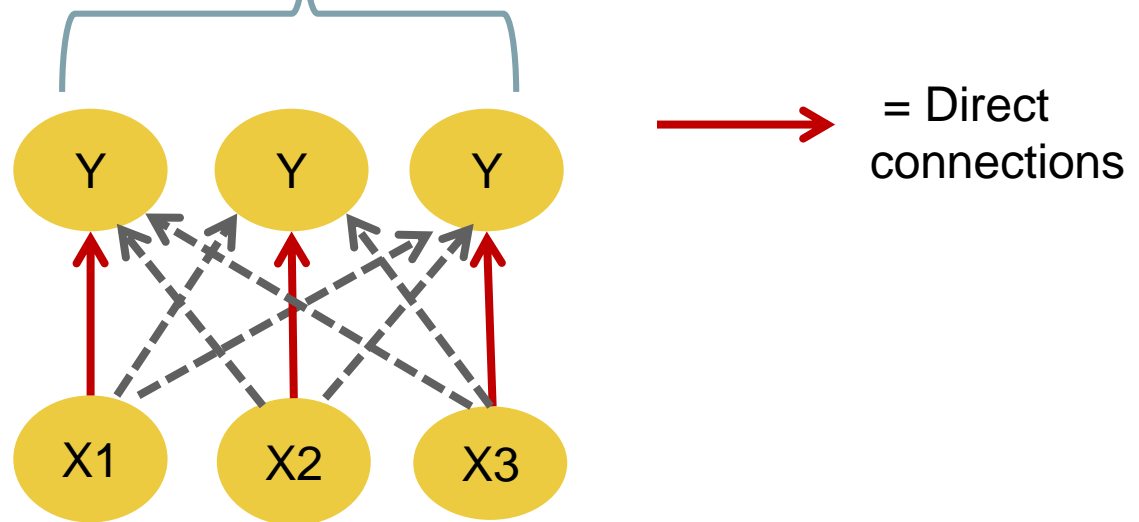
--> contain the **covariance.**

Our latent state variable model

$$Y = w_{11} * X_1 + w_{21} * X_2 + w_{31} * X_3$$

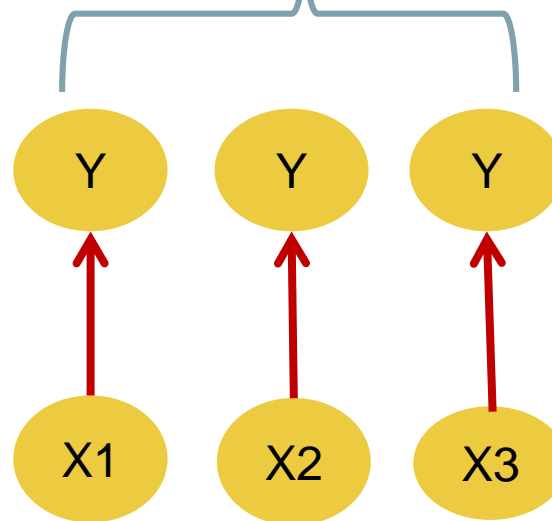
$$Y = w_{12} * X_1 + w_{22} * X_2 + w_{32} * X_3$$

$$Y = w_{13} * X_1 + w_{23} * X_2 + w_{33} * X_3$$



Our latent state variable model

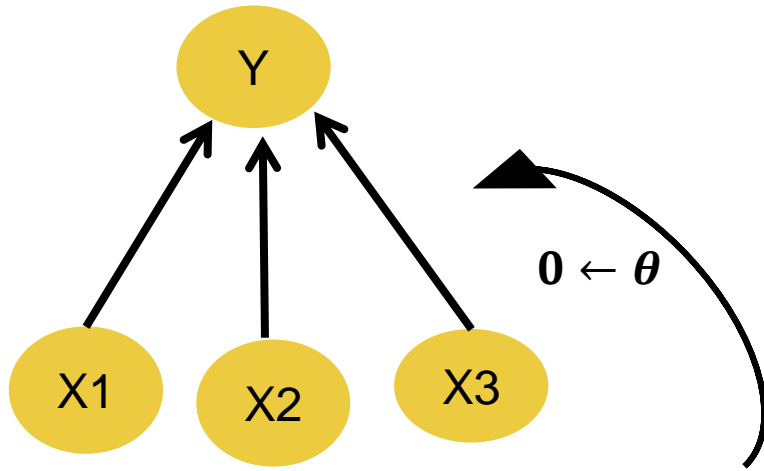
$$\begin{aligned}
 Y &= w_{11} * X_1 + w_{21} * X_2 + w_{31} * X_3 \\
 Y &= w_{12} * X_1 + w_{22} * X_2 + w_{32} * X_3 \\
 Y &= w_{13} * X_1 + w_{23} * X_2 + w_{33} * X_3
 \end{aligned}$$



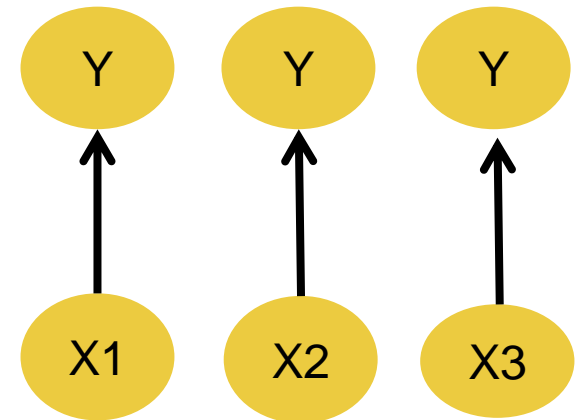
→ = Direct connections

→ **no covariance** estimated!

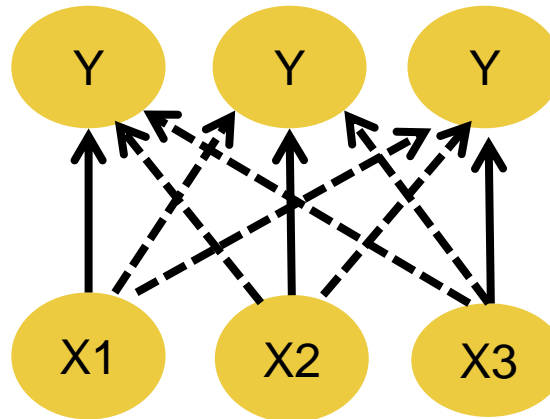
Our Bayesian model



Linear Regression
(high covariance)



Cue validities
(no covariance)



Multivariate Linear Regression



Our Bayesian model

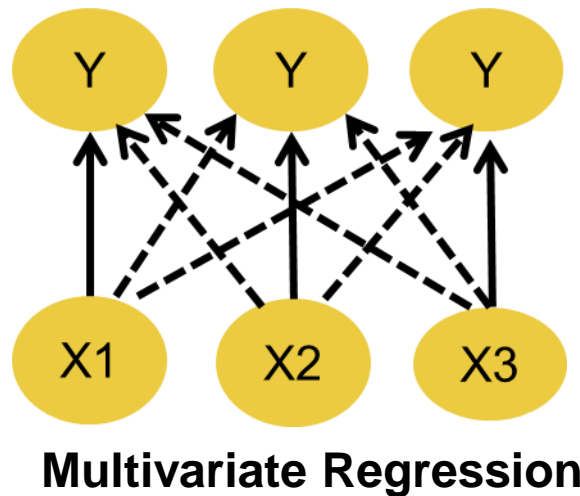
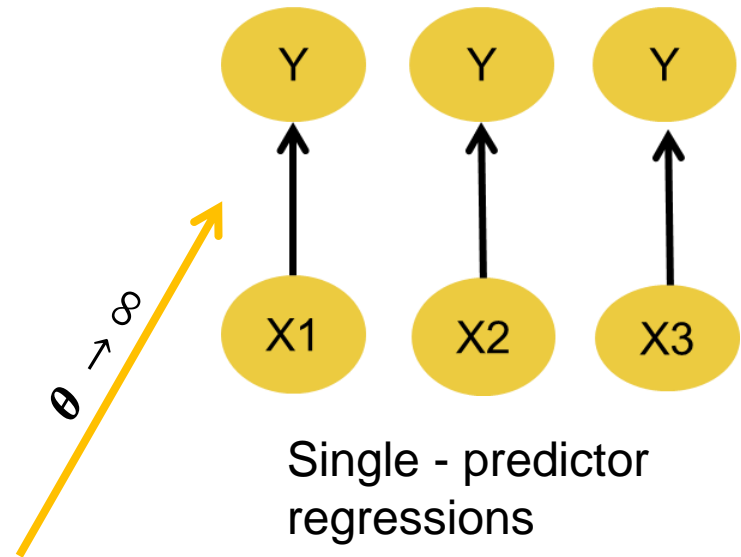
Prior

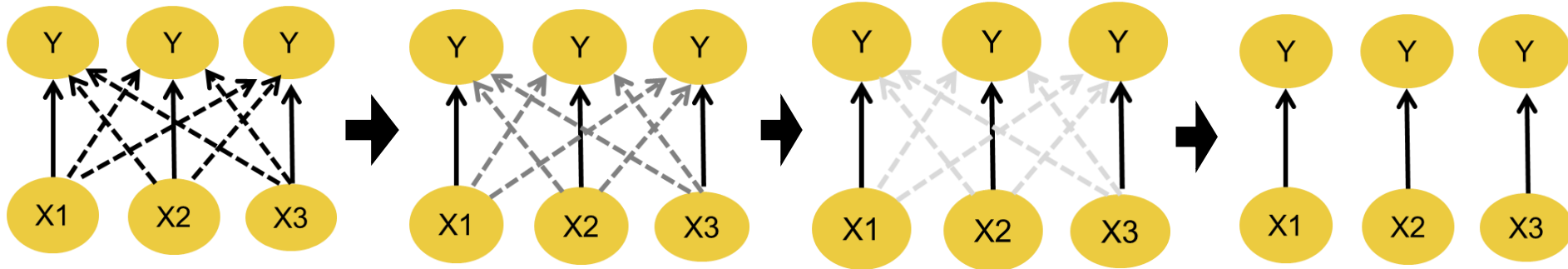
- In analogy to ridge regression:

$$\text{Prior} = -\theta * \left[\sum_{i=1}^m \sum_{j=1}^m |w_{ij}|^2 - \text{tr}(W^2) \right]$$

Log Likelihood: Multivariate Normal

$$\ln P_{X,W}(Y_i) \propto -\frac{1}{2} \sum_{i=1}^n (Y_i - XW)^T C^{-1} (Y_i - XW)$$





$\theta = 0$

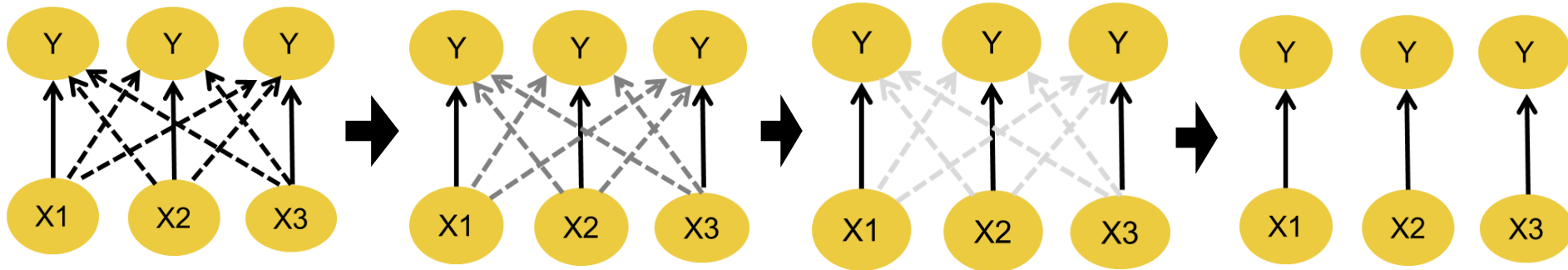
$\theta = 10$

$\theta = 50$

$\theta = 100$



penalty parameter θ



$\theta = 0$

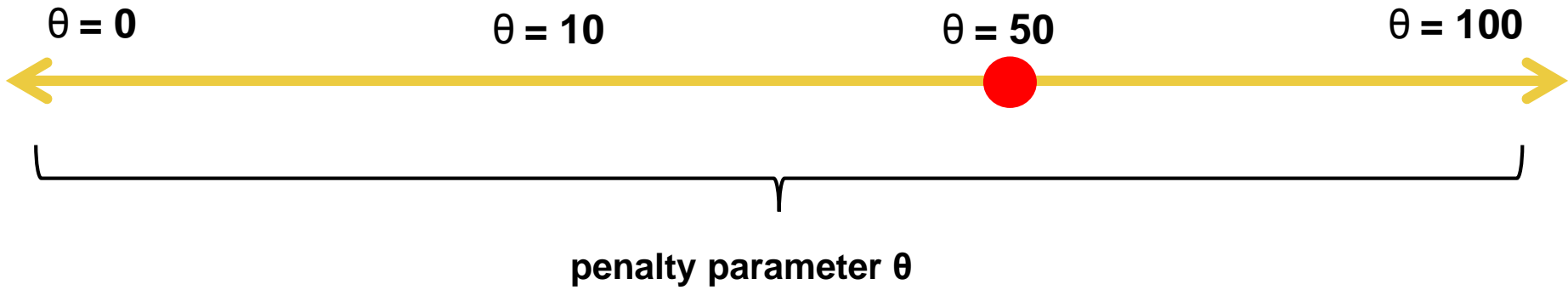
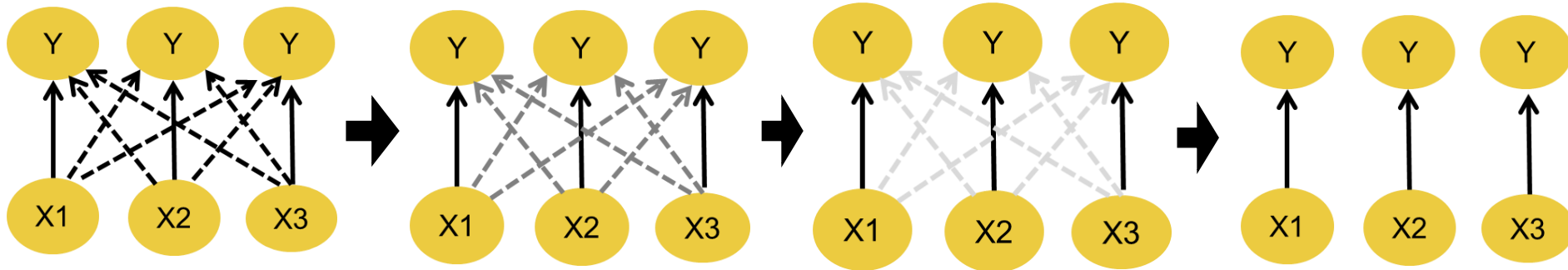
$\theta = 10$

$\theta = 50$

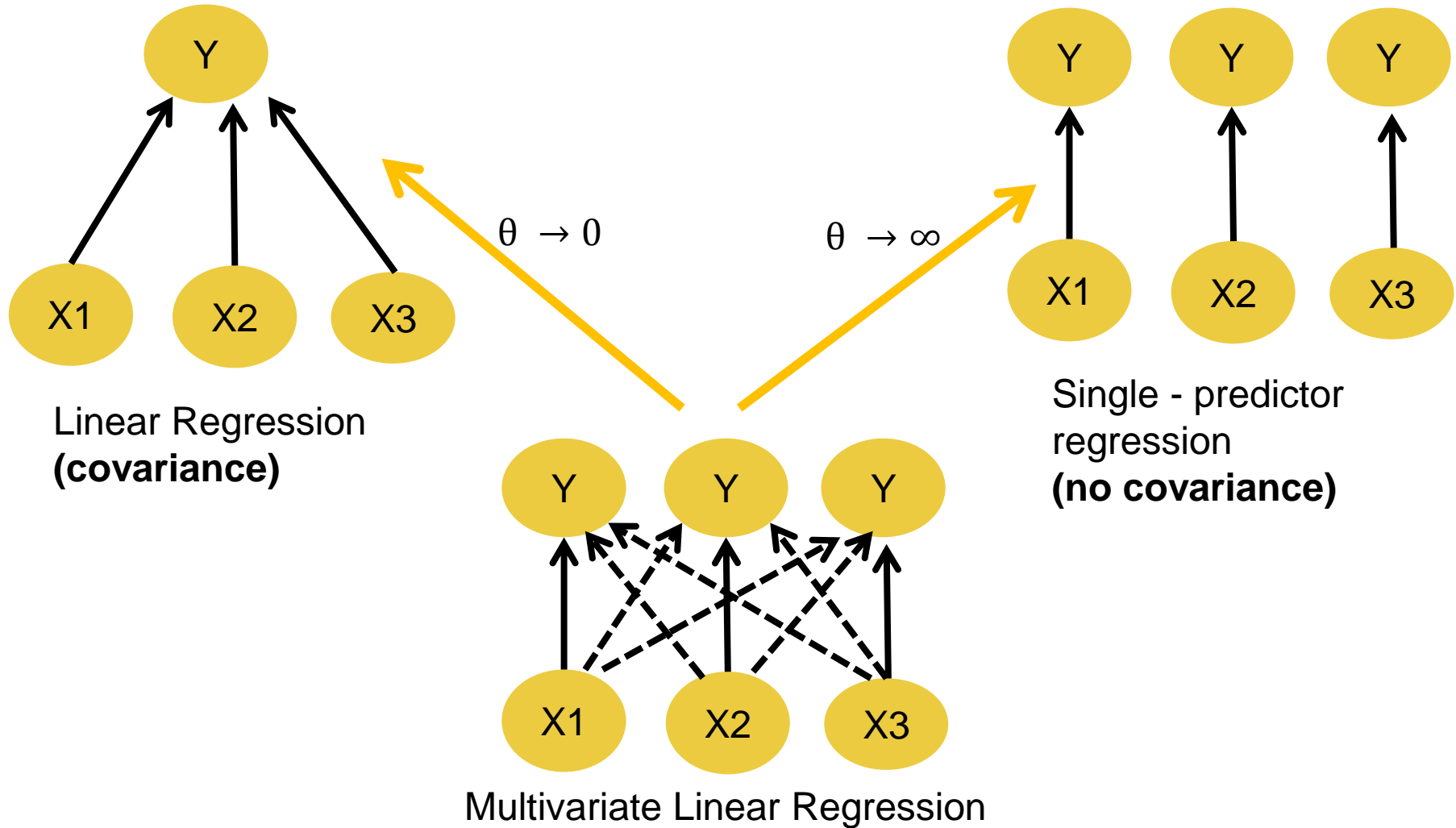
$\theta = 100$



penalty parameter θ

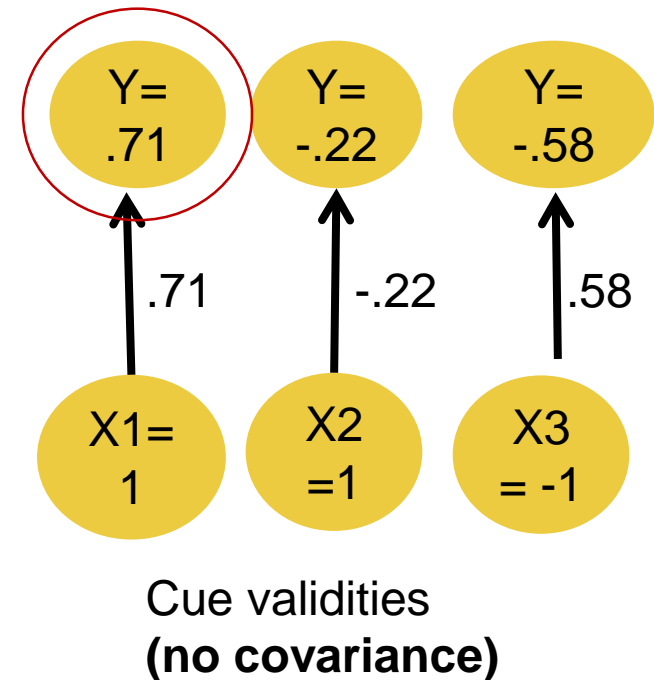


Our Bayesian model



TTB decision rule

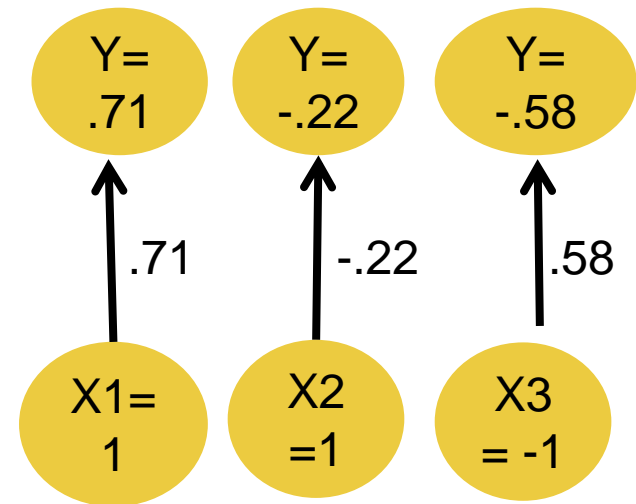
- **Single predictor weights = cue validities.**
- Find the $\max(\text{absolute}(Y))$, and take the sign.



Tallying decision rule

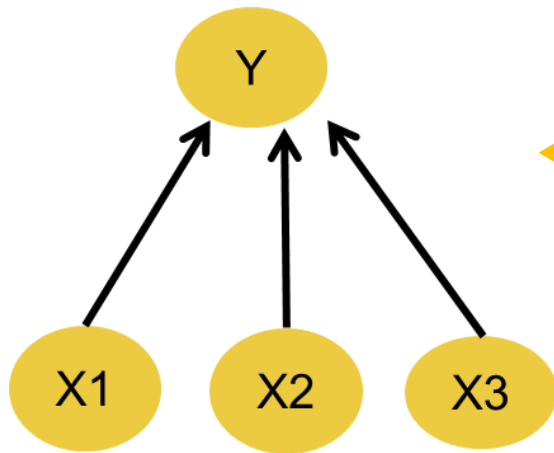
- Count the signs of the outputs Y .

$$= \text{sign}(\text{sum}(\text{sign}(Y)))$$
- Tallying would count: $+1-1-1 = -1$.



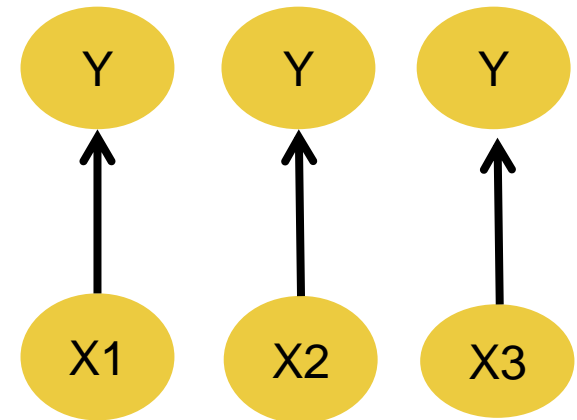
Cue validities (**no covariance**)

What is linear regression?



Linear Regression

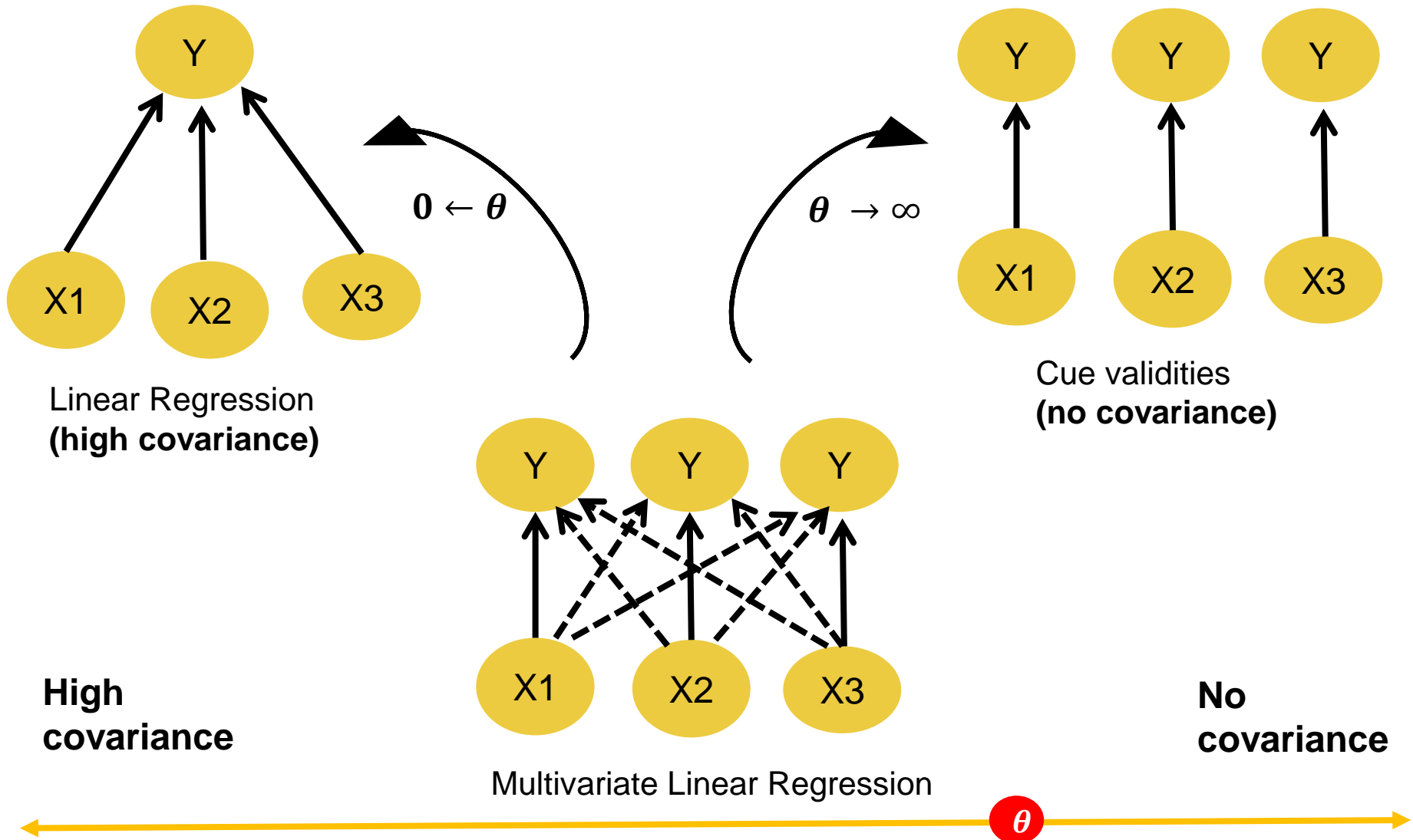
= when $\theta \rightarrow 0$



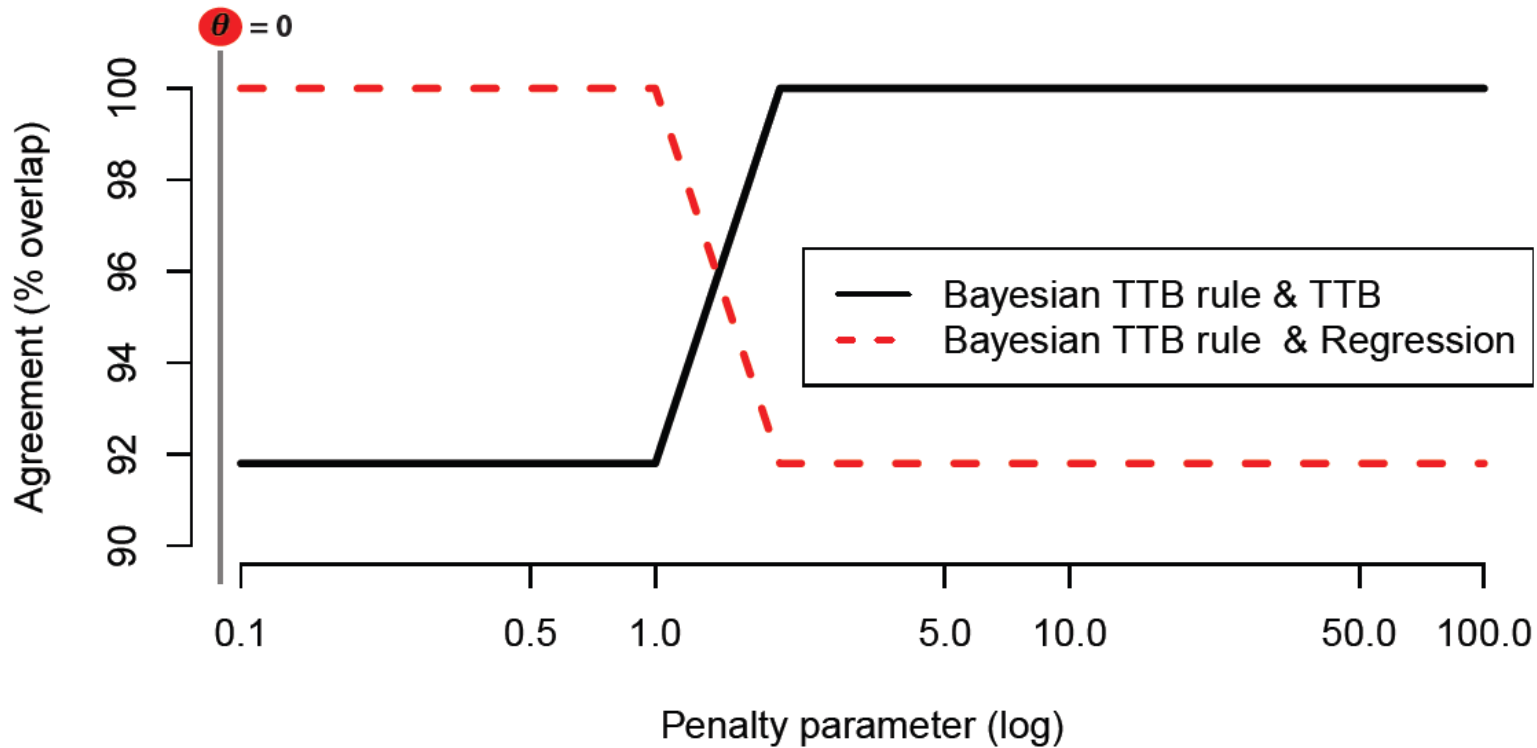
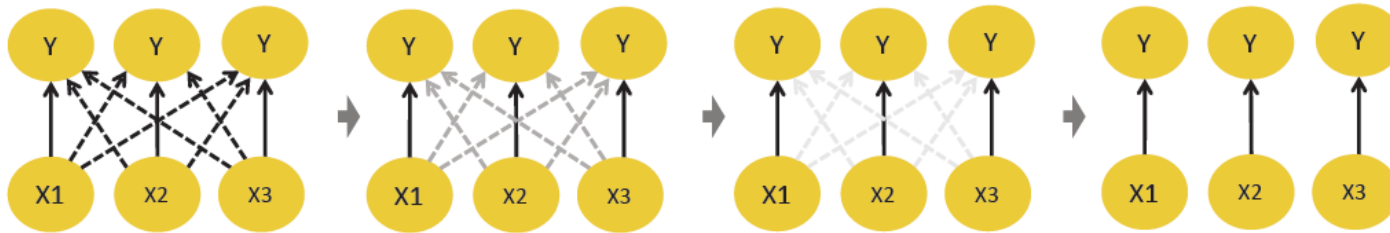
Single - predictor regression

- Regular Linear Regression is either heuristic decision rule (TTB or Tallying decision rule) when the penalty term is zero.

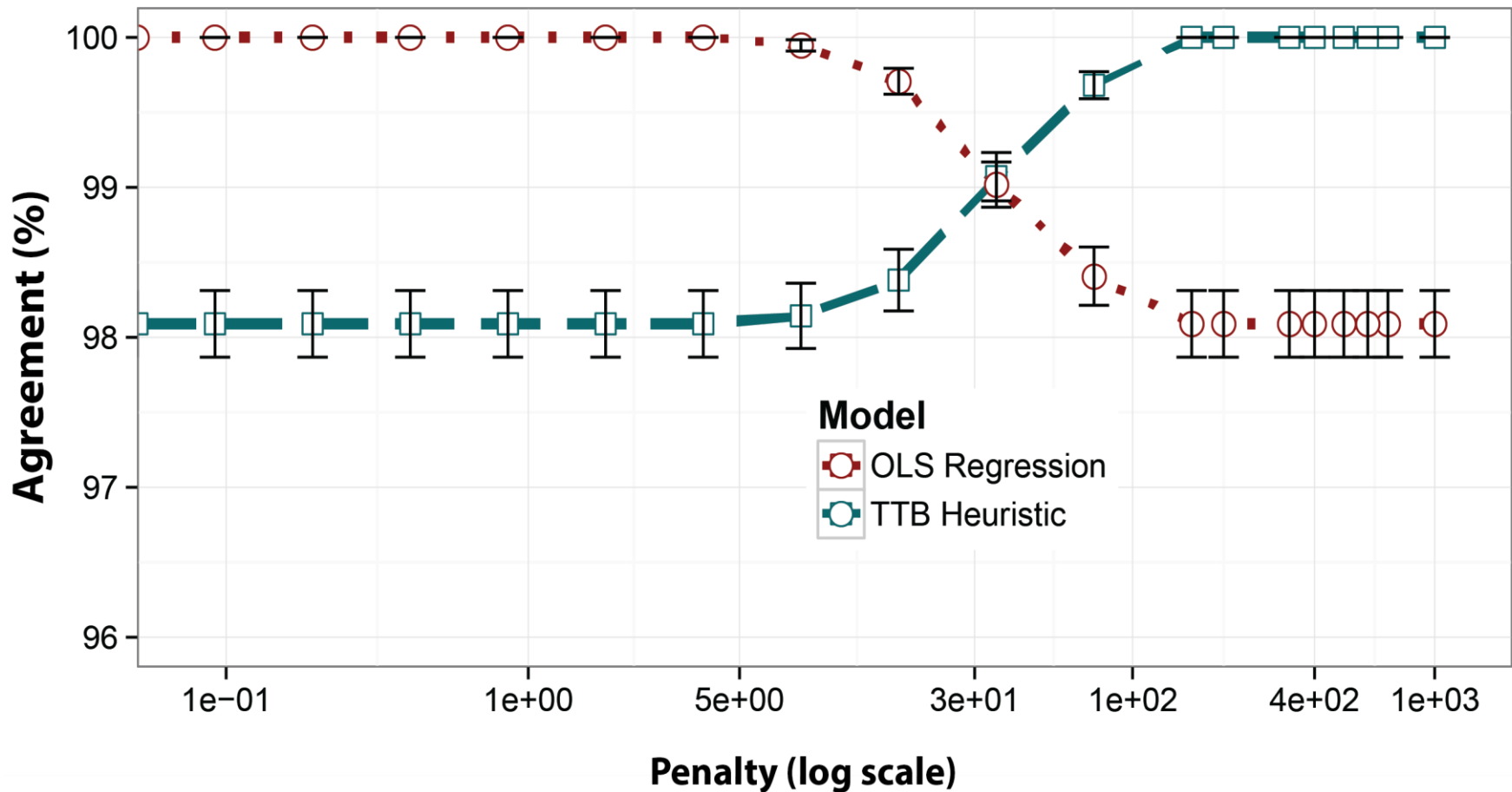
Continuum between heuristics and LR



Convergence of Bayesian model with Take-the-Best and Linear Regression



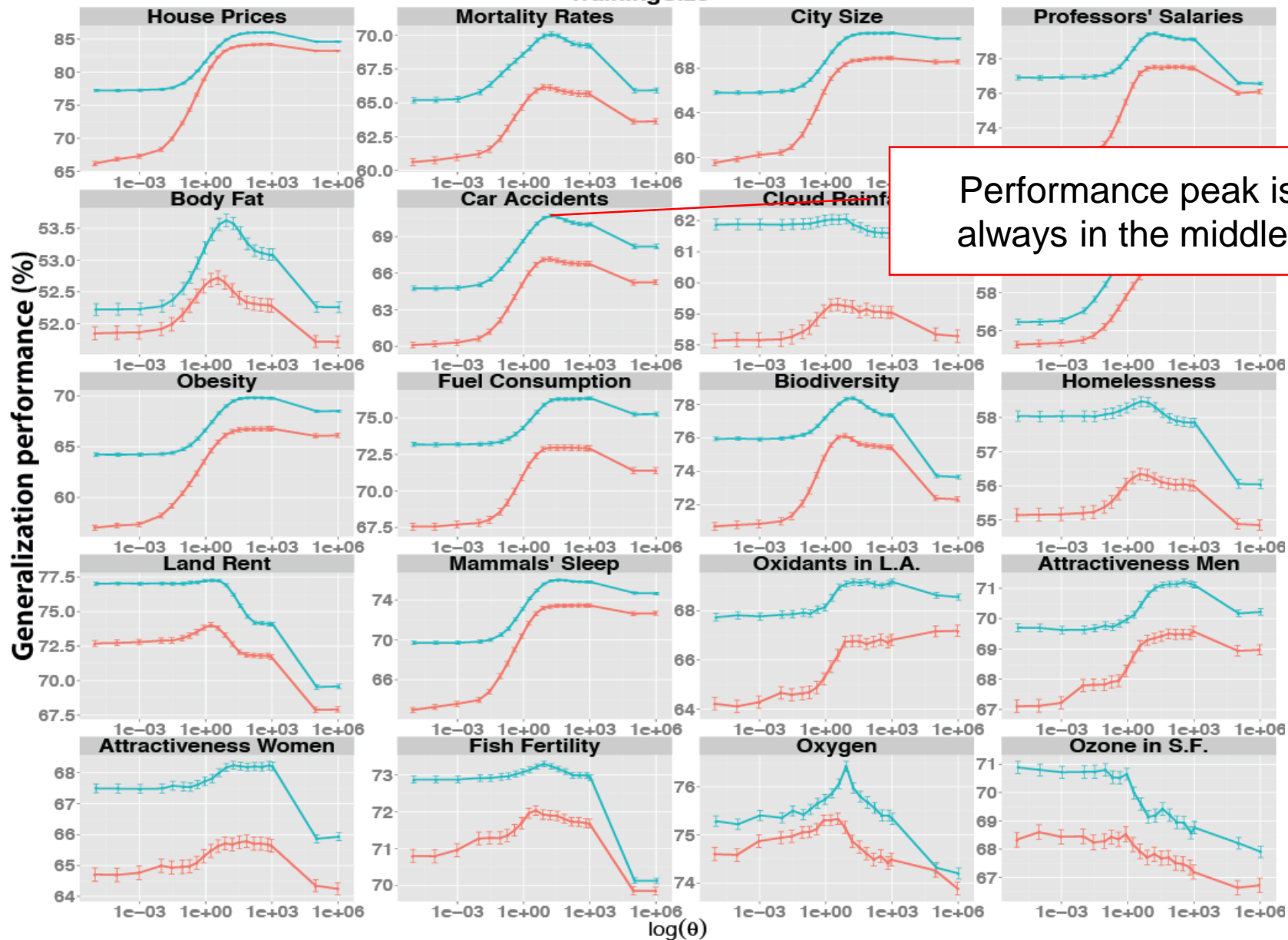
Agreement between Bayesian model and Take-The-Best Heuristic



How does the Bayesian COR model perform compared to heuristics and linear regression?

We test this on the famous twenty ABC datasets (Gigerenzer et al., 1999)

TrainingSize = 10-20



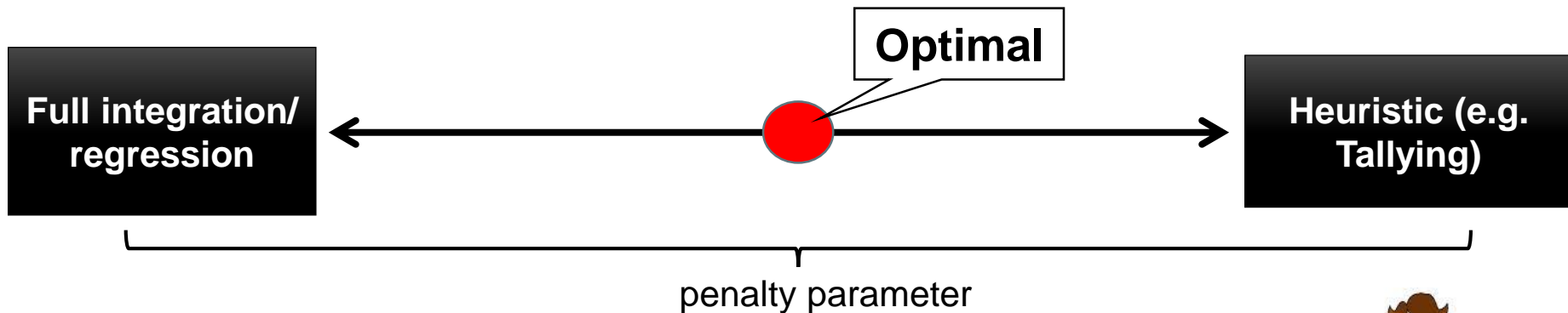
What happens in between (for moderate penalty)?

- Of course, everything in between will occur.
- The optimum is often in the middle, i.e. **not** zero covariance or high covariance estimation, but a little bit.



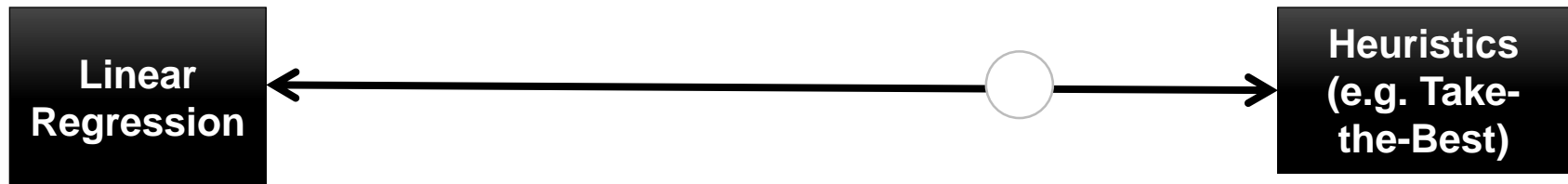
Optimal strategy depends on the environment

- Peak in the middle suggests that true environmental structure and potentially psychological processing often lies somewhere between the assumptions of heuristic and standard regression approaches.



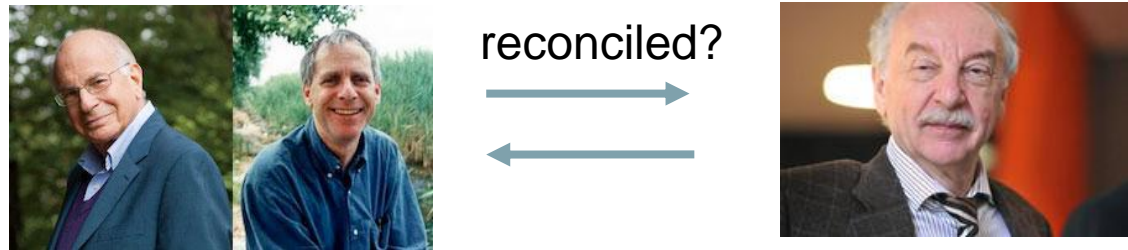
CONCLUSIONS

Heuristics AND traditional linear regression are a special case of a Bayesian inference model. They can be seen as two extreme positions on a continuum of decision strategies:



CONCLUSIONS

1. Heuristics represent extreme cases on a Bayesian inference model.
→ Heuristics = Bayesian Inference.



2. **We showed that less is never more.** → Heuristics are outperformed by a prior of finite strength that learns from the training data but nonetheless down-weights that information.
 - The strongest form of less-is-more, i.e., that one can do better with heuristics by throwing out information, is false.

Reconciled?

...maybe Kahneman is pleased to see that heuristics end up as a special case of a probabilistic inference model.



Daniel Kahneman & Amos Tversky (1974, 1981, 2003)

...maybe Gigerenzer and colleagues are pleased to find that provably the best strategy in some environments is a heuristic.

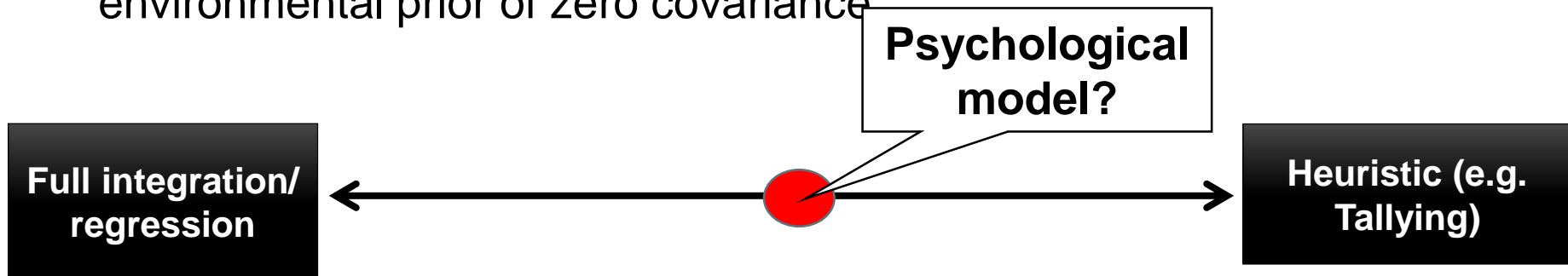


Gerd Gigerenzer, Peter Todd & abc research group (1999)

IMPLICATIONS

This provides an explanation for why and *when* heuristics work.

- A central message of this work is that ignoring information is rarely optimal.
- Heuristics may work well in practice because they correspond to an environmental prior of zero covariance



- One question for future research is whether heuristics give an accurate characterization of psychological processing, **or whether actual psychological processing is more akin to these more complex intermediate models.**

DISCUSSION

- What could this mean on a psychological level?
(→ Note that the framework presented here is merely on a formal, computational theory.)
- If you were the scientist, what would you do next?
- What are the big implications of this research?

THANK YOU!

DISCUSSION

What could this mean on a psychological level?

1. On the one hand, it could be that implementing the intermediate models is computationally intractable, and thus the brain uses heuristics because they efficiently approximate these more optimal models. → Heuristics- and- biases approach
2. On the other hand, it could be that the brain has tractable means for implementing the intermediate models (i.e., for using all available information but down-weighting it appropriately).
→ This case would be congruent with the view from ecological rationality where the brain's inferential mechanisms are adapted to the statistical structure of the environment.

Marr's Levels of Analysis (1982)

- Marr 's (1982) 3 levels of analysis:
 - computational level
 - process/algorithmic level
 - neuronal/implementational level
- Modeling takes place only at computational level
→ will be integrated with process level research later

Implications

- The current model can help with a prescriptive analysis:
When should people rely on a given heuristic rather than a complex strategy?
- Now we can answer the question of when it is helpful and harmful to use simple shortcuts, because we have a Bayesian model that tells us what strategy is optimal in what environment (it answers the question of ecological rationality).

When is it helpful and harmful to rely on heuristics?



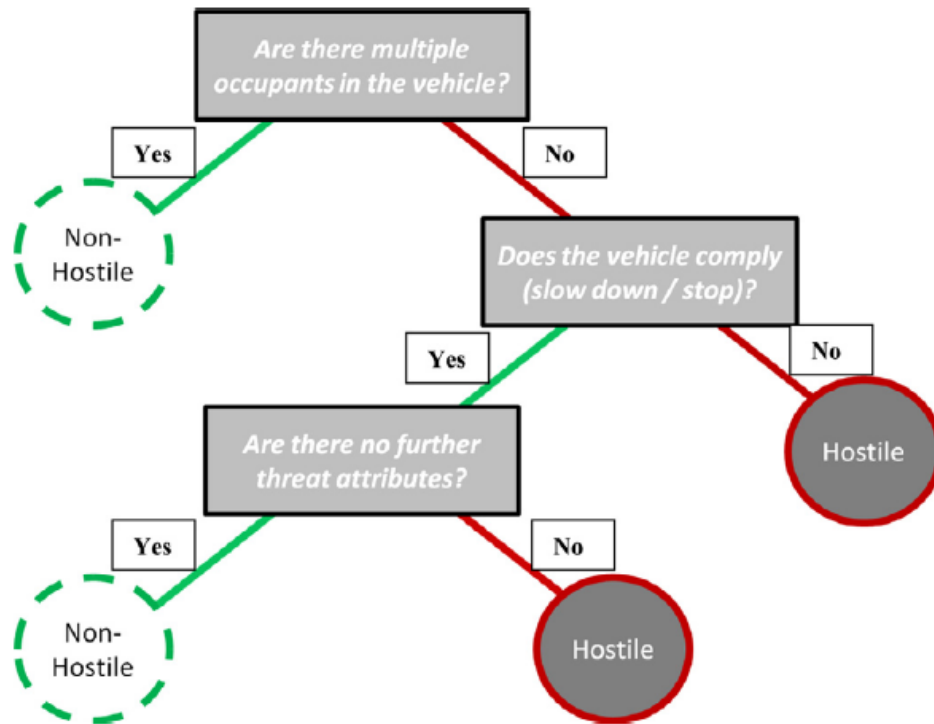
- Heuristics can reduce the complexity of decisions by a lot, which is required when decisions have to be made quickly.

When is it helpful and harmful to rely on heuristics?



- Heuristics at war.
(see Keller & Katsikopoulos, 2016)

When is it helpful and harmful to rely on heuristics?



- Heuristic Decision Trees for high stress situations. (Keller & Katsikopoulos, 2016)

Fig. 4. A fast and frugal tree for classifying oncoming traffic as hostile or nonhostile. The tree was constructed before obtaining the reports on the Afghanistan incidents.

Examples: Where are these heuristics used?

- Take-The-Best
 - ❖ Predicting consumer choices: Hauser et al. (2009), decisions between computers (Kohli & Jedidi, 2007); smartphones (Yee et al., 2007)
 - ❖ Literature Search (Lee et al., 2002): TTB performed as well as a Bayesian search algorithm
- Tallying
 - ❖ Detecting Strokes: Bedside eye exam could outperform MRI scans (Kattah et al. 2009)
 - ❖ Avoiding avalanche accidents: check how many out of seven cues have been observed en route or on the slope (McCammon & Haegeli 2007). When > 3 cues are present, the situation is considered dangerous. 92% of historical accidents could have been prevented with this strategy.
- Recognition
 - ❖ Predicting elections (Gaissmaier & Marewski, 2010)
 - ❖ Investment (recognition-based portfolios) (Ortman et al., 2008)
 - ❖ Predicting Wimbledon (Serwe & Frings, 2006)

References

1. Simon HA (1990) Invariants of human behavior. *Annual review of psychology* 41:1-19.
2. Czerlinski J, Gigerenzer G, & Goldstein DG (1999) How good are simple heuristics? *Simple heuristics that make us smart*, eds Gigerenzer G, Todd PM, & Gigerenzer AR (Oxford University Press, New York), pp 97–118.
3. Tversky A & Kahneman D (1974) Judgment under Uncertainty: Heuristics and Biases. *Science* 185(4157):1124-1131.
4. Gigerenzer G, Todd PM, & Gigerenzer AR (1999) *Simple heuristics that make us smart* (Oxford University Press).
5. Gigerenzer G & Goldstein DG (1996) Reasoning the fast and frugal way: models of bounded rationality. *Psychological review* 103(4):650-669.
6. Tversky A (1972) Elimination by aspects: A theory of choice. *Psychological review* 79(4):281.
7. Dawes RM (1979) The robust beauty of improper linear models in decision making. *American psychologist* 34(7):571.
8. Dawes RM & Corrigan B (1974) Linear models in decision making. *Psychological bulletin* 81(2):95.
9. Kahneman D (2003) A perspective on judgment and choice: mapping bounded rationality. *The American psychologist* 58(9):697-720.
10. Goldstein DG & Gigerenzer G (2002) Models of ecological rationality: the recognition heuristic. *Psychological review* 109(1):75-90.
11. Gigerenzer G & Brighton H (2009) Homo heuristicus: why biased minds make better inferences. *Topics in cognitive science* 1(1):107-143.
12. Einhorn HJ & Hogarth RM (1975) Unit weighting schemes for decision making. *Organizational Behavior and Human Performance* 13(2):171-192.
13. Chater N, Oaksford M, Nakisa R, & Redington M (2003) Fast, frugal, and rational: How rational norms explain behavior. *Organizational behavior and human decision processes* 90(1):63-86.
14. Katsikopoulos KV, Schooler LJ, & Hertwig R (2010) The robust beauty of ordinary information. *Psychological review* 117(4):1259.
15. Gigerenzer G & Gaissmaier W (2011) Heuristic decision making. *Annual review of psychology* 62:451-482.
16. Martignon L & Hoffrage U (1999) Why does one-reason decision making work. A case study in ecological rationality. *Simple heuristics that make us smart*, (Oxford University Press, New York), pp 119-140.
17. Hogarth RM & Karelaia N (2007) Heuristic and linear models of judgment: matching rules and environments. *Psychological review* 114(3):733-758.

18. Gigerenzer G (2008) Why heuristics work. *Perspectives on psychological science* 3(1):20-29.
19. Kohavi R (1995) A study of cross-validation and bootstrap for accuracy estimation and model selection. *Ijcai*, pp 1137-1145.
20. Pitt MA & Myung IJ (2002) When a good fit can be bad. *Trends in cognitive sciences* 6(10):421-425.
21. Geman S, Bienenstock E, & Doursat R (1992) Neural networks and the bias/variance dilemma. *Neural computation* 4(1):1-58.
22. Dieckmann A & Rieskamp J (2007) The influence of information redundancy on probabilistic inferences. *Memory & cognition* 35(7):1801-1813.
23. Rieskamp J & Dieckmann A (2012) Redundancy: Environment structure that simple heuristics can exploit. *Ecological rationality: Intelligence in the world*, (Oxford University Press, New York), pp 187-215.
24. Hoerl AE & Kennard RW (1970) Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics* 12(1):55-67
25. Gelman A, Carlin JB, Stern HS, & Rubin DB (2014) *Bayesian data analysis* (Taylor & Francis).
26. Todd PM & Gigerenzer G (2012) *Ecological rationality: Intelligence in the world* (Oxford University Press).
27. Marr D (1982) *Vision: A computational investigation into the human representation and processing of visual information*. (Freeman, San Francisco).
28. Brown SD & Steyvers M (2009) Detecting and predicting changes. *Cognitive psychology* 58(1):49-67.
29. Daw N & Courville A (2008) The Pigeon as Particle Filter. in *Advances in neural information processing systems*, eds Platt J, Koller D, Singer Y, & Roweis S (MIT Press), pp 369–376.
30. Jones M & Love BC (2011) Bayesian Fundamentalism or Enlightenment? On the explanatory status and theoretical contributions of Bayesian models of cognition. *The Behavioral and brain sciences* 34(4):169-188; discussion 188-231.
31. Lee MD & Cummins TD (2004) Evidence accumulation in decision making: unifying the "take the best" and the "rational" models. *Psychonomic bulletin & review* 11(2):343-352.
32. Sanborn AN, Griffiths TL, & Navarro DJ (2010) Rational approximations to rational models: alternative algorithms for category learning. *Psychological review* 117(4):1144.
33. Scheibehenne B, Rieskamp J, & Wagenmakers E-J (2013) Testing adaptive toolbox models: A Bayesian hierarchical approach. *Psychological review* 120(1):39.
34. van Ravenzwaaij D, Moore CP, Lee MD, & Newell BR (2014) A hierarchical bayesian modeling approach to searching and stopping in multi-attribute judgment. *Cognitive science* 38(7):1384-1405.
35. Griffiths TL, Lieder F, & Goodman ND (2014) Rational use of cognitive resources: Levels of analysis between the computational and the algorithmic. *Topics in Cognitive Science*. forthcoming.