# Title: Everyday Heuristics as Bayesian Inference

**Authors:** Paula Parpart[1]*, Matt Jones[2], Bradley C. Love[1]

**Affiliations:**

[1]Department of Experimental Psychology, University College London, United Kingdom.

[2]Department of Psychology and Neuroscience, University of Colorado, Boulder, USA.

*Correspondence to: E-mail: paula.parpart.10@ucl.ac.uk, b.love@ucl.ac.uk

**Abstract:** Simple heuristics are often regarded as plausible decision strategies in light of the limitations of the human mind. Many regard heuristics as inherently irrational and suboptimal as they do not follow a normative Bayesian account that relies on probability calculus, whereas others note that heuristics perform well in practice. Resolving this tension, we prove that two prominent decision heuristics, take-the-best and tallying, are in fact special cases of Bayesian inference. Inspired by regularization approaches from machine learning, we construct a Bayesian prior reflecting co-variation sensitivity. Parametric variation of this prior generates a continuum of models, with linear regression (often viewed as rational) at one extreme and frugal heuristics, which completely ignore covariance among predictors, at the other extreme. This unifying framework explains *why* heuristics can be so successful: They implement rational inference under a particular prior, which is optimal in the corresponding task environment.

**One Sentence Summary**: We prove that everyday decision-making heuristics and economic theories based on probability calculus are compatible and are limiting cases of the same rational framework.

**Main Text:**

*Homo economicus* is a rational creature with the ability to make decisions which fully maximize its utility. Unfortunately, *Homo economicus* sightings are as rare and difficult to verify as Bigfoot sightings. Humans fail to rise to *Homo economicus's* standard because cognitive capacities are limited (*1*). Rather than function as a rational animal, one view is that humans use *heuristics*, which are decision procedures that are quick, require little effort, and neglect a large portion of the information (*2*, *3*). To illustrate, consider deciding whether to purchase a sedan or an SUV. One heuristic procedure is to first order the available cues, such as gas mileage, price, safety, and seating space by their importance, i.e., cue validity, then work down the list until a cue favors one option (*4, 5*) (Fig. 1).

Notice that this decision heuristic, known as *Take-The-Best* (TTB), will consider only one cue when the most important cue distinguishes the vehicles. In contrast, a seemingly more rational procedure such as linear regression weights and linearly adds all cues, while taking into account the co-variation among cues. Other heuristics, such as *tallying (6, 7)*, use the directions of all cues, while neglecting cue validity magnitudes. In deciding between the sedan and SUV, tallying would simply choose the vehicle that bested the other across the most cues. A fundamental difference between linear regression and heuristics is that heuristics completely ignore covariance among cues. For example, seating capacity and safety co-vary because very small cars do not fare well in collisions. The cue validities used by heuristics are not sensitive to these kinds of dependencies. Instead, heuristics estimate how good each cue is in isolation at predicting the better option.

Two influential schools of thought in decision making both claim that humans rely on heuristics. Kahneman and Tversky's *heuristics and biases* program (*2, 8*) emphasizes peoples' deviation from rational norms. These rational norms, which we refer to as *Bayesian rationality*, were adopted from traditional economic theories where rationality is construed as the degree of compliance with logic and the axioms of probability theory. In contrast, the *fast and frugal heuristics* program, championed by Gigerenzer and colleagues, emphasizes *ecological rationality* wherein heuristics are evaluated by how well adapted they are to the structure of the environment (*3*). Rather than focus on human failings, this program catalogs cases in which humans excel by using simple heuristics in everyday decisions. The rare point of agreement between these two views is that heuristics and Bayesian rationality are distinct and incompatible.

In this contribution, we prove that two popular heuristics, Take-The-Best and tallying, are equivalent to Bayesian inference. Along with linear regression, which is often used as a stand-in for the rational view because it integrates and weighs all cues, the two heuristics naturally arise as special cases within a single Bayesian framework. These specific models correspond to different prior beliefs about co-variance among cues in the environment, which echoes ideas from ecological rationality. Conversely, echoing the rational yardstick used in the heuristics-and-biases program, inference in our framework follows Bayesian rationality.

| | Cue validities | Sedan or | SUV | Cue coding |
|---|---|---|---|---|
| **(1) Gas Mileage** | **.90** | ☺ | ☹ | **+1** |
| **(2) Price** | **.81** | 😐 | 😐 | **0** |
| **(3) Safety** | **.73** | ☹ | ☺ | **-1** |
| **(4) Seating capacity** | **.54** | ☹ | ☺ | **-1** |

**Fig. 1.** Deciding between a sedan and a SUV on the basis of four cues, i.e., gas mileage, price, safety, seating capacity. A positive icon indicates that the vehicle has a higher value on the cue than the alternative, whereas a negative icon refers to a lower value on that cue in comparison to the alternative. The rightmost column explains how cues are coded: a cue is coded +1 when it favors the left vehicle, -1 when it favors the right vehicle, and 0 when both vehicles have the same value on that cue (i.e., the cue does not discriminate). The outcome variable about which vehicle was better in the end is coded in the same way, with +1 representing the vehicle on the left and -1 representing the vehicle on the right.

## When and why heuristics perform well

Heuristics can surprisingly outperform more sophisticated rational strategies in certain cases (*6*, *7*, *9*, *10*). Following the tenets of ecological rationality, these less-is-more effects may arise because heuristics are best-suited to certain environments (*11-14*). From a machine-learning perspective, this conclusion is sensible because every model has an inductive *bias*, akin to a Bayesian prior, that makes the model best-suited to certain learning problems.

A models' bias is responsible for what it can learn from a set of training cases (i.e., the training sample), which determines how the model classifies novel test cases. How well the model classifies test samples is crucial - a model's performance is usually judged by its generalization performance (*15*). In addition to differing in bias, models can also differ in how sensitive they are to the *variance* in the training sample. More complex models can be 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 (Fig. 2A) (*16*).

Bias and variance trade-off with one another such that models with low bias suffer from high variance and vice versa (*17*), which implies that more complex (i.e., less biased) models will overfit small training samples and can be bested by simpler (i.e., more biased) models, such as heuristics. However, as the size of the training sample increases, more complex models should fare better. Indeed, in a reanalysis of a dataset favoring heuristics over linear regression, we find that the advantage for heuristics disappears when training sample size is increased (Fig. 2B). This result underscores the utility of adopting a machine-learning perspective to understand why and when psychological models perform well.

In summary, one possible explanation for the success of heuristics is that they have the appropriate bias (akin to a Bayesian prior) for many ecologically relevant problems. We formalize this idea in the remainder of this paper. Our treatment of model complexity centers on sensitivity to co-variation among predictors, a characteristic important for the performance of heuristics (*18, 19*). Along the continuum of complexity (i.e., a models' flexibility in capturing a variety of possible data patterns), heuristics are the simplest and most biased models because they are only being sensitive to cue validities computed in isolation from one another. On the other extreme, linear regression is the most complex and least biased model because it is sensitive to the full co-variance structure among cues. The best performing model for a given learning and decision problem may lie anywhere along this continuum of co-variance sensitivity.

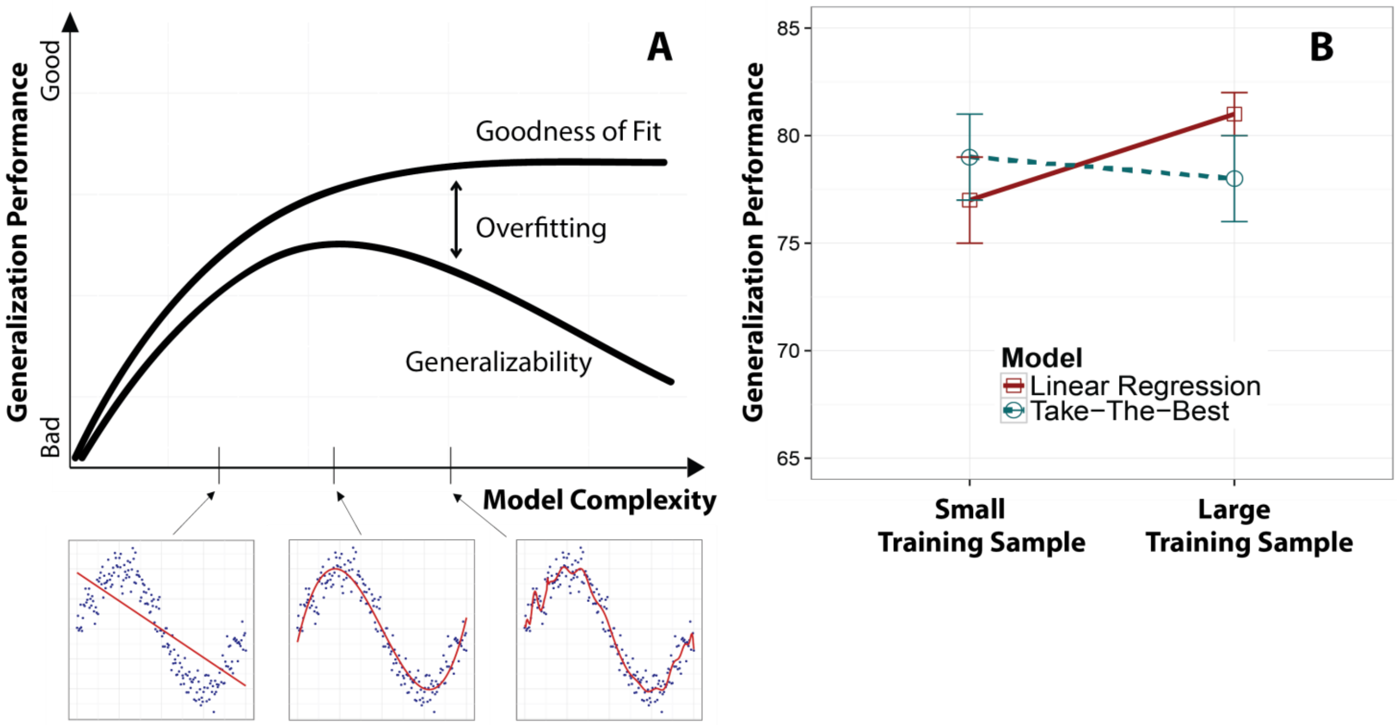


**Fig. 2. (A)** More complex models use their flexibility to fit the training sample better, accounting for most of the variance**.** However, these models can fare poorly in generalization tasks that test on novel samples. (**B)** Reflecting previous findings by Gigerenzer and colleagues, a re-analysis of a heuristic dataset (ozone levels) (*10*) finds that Take-The-Best can outperform linear regression at generalization when the training sample is small (50% of the data). However, the pattern reverses when the training sample is increased (90% of the data).

**Bayesian Inference Model for Heuristics**

The Bayesian approach we develop is conceptually related to ridge regression (*20*), a successful regularized regression approach in machine learning. Ridge regression contains a penalty term that adjusts model complexity to improve weight estimates and avoid overfitting training data (Fig. 2A). When the penalty parameter is 0, ridge regression is only concerned with model fit (i.e., minimizing squared error on the training set). For this special case, ridge regression is equivalent to standard linear regression, which is highly sensitive to the variance of a training sample. Formally, the weights estimated by ridge regression are defined by

$$\hat{w}_{ridge} = \arg\min_{w}\left\{\underbrace{\left\|\boldsymbol{y}-\boldsymbol{X}w\right\|^2}_{\text{Goodness of Fit}} + \theta\underbrace{\left\|w\right\|^2}_{\text{Penalty Term}}\right\} \tag{1}$$

where the penalty parameter $\theta \geq 0$. In the fit term, $\|.\|^2$ denotes the square of the Euclidean norm, $y=\left[y_{(1)},...,y_{(n)}\right]^T \in \mathbb{R}^n$ is the outcome variable defined over the *n* training stimuli, and $\boldsymbol{X}$ is a $n \times m$ matrix with one column for each of the *m* predictor variables $x_{(i)}$. The goodness-of-fit term reflects ordinary least-squares regression, which finds the weights that best fit the data set. A training set consists of a set of input-output pairs $\left(\boldsymbol{x_{(1)}}, y_{(1)}\right),...,\left(\boldsymbol{x_{(n)}}, y_{(n)}\right)$ where each vector $\boldsymbol{x_{(i)}} \in \mathbb{R}^m$ and $y_{(i)} \in \mathbb{R}$. An example of a regression problem is the decision between two automobiles (Fig.1), where the explanatory variables $x_{(i)}$ correspond to the cues, while the outcome variable $y$ contains information about which car (e.g., sedan or SUV) was chosen. The goal in a regression problem is to estimate the weights, i.e., a vector of regression coefficients $w=\left[w_1,...,w_m\right]^T$ that minimizes the prediction error. What differentiates ridge regression from ordinary least-squares regression is the penalty term: As the penalty parameter $\theta$ increases in Eq. 1, the pressure to shrink the magnitude of the weights increases, reducing to zero as $\theta \to \infty$. This inductive bias can reduce overfitting, though the optimal setting of $\theta$ will always depend on the environment.

The ridge penalty term is mathematically equivalent to a Gaussian Bayesian prior on the weights, with larger penalty parameters resulting in priors that more strongly favor (are more biased toward) smaller weights. This prior distribution over the weights is combined with current observations (i.e., the training sample) to form a posterior distribution (also Gaussian) over the cue weights. However, like standard linear regression, ridge regression provides a point estimate, rather than the full Bayesian posterior. Nevertheless, the conceptual relationship to Bayesian rationality is clear (*21*) (supplementary text 4).

Our Bayesian approach differs from ridge regression in that we use a prior that modulates expectation for co-variation among cues. Another difference from ridge regression is that we express the regression problem in multivariate terms in order to implement the priors' function and the sequential nature of Take-The-Best. The penalty parameter $\theta$ in the prior yields a continuum of models (Fig. 2) that smoothly vary in their mean weight estimates from those of linear regression to those of the heuristics.

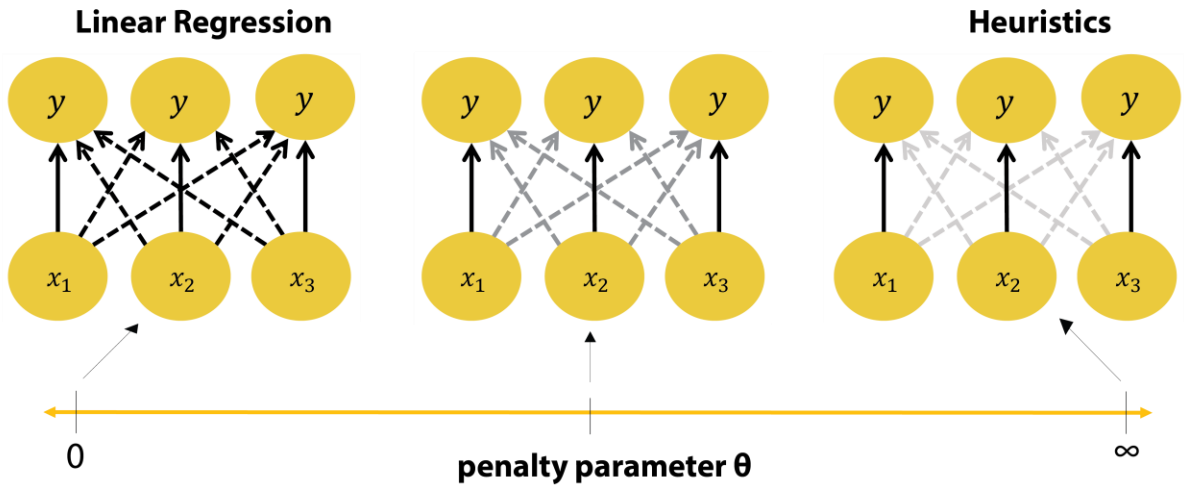


**Fig. 2.** The prior of the Bayesian inference model influences the posterior solution such that it can encompass linear regression and the heuristics as extreme cases. In this example, there are $m = 3$ cues $\boldsymbol{x_1}, \boldsymbol{x_2}, \boldsymbol{x_3}$, which represent the explanatory variables that predict the output $y$ . As in Fig. 1, cues can take values of 1, -1, or 0, depending on whether the left or right option is superior, or neither, and likewise the $y$ output contains values of +1 or -1, depending on the correct choice. The output variable $y$ gets replicated as many times as there are cues, in order to formulate the choice problem as a multivariate regression problem. Note that due to this replication, the model architecture at any given penalty parameter reflects three regression equations, where all $m$ cues are regressed onto the same output variable $m$ times in a row. We refer to the dashed arrows as *cross-weights*, and the solid arrows as *direct weights*. The cross-weights are necessary for the model to estimate co-variation among cues, as without the cross-weights, the model would only contain $m$ single-predictor regression weights that are computed independent of each other (*direct weights*). When $\theta = 0$, the prior does not penalize the cross-weights, and the set of mean posterior weights to each output variable $y$ are equivalent to the regular linear regression solution. At the other extreme, as $\theta \to \infty$, the mean of cross-weights is shrunk to zero, and the knowledge captured in the direct posterior weights becomes equivalent to that embodied by cue validities in heuristics, i.e., co-variation information is ignored (rightmost panel). Between these two extreme values of $\theta$ lie models that are sensitive to co-variation to varying degrees.

As shown in Fig. 2, a sample from the posterior distribution can be represented as a $m \times m$ weight matrix $W$, where $m$ is the number of cues. The posterior distribution of weights is the model's knowledge representation. Unlike ridge regression, only the cross-weights are penalized (see Fig.2). Therefore, in the limiting case, when $\theta \rightarrow \infty$, the cross-weights reduce to zero and the posterior weights will only reflect cue validity magnitudes and valence (neglecting co-variation information) as used by the heuristics. For example, Take-The-Best utilizes both cue direction and cue validity, whereas tallying's decision procedure only utilizes cue direction. At the other extreme, when $\theta = 0$, all sets of mean weights relating to each $y$ are equivalent to the linear regression solution. For $\theta = 0$, cue directions, cue validities, and co-variance information are reflected in the posterior weights as in linear regression. We refer to this formalism as Covariance Orthogonalizing Regularization (COR) because the model can vary its complexity along a continuum of co-variance sensitivity.

Model weights are paired with a decision procedure to classify a test item, e.g., $\boldsymbol{x_i} = \left[ \boldsymbol{x}_{\boldsymbol{i}_1}, \boldsymbol{x}_{\boldsymbol{i}_2}, \boldsymbol{x}_{\boldsymbol{i}_3} \right]$, that is described on all cues. The vector $\boldsymbol{x_i}$ is multiplied by the mean posterior weight matrix $W^*$ to generate an output vector $\overline{\boldsymbol{y}}_{\boldsymbol{i}} = \left[ \boldsymbol{y}_{\boldsymbol{i}_1}, \boldsymbol{y}_{\boldsymbol{i}_2}, \boldsymbol{y}_{\boldsymbol{i}_3} \right]$. Then, a decision rule is applied to the resulting $\overline{\boldsymbol{y}}_{\boldsymbol{i}}$.

The *tallying decision rule,*

$$z_i = \sum_j sgn\left(\overline{\boldsymbol{y}}_{\boldsymbol{ij}}\right) \tag{2}$$

$$choice_i = \begin{cases} +1\,(left) & if \quad z_i \succ 0 \\ -1\,(right) & if \quad z_i \prec 0 \\ \ \ 0\,(guess) & if \quad z_i = 0 \end{cases}$$

chooses the option with a majority of outputs in its favor (conveyed by the sign), irrespective of the magnitudes of the outputs. This is in line with the tallying heuristic counting up the evidence in favor of either option in a binary choice.

The *Take-The-Best decision rule,*

$$z_i = sgn\left( y_{\arg\max_j |y_{ij}|} \right) \tag{3}$$

$$choice_i = \begin{cases} +1\,(left) & if \quad z_i \succ 0 \\ -1\,(right) & if \quad z_i \prec 0 \\ \ \ 0\,(guess) & if \quad z_i = 0 \end{cases}$$

selects the maximum absolute output (corresponding to most valid cue) and takes the sign of that output as its choice. Notice that the decision rule naturally exhibits the sequential nature of the Take-The-Best heuristic, because whenever a cue cannot discriminate between two alternatives, its cue value is coded 0 (Fig. 1) and therefore its output $y$ will also be 0. In such cases, the next-most valid cue will guide the decision. When the decision rules are adopted by the model, as $\theta \to \infty$, the tallying decision rule converges to the tallying heuristic, and the Take-The-Best decision rule converges to the Take-The-Best heuristic. Lastly, with $\theta = 0$, either decision rule will yield decisions equivalent to linear regression. Thus, we have illustrated how linear regression and the two heuristics are special cases along a continuum defined by expectation of cue covariance.

**Simulation Results**

We ran simulations to demonstrate convergence of the Bayesian inference model with the heuristics as a function of the penalty parameter, and secondly to explore the performance of the Bayesian inference model as a function of environmental conditions such as the number of observations in the training set. All simulations were run on artificial datasets created by generating choice pairs defined on three or five cues of the type illustrated in Fig. 1. The coding scheme followed the pattern in Fig. 1 (see supplementary methods for details).

Firstly, we demonstrate convergence with the Take-The-Best decision rule, though parallel results hold for the tallying decision rule (see supplementary figures for tallying results). The simulation explored how the Bayesian model agrees in choices with linear regression and with the Take-The-Best heuristic as a function of the penalty parameter $\theta$. As expected, the agreement between the predictions of the Bayesian model and Take-The-Best increased with the penalty parameter, reaching an asymptote of perfect agreement as the parameter approached infinity (Fig. 3A). The opposite pattern held for linear regression, with agreement being perfect at $\theta = 0$ and declining as the penalty parameter increased.

One exciting aspect of our Bayesian approach is that it specifies a multitude of models between the extremes of linear regression and the heuristics. For many environments, the best performing model should lie somewhere between these two extremes of co-variance expectation. Another prediction is that, for a given environment, the size of the training sample should influence the optimal penalty parameter. With more data, more complex models (with smaller penalty parameters) that freely estimate co-variance usually fare better (*14, 22*). Both of these predictions held in a second simulation (Fig. 3B), illustrating that lower penalties perform better with more training data, while the opposite holds for smaller training samples. As expected, for smaller training samples, the performance peak was often found in the middle, i.e., for a medium penalty.

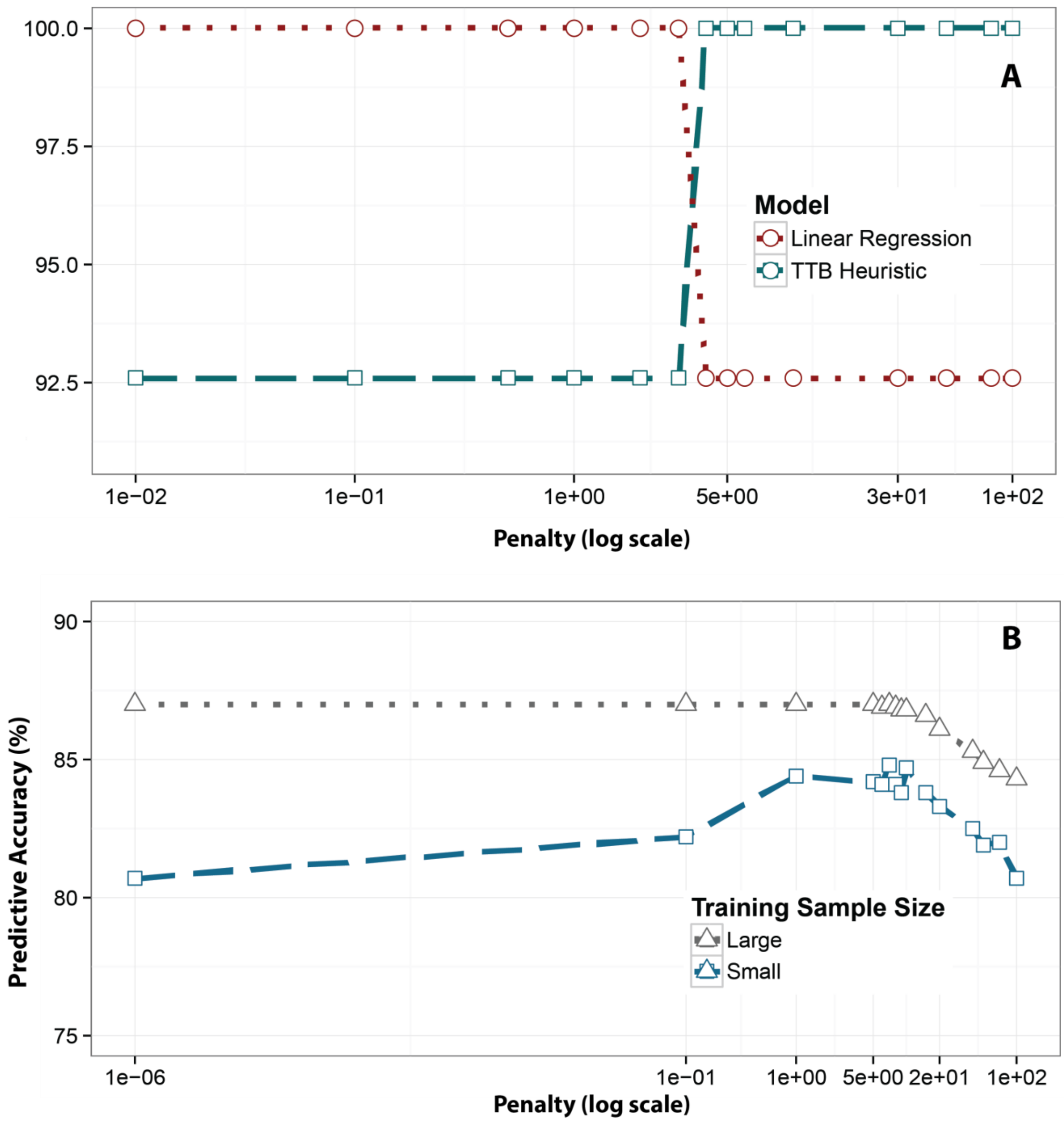


**Fig. 3. (A)** Agreement between the Bayesian models' Take-The-Best decision rule and the Take-The-Best heuristic, as well as linear regression, as a function of penalty parameter. The ordinate indicates the percentage agreement on test item choices in an artificial dataset. **(B)** Generalization performance of the Bayesian Take-The-Best decision rule as a function of training sample size: The Bayesian Take-The-Best performs better with smaller penalties when training samples are large, and it performs best for intermediate penalties when training samples are small.

## Implications

We find that probabilistic rational models, which are widely used across economics and cognitive science, include everyday heuristics. Thus, heuristics are not irrational, deficient processes, but are instead part of Bayesian rationality and arise from a particular prior expectation about co-variance. This formal result helps explain why heuristics work so well in practice – heuristics correspond to certain assumptions about the structure of the task environment, and when there is a match between the environment and the heuristic strategy, heuristics can surpass the performance of more complex algorithms. Despite the algorithmic simplicity of heuristics, they are provably optimal in the appropriate environments.

For the heated rationality debate in the literature, we can conclude that the core question is no longer whether heuristics are rational or irrational. Instead, the key factor is whether the task environment matches the assumptions made by the heuristic, which can be formalized as a Bayesian prior. Previous work, such as that by Kahneman and Tversky, that has cataloged the "irrationalities" of heuristics focused on cases where the heuristics' prior assumptions about the environment were incorrect, whereas Gigerenzer and colleagues focused on cases where the heuristics' prior assumptions where correctly matched to the environment. Heuristics are highly biased (in a machine learning sense), but this bias reflects a strong Bayesian prior that can lead to poor or optimal performance depending on its match to the task environment.

With the current framework it becomes possible to predict *when* heuristics excel, which has strong implications for various real-world domains where quick actions are required. For example, soldiers at a checkpoint must rapidly decide whether an approaching car is a threat. Likewise, doctors need to quickly decide whether to assign a patient to a coronary care unit or a regular nursing bed. In light of people's limited capacities, heuristics can be valuable in these scenarios because they reduce the algorithmic complexity of the decision procedure. However, it is crucial to know when these shortcuts are helpful and harmful, which the present work addresses. As our analyses indicate, the best performing model will usually lie between the extremes of linear regression and fast-and-frugal heuristics. Between these extremes, lie a host of models along a continuum of sensitivity to co-variation information in the environment. To the extent that people are tuned to the structure of the environment, intermediate solutions may reflect the functioning of psychological processes, though these intermediate strategies are yet to be discovered.

The current research helps to move closer the different levels of analysis in cognitive science, which are usually referred to as the computational level, and the algorithmic, i.e., process level (*23*). We show that a computational Bayesian model can incorporate simple heuristics which are often seen as mechanistic, process level theories. There have been many recent efforts looking at the compatibility between psychologically plausible mechanisms and rational models of cognition (*23*-*31*). These studies are interlinked with our own, and we hope this line of research will flourish in the future.

Regularized regression approaches are used extensively across fields ranging from genetics to neuroscience, neuroimaging techniques, to financial forecasting. While ridge regression can reduce overfitting, many problems researchers face may be better addressed by COR's regularization-by-co-variation approach. An efficient solution to COR (*32*) that exploits the convexity of the error surface can find solution weights to large problems rapidly.

Our results underscore the utility of applying bias-variance concepts from machine learning to psychological questions. Previous bias-variance analyses in psychology have focused on the variance side (*14*). In contrast, we focused on the contribution of bias, and showed that increasing bias can result in superior outcomes when the environment conforms to that bias. Relying on the fundamental bias-variance concept made it possible to organize past psychological theory, and develop a new formal account for various decision strategies.

## References and Notes:


1. H. A. Simon, Invariants of human behavior. *Annual review of psychology* **41**, 1-19 (1990).
2. A. Tversky, D. Kahneman, Judgment under Uncertainty: Heuristics and Biases. *Science* **185**, 1124-1131 (1974).
3. G. Gigerenzer, P. M. Todd, A. R. Group, *Simple heuristics that make us smart*. (Oxford University Press, 1999).
4. G. Gigerenzer, D. G. Goldstein, Reasoning the fast and frugal way: models of bounded rationality. *Psychological review* **103**, 650-669 (1996).
5. A. Tversky, Elimination by aspects: A theory of choice. *Psychological review* **79**, 281 (1972).
6. R. M. Dawes, The robust beauty of improper linear models in decision making. *American psychologist* **34**, 571 (1979).
7. R. M. Dawes, B. Corrigan, Linear models in decision making. *Psychological bulletin* **81**, 95 (1974).
8. D. Kahneman, A perspective on judgment and choice: mapping bounded rationality. *The American psychologist* **58**, 697-720 (2003).
9. H. J. Einhorn, R. M. Hogarth, Unit weighting schemes for decision making. *Organizational Behavior and Human Performance* **13**, 171-192 (1975).
10. J. Czerlinski, G. Gigerenzer, D. G. Goldstein, in *Simple heuristics that make us smart,* G. Gigerenzer, P. M. Todd, A. R. Group, Eds. (Oxford University Press, New York, 1999), pp. 97–118.
11. G. Gigerenzer, W. Gaissmaier, Heuristic decision making. *Annual review of psychology* **62**, 451-482 (2011).
12. L. Martignon, U. Hoffrage, in *Simple heuristics that make us smart*. (Oxford University Press, New York, 1999), pp. 119-140.
13. R. M. Hogarth, N. Karelaia, Heuristic and linear models of judgment: matching rules and environments. *Psychological review* **114**, 733-758 (2007).
14. K. V. Katsikopoulos, L. J. Schooler, R. Hertwig, The robust beauty of ordinary information. *Psychological review* **117**, 1259 (2010).
15. R. Kohavi, in *Ijcai*. (1995), vol. 14, pp. 1137-1145.
16. M. A. Pitt, I. J. Myung, When a good fit can be bad. *Trends in cognitive sciences* **6**, 421-425 (2002).
17. S. Geman, E. Bienenstock, R. Doursat, Neural networks and the bias/variance dilemma. *Neural computation* **4**, 1-58 (1992).
18. J. Rieskamp, A. Dieckmann, in *Ecological rationality: Intelligence in the world*. (Oxford University Press, New York, 2012), pp. 187-215.
19. H. Brighton, G. Gigerenzer, in *The Probabilistic Mind: Prospects for Bayesian Cognitive Science,* N. Chater, M. Oaksford, Eds. (Oxford University Press, New York, 2008), pp. 189–208.
20. A. E. Hoerl, R. W. Kennard, Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics* **12**, 55-67 (1970).
21. A. Gelman, J. B. Carlin, H. S. Stern, D. B. Rubin, *Bayesian data analysis*. (Taylor & Francis, 2014), vol. 2.
22. N. Chater, M. Oaksford, R. Nakisa, M. Redington, Fast, frugal, and rational: How rational norms explain behavior. *Organizational behavior and human decision processes* **90**, 63-86 (2003).
23. D. Marr. (Freeman, San Francisco, 1982).
24. S. D. Brown, M. Steyvers, Detecting and predicting changes. *Cognitive psychology* **58**, 49-67 (2009).
25. N. Daw, A. Courville, paper presented at the Advances in neural information processing systems, 2008.

26. M. Jones, B. C. Love, Bayesian Fundamentalism or Enlightenment? On the explanatory status and theoretical contributions of Bayesian models of cognition. *The Behavioral and brain sciences* **34**, 169-188; disuccsion 188-231 (2011).
27. M. D. Lee, T. D. Cummins, Evidence accumulation in decision making: unifying the "take the best" and the "rational" models. *Psychonomic bulletin & review* **11**, 343-352 (2004).
28. A. N. Sanborn, T. L. Griffiths, D. J. Navarro, Rational approximations to rational models: alternative algorithms for category learning. *Psychological review* **117**, 1144 (2010).
29. B. Scheibehenne, J. Rieskamp, E.-J. Wagenmakers, Testing adaptive toolbox models: A Bayesian hierarchical approach. *Psychological review* **120**, 39 (2013).
30. D. van Ravenzwaaij, C. P. Moore, M. D. Lee, B. R. Newell, A hierarchical bayesian modeling approach to searching and stopping in multi-attribute judgment. *Cognitive science* **38**, 1384-1405 (2014).
31. T. L. Griffiths, F. Lieder, N. D. Goodman, Rational use of cognitive resources: Levels of analysis between the computational and the algorithmic. *Topics in Cognitive Science. forthcoming*, (2014).
32. T. Noguchi, P. Parpart, B. C. Love, Efficient Regularization of covariance information. *in preparation*, (2015).

**Acknowledgments:** We thank T. Noguchi, N. Bramley, S. Solomon, E. Schulz, and M. Speekenbrink for helpful discussions and comments on the manuscript. P.P. acknowledges support from the UK Centre for Doctoral Training in Financial Computing & Analytics. We also acknowledge support by the Leverhulme Trust grant RPG-2014-075 to B.C.L., and the AFOSR grant FA9550-14-1-0318 to M.J.. The data reported in this paper are archived at UCL Discovery, http://discovery.ucl.ac.uk/.