Heuristics as Bayesian Inference

Paula Parpart

29.02.2016, KLI Lecture

Collaborators:

Prof Brad Love – UCL

Prof Matt Jones – University of Colorado

Takao Noguchi – UCL

UK PhD Centre in Financial Computing & Analytics

Overview

1. Fast and frugal Heuristics: what defines them?
2. Rational Models of Cognition
3. Are heuristics compatible with Bayesian inference?
4. Our Bayesian Model
5. Simulation Results
6. Discussion: Q & A
7. Implications

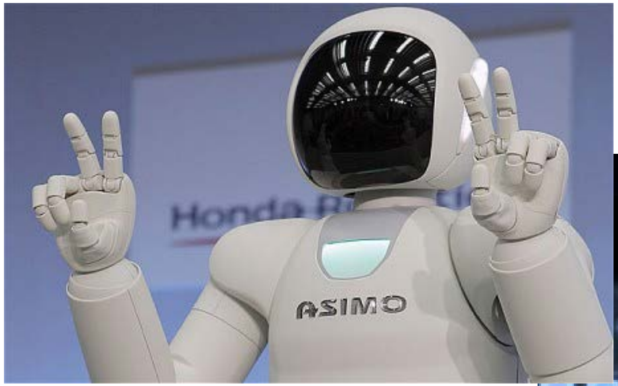
Heuristics

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

Wübben and Wangenheim (2008)

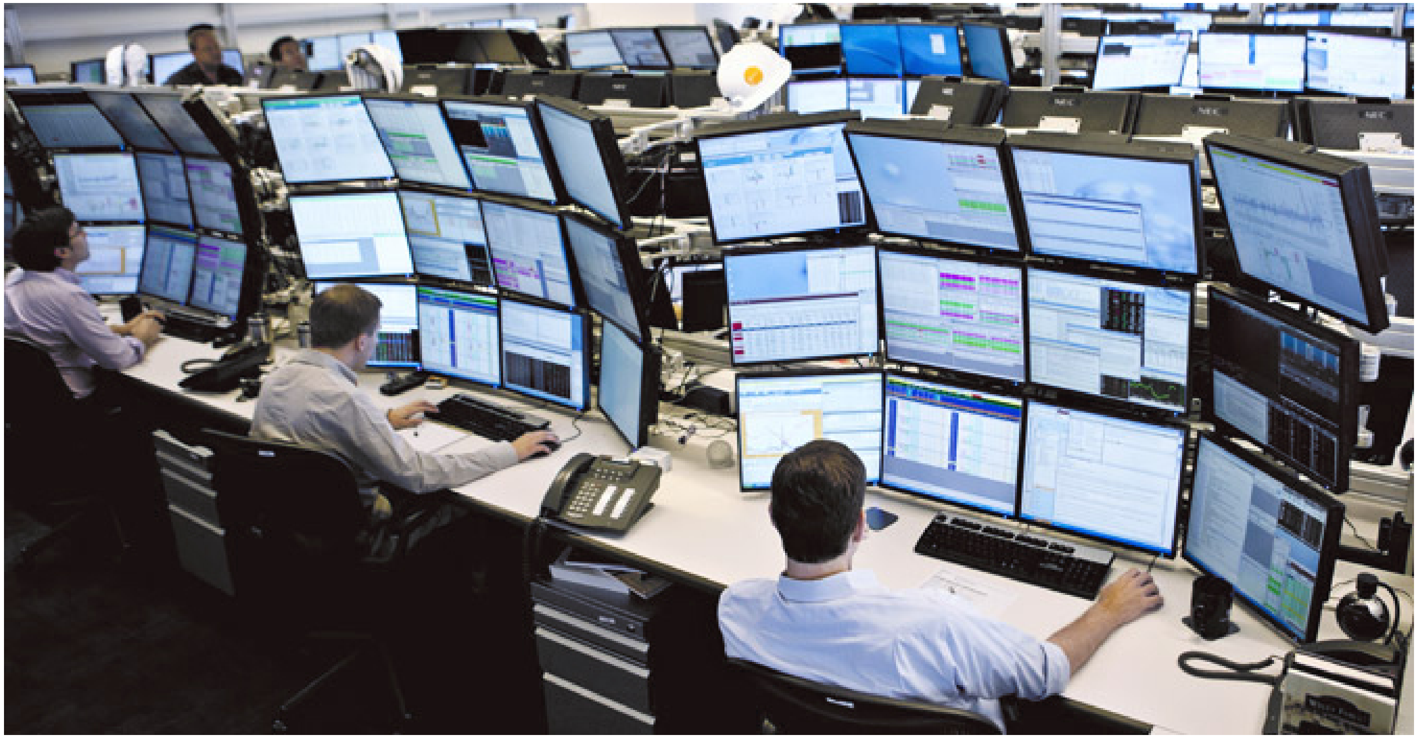
Why are Heuristics so important in AI and Computer Science?

- 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)

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

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

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 - 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)

Heuristics are often contrasted with *rational* decision procedures that make full and proper use of available information.

- Rational Models of Cognition
  - Bayesian inference models

  → Optimal inference as benchmark to compare human behaviour against.

  - Linear Regression aka. WADD = weighted linear additive (e.g., Czerlinski et al., 1999): It is debatable whether this really a “rational” model of cognition?

Take-The-Best Heuristic

What team will win the game?

Cues	*v*	DEUTSCHER FUSSBALL-BUND	ENGLAND	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 co-variance among cues

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

Regression uses regression weights (OLS). Heuristics use cue validities as weights.

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

R = number of correct predictions,

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

A

Cues	*v*	DEUTSCHER FUSSBALL-BUND	ENGLAND	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

Fundamental difference between Heuristics and Linear Regression

- Cue validities ignore co-variance among cues!
- Cue validities are computed in isolation of one another.

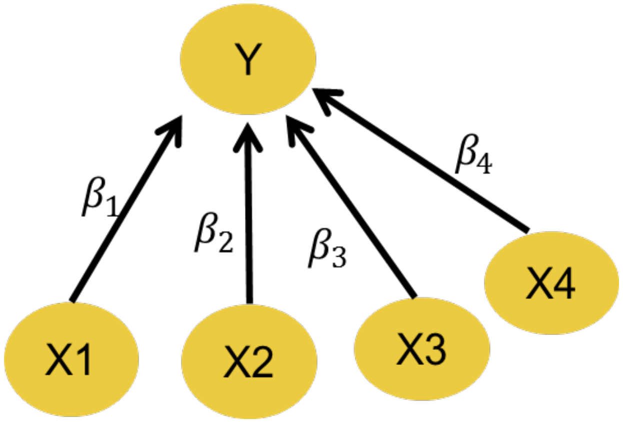
A

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

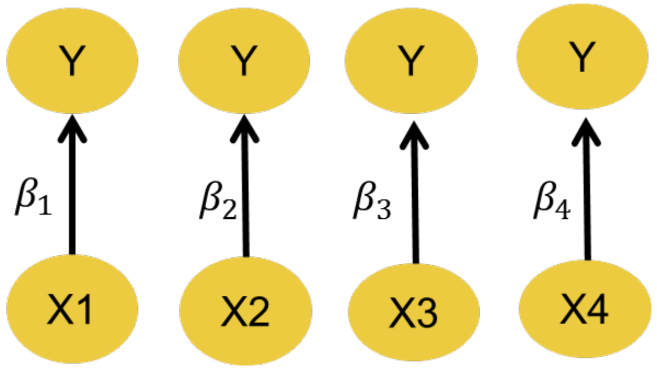
The difference between cue validities and regression weights:

<u>Multiple Regression weights:</u>



β1	β 2	β 3	β 4
-0.29	1.16	-0.11	0.08

<u>Single - predictor regressions weights/Cue validities</u>



β 1	β 2	β 3	β 4
0.00	1.00	0.25	0.71

Note: cue validities are a linear transformation of single predictor regression coefficients. They ignore any dependencies among cues.

Prominent notions of heuristics

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

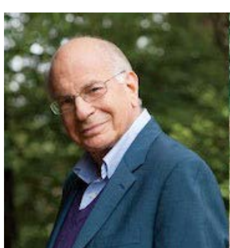
Gerd
Gigerenzer &
the ABC
research
group (1999)

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

Ecological Approach

I. We combine parts of the theories of Kahneman and Gigerenzer into a more complete view.

I. Thereby, we create a third framework, relying on
- Bayesian Framework: Probability axioms as used by Kahneman & Tversky
- Ecological rationality approach by Gigerenzer et al.,(1999)

II. When combining these rivalry approaches in a Bayesian rationality framework, **you find that heuristics can in fact be seen as special cases of a Bayesian inference model.**

Less-is-more effects: Heuristics can outperform "rational" models

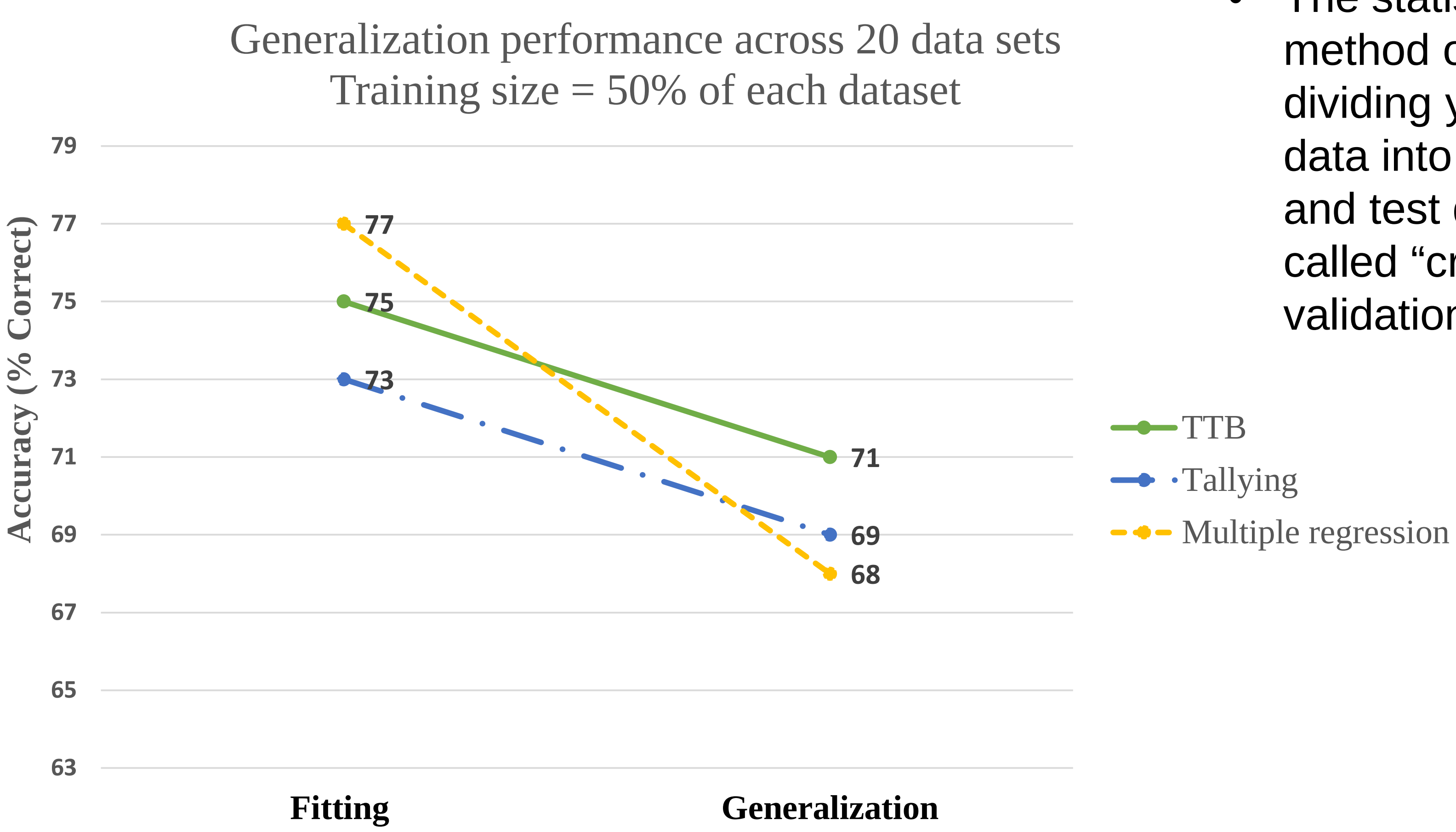
- Czerlinski et al. (1999) showed that heuristics can sometimes outperform "rational" linear multiple regression

- …as well as a three-layer feed-forward connectionist 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 rational methods make full use of the data.

Heuristics vs. "rational" accounts

- The statistical method of dividing your data into training and test data is called "cross-validation"



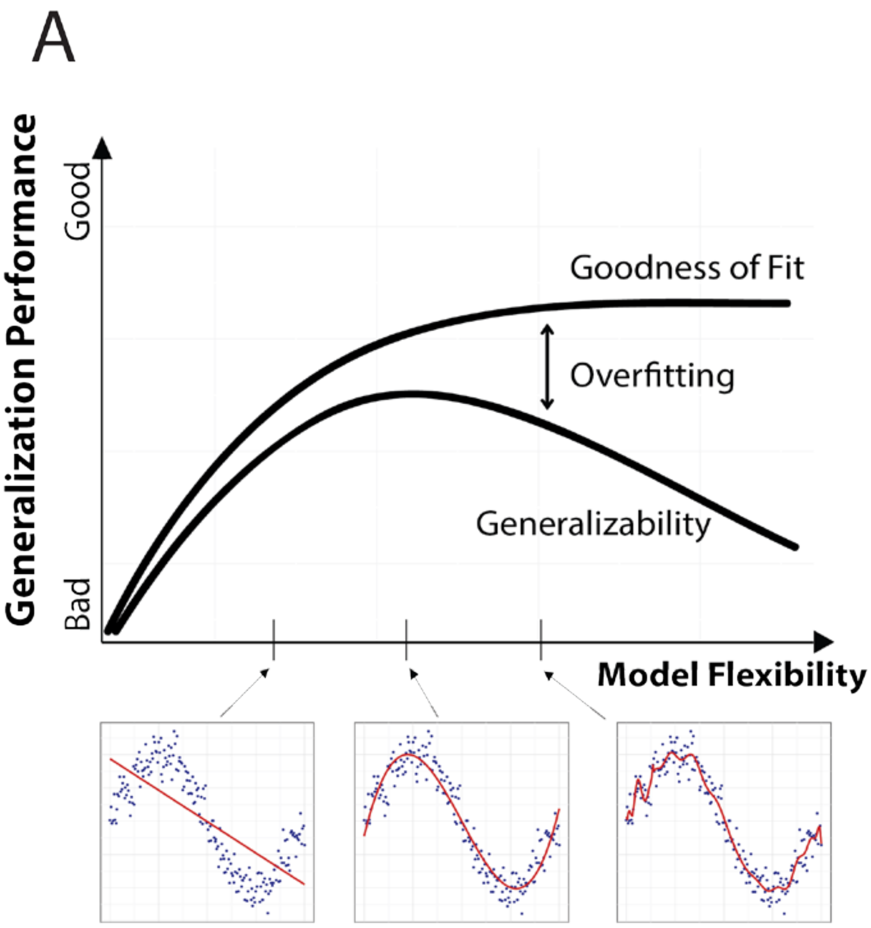
Original data sets (Czerlinski et al., 1999): City size task, professors salaries, High school dropout rates, Homelessness House price, Mortality, Land rent, Car accidents, Fuel consumption, Obesity at age 18, Body Fat, Fish fertility, Mammals' sleep, Cow manure, Biodiversity, Rainfall from cloud, Oxidant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

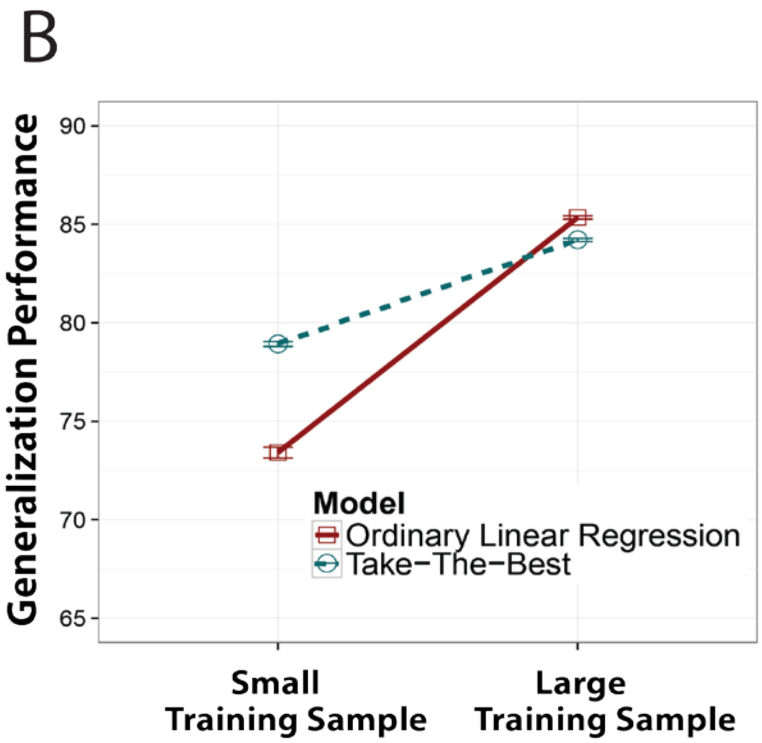
How do Less-is-more effects work?

- According to the ecological rationality framework, these **less-is-more effects (i.e., heuristic performance > rational model's performance)** occur because heuristics are best tuned to certain environments (Gigerenzer, Todd & the ABC research group, 1999)

- From a machine-learning perspective, this conclusion is sensible because every model, including heuristics, has an **inductive *bias*, akin to a Bayesian prior**, that makes the model best-suited to certain learning problems.

Why can simple heuristics sometimes outperform more complex algorithms?





<u>bias-variance tradeoff:</u>

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

(Adopted from Pitt & Myung, 2002)

How do Less-is-more effects work?

- 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.

- 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)

bias-variance tradeoff:

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

Overfitting

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

<u>bias-variance dilemma</u>

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

Overfitting

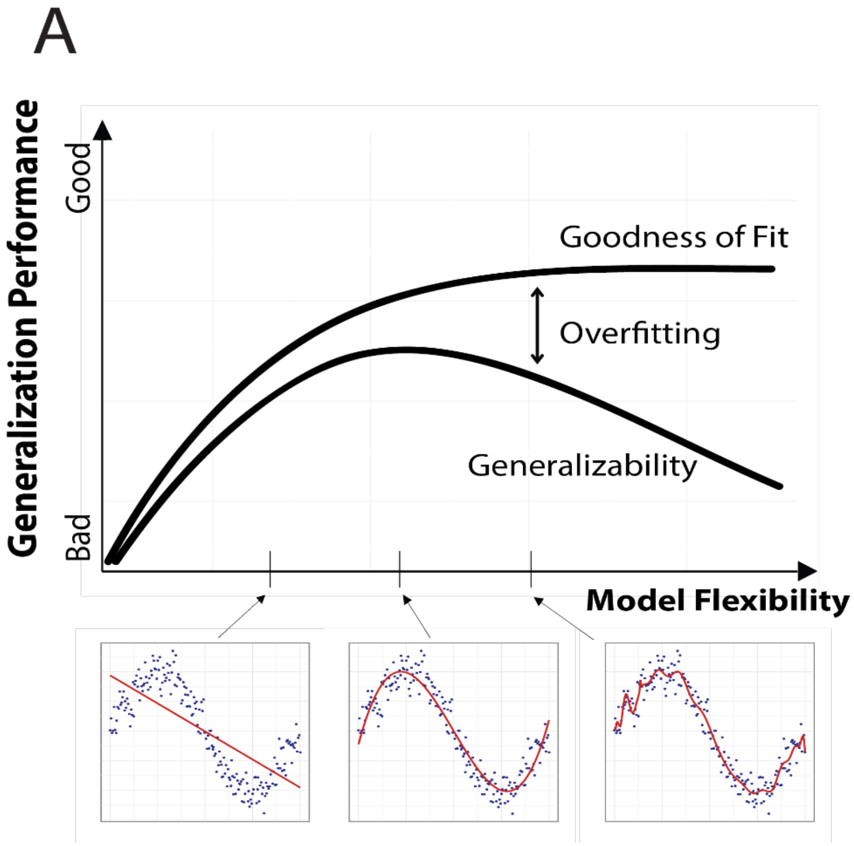
- Bias and variance trade off with one another such that models with low bias suffer from high variance and vice versa

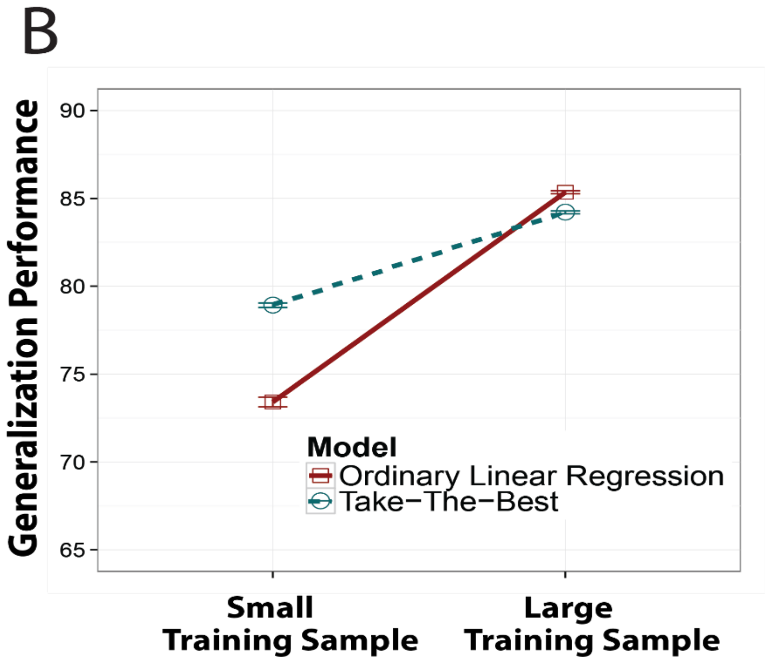
→ implies that more flexible (i.e., less biased) models will overfit small training samples and can be bested by simpler (i.e., more biased) models, such as heuristics.

<u>bias-variance dilemma</u>

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

Why can simple heuristics sometimes outperform more complex algorithms?





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

- However, as the size of the training sample increases, more complex models should fare better. → in a reanalysis of a dataset favoring a heuristic over linear regression, we find that the advantage for the heuristic disappears when training sample size is increased (Figure B)

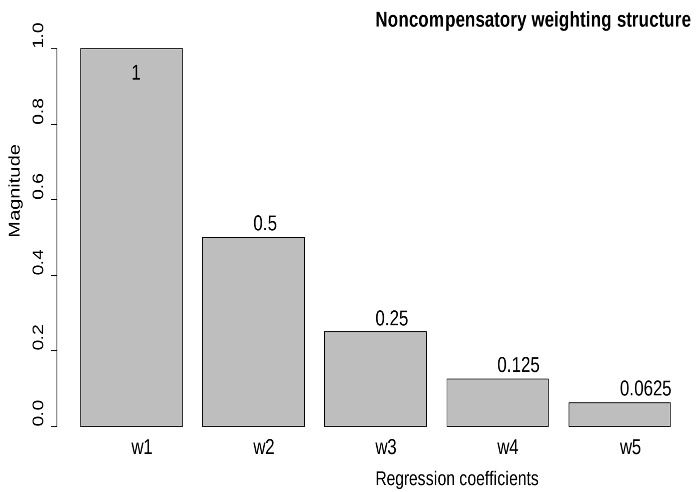
To summarize:

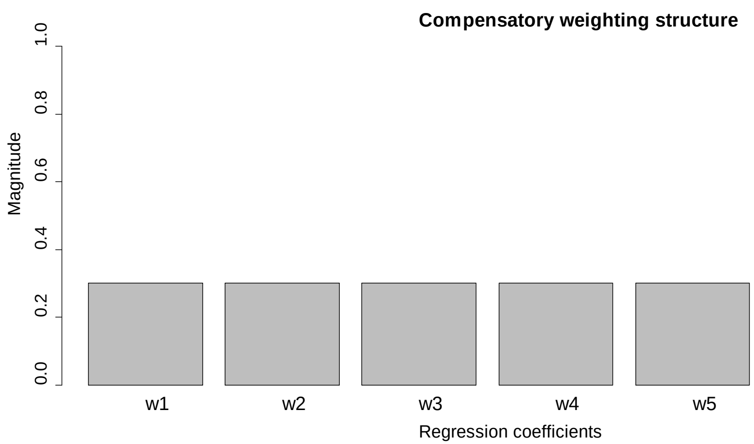
- Under some conditions, heuristics outperform multiple regression.
- Under some conditions, multiple regression outperforms heuristics.

> ➢ **We need to move beyond demonstrations like these, and get a deeper, formal understanding that is general and powerful.**
>
> ➢ **When** and **why** do heuristics work? (Ecological Rationality)

When do heuristics work and *why*?

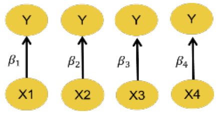
TTB Heuristic	Tallying Heuristic
• Noncompensatory weights (Martignon & Hoffrage, 1999, 2002) • With smaller sample sizes • When predictability (variance explained) is skewed among cues	• Compensatory weights • With smaller sample sizes • when the linear predictability of the criterion was moderate or small (Hogarth & Karelaia, 2007, Hogarth & Karelaia, 2005)





When do heuristics work and *why*?
The Role of Covariance

- ***Heuristics***: with more covariance, they do better. (Hogarth and Karelia, 2007; Brighton, 2006; Gigerenzer & Brighton, 2009, Dieckmann & Rieskamp (2007))
- ***Complex models***: with more covariance, they usually do worse. (Hogarth and Karelia, 2007)



Why?

- ***Heuristics***: they ignore covariance in their cue validity estimates. (no overfitting)
- ***Complex models***: They need to estimate covariance from the data in the learning phase, and this hurts at generalization phase (overfitting).

Regularized regression (L2): ridge regression

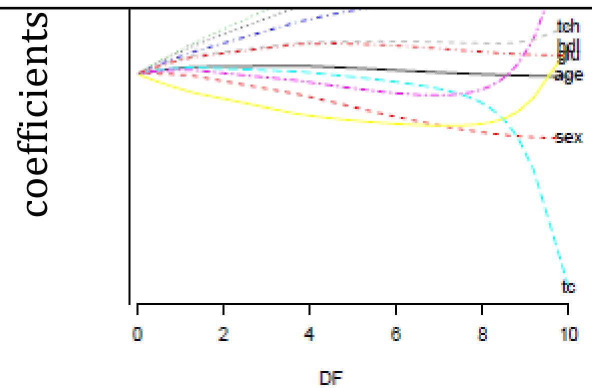
Least squares (OLS) — Penalty term

$$\boldsymbol{w}^* = \arg\max_{w} \left\{ -\sum_i (y_i - x_i w)^2 - \theta \sum_k w_k^2 \right\}$$

Why am I talking about this ridge regression?

Because we developed a model similar to this.

- Applying the ridge regression penalty has the effect of shrinking the estimates w^* toward zero.



$\infty \leftarrow$ penalty θ

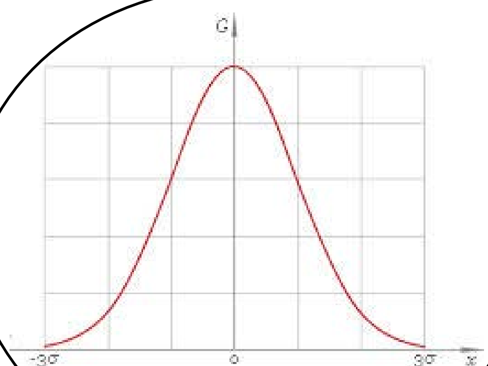
- As $\theta \rightarrow \infty,\ \widehat{w}^{ridge} \rightarrow 0$
- As $\theta \rightarrow 0, \widehat{w}^{ridge} \rightarrow \widehat{w}^{OLS}$

Penalty term is like a Bayesian prior

Penalty term

$$\boldsymbol{w}^* = \arg\max_{w} \left\{ -\sum_i (y_i - x_i w)^2 - \theta \sum_k w_k^2 \right\}$$

Possible prior distributions:

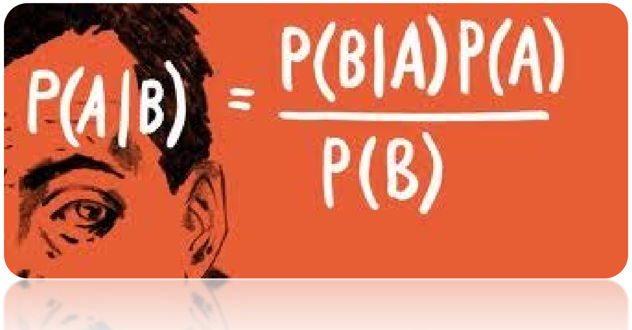


$$f(w) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(\mu - w)^2}{2\sigma^2}}$$

Or a prior that reflects covariance information in environment!?

Bayesian Rationality

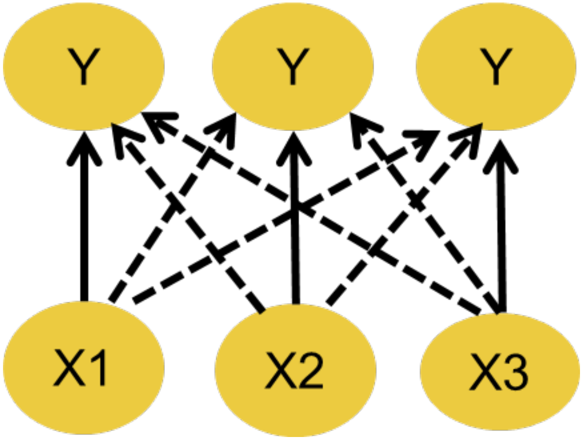
- A Bayesian framework is ideal to formalize the ideas
- Prior = reflecting the amount of **covariance** in the environment
- Likelihood = a latent state variable model that enables us to smoothly move between linear regression and the heuristics (tallying and TTB heuristic).



Overview

1. Fast and frugal Heuristics: what defines them?
2. Rational Models of Cognition
3. Are heuristics compatible with Bayesian inference?
4. Our Bayesian Model
5. Simulation Results
6. Discussion: Q & A
7. Implications

Our latent state variable model



Multivariate (= multiple DV's) Regression

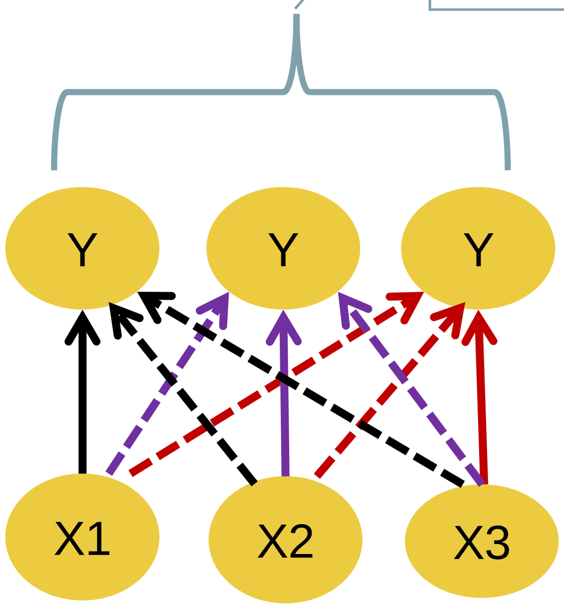
Our latent state variable model

Y = w11*X1 + w21*X2 + w31*X3

Y = w12*X1 + w22*X2 + w32*X3

Y = w13*X1 + w23*X2 + w33*X3

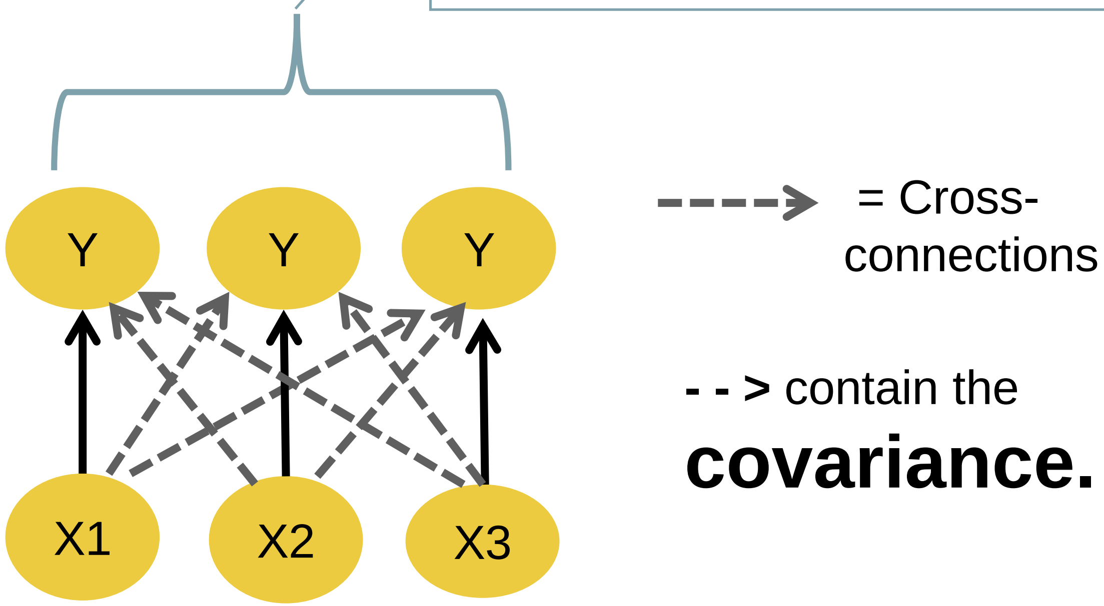
-> like doing linear regression 3 times!



Multivariate (= multiple DV's) Regression

Our latent state variable model

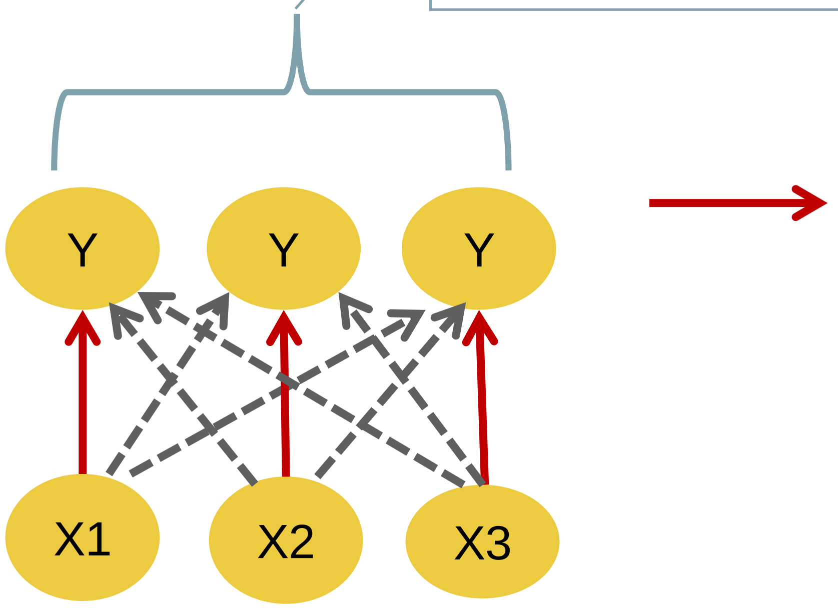
Y = w11*X1 + w21*X2 + w31*X3
Y = w12*X1 + w22*X2 + w32*X3
Y = w13*X1 + w23*X2 + w33*X3



- - > contain the **covariance.**

Our latent state variable model

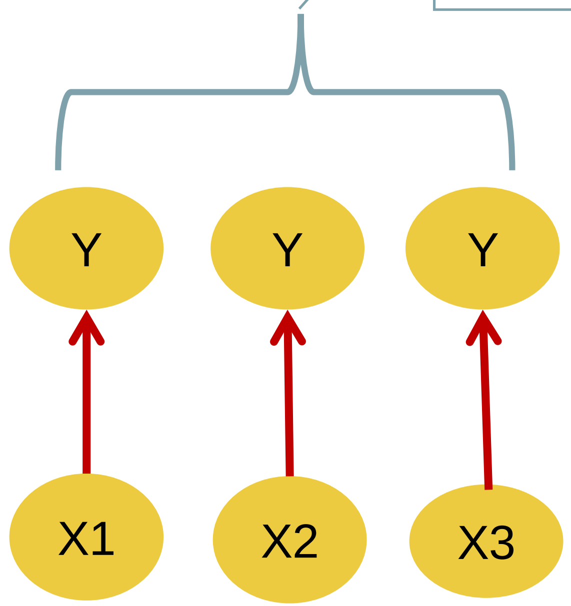
Y = w11*X1 + w21*X2 + w31*X3
Y = w12*X1 + w22*X2 + w32*X3
Y = w13*X1 + w23*X2 + w33*X3



= Direct connections

Our latent state variable model

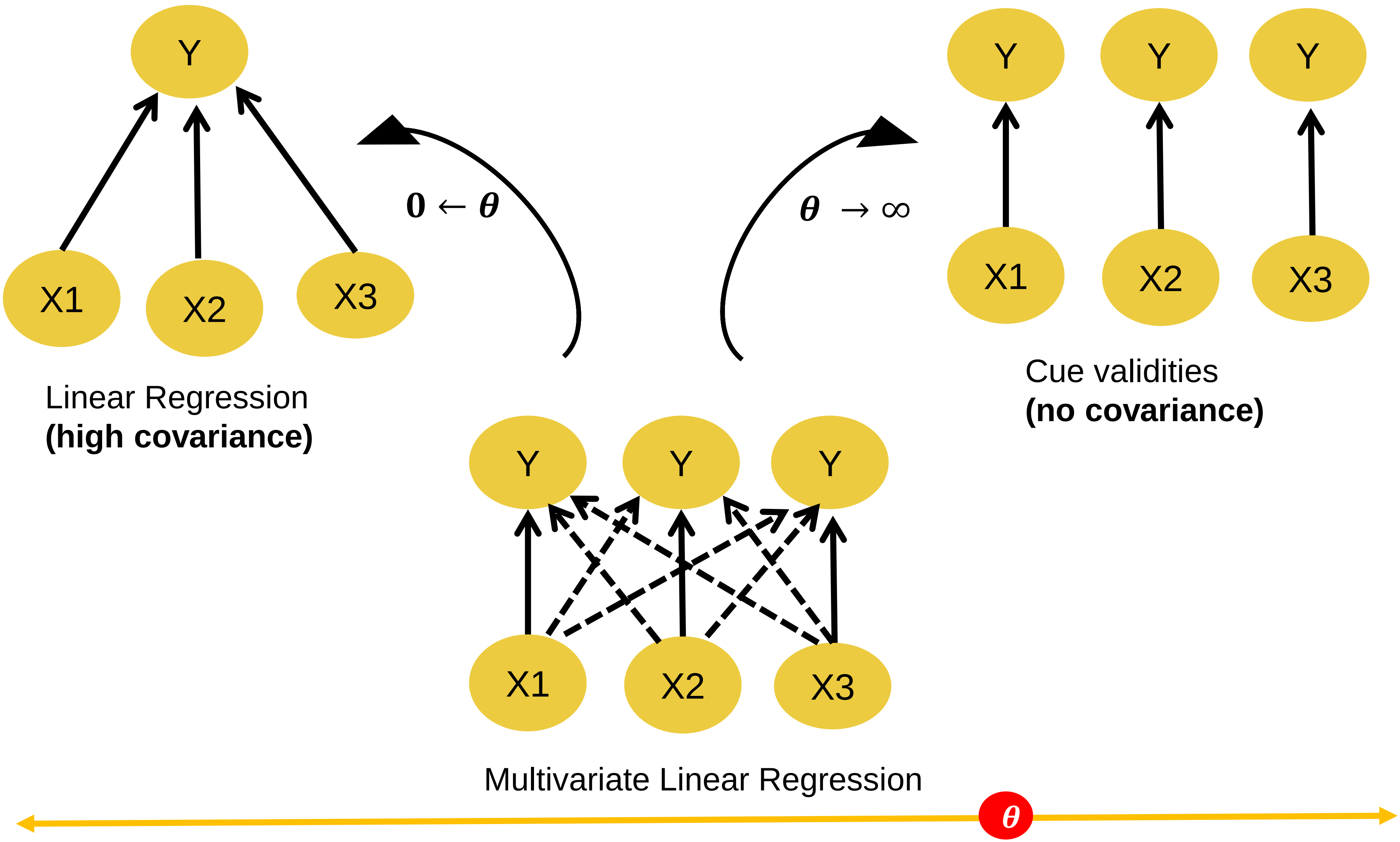
Y = w11*X1 + w21*X2 + w31*X3
Y = w12*X1 + w22*X2 + w32*X3
Y = w13*X1 + w23*X2 + w33*X3




= Direct connections

→ **no covariance** estimated!

Our Bayesian model



Our Bayesian model

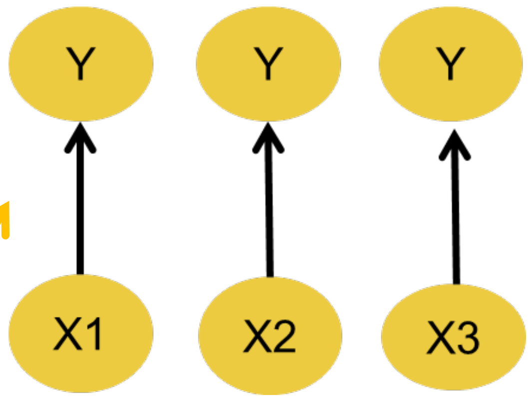
Prior

- In analogy to ridge regression:

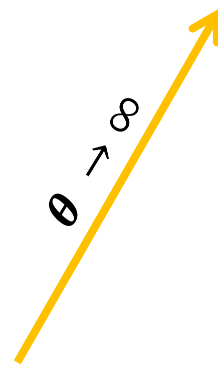
Prior = $-\theta * \left[\sum_{i=1}^{m}\sum_{j=1}^{m}\left|w_{ij}\right|^{2} - tr\left(W^{2}\right)\right]$

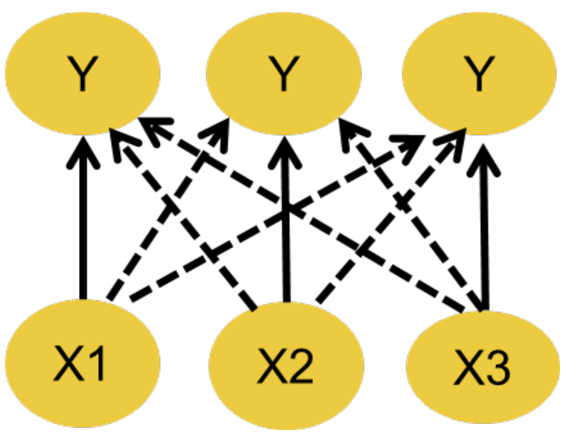
Log Likelihood: Multivariate Normal

$$\ln \boldsymbol{P}_{X,W}(\boldsymbol{Y}_i) \propto -\frac{1}{2}\sum_{i=1}^{n}(Y_i - XW)^{T}\mathrm{C}^{-1}(Y_i - XW)$$

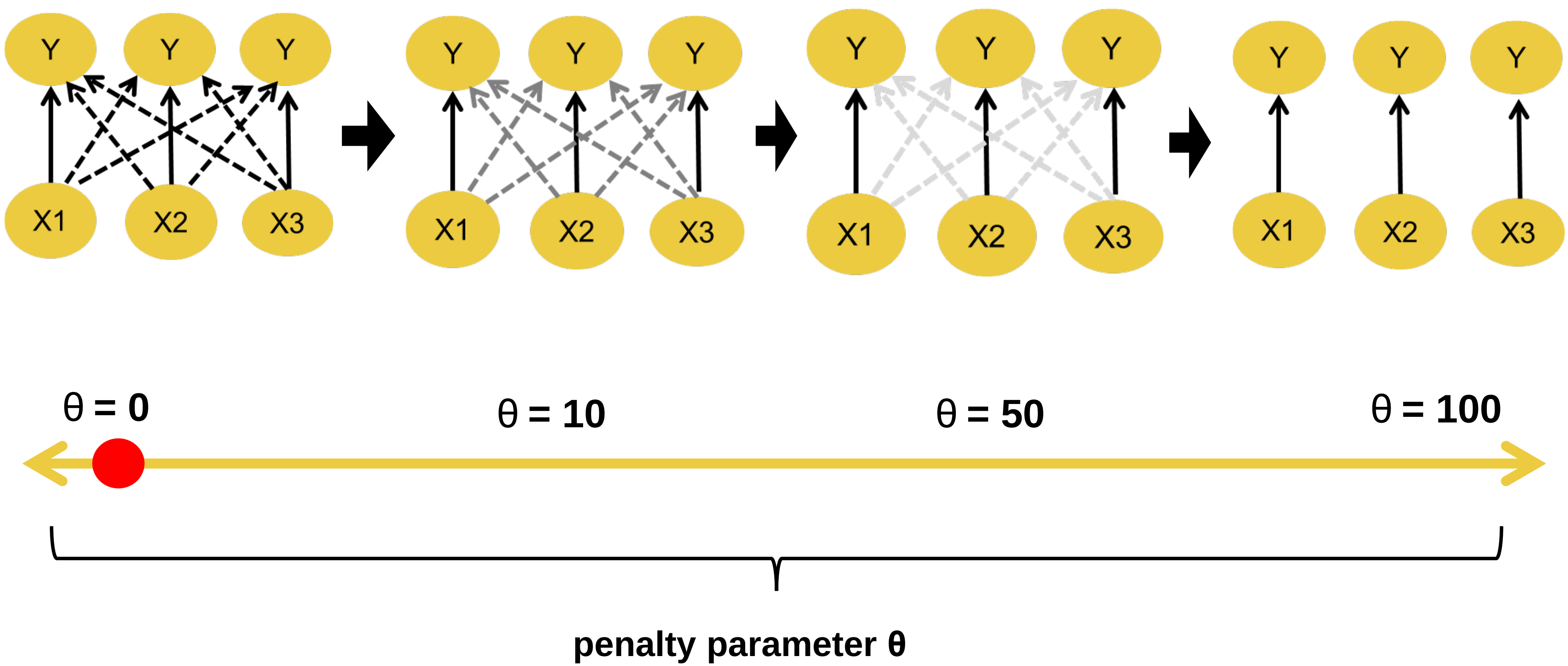


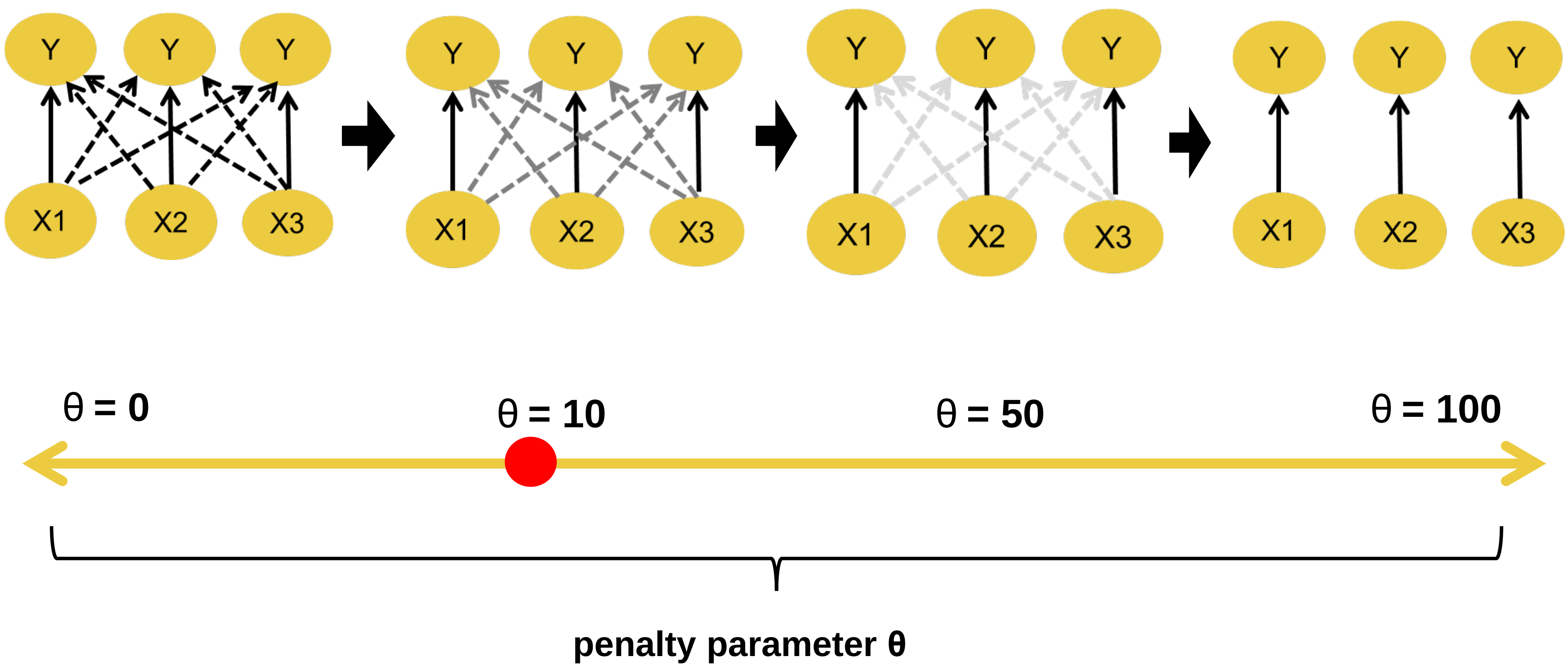
Single - predictor regressions

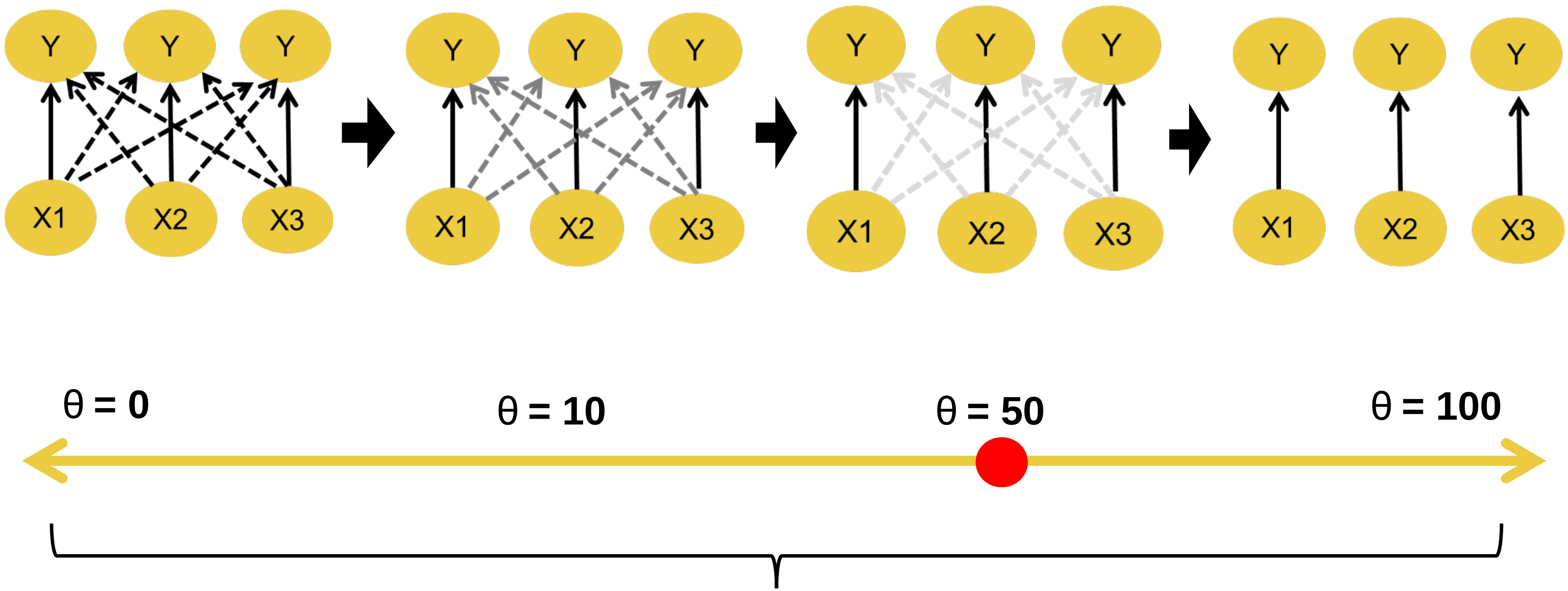




Multivariate Regression

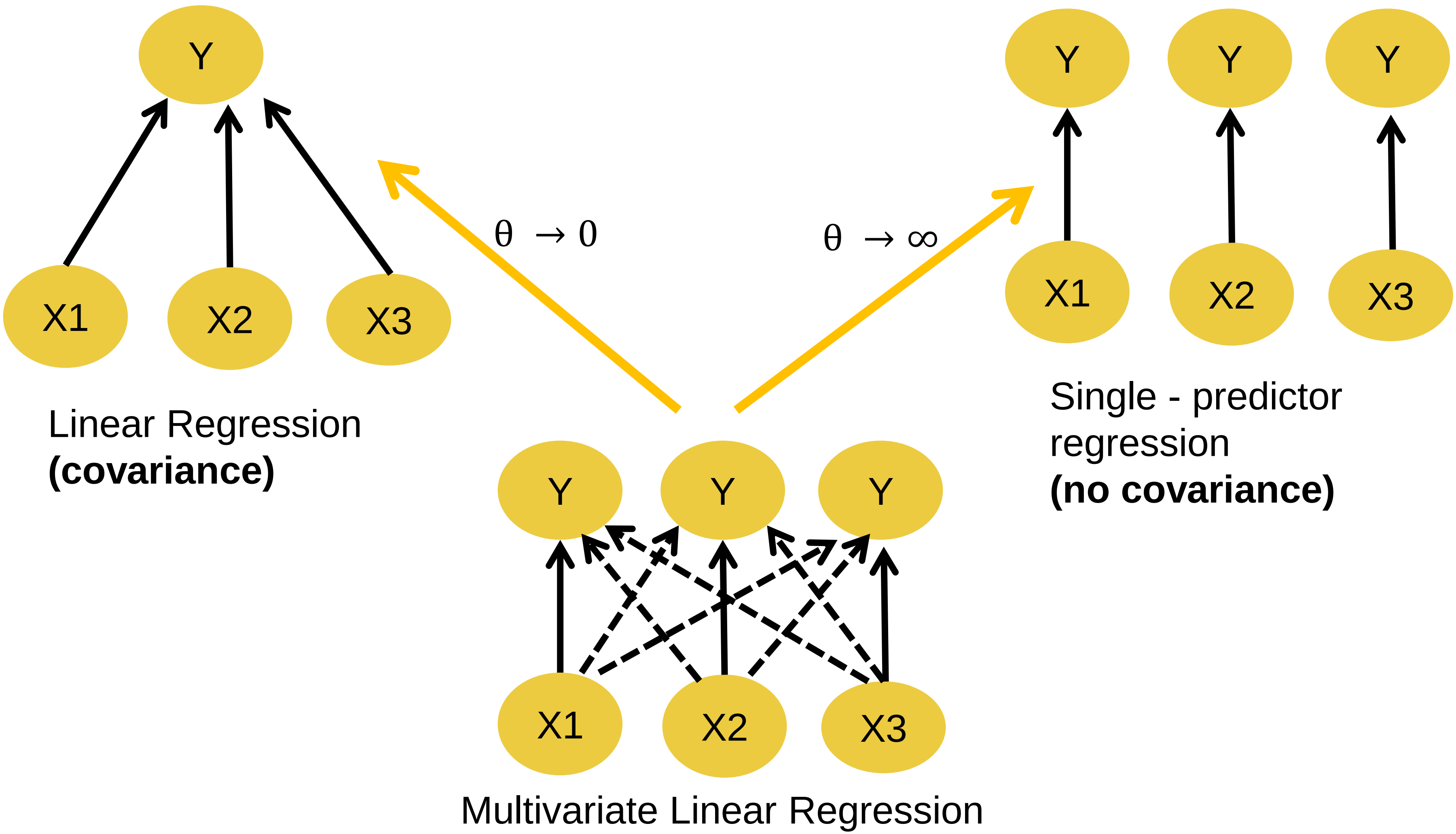

Y
Y
Y
X1
X2
X3
Y
Y
Y
X1
X2
X3
Y
Y
Y
X1
X2
X3
Y
Y
Y
X1
X2
X3
θ = 0
θ = 10
θ = 50
θ = 100
penalty parameter θ


Y
X1
X2
X3
θ = 0
θ = 10
θ = 50
θ = 100
penalty parameter θ



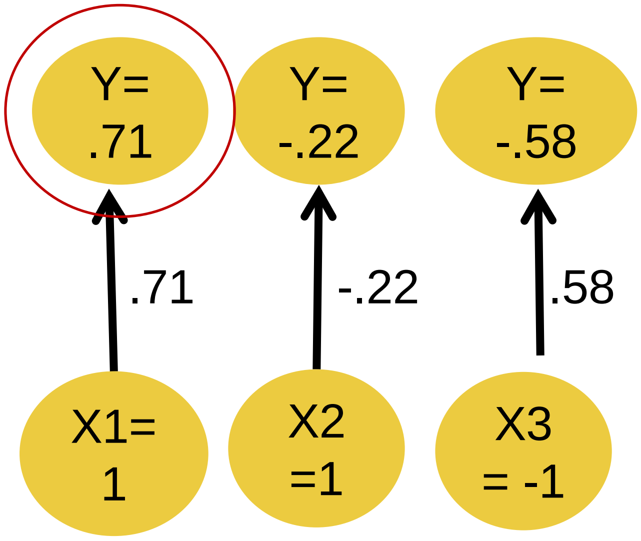
penalty parameter θ

Our Bayesian model



TTB decision rule

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

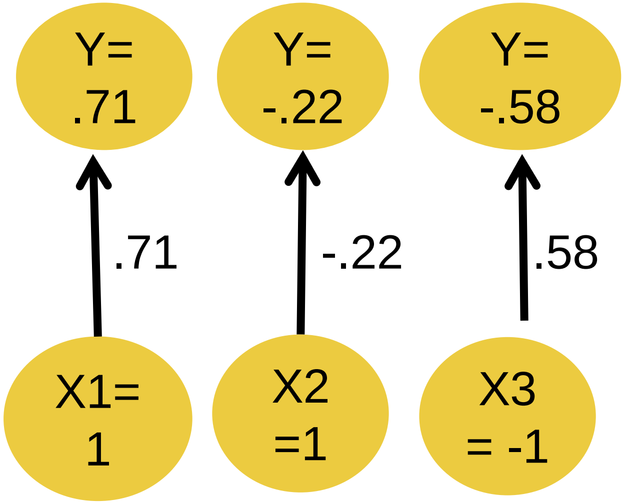


Cue validities
(no covariance)

Tallying decision rule

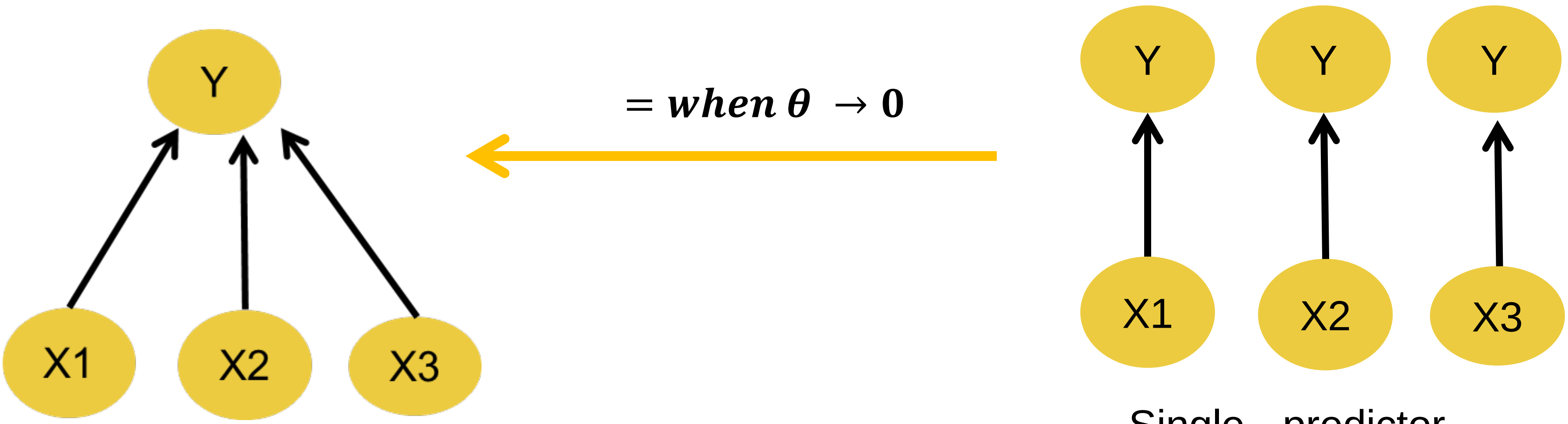
- Count the signs of the outputs Y.
  = sign(sum(sign(Y))

- Tallying would count: +1-1-1 = -1.



Cue validities (**no covariance**)

What is linear regression?

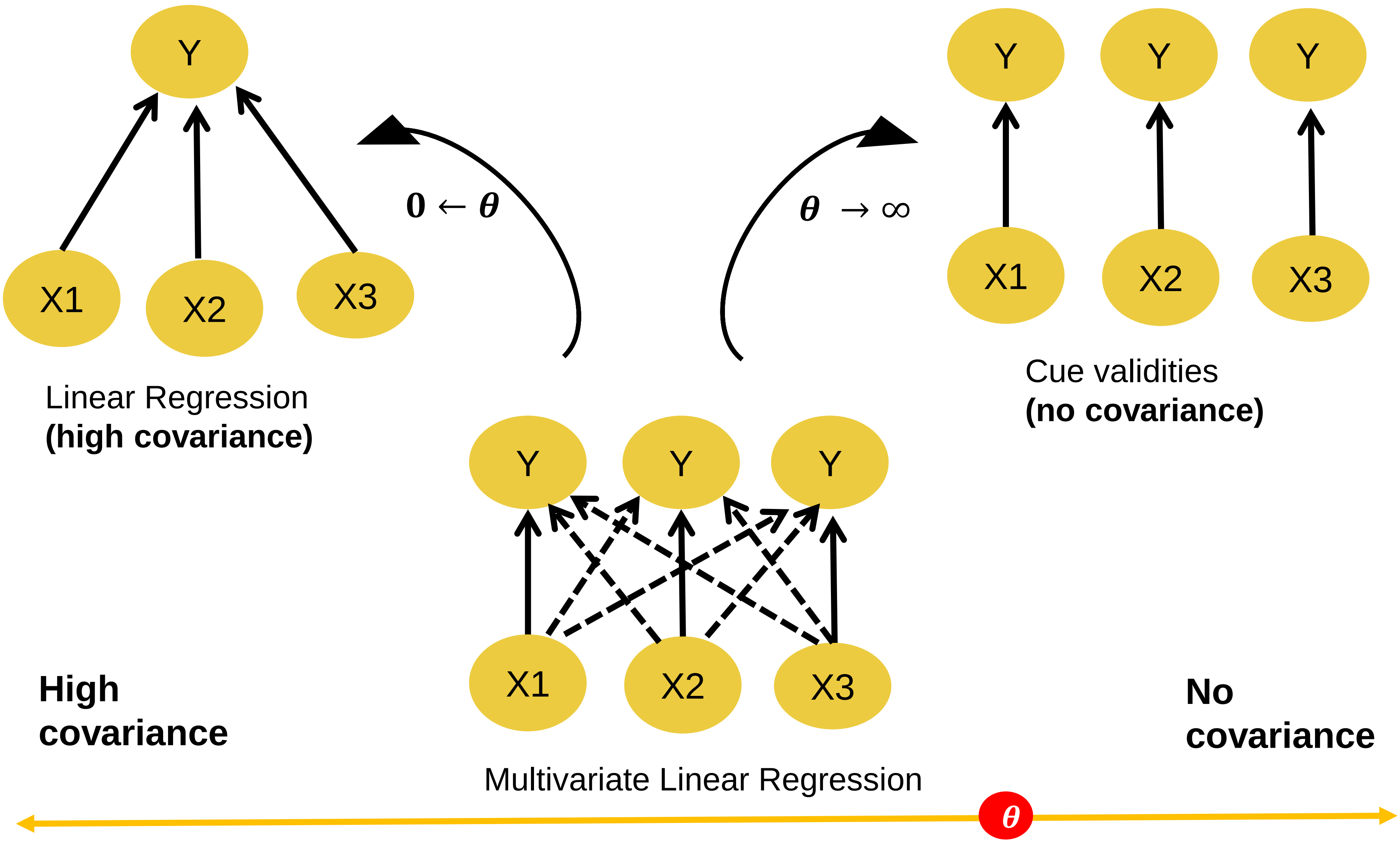


Linear Regression

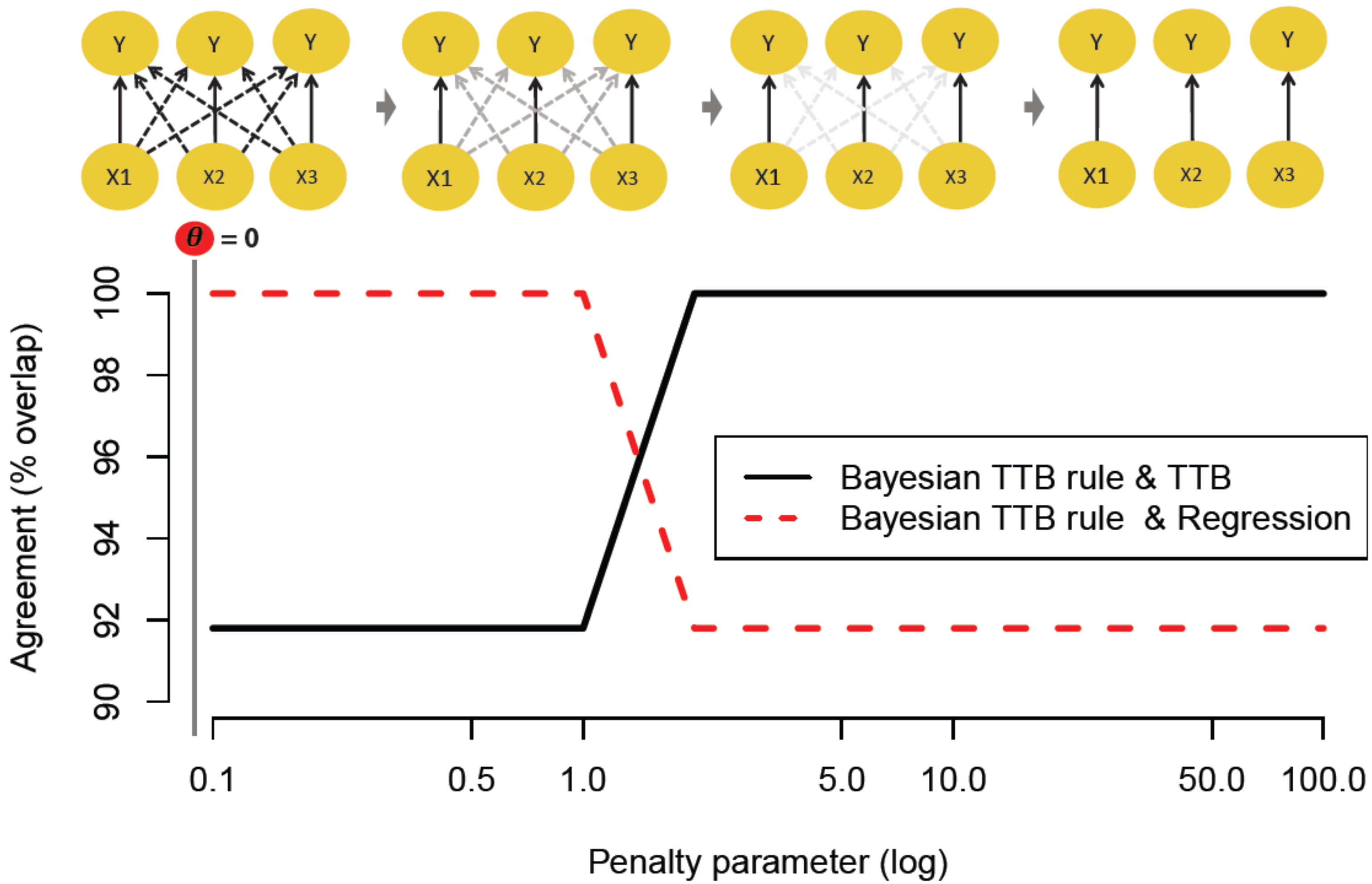
Single - predictor regression

- It is either heuristic decision rule when the penalty term is zero.

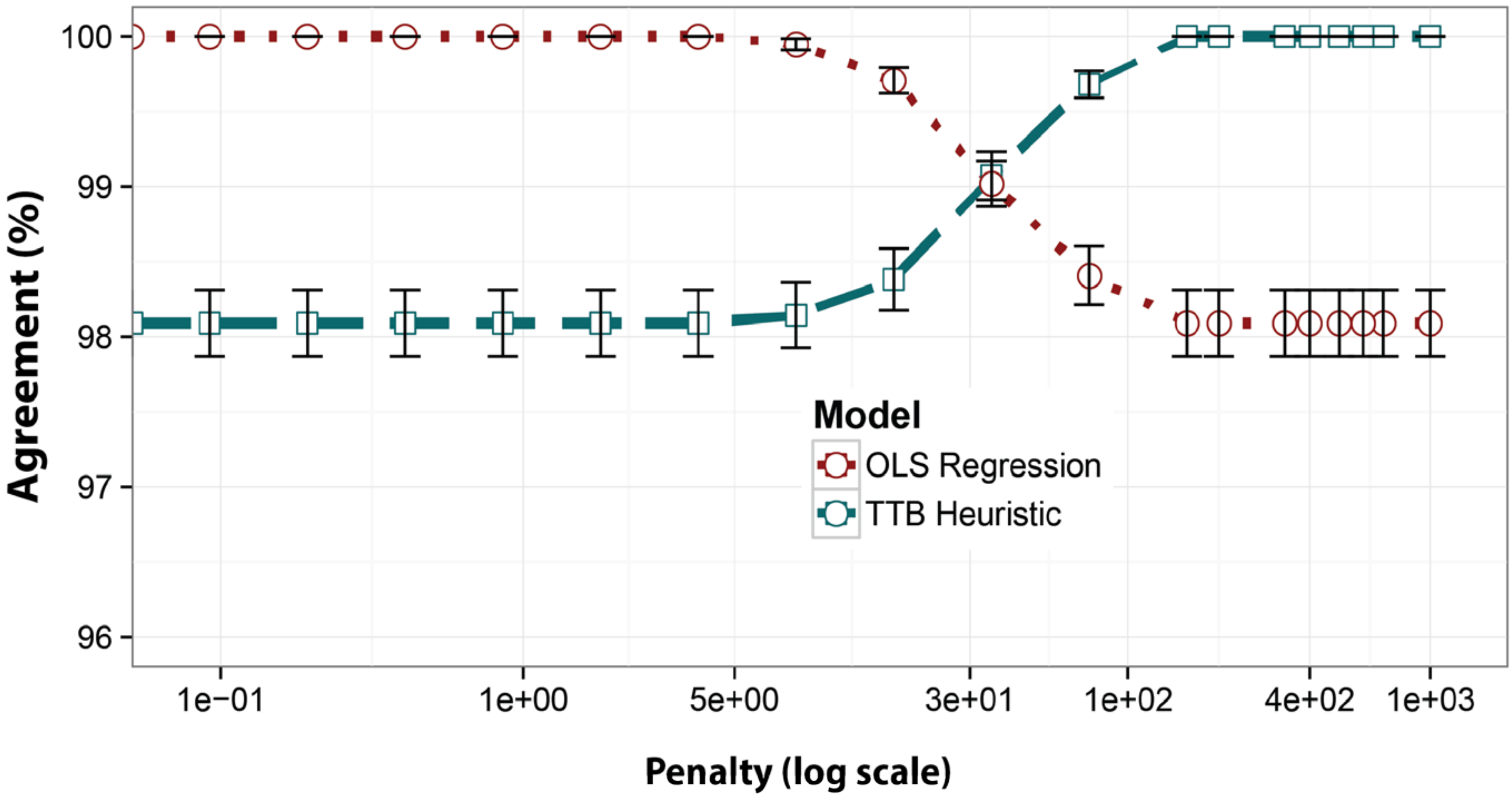
Continuum between heuristics and LR



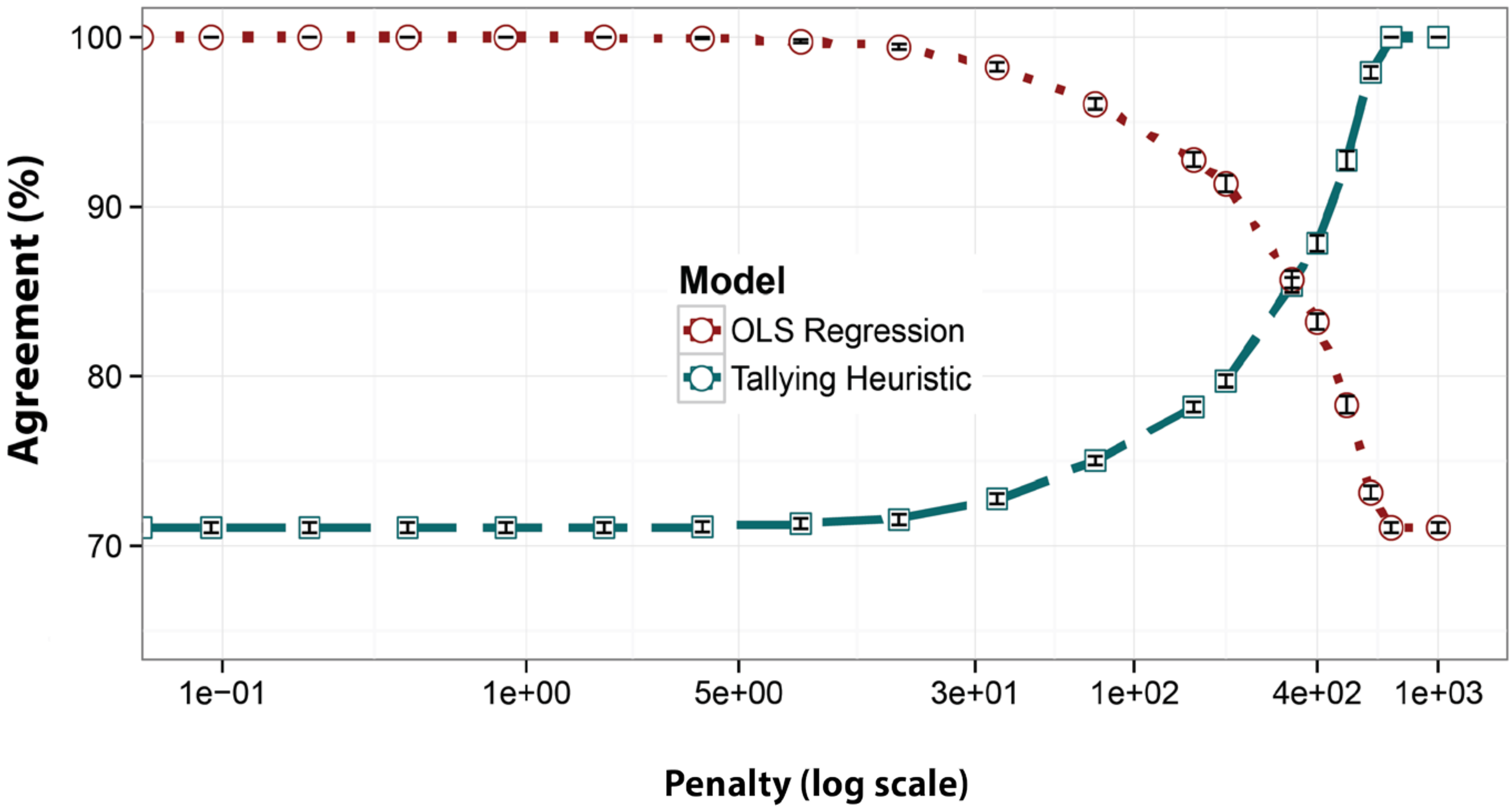
Convergence of Bayesain model with Take-the-Best and Linear Regression



Agreement between Bayesian model and Take-The-Best Heuristic



Agreement between Bayesian model and Tallying Heuristic

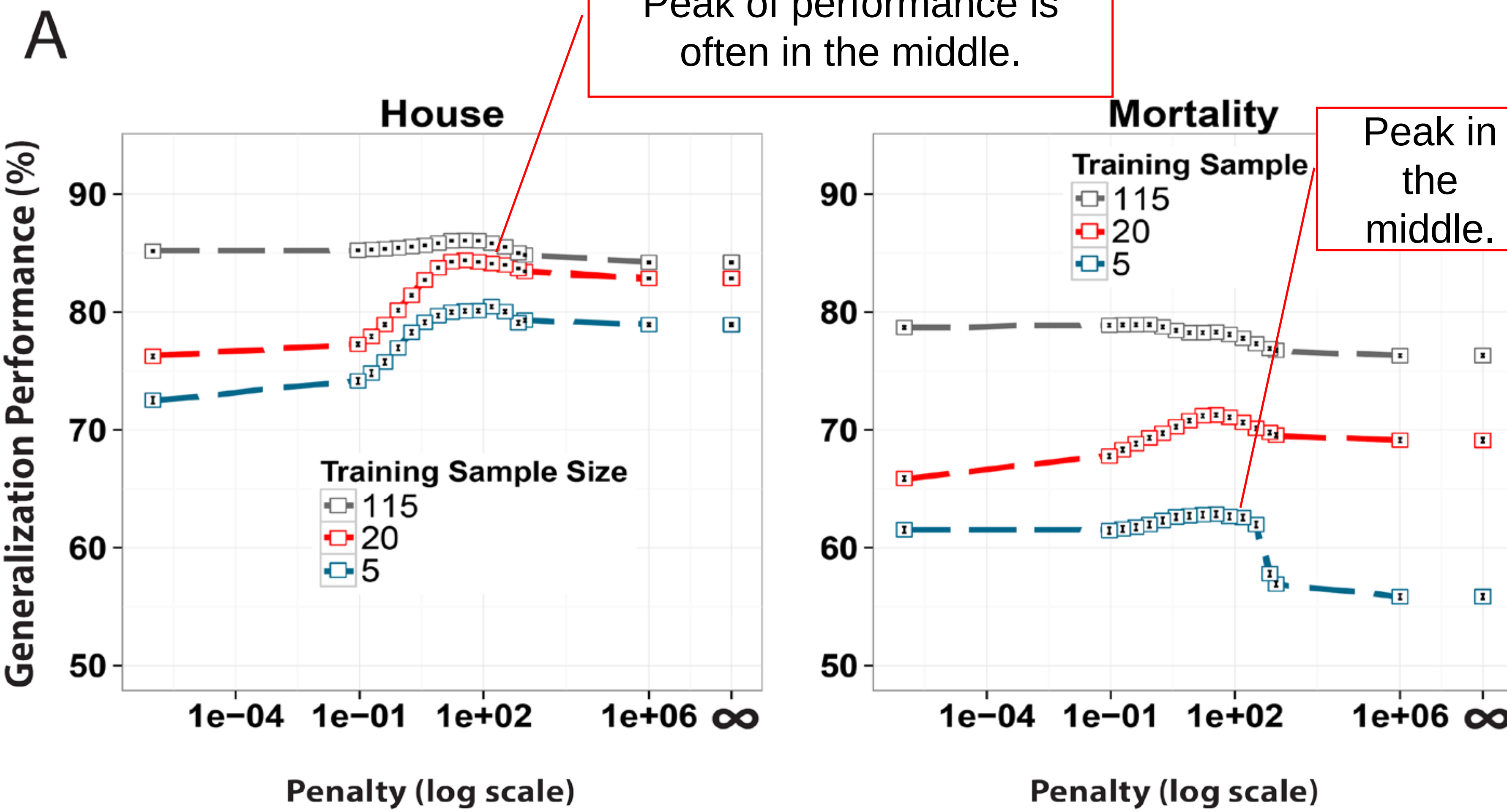


- **Avg. covariance of dataset = 0.55 (high)**

Performance of the Bayesian model compared to heuristics and linear regression

Peak of performance is often in the middle.

Peak in the middle.



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, 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.

Full integration/ regression ⟵ **Optimal** ⟶ **Heuristic (e.g. Tallying)**

penalty parameter

CONCLUSIONS

1. We find that probabilistic rational models (i.e., Bayesian inference models) include simple heuristics as special, or limiting, cases.

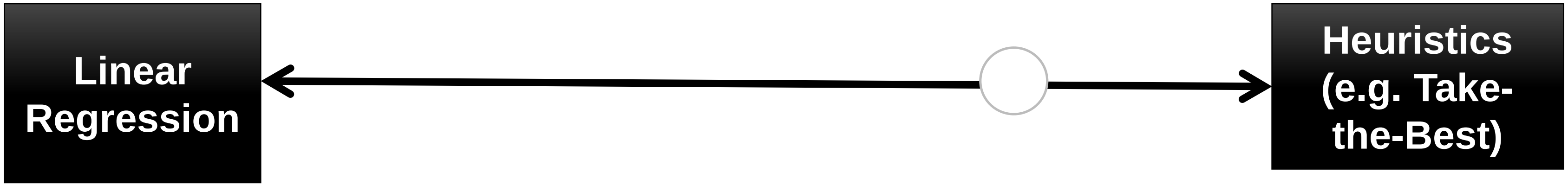
2. We provide a new formal explanation for less-is-more phenomena.

    (The novel explanation that follows from the current formal approach is that less-is-more phenomena occur because ignoring information is tantamount to an extreme prior that some environments approximately satisfy.)

3. The strongest form of less-is-more, i.e., that one can do better with heuristics by throwing out information, is false. The optimal solution always uses all of the information (a finite value of $\boldsymbol{\theta}$) but merely down-weights it.

CONCLUSIONS

4. 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:



5. We developed a new regularization tool – that formally links two opposing theories, i.e. Bayesian and heuristic models of cognition.

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

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)

DISCUSSION

- What could this mean on a psychological level?

(→ Note that the framework presented here is merely on a computational level of analyses. If you are not clear on the different levels of analysis in cognitive science, see slide below on Marr's levels)

- If you were the scientist, what would you do next?

- What are the big implications of this research?

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

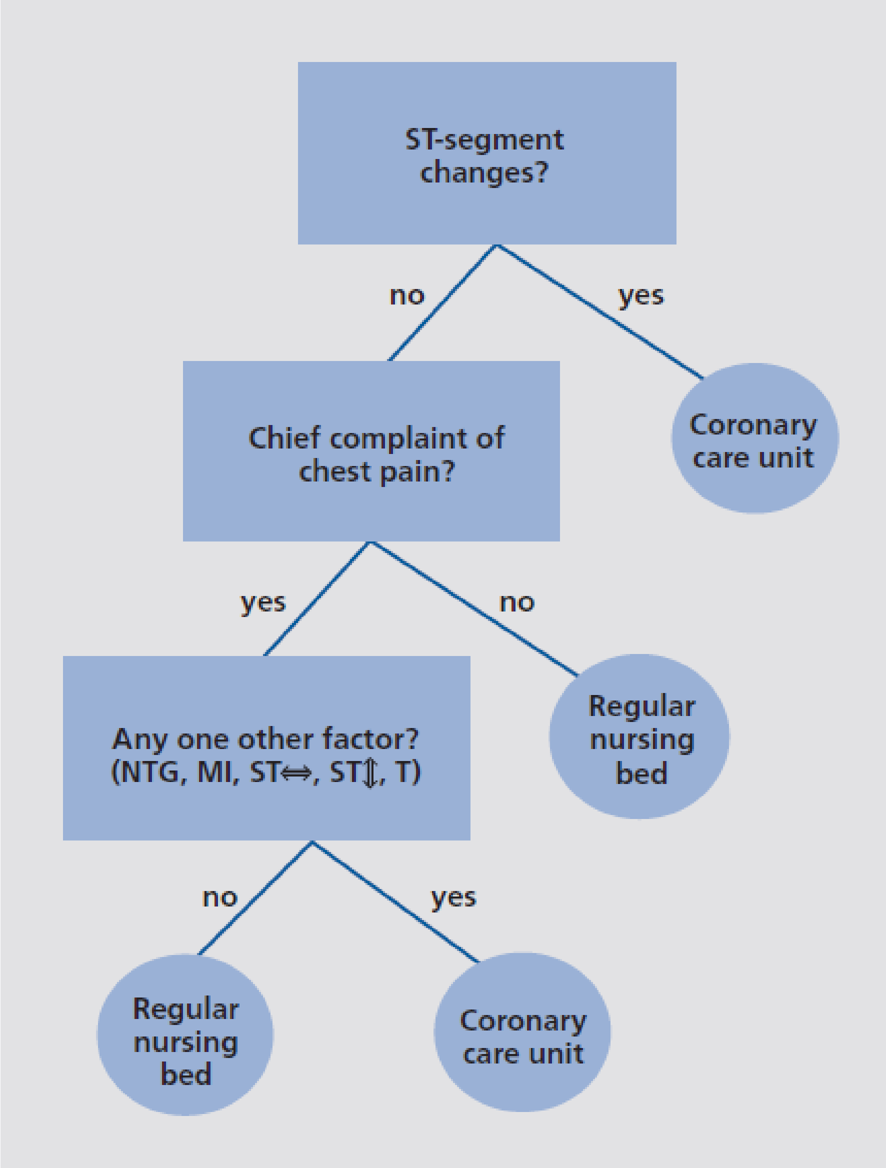
Implications

- 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.

You are probably familiar with this example…



- A simple heuristic for deciding whether a patient should be assigned to the coronary care unit or to a regular nursing bed.
- If there is a certain anomaly in the electrocardiogram (ST-segment), the patient is immediately sent to the coronary care unit
- Otherwise a second predictor is considered etc.
- Adopted from Marewski & Gigerenzer (2012)

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?

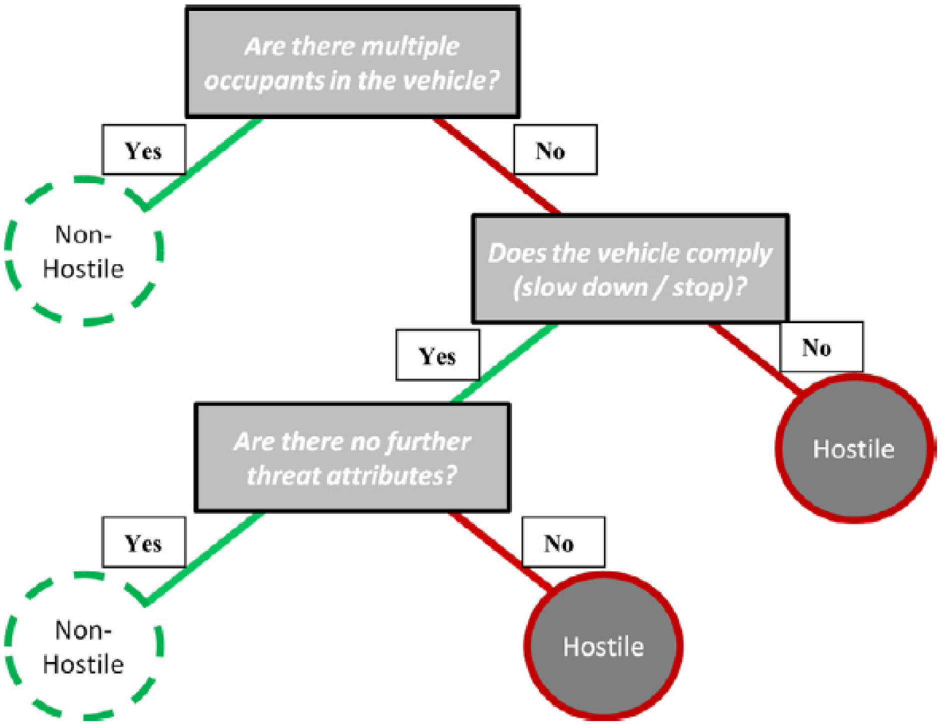


Fig. 4. A fast and frugal tree for classifying oncoming traffic as hostile or nonhostile. The tree was constructed before obtaining the reports on the Afghanistan incidents.

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

Implications

- The current research helps to move closer the different levels of analysis in cognitive science, e.g., in particular the computational level, e.g., Bayesian inference models, and the algorithmic, or process level (e.g., Jones & Love, 2011)

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, & Group 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, & Group 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; disuccsion 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*.