# A Bayesian Antidote Against Strategy Sprawl

**Benjamin Scheibehenne (benjamin.scheibehenne@unibas.ch)**
University of Basel, Missionsstrasse 62a
4055 Basel, Switzerland
&
**Jörg Rieskamp (joerg.rieskamp@unibas.ch)**
University of Basel, Missionsstrasse 62a
4055 Basel, Switzerland

## Abstract

Many theories in cognitive science assume that people possess a repertoire of strategies or a "toolbox" from which they choose depending on the situation. This approach suffers from the problem that the number of assumed strategies is often not constrained and may be extended post-hoc to improve the fit to the data. This makes it difficult to rigorously test and compare strategy repertoire models. To prevent this "strategy sprawl", a criterion is necessary to decide how many strategies a toolbox should include. Here, Bayesian statistics provide a powerful tool to evaluate toolboxes of different sizes based on their marginal likelihoods. The present work illustrates how such a Bayesian approach can be implemented and demonstrates its applicability by means of parameter recovery studies. Our approach also makes the novel contribution of showing how Bayesian statistics allow testing the strategy repertoire theory against alternative decision theories.

**Keywords:** Strategy repertoire theories, Bayes factor, model selection, simulation, Bugs.

## The problem of strategy sprawl

A common assumption within many research areas in cognitive science is that people possess a repertoire of cognitive strategies to solve the problems they face. For example, people use different strategies for making consumer decisions (Payne, Bettman, and Johnson, 1993), for organizational memory tasks (Coyle, Read, Gaultney, & Bjorklund, 1998), for estimations of frequencies (Brown, 1995), for categorization problems (Patalano, Smith, Jonides, & Koeppe, 2001), for the development of mathematical skills (Siegler, 1991), or for inference problems (Gigerenzer, Todd, and the ABC Research Group, 1999). The strategy repertoire approach provides a fruitful way to explain intra- and inter-individual differences in cognitive processes. This approach has also been described by the metaphor of an adaptive toolbox according to which individual decision makers select between different cognitive strategies to solve specific tasks just as a craftsman selects tools from a toolbox.

Despite its undisputed success in explaining a wide range of human behavior, the idea of a toolbox raises the question of how many different strategies the mental toolbox should contain in the first place. A larger number of possible strategies will always lead to a better description of the data but not necessarily to greater insight. For example, by assuming a specific tool for each possible task, the toolbox should provide a good description of observed behavior due to its great flexibility. Along the same lines, Dougherty, Thomas, and Franco-Watkins (2008) criticized that in a situation in which no strategy out of a set is able to describe a person's choices, an unconstrained toolbox could be enlarged by a new strategy to describe the data. On the other hand, if the toolbox is restricted to only a few or to a single strategy it would lose its ability to describe different cognitive processes.

In the following, we outline a possible solution to the question of how many tools a toolbox should contain based on a Bayesian approach. Having a criterion for determining how many strategies to include keeps strategy sprawl at bay and is also a necessary pre-condition for rigorously comparing different toolbox models with competing cognitive theories that do not assume different strategies but rely on the idea of an "all-purpose" process (Newell, 2005).

## Example of a cognitive toolbox

As an illustrative example of a cognitive toolbox, imagine a situation in which a person tries to determine which of two used cars is a better deal. To make this decision, a person could use different pieces of information (i.e., cues) such as mileage, number of previous owners, or accident history of the cars. In such a situation, each single cue provides a hint of which car might be better, but none of the cues provide an indisputable prediction, because it could be that a car with many previous owners still turns out to be superior overall. In other words, the cues are probabilistically related to the criterion, so that even an object with positive cue values for all cues could sometimes be inferior compared to an object with negative cue values. Probabilistic inferences can be complicated because it is not always clear which information is relevant and how and whether the different pieces of information should be combined.

To make probabilistic inferences, such as choosing the better of two cars, people may choose from a variety of cognitive strategies, that is, from their adaptive toolbox (Gigerenzer, Todd & the ABC Research Group, 1999). For instance, when choosing between two options, people could use a simple non-compensatory decision strategy called take the best (TTB) that only focuses on the single most important or valid cue that discriminates between the two

options. If the most valid cue discriminates between the alternatives, TTB chooses the object with the positive cue value. Only in cases where the most valid cue does not discriminate does TTB then consider the second most valid cue, and so on. An alternative strategy in the toolbox could be a weighted additive (WADD) rule. This model adds up all available cue values weighted by their respective validities and then selects the alternative with the largest score. The WADD rule is compensatory because a highly valid cue may be compensated for by a number of other cues that point to the opposite choice.

This example illustrates that different decision strategies can be applied to make a choice between two options that are described by several attributes. Here, proponents of a toolbox approach could argue that people use either TTB or WADD depending on the decision situation and the characteristics of the decision maker.

## Preventing strategy sprawl

When examining how people solve an inference problem researchers aim to identify the model that best describes the cognitive processes, that is, the one that most likely generated the observed data. Under the assumption that people have a repertoire of strategies, the goal is to identify the strategy that a person has selected. Possible strategies are not known a priori. Therefore, a researcher may add more and more strategies to the toolbox to increase chances that one strategy provides a good fit to the data. From a model comparison perspective, a given toolbox becomes more flexible and complex when more strategies are added. Accordingly, it is not surprising when it provides a better description of the observed data. Therefore, the question of how many strategies to include in the toolbox essentially becomes a trade-off between the complexity of the toolbox and its fit in describing observed data. Adding another strategy is only justified if it increases the fit substantially. Bayesian techniques offer a valuable theoretical framework to make this trade-off and to identify a toolbox that fits the data well. More precisely, the probability of a specific dataset given a specific toolbox model (referred to as the evidence or marginal likelihood of that model) can be used as a criterion of how many tools to include.

### Bayes' theorem

In a Bayesian framework, the marginal likelihood $p(D)$ is a measure of how well a given model $M$ describes the observed data $D$ across all possible parameter values of that model (Kass & Raftery, 1995; Shiffrin, Lee, Kim, & Wagenmakers, 2008):

$$p(D) = \int p(D|M) \times p(M) \,|\, \mathrm{d}M \qquad (1)$$

where $p(D|M)$ is the likelihood of the observed data given the model and $p(M)$ is the prior probability of the model. The evidence provides a viable metric to compare different models against each other. However, to eventually apply this criterion the possible toolbox models and the strategies

within each toolbox need to be specified in a Bayesian framework. In particular, this requires the specification of prior distributions and likelihood functions.

In the following, we lay out the necessary specifications in detail. To illustrate the basic principle behind this approach, we start with the simple example of comparing two toolboxes consisting of only one strategy each, before proceeding to the more complex case of comparing toolboxes of different sizes.

## Model specification for a simple toolbox

As a first step, we compare two toolboxes that only consist of one single strategy each. For illustrative purposes, we assume that the first toolbox consists of TTB, described above. In its basic form, TTB is a deterministic strategy that assumes people make no errors. This is a rather unrealistic assumption because if someone uses TTB but occasionally makes an error, strictly speaking the resulting choice data would contraindicate the application of the strategy. Therefore we allow for the possibility of inconsistent choices due to application errors or "unsystematic noise". We extended the deterministic model with a simple error theory, so that a parameter $\alpha_{\mathrm{TTB}}$ indicates the probability that a decision maker will chose the alternative that was not predicted by TTB in a pairwise choice. In the following, we refer to this probabilistic version of TTB as $\mathrm{TTB}_\alpha$. Other deterministic choice strategies such as WADD can be extended by similar error terms, in an analogous manner.

### Specifying the prior distribution

In the case of a toolbox that only consists of $\mathrm{TTB}_\alpha$ as a single strategy, the only free model parameter is the application error $\alpha'_{\mathrm{TTB}}$. A reasonable prior on $\alpha'_{\mathrm{TTB}}$ may be to assume an average application error of about 10%. Of course, other values are also possible. In any case, the application error will probably vary depending on the situation and the type of experiment. Therefore, a moderate degree of uncertainty concerning the prior distribution seems justified. As $\alpha'_{\mathrm{TTB}}$ may fall within a range from 0 (no implementation error) to 1 (100% implementation error), we choose the prior to be beta distributed. For illustrative purposes, we set the rate and shape parameters of this beta distribution to 1 and 9, resulting in a mean of 0.1 and a standard deviation of 0.09.

$$\alpha'_{\mathrm{TTB}} \sim \mathrm{beta}(1, 9) \qquad (2)$$

### Specifying the likelihood function

Next, a likelihood function needs to be defined that indicates the probability of the observed data given the model. The predictions of a deterministic choice model like TTB are readily available as long as the attributes of options in the experiment are known. In this case, the likelihood is just a function of the implementation error $\alpha$. If a single choice between a pair of options is in line with the deterministic predictions of the model for that pair, then likelihood of that choice equals $1 - \alpha$, otherwise, the

likelihood equals α. Hence, the likelihood of a series of $N$ choices in an experiment is the product of likelihood values for all pairwise choices:

$$p(D|M) = \Pi \ [d_n \times (1- \alpha) + (1 - d_n) \times \alpha] \qquad (3)$$

where $d_n$ is 1 if the decision for the $n^{th}$ pair of options is in line with the deterministic prediction of the model for that pair and 0 otherwise.

For illustrative purposes, suppose a participant in an experiment made $z$ pairwise choices that were inconsistent with TTB's deterministic prediction and $N - z$ choices that were consistent with it. Following Equation 3, the likelihood of this data can be expressed as a Bernoulli function:

$$p(D|TTB_\alpha) = \alpha'_{TTB}{}^{(z)} \times (1 - \alpha'_{TTB})^{(N - z)} \qquad (4)$$

### Deriving $p(D)$ for a single strategy

Once the prior and the likelihood function of a model are specified, the evidence for the observed data $p(D)$ can be estimated. In the present case of a single strategy with one free parameter α, a closed-form solution exists:

$$p(D) = B(a + z, \ b + N - z) \ / \ B(a,b) \qquad (5)$$

where $B$ is the beta function, $a$ and $b$ are the rate and shape parameter of the beta distribution that defines the prior, and $z$ quantifies how many out of a total of $N$ choices are in line with the prediction of the deterministic model.

### Specification for WADD

The outlined approach for $TTB_\alpha$ can be conveyed to other deterministic strategies like WADD. Like TTB, WADD can be extended with a similar beta-distributed error term leading to $WADD_\alpha$. The prior distribution and likelihood function for $WADD_\alpha$ can be specified analogously to $TTB_\alpha$ with the only difference being different deterministic predictions of WADD.

### Comparison between two simple toolboxes

Now that two toolboxes are specified, they can be compared with regard to a given set of data. This case is analogous to a model selection task in which an individual decision maker can be classified as a $TTB_\alpha$ or $WADD_\alpha$ user.

To lay out the approach in a concrete way, we assume a hypothetical decision experiment in which a single participant made 40 choices among pairs of options described on a number of attributes, as in the used-car example outlined in the introduction. We further assume that the options were chosen to differentiate TTB from WADD, such that both strategies would make opposing predictions. In this example, a decision maker chooses option $A$ 30 times and $B$ 10 times. If we set the prior as beta(1,9) for both models, we can calculate the respective marginal likelihoods according to Equation 2. For $TTB_\alpha$, this yields $p(D) = 2.8 \times 10^{-11}$ for $TTB_\alpha$ as compared to

$p(D) = 2.5 \times 10^{-14}$ for $WADD_\alpha$. The ratio of the marginal likelihoods, also known as the Bayes factor (Kass & Raftery, 1995), is 1118:1 in favor of $TTB_\alpha$. Therefore, Bayes' rule clearly indicates that the decision maker should be classified as a $TTB_\alpha$ user.

Next, we outline how this procedure can be extended to toolbox comparisons that include more than one decision strategy.

## Specifying toolboxes with more than one strategy

The concept of a cognitive toolbox relies on the idea that decision makers have several decision strategies available to them. To account for this assumption, we extend the Bayesian approach to toolboxes that contain more than one strategy. Again, precise model specifications are required so that Bayesian techniques can be applied.

### Model specification

For illustrative purposes, we assume toolbox $TB_{TTB,WADD}$ contains two strategies, $TTB_\alpha$ and $WADD_\alpha$. We further assume that an individual decision maker who uses this toolbox will choose according to $TTB_\alpha$ with probability β and according to $WADD_\alpha$ with the complementary probability $(1 - \beta)$. Thus, $TB_{TTB,WADD}$ has three free parameters: The implementation error for TTB in the toolbox ($\alpha_{TTB}$), the implementation error for WADD in the toolbox ($\alpha_{WADD}$), and the probability of selecting $TTB_\alpha$ (β). The likelihood function of this toolbox is simply a function of the likelihood for each single strategy weighted by the probability of selecting it:

$$p(D|TB_{TTB,WADD}) = \beta \times p(D|TTB_\alpha) + (1-\beta) \times p(D|WADD_\alpha) \quad (6)$$

Next, a prior distribution for each of the three parameters needs to be specified. Without any prior knowledge about the probability of selecting $TTB_\alpha$ over $WADD_\alpha$ all possible values between 0 and 1 seem equally likely a priori. Accordingly, we assume that the prior on β is uniformly distributed:

$$\beta \sim uniform(min = 0, \ max = 1) \qquad (7)$$

Likewise, in this example we do not make any a priori assumptions about the probability of particular implementation errors. Thus, we assume priors for $\alpha_{TTB}$ and $\alpha_{WADD}$ are uniformly distributed. Based on these specifications, the marginal likelihood of $TB_{TTB,WADD}$ can be estimated by integrating out all three parameters in the model analogous to Equation 1:

$$p(D) = \iiint p(D| \alpha_{TTB}, \alpha_{WADD}, \beta) \times p(\alpha_{TTB}, \alpha_{WADD}, \beta)$$
$$| \ d\alpha_{TTB}, \ d\alpha_{WADD}, \ d\beta \qquad (8)$$

While this approach is conceptually similar to the case with only one free parameter, it becomes more elaborate to

estimate the integral of Equation 1 as the number of free parameters increases.

## MCMC methods to estimate the evidence

Fortunately, a closed-form mathematical solution, is not mandatory to estimate marginal likelihood because all that is needed is a representative sample from the integral of Equation 1 that is large enough to draw reliable conclusions on the shape of the distribution. Such a sample may be obtained by means of Monte Carlo Markov Chain (MCMC) methods for which statistical packages are readily available (Gilks, Richardson, & Spiegelhalter, 1996). For the present analysis, we utilized utilized the *OpenBugs* software implemented in the *BRugs* package, version 0.51, that can be integrated into the statistics software *R*.

Similar to the mathematical solution outlined above, the implementation in *BRugs* requires the specification of prior distributions and likelihood functions. Provided it is properly implemented, the software returns the evidence as well as a representative sample of the full posterior distribution across all parameters.

## Comparison between a small and a large toolbox

To illustrate the principle of comparing toolboxes that differ in the number of cognitive strategies they contain, we will compare a simple toolbox $TB_{TTB}$ that only consists of $TTB_\alpha$ as a single strategy to a more complex toolbox $TB_{TTB,WADD}$ that contains both $TTB_\alpha$ and $WADD_\alpha$. Thus, $TB_{TTB}$ is nested within $TB_{TTB,WADD}$.

### Transdimensional prior

Instead of calculating the evidence for both toolboxes separately, we directly compared the two models by means of a transdimensional prior θ. This prior acts like a model indicator, controlling which of two models most likely generated the observed data. Thus, θ immediately informs us which of the two toolboxes best describes the choices of an individual decision maker. The parameter θ is Bernoulli-distributed with a prior that assigns equal probabilities to both models (Han & Carlin, 2001; Shiffrin, et al., 2008). Like any other estimated parameter in the model, θ is updated during the course of MCMC simulation. The Bayes factor (*BF*) is simply the odds ratio of this probability, that is, $BF = \theta/(1 - \theta)$. Figure 1 graphically depicts the model's implementation in *OpenBugs*. The graph follows the notation used by Lee and Wagenmakers (2009) and Shiffrin et al. (2008).
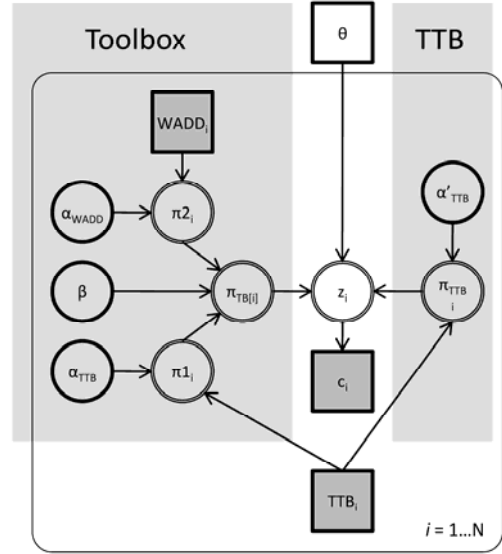


Figure 1: Graphical representation of the model comparison. $TTB_i$ and $WADD_i$ depict the deterministic predictions for each choice *i* out of *N* choices. The notation *c* indicates the actual choice (*A* or *B*) and π depicts the probability of choosing *A* over *B* as predicted by the particular model.

## Model recovery

To test this approach, we set up a model recovery simulation based on 1,000 pairs of options described on 30 attributes to ensure that the results would not be influenced by an extreme constellation of binary attribute values that were randomly drawn from a Bernoulli distribution with *p* = .5. The importance weights for each attribute were set to increase linearly from the least to the most important attribute. Next, we simulated hypothetical decision makers who repeatedly chose among the pairs of options according to either $TTB_\alpha$ or $WADD_\alpha$ (i.e., ß was set to either 1 or 0). The implementation error α for both decision strategies varied from 0 to 0.5 in steps of 0.1 across decision makers. For each value of α in the simulation, we estimated θ by sampling from three separate MCMC chains in OpenBugs that were run for 2,000 steps each with a thinning of 10.

For $TB_{TTB,WADD}$ the priors were set to uniform distributions ranging from 0 to 0.5 for $\alpha_{TTB}$ and $\alpha_{WADD}$ and from 0 to 1 for β. For $TB_{TTB}$, the prior on $\alpha'_{TTB}$ was set to a uniform distribution ranging from 0 to 0.5.

## Predictions

If a comparison based on Bayesian evidence is a feasible way to solve the problem of strategy sprawl, the method should assign more evidence to the model that generated the data. Thus, if the data was generated by choices according to $TTB_\alpha$, the evidence for a simple toolbox $TB_{TTB}$ should be higher than that of a toolbox $TB_{TTB,WADD}$ even though the latter one contains $TTB_\alpha$ as a special case.

## Results

The samples from the three estimated chains in OpenBugs provided representative samples as indicated by a visual inspection of the trace plot, the autocorrelation and the Gelman−Rubin statistic. The results clearly indicate that the marginal likelihood for a smaller toolbox can indeed be higher than that for a larger toolbox. Figure 2 plots the actual implementation error $\alpha_{TTB}$ for choices based on $TTB_{\alpha}$ (Figure 2a) and the actual implementation error $\alpha_{WADD}$ for choices based on $WADD_{\alpha}$ (Figure 2b) against the estimated $\theta$. Here, $\theta$ indicates the probability of the more complex $TB_{TTB,WADD}$ over $TB_{TTB}$. As can be seen from Figure 2a, $\theta$ increases as $\alpha_{TTB}$ increases. This indicates that a decision maker who uses $TTB_{\alpha}$ with a small implementation error is better described by $TTB_{\alpha}$ as compared to $TB_{TTB,WADD}$. Likewise, a decision maker who chooses according to $WADD_{\alpha}$ is clearly better described by $TB_{TTB,WADD}$ even if the implementation error $\alpha_{WADD}$ is large. This relationship seems plausible because the larger toolbox contains $WADD_{\alpha}$ whereas the smaller toolbox does not.

For an $\alpha$ of 0.5, which indicates random choice, Bayes' rule tells us to favor $TTB_{\alpha}$, because in the case of very noisy data a simpler model is favored over a more complex one. Together, the results show that a small toolbox with only one strategy should be preferred over a more complex one provided that the small toolbox contains the appropriate tool to describe the initial decision-making process.
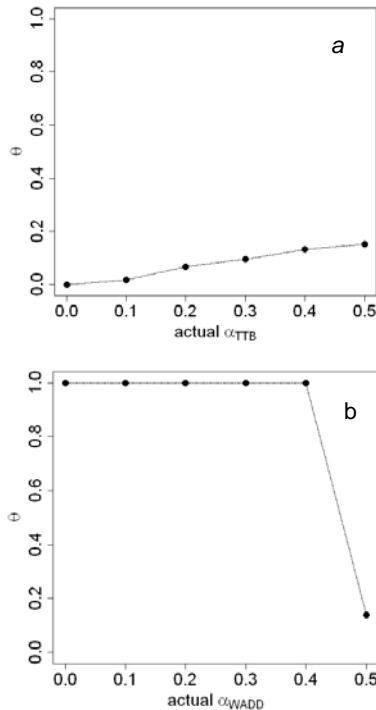


Figure 2: Plot of $\theta$ against the implementation error $\alpha_{TTB}$ (Figure 2a) and $\alpha_{WADD}$ (Figure 2b) used in the simulation. $\theta$ indicates the evidence in favor of $TB_{TTB,WADD}$ relative to $TB_{TTB}$.

## Parameter recovery

The results so far show that Bayesian methods can be fruitfully applied to solve the question of how many strategies a toolbox should contain. However, we have implicitly assumed that the estimation methods accurately estimate the free parameters of the choice models within the toolboxes. To test if this condition really holds, the actual parameters used in the choice simulation can be compared with their respective posterior estimates. Figure 3a shows the marginal of the posterior distribution for $\alpha_{TTB}$ in the toolbox plotted against the actual parameters used in the simulation. As can be inferred from Figure 3, the posterior distributions of $\alpha_{TTB}$ match the actual values of $\alpha_{TTB}$ used in the simulation quite well. The parameter recovery for the $\alpha_{TTB}$ parameter of the simple $TTB_{\alpha}$ model appears similar.

Figure 3b shows the estimated $\beta$ values plotted against $\alpha_{TTB}$ parameters used in the simulation. For low values of $\alpha_{TTB}$, estimated $\beta$ values are high. As $\beta$ indicates the probability of using $TTB_{\alpha}$ over $WADD_{\alpha}$ within the toolbox, this relationship also seems plausible.

Figure 3 also shows the actual values for $\alpha_{WADD}$ from the choice simulation plotted against the posterior of $\alpha_{WADD}$ (Figure 3c) and $\beta$ (Figure 3d) in the toolbox. Again, the estimated values match up with the actual values, indicating that the parameters were recovered across the whole parameter space. For high values of $\alpha_{WADD}$ the Bayesian model slightly underestimates the implementation error. Presumably this is the case because the prior distribution constrains the parameter space between 0 and 0.5. If the prior distribution is extended to range from 0 to 1, the estimated parameters match more closely.
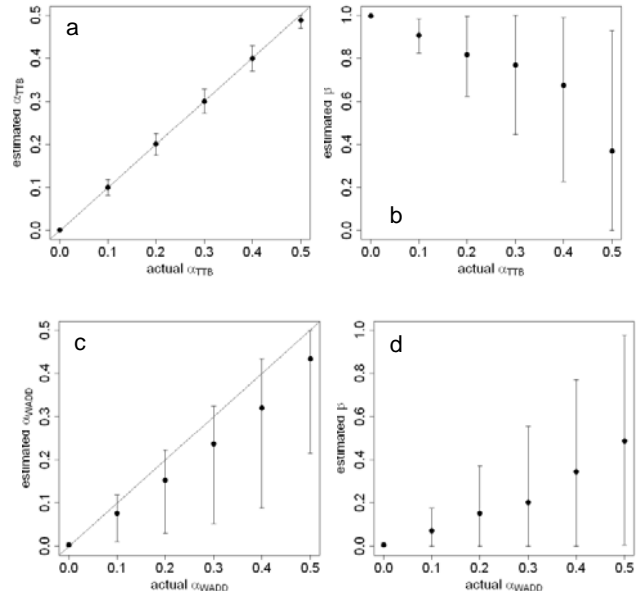


Figure 3: Parameter recovery $TB_{TTB,WADD}$ for choices according to $TTB_{\alpha}$ (Figure 3, a & b) and $WADD_{\alpha}$ (Figure 3, c & d). Error bars indicate the 95% highest probability density region of the posterior distribution.

## Discussion

Due to higher degrees of freedom, a cognitive toolbox that includes many strategies will always provide a better fit to the data, but it will not necessarily provide the highest evidence or marginal likelihood. The results of our analyses indicate that the marginal likelihood within a Bayesian framework can be fruitfully used to determine the number of tools to include in a toolbox. Thus, Bayesian statistics are well suited to prevent strategy sprawl.

As outlined above, the marginal likelihood can only be estimated with regard to a specific set of choice data and to precisely defined cognitive strategies. Thus, the number of tools to include may vary depending on cognitive strategies included in the toolbox. Yet, within these boundaries, the approach indicates that a small toolbox may be preferred over a large toolbox if the small toolbox contains a tool that describes the data well—even if the small toolbox is nested within the larger one.

The reason why the marginal likelihood provides a common comparison metric is because it implicitly accounts for differences in model complexity. This happens because the prior probability of each possible combination of parameters decreases with an increase in parameters. This carries over to the marginal likelihood that weights the likelihood of the data by the probability of each combination of possible parameter values. Thus, even though the likelihood of the data can be expected to increase with more free parameters, this increase is counteracted by a lower prior probability for each possible combination of parameters.

A comparison of the prior distributions of $TB_{TTB}$ and $TB_{TTB,WADD}$ illustrates this mechanism. The prior probability of $TB_{TTB}$ follows a beta-distribution around a single parameter $\alpha$. On the other hand, the prior probability of the parameters in $TB_{TTB,WADD}$ embrace a total of three parameters. As prior distributions are probability distributions, they must integrate to 1. With more parameters in the model, the probability of each specific combination of parameter values should decrease because parameter space is more spread out.

### Limitations and future research

Here we demonstrated that the Bayesian approach provides a powerful statistical tool for comparing and evaluating cognitive toolboxes that contain rather few strategies. In principle, the same approach can also be used for more complex scenarios. The only constraint of this methodology lies in the potential difficulties of implementing efficient MCMC sampling for vastly more complex models. As long as these computational challenges are met, the approach is not constrained to comparing toolboxes but can also be extended to compare different toolboxes against alternative cognitive models that do not conform to the notion of a repertoire of strategies.

## References

Brown, N. R. (1995). Estimation strategies and the judgment of event frequency. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*, 1539–1553.

Coyle, T. R., Read, L. E., Gaultney, J. F., & Bjorklund, D. F. (1998). Giftedness and variability in strategic processing on a multitrial memory task: Evidence for stability in gifted cognition. *Learning & Individual Differences, 10*, 273–290.

Dougherty, M. R., Thomas, R., Franco-Watkins, A. M. (2008). Postscript: Vague Heuristics Revisited. *Psychological Review,115*, 211-213.

Gigerenzer, G., Todd, P. M., & the ABC Research Group. (1999). *Simple Heuristics That Make Us Smart*. New York: Oxford University Press.

Gilks, W. R., Richardson, S. and Spiegelhalter, D. J. (1996) *Markov Chain Monte Carlo in Practice*. Chapman and Hall, London.

Han, C. & Carlin, P. (2001). Markov Chain Monte Carlo methods for computing Bayes factors: a comparative review. *Journal of the American Statistical Association, 96*, 1122-1132.

Kass, R. E. & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association, 90*, 773-795.

Lee, M. D. & Wagenmakers, E.J. (2009). *A Course in Bayesian Graphical Modeling for Cognitive Science*. Unpublished lecture notes. http://users.fmg.uva.nl/ewagenmakers/BayesCourse/BayesBook.pdf.

Newell, B. R. (2005). Re-visions of rationality. *TRENDS in Cognitive Sciences, 9*, 11-15.

Payne, J. W., Bettman, J. R., & Johnson E. J. (1993). *The Adaptive Decision Maker*. Cambridge: Cambridge University Press

Patalano, A. L., Smith, E. E., Jonides, J., & Koeppe, R. A. (2001). PET evidence for multiple strategies of categorization. *Cognitive, Affective & Behavioral Neuroscience, 1*, 360–370.

Shiffrin, R. M., Lee, M. D., Kim, W., & Wagenmakers, E.-J. (2008). A survey of model evaluation approaches with a tutorial on hierarchical Bayesian methods. *Cognitive Science, 32*, 1248-1284.

Siegler, R. S. (1991). Strategy choice and strategy discovery. Learning and Instruction, Volume 1, Issue 1, 1991, 89-102.