Variability in behavior that cognitive models do not explain can be
linked to neuroimaging data

Sebastian Gluth & Jörg Rieskamp

Department of Psychology, University of Basel

Correspondence:

Sebastian Gluth

Department of Psychology, University of Basel, Missionsstrasse 62a, 4055 Basel, Switzerland

sebastian.gluth@unibas.ch

Tel.: 0041 61 267 0606

Abstract

It is known that behavior is substantially variable even across nearly identical situations. Many cognitive models are not able to explain this *intraindividual* variability but focus on explaining *interindividual* differences captured in model parameters. In sequential sampling models of decision making, for instance, one single threshold parameter value is estimated for every person to quantify how much evidence must be accumulated for committing to a choice. However, this threshold may vary across trials even within subjects and experimental conditions. Neuroimaging tools such as functional magnetic resonance imaging (fMRI) or electroencephalography (EEG) can reveal moment-to-moment fluctuations in the neural system that are likely to contribute to fluctuations in behavior. We propose that neural and behavioral variability could be linked to each other by assuming and estimating trial-by-trial variability in model parameters. To illustrate our proposal, we first highlight recent studies in model-based cognitive neuroscience that have gone beyond correlating model predictions with neuroimaging data. These studies made use of variance in behavior that remained unexplained by cognitive modeling but could be linked to specific fMRI or EEG signals. Second, we specify in a tutorial a novel and efficient approach, how to extract such variance and to apply it to neuroimaging data. Our proposal shows how the variability in behavior and the neural system can provide a fruitful source of theory development in cognitive neuroscience.

Keywords: cognitive modeling; fMRI; EEG; intraindividual differences; Bayes; decision making

1. Introduction

Cognitive neuroscience and cognitive modeling pursue a common goal, which is to identify the various mechanisms that give rise to observable behavior. It is therefore not surprising that their combination has become an extremely productive research agenda. For example, model-based functional magnetic resonance imaging (fMRI; O'Doherty, Hampton, & Kim, 2007) has been applied to link activation in the brain's reward circuitry, including the ventral striatum and the ventromedial prefrontal cortex, to the parametric encoding of reward expectations and reward prediction errors (Bartra, McGuire, & Kable, 2013; Clithero & Rangel, 2014; Garrison, Erdeniz, & Done, 2013). Expected values and prediction errors are particularly suitable for neuroimaging data analysis, because they vary on a trial-by-trial basis when using reinforcement learning (RL) paradigms. The predictions of an RL model—that has been fitted to the choice data—are simply regressed against the fMRI data, revealing which brain areas track the development of expected values or prediction errors over the course of the experiment (e.g., Daw, O'Doherty, Dayan, Seymour, & Dolan, 2006; Gläscher, Hampton, & O'Doherty, 2009; Gluth, Rieskamp, & Büchel, 2014; O'Doherty et al., 2004; Pessiglione et al., 2008).

Importantly, a person will perceive an identical decision situation differently after receiving feedback, which leads to learning and consequently to different behavior across trials. In contrast to learning situations, however, many psychological experiments use static environments, and potential changes in cognitive processes from trial to trial are not accounted for. Even an RL model does not predict any trial-by-trial variability, as long as no learning takes place (e.g., if the feedback is fully predictable). In the worst case, some models make deterministic predictions and do not allow accounting for any form of variability. Traditional models of decision making under uncertainty, including expected utility theory (Von Neumann & Morgenstern, 1947), prospect theory (Kahneman & Tversky, 1979), but also some choice heuristics (e.g., Brandstätter, Gigerenzer, & Hertwig, 2006), belong to this

category. Applying a naïve error theory allows formulating choice predictions in terms of a probability distribution (Hey, 1995; Loomes & Sugden, 1995; McFadden, 2001; Rieskamp, 2008). Beyond that, sequential sampling models of decision making assume that stochastic sampling of evidence underlies the decision process, which allows probabilistic predictions of response times (Busemeyer & Townsend, 1993; Smith & Ratcliff, 2004). However, even probabilistic choice models make identical probabilistic predictions for identical choice situations, so that if the probability of choosing option A over alternative option B is predicted to be, say, 60% in one trial, it will also be 60% when the same situation is encountered a few trials later. That the subject might actually choose A in the first but B in the second situation (instead of B and then A, or A in both trials, or B in both trials) has to be attributed to unsystematic noise the model cannot explain.

More precisely, when model parameters are assumed to be fixed for specific conditions or persons, variability in behavior is difficult to explain (Lee & Wagenmakers, 2014; Lewandowsky & Farrell, 2011). In sequential sampling models of decision making, for instance, a single decision threshold parameter value is estimated that reflects the amount of evidence required to make a choice (Bogacz, Brown, Moehlis, Holmes, & Cohen, 2006; Smith & Ratcliff, 2004). This parameter might vary across different conditions, for example, when either speed or accuracy is emphasized. But within a single condition the threshold is assumed to be fixed, and its value is usually estimated based on hundreds or even thousands of trials (e.g., Ratcliff & Rouder, 1998). Naturally, the behavior varies across these trials, and the sequential sampling model “explains” this by the variability of the stochastic process of sampling information. However, we argue that the variability in behavior across trials could also be due and explained by the variability of model parameters such as a decision threshold (see also Craigmile, Peruggia, & Van Zandt, 2010; Zandbelt, Purcell, Palmeri, Logan, & Schall, 2014). Moreover, behavioral variability should in principle be reflected in variability of the recorded brain data (see also Glimcher, 2005). We suggest that the variability of

behavior and the neural system could be matched to each other and might offer important insights about the neurocognitive system and the models trying to explain it.

Indeed, recent studies in model-based cognitive neuroscience have tried to exploit this trial-by-trial variability by first contrasting a model's average predictions with the participants' actual behavior in single trials (i.e., response times [RTs] and/or choices) and then using this prediction–observation discrepancy to inform the analysis of neuroimaging data (Gluth, Rieskamp, & Büchel, 2012; Gluth, Sommer, Rieskamp, & Büchel, 2015; van Maanen et al., 2011). The central aim of this review and tutorial is to present a generalization of this approach to illustrate its applicability to (almost) any form of cognitive model and (almost) any form of neural or physiological data.

The rest of the article is structured as follows: In Section 2, we summarize neuroimaging studies of human decision making that captured trial-by-trial deviations from model predictions to analyze the brain data as outlined above; in Section 3, we provide a general formalization of this approach rooted in Bayesian theory, as well as an accompanying example to facilitate understanding; in Section 4 we discuss and exemplify parameter recovery simulations as a precondition for the successful application of our technique; in Section 5 we specify further requirements and recommendations for models and neuroimaging data; Section 6 provides a comparison with alternative ways to test for trial-by-trial effects; finally, Section 7 concludes with a discussion of potential results that can be obtained with this technique and their implications for further developments of theories in cognitive neuroscience.

2. Model-based neuroimaging studies of parameter variability

Whereas reinforcement learning studies exploited the fact that many learning components such as reward expectations, prediction errors, but also learning rates and various forms of uncertainty are naturally changing (Behrens, Woolrich, Walton, & Rushworth, 2007;

Boll, Gamer, Gluth, Finsterbusch, & Büchel, 2013; Krugel, Biele, Mohr, Li, & Heekeren, 2009; McGuire, Nassar, Gold, & Kable, 2014; Payzan-LeNestour, Dunne, Bossaerts, & O'Doherty, 2013), models of decision making (including sequential sampling models) face the problem that predictions often remain constant from trial to trial.¹ In fact, observations are assumed to be generated by the same probability distribution (e.g., a Wiener diffusion process) and to be mutually independent; that is, they are independent and identically distributed (i.i.d.) random variables (Luce, 1986). Accordingly, early model-based fMRI studies on perceptual decision making restricted the association of cognitive modeling and brain data to correlations across subject (Forstmann et al., 2008). In particular, Forstmann and colleagues used the linear ballistic accumulator (LBA) model (Brown & Heathcote, 2008) to model choices and RTs in a random dot motion task with different conditions that prioritized speed, accuracy, or neither. The LBA was estimated so as to allow specifying the level of response caution (i.e., the distance between the decision threshold and the upper end of the starting point distribution) in speed and accuracy conditions separately. As expected, subjects were less cautious in the speed condition. At the time when people were instructed to respond quickly or accurately in the upcoming trial, the authors recorded the fMRI signal in the pre-supplementary motor area (pre-SMA) and the caudate nucleus (CN). They reported a negative correlation between the difference in response caution in the speed and accuracy conditions and the difference in fMRI signal in pre-SMA and CN. Their conclusion was that a higher increase in baseline activity in the pre-SMA and CN mediated particularly fast decisions in the speed condition.

Forstmann et al.'s (2008) results clearly indicate an involvement of the pre-SMA and CN in response caution. However, if we consider the possibility that the decision threshold

¹ Ratcliff's drift diffusion model assumes across-trial variability in the drift rate, starting point, and non-decision time (Ratcliff, 1978; Ratcliff & McKoon, 2008). However, the model only estimates the "average" amount of across-trial variability and does not specify the amount or direction of variability in single trials. Thus, it makes no trial-by-trial predictions of how the parameter values vary. More recent modeling techniques have been developed to address this point (Turner, van Maanen, & Forstmann, 2015; Wiecki, Sofer, & Frank, 2013), which we discuss in more detail in Section 6.

does not stay constant across trials within a specific condition, then the study leaves open whether these brain structures could also be important for threshold adjustments on a trial-by-trial level. The following hypothetical example illustrates the point: The average “brain activity” (across all trials and subjects) in the pre-SMA/CN could be $x_a = 10$ in the accuracy condition and $x_s = 20$ in the speed condition; the average value of response caution parameter could be $\pi_a = 20$ in the accuracy condition and $\pi_s = 10$ in the speed condition. Thus, in this example there is a negative relationship between brain activation x and model parameter π , and they might further be negatively correlated across subjects (as in Forstmann et al., 2008). Such “results” would indicate that whenever activity in the pre-SMA/CN is high, response caution is low. Yet, it is possible that on a trial-by-trial level, just the opposite is true: In one trial within the accuracy condition with a particularly low pre-SMA/CN activation of $x_{a,t=1} = 8$, response caution was actually not that high ($\pi_{a,t=1} = 18$) compared to another trial with a stronger brain signal ($x_{a,t=1} = 12$; $\pi_{a,t=1} = 22$). Similarly, within the speed condition response caution could be higher in a trial with a particularly strong pre-SMA/CN signal ($x_{s,t=1} = 22$; $\pi_{s,t=1} = 12$) compared to another trial with weaker activation ($x_{s,t=1} = 18$; $\pi_{s,t=1} = 8$). In other words, within conditions there could be a positive correlation between x and π and not the negative correlation that the aggregate data suggests. Such a reverse pattern might be unlikely, but cannot be ruled out. Furthermore, it is possible that simply no pattern at the trial-by-trial level exists. Hence, it appears unjustified to infer cognitive and neural processes in single trials from brain–behavior correlations across subjects (cf. Robinson, 1950).

Van Maanen et al. (2011) have recently addressed the idea of trial-by-trial adjustments of response caution in perceptual decision making (Fig. 1A). The authors extended the LBA model to allow for single-trial variability of the starting point (and the drift rate). To do this, they first fitted the standard LBA to the entire RT distribution and then derived the maximum likelihood estimates of starting points for each trial based on the trial-specific RTs. In other

words, they asked: “Keeping all other parameters fixed (to the values derived from LBA), what is the most likely parameter value for the starting point in trial t given the RT in t ?” Intuitively, faster RTs should indicate higher starting points, but RTs are also affected by other parameters (i.e., drift rate, decision threshold, non-decision time). The single-trial starting point estimates were then correlated with single-trial estimates of the fMRI hemodynamic response as derived from an independent component analysis approach (see Eichele et al., 2008). Van Maanen et al. found that the trial-by-trial signal of an independent component comprising the pre-SMA was linked to the trial-wise estimates of starting points. Interestingly, a component that included the CN was also related to starting points but in a different way from the pre-SMA (only in speed-emphasis trials that directly followed accuracy-emphasis trials), illustrating how single-trial analyses can provide new insights into brain–behavior relationships. The single-trial LBA by van Maanen et al. was also used in combination with EEG data to link variability in response caution with the amplitude of the contingent negative variation (Boehm, van Maanen, Forstmann, & van Rijn, 2014) as well as with fMRI data to link variability in the drift rate with tuning curves in the primary visual cortex (Ho et al., 2012).

Further evidence for a trial-wise association between response caution and activation in the pre-SMA and CN comes from an fMRI study on the emergence of value-based decisions (Gluth et al., 2012) (Fig. 1B). In this study, participants had to decide whether to take or reject a financial offer of unknown value based on sequentially revealed probabilistic cues. A sequential sampling model was developed to predict which decision was made and how many cues were sampled. One step in the analysis of the fMRI data addressed whether the trial-wise discrepancy between the number of sampled cues predicted by the best model and the actual number of sampled cues by the participants was forecasted by fMRI signals at the onset of each trial. The authors found a network of brain areas including the pre-SMA and CN to be related to this discrepancy estimate, meaning that when activity in the pre-SMA and

CN was high at the beginning of a trial, subjects tended to sample fewer cues (i.e., behaved less cautiously) than predicted by the model. Importantly, the discrepancy does not reflect merely the number of sampled cues, because the model's predictions varied from trial to trial based on what information was conveyed by the respective sequences of cues.

Fig. 1. Neuroimaging studies testing for trial-by-trial variability in model estimates. (A) The single-trial LBA model estimates the most likely height of the starting point distribution in each trial based on trial-wise RTs in perceptual tasks such as the random dot motion paradigm (left). The related response caution is (negatively) linked to fMRI signals in the pre-SMA (right, red cluster). (B) In a sequential, value-based choice task (left), a sequential sampling model predicted the cumulative sum of probability (CSP) of making a choice after 1 to 6 samples (middle). However, only one choice per trial can be observed (the example shows two possible observations at 3 or 5 samples). Similar to (A), the CSP at the actual choice point approximates response caution and is also (negatively) linked to pre-SMA activation (right). (C) In a memory-based choice paradigm, subjects first learned to associate (food snack) options with positions on the screen and then had to choose between two options by recalling

these associations (left). A sequential sampling model was used to describe decisions and to derive the posterior probability of remembering a snack at the decision stage (middle). In this example, snack #37 is not chosen even though it has a higher subjective value than snack #1. Accordingly, the memory probability for snack #37 ($Posterior_x$) is decreased compared to the initial estimate (P_x) that is based on modeling all trials. A higher value of $Posterior_x$ is linked to higher activation in the right hippocampus at the encoding stage (right).

Of course, mapping trial-by-trial fluctuations in behavior with variability in parameter values of cognitive models is not restricted to a specific process such as response caution in sequential sampling models but can be generalized to various processes of other models. For instance, in a recent study, Gluth and colleagues (2015) again used the discrepancy between average model predictions and single-trial behavior to infer the strength of episodic memory on the basis of memory-guided decisions (Fig. 1C). While lying in the MR scanner, their participants first had to learn associations between different food snacks and different positions on the screen (encoding phase). Later, two locations were highlighted, and participants had to recall the associated snacks and to choose the snack they preferred (decision phase). Gluth and colleagues (2015) modeled decisions by means of a sequential sampling model tailored to the mnemonic demands of the task. More specifically, the model included a memory parameter to account for interindividual differences in the ability to recall the snacks in the decision phase. Critically, this parameter reflects a person's average memory capacity over all decision trials and does not tell whether a particular snack has been remembered or forgotten. On the other hand, observing that a person does not choose a very good snack (on several occasions) indicates that the subject might have forgotten about it. The authors followed this reasoning and used a Bayesian approach to derive the posterior probability that a snack S had been remembered based on the choices and choice sets that included S . Technically, the cognitive model was re-run over these choice sets under the assumptions that S had been remembered or forgotten to obtain the likelihood of the observed behavior. This likelihood and the average "memory" parameter, which served as the prior, were then used to derive the posterior estimate of the "memory" parameter for the particular

snack S according to Bayes's theorem. The resulting probability was then used in the fMRI data analysis as a parametric modulator (Büchel, Holmes, Rees, & Friston, 1998) of the hemodynamic response function during encoding. The rationale of this analysis was that the model-based probability of remembering a choice option is similar to the well-established subsequent memory effect, which refers to a stronger fMRI signal in the hippocampus for (later) recalled versus forgotten items at the time when items are stored in memory (Kim, 2011; Paller & Wagner, 2002). Indeed, Gluth and colleagues (2015) found that their snack-specific estimate of memory strength was correlated with activation in the right hippocampus in the encoding phase, showing that we can make (trial-by-trial) inferences about related cognitive processes such as memory or attention on the basis of observing and modeling decisions.

3. A general framework for capturing variability

In this section, we introduce a general approach for capturing trial-by-trial variability by assuming variability of the values of model parameters that represent specific mechanisms of the model. Our method is most similar to the approach we followed in Gluth et al. (2015), in which we determined trial-by-trial posterior estimates of model parameters using a Bayesian approach. For better comprehensibility, we supplement the general formalization with a concrete example taken from Gluth et al. (2012). As described above, in this study trial-by-trial estimates of response caution were based on the probability that a choice has been made until the time point at which it was actually made (see middle panel of Fig. 1B). This is, however, just an approximation of the most likely decision threshold parameter, and we show how to arrive at more precise (Bayesian optimal) trial-by-trial thresholds in this task.

Let us start with the general framework. We suppose that after estimating a model M that contains parameter π we still observe some variability of behavior across nearly identical

situations that cannot be explained by the model. We suggest that the variability of the behavior could be explained by the model M if one assumes that parameter π is not fixed but also varies across these identical situations. To explain the variability of behavior by variability of this parameter we are interested in estimating the exact value of π for a specific situation or trial. More specifically, we are aiming at deriving the value for $\pi_{s,t}$ of subject s in trial t that has the highest posterior probability given the data D at t (i.e., choice and/or response time at t) and the prior probability of different values for π with respect to the subject's data D_s in all trials. The posterior probability $P(\pi_{s,t} | D_{s,t})$ is given by Bayes's rule:

$$P(\pi_{s,t} | D_{s,t}) = \frac{P(D_{s,t} | \pi_{s,t}) \times P(\pi_s)}{\int P(D_{s,t} | \pi_{s,t}) \times P(\pi_s) d\pi_s}, \quad (1)$$

where $P(D_{s,t} | \pi_s)$ is the likelihood of the data of s at t given π , and $P(\pi_s)$ is the prior probability of π 's value (over the entire time series). The integral sign shows that one has to marginalize over all possible values of π_s . The procedure can thus be divided into five steps:

1. First estimate the model M across all trials on the basis of the entire data set D_s to get the most likely parameter set θ_s for subject s (which includes a single estimate of π_s).
2. Define a reasonable range of parameter values for π_s (if this has not already been done in Step 1). For example, the value of a parameter representing the probability of remembering information should lie between 0 and 1.
3. Get $P(\pi_s)$ for all possible values of π_s by re-running model M over all trials, keeping the parameter set θ_s constant except for π_s and extracting the model's overall fits (i.e., the likelihoods). Although $P(\pi_s)$ is estimated on the basis of the data and the estimated parameter set θ_s , it will function as the prior for Equation 1. Therefore, $P(\pi_s)$ has to be normalized so that it sums up to 1.

4. Get $P(D_{s,t} | \pi_s)$ for all possible values of π_s by re-running the model M only over trial t and extracting the probability that M predicts $D_{s,t}$ correctly. $P(D_{s,t} | \pi_s)$ will serve as likelihood in Equation 1 (and thus does not have to be normalized).
5. Compute Equation 1 for all possible values of π_s and take the $\pi_{s,t}$ that has the highest posterior probability.

Steps 4 and 5 have to be repeated for every trial (and every subject) to get the entire series of trial-by-trial variability in π . These values can then be subjected to various neuroimaging analyses; for instance, they might be implemented as a parametric modulator in a general linear model analysis of fMRI data (see also Section 6).

As a practical example, we now apply this general framework to derive trial-by-trial estimates of the decision threshold of the sequential sampling model proposed by Gluth et al. (2012). As stated above, this model predicts how many pieces of information a subject samples before buying or rejecting an offer. Fig. 2A shows an example trial of the experiment. In each trial, the subject could sample up to six cues to decide whether to accept or to reject a stock offer. Depending on the performance of the stock the subject received a positive or negative payoff. Possible cue values were +2, +1, -1, and -2, where positive (negative) values increased the probability that accepting the current offer leads to a gain (loss) of money. Cues appeared/disappeared consecutively after 2 to 4 seconds until a decision was made, and acquiring a cue led to a fixed amount of information costs. If a choice was made before presenting the last cue, the remaining cues were not revealed. Participants were not informed about whether their choice was correct or incorrect. Generation of cue sequences was randomized. The sequential sampling model tracked the development of evidence for/against accepting the offer by means of the log-evidence. However, we assumed that this estimate of evidence was affected by Gaussian noise, with the amount of noise being a free parameter. The model predicted a choice when an upper (accept) or a lower (reject) decision threshold was crossed. The decision variable (i.e., the accumulated evidence) was

quantified as the log-evidence for buying the offer being the correct response. Here, we are interested in the initial threshold and in deriving trial-by-trial estimates for it (so we keep the decrease of the threshold and the noise parameter constant).

In our example, we focus on a single trial (#103) of a real subject (#24) from the study who behaved close to the group average with respect to how many cues were sampled (group average: 4.16 ± 0.70 *SD*; Subject 24: 4.25). The cues from 1 to 6 in this trial were +2, -2, -2, +1, +1, and +1, so after suggesting accepting the offer at Cue 1 the evidence switched to rejecting at Cues 2 and 3 and then turned back slowly to accepting at Cue 6 (Fig. 2A). The subject rejected the offer after seeing the first three cues. To illustrate the estimation of trial-by-trial variability, we imagine that this trial was repeated two times later in the experiment, but now the subject did not reject after three cues but accepted the offer after one observed cue or after six observed cues. We now examine how the posterior estimate of the threshold changes across these three trials.

Fig. 2. Example trial and estimated choice probability. (A) Sequence of sequentially presented cues. (B) Choice probability for accept and reject decisions at each cue for Subject 24 based on fitting the best performing sequential sampling model in Gluth et al. (2012) to all trials of this subject. The probabilities sum up to 1. As can be seen, the model predicts that choices after sampling only one or two cues are very unlikely.

As mentioned, the first step is to estimate parameter set θ_s based on all trials. The best fitting (initial) threshold value is 1.729 (unlike in the original study, we use a single value for accept and reject thresholds for the sake of simplicity). Fig. 2B presents the predictions of the model for our trial of interest. These predictions are derived by estimating the probability that the decision variable (i.e., the sampled log-evidence) crosses either the buy or the reject threshold after 1 to 6 cues, with the thresholds decreasing from cue to cue (see Gluth et al., 2012, for details). Interestingly, an accept decision after 6 cues is deemed more likely than what the subject actually did (rejected after 3 cues), indicating the subject might have behaved less cautiously in this particular trial than on average.

In Step 2, we specify the range between 0.001 and 3 (with a step size of 0.001) as our range of reasonable, possible estimates of the threshold parameter. This is somewhat arbitrary, but a threshold of 0 or lower makes little sense because the model would always predict immediate responses, and with a threshold of higher than 3 (while keeping all other parameters constant) the model would always predict responses at the last cue. For this parameter range, the model is run again over all trials to obtain the likelihood values at each possible threshold (Step 3). Because these likelihoods will be used as the prior probability of threshold values [i.e., $P(\pi_s)$], they must be normalized so that they sum to 1 (see above). As can be seen in Fig. 3A, there is only a very narrow range of probable values for the threshold parameter. Re-running the model is again used in Step 4, but now this is restricted to our trial of interest to obtain the likelihood of the current behavior given the possible threshold values

[(i.e., $P(D_{s,t} | \pi_s)$].² Fig. 3B depicts these likelihoods for accepting after 6 (dark blue), accepting after 1 (light blue), and rejecting after 3 (red) cues. As expected, later choices are associated with higher thresholds. Notably, the likelihood is maximal for many low (high) thresholds in the case of accepting at the beginning (end), because any of these thresholds predict that a choice is made for sure at the first (last) cue. Hence, uncertainty about the single most likely threshold value remains in such cases. This problem dissolves after applying Bayes's rule as given by Equation 1 (Step 5), where priors and likelihoods are combined (Fig. 3C). The fact that probable threshold values are restricted to a very narrow range is determined by the narrow range of priors (see Fig. 3A). Nevertheless, the peak values differ between the three possible observations, showing that our method can effectively capture trial-by-trial variability with the parameter (only the relative and not the absolute distance of peak values matters). If the peak values should all be the same, one could increase the granularity of the search for possible parameter values in Steps 2 to 4 (e.g., using a step size of 0.0001 or smaller).

² Actually, Step 4 can be realized during Step 3 by saving each likelihood value per trial and possible threshold value.

Fig. 3. Estimation of the posterior probability of the initial threshold values in the example trial with 3 possible observations. (A) Prior probability of possible initial threshold values according to fitting the model to all trials. (B) Likelihood of initial threshold values in the example trial given each observation. (C) Posterior probability of initial threshold values (modes are labeled). Notably, the modes of the two likely observations (accept at 6, reject at 3) are close to the estimated threshold of 1.729 (indicated by the dashed line), whereas the mode for the unlikely decision after 1 cue lies far off to the left side. The narrower range of the x -axis in (C) is chosen to highlight differences between the three distributions.

This procedure has to be done for all trials to obtain the entire time course of trial-by-trial variability in the experiment. Fig. 4A shows the posterior estimates of thresholds together

with the choice points for all trials of Subject 24. A comparison of thresholds and choice points is important because on the one hand there should be a positive relationship between them (earlier choices should imply lower thresholds), but on the other hand a perfect correlation is undesirable, as it would render the trial-by-trial estimation of thresholds redundant. In fact, the correlation of thresholds and choice points for this subject is .58, so there is a substantial dependency but also sufficient unexplained variance (implying that thresholds and choice points may be differently related to the neuroimaging data). This is reasonable because choosing early or late in this task heavily depends on how congruent the cues are: If the cues are all positive or all negative, the subject might decide early even with a relatively high threshold; if consecutive cues are constantly canceling out, the subject might decide late even with a rather low threshold. Looking at the time course is also helpful to see whether there is enough variance in itself (if not, one could try out a smaller step size) as well as to detect outliers and slow drifts in the development of the parameter. Outliers could distort the neuroimaging analysis, and slow drifts are usually not measurable with fMRI or EEG due to high-pass filters and baseline corrections.

Fig. 4. Comparison of trial-by-trial thresholds and choice points. (A) Development of threshold values (upper panel) and choice points (lower panel) over the course of the experiment. (B) As expected, there is a substantial but not perfect correlation between these two variables.

4. Parameter (variability) recovery

Fluctuations in behavior are not necessarily due to systematic changes in the neurocognitive system but might just be random noise. For instance, if you generate simulations of a sequential sampling model with a noisy evidence accumulation process, you will observe variability in response times and accuracy rates even without parameter variability. Apart from the more general question of whether parameter values are always variable (which we discuss in Section 7), this raises the issue of parameter recovery, or in our case of “parameter variability recovery”: Given that there is variability in the parameter of interest, does this variability produce such (fluctuating) behavior that we are able to capture it with our method? We think that researchers who seek to apply our method must conduct parameter recovery tests (preferably before data collection) to evaluate whether it is possible to identify the variability in parameter values that might cause variability in behavior. Thus a recovery study is able to evaluate whether a cognitive model and a planned experiment are suited to identify the data generating trial-by-trial variability.

Here, we again use the model and experiment from Gluth et al. (2012) to exemplify this parameter variability recovery test. In a first step, we generate variability in the threshold parameter from trial to trial, simulate the model, and apply our technique to test whether this variability can be recovered. We then keep the threshold parameter constant but generate variability in another parameter, the noise parameter of the accumulation process, to examine to what extent this “noise” variability tends to be misattributed to “threshold” variability. Finally, we generate variability in both parameters to investigate potential interactions between them (for example, it could be that variability in the threshold can only be recovered as long as variability in the noise parameter is low). For all recovery analyses, we take the model parameters of all 26 subjects from Gluth et al., added trial-wise variability (i.e., Gaussian noise) to the respective parameter(s), generated synthetic data for the 120 trials of

each participant, fitted the model to these data, and applied the technique described in Section 3 to estimate trial-by-trial variability in the decision threshold. By using the parameters of all subjects we ensured having reasonable and reasonably different parameter sets (cf. van Ravenzwaaij & Oberauer, 2009). The critical test is the strength of the correlation between true and estimated trial-by-trial parameter values (van Maanen et al., 2011). To investigate how the amount of variability affects the recovery performance, we simulated 10 different levels of variability: the standard deviation of the Gaussian distribution, which was used to create the trial-by-trial variability in a parameter, ranged from 10 to 100% of the standard deviation of this parameter across subjects. We ran 100 simulations per subject and variability level.

The first step was to test whether our method can recover trial-by-trial variability in the decision threshold. The left panel of Fig. 5A shows the correlation between data-generating and recovered parameters for a randomly selected subject and simulation with the highest level of variability (i.e., 100% of the group SD). The correlation is substantial ($r = .65$). However, the range of recovered parameters is much smaller than the range of data-generating parameters, which is due to the narrow prior distributions (see also Fig. 3A). The right panel of Fig. 5A depicts the average correlations over all subjects and simulations as a function of the level of parameter variability. If the variability is high enough, the correlations are sufficiently high (for the two highest levels of variability, more than 90% of the correlations are above .5). Taken together, the range of recovered parameters is reduced compared to the range of true parameters, but more importantly recovered and true parameters are highly correlated with each other, so that linking the variability estimate to neuroimaging data is warranted.

When we do not induce variability in the threshold but in the noise parameter, this variability is picked up by our technique to some extent in form of a small negative correlation between the estimated value of the threshold parameter and the data-generating

noise parameter (Fig. 5B). This is due to the fact that simulations with high noise in the decision variable tend to generate decisions that are very fast, which is then attributed to a low decision threshold (note that this might not have been the case, if we had modeled variability in the noise parameter, too). Importantly, the correlation seems to be so small that significant correlations with neuroimaging data are not to be expected. Thus, it is unlikely that (potential) brain signals reflecting variability in the noise parameter could be misattributed to variability in the threshold parameter.

The small negative correlation might further suggest that recovery of threshold variability could be hampered by variability in the noise parameter. However, when variability is induced in both parameters, the correlation between true and estimated threshold parameters appears to be unaffected, with the average correlations (for high variability levels) being again above .6 (Fig. 5C). The same is true for the negative correlation between the noise parameter and the estimated thresholds. In summary, our method of estimating trial-by-trial variability in model parameters tends to underestimate the variability on an absolute scale but recovers the (relative) differences from trial to trial. Thus, it seems to be suited for correlational analyses with neuroimaging data. However, generalizing these results to other models and tasks is very limited, and researchers must run separate recovery analyses for their own work.

Fig. 5. Parameter recovery results. (A) Correlation between true (data-generating) and estimated threshold parameters for an example subject and simulation with the highest level of variability (left panel) and averaged across all subjects and simulations as a function of the level of variability (right panel). Notably, the correlations increase with increasing parameter variability. The intuition for this effect is that there has to be enough (true) variability in the parameter to generate systematic changes in behavior, which are captured by the recovery analysis. (B) The same as in (A) but for variability generated in the noise parameter. The correlation in the example on the left is -.23. (C) Correlations between true and estimated thresholds (left panel) and between true noise parameter values and estimated thresholds (right panel), if variability is induced in both parameters. The correlations are unaffected by the degree of variability in the other parameter.

5. Other requirements and recommendations

In addition to running parameter recovery tests, some more requirements with respect to both behavioral and brain data and the cognitive model should be fulfilled, if one wants to apply our approach successfully. Ideally, the researcher should take care of these arrangements before conducting the study, because at least some of them are irreversible. First, the experimental paradigm should allow researchers to track behavioral measures during fMRI or EEG recordings for every trial. This seems trivial, but especially in fMRI studies behavior is often recorded in just a small portion of trials (e.g., Hindi Attar, Finckh, & Büchel, 2012), only after scanning (e.g., Singer et al., 2004), or even not at all (e.g., Park et al., 2012). Similarly, the neuroimaging set-up must allow capturing signal differences on a trial-by-trial level, so positron emission tomography or fMRI block designs are not suitable for our approach. The same is true for analyses that remove trial-specific variability such as event-related potentials of EEG.

Besides identifying these requirements, we also give some recommendations to facilitate the application and increase its efficiency. Our first piece of advice is that a rich set of behavioral data is helpful when one wants to explain variability across trials: If model predictions are based not only on choices but also on response times (such as in sequential sampling models), there are more opportunities to observe trial-by-trial fluctuations in behavior. This does not necessarily imply more trial-by-trial variability in parameters, but given that such variability does exist, the likelihood of capturing it is increased by including response times in the analyses. In perceptual decision making, for instance, people often achieve very high accuracy rates of above .9, which means that for about 90% of the trials parameter variability could not be estimated, if a model takes only accuracy into account. Alternatively, one could arrive at more precise predictions of variability by combining multiple choices to specify the posterior distribution. For example, in Gluth et al. (2015) the

probability of successful encoding of a single item was based on three observations in the decision phase (Fig. 1B).

Second, a Bayesian modeling approach (Lee & Wagenmakers, 2014) is highly recommended because it directly addresses and solves several practical issues of our approach: It forces the modeler to specify the range of possible parameter values beforehand (though this is usually also done for maximum likelihood estimation), and it can directly provide posterior distributions of model parameters—similar to the prior distributions generated in our approach (Fig. 2A). In fact, our practical example above used an approximation to the prior distribution of the threshold parameter that suffers from two shortcomings. First, the likelihood of threshold values is calculated while keeping all other parameters constant (at the values taken from the initial estimation). A hierarchical Bayesian modeling approach can handle this more elegantly by simultaneously drawing entire parameter sets during the sampling process (though hierarchical modeling is also possible with maximum likelihood estimation; e.g., Farrell & Ludwig, 2008). This might even alleviate the problem of the reduced range of parameter variability, that we observed in our recovery analyses (Fig. 5A): This reduction results from the very narrow prior distribution (Fig. 3A), which in turn is determined by keeping all other parameters constant. Second, we followed a grid-search algorithm with a specific step size for trying out different parameter values, which runs the risk of being too imprecise for capturing variability on the one hand and of requiring much computation time on the other. This becomes irrelevant with directly estimating the parameter distribution via Bayesian modeling (though this difficulty can be reduced by applying more efficient optimization algorithms).

Intriguingly, Bayesian modeling could be used to estimate trial-by-trial variability in a model parameter in a single step: The hierarchical Bayesian model could assume one parameter value for every trial that is drawn from a higher subject-level distribution (just as subject-specific parameter values can be drawn from higher group-level distributions with this

technique; Kruschke, 2010; Lee & Wagenmakers, 2014). This would even allow testing whether the increased model complexity that comes with assuming trial-specific parameter estimates is warranted, for instance, by comparing the trial-unspecific and the trial-specific model by means of the deviance information criterion (Spiegelhalter, Best, Carlin, & van der Linde, 2002, 2014). Stated differently, one could test whether assuming trial-by-trial variability in a model parameter leads to a substantial improvement of model fit (that outweighs the loss of parsimony).³ Despite all these advantages of a hierarchical Bayesian modeling approach, it is important to note that one does not have to be proficient in these techniques to apply our method, but that it also works with maximum likelihood estimation as shown in our example above.

Our final recommendation pertains to the quality of the cognitive model itself. The model should provide a very good description of the data, so that remaining variability can be clearly attributed to specific components of it and not just to its general insufficiency. Our approach is only optimal within the limits of the underlying model. If the model itself makes little sense, then attributing behavioral fluctuations to one of its components cannot be very meaningful, either. Hence, it is advisable to subject multiple models to a rigorous model comparison procedure before extracting trial-by-trial predictions on the basis of the winning model.

6. Comparison with other trial-by-trial techniques

How is our approach related to the studies and techniques discussed in Section 2 and to alternative approaches? As stated above, the approach is basically a generalization of the method used in Gluth et al. (2015), where memory-based choices were used to derive the posterior probability that a choice option was remembered. It is also similar to the methods

³ We would therefore recommend testing for trial-specific parameter values only if there is a strong hypothesis that this variability actually exists and causes fluctuations in behavior. Note, however, that improving a model's prediction of behavior is not the primary goal of our approach.

used in the two other studies (Gluth et al., 2012; van Maanen et al., 2011)—but with some notable differences. Compared to Gluth et al. (2012), the new framework has a stronger theoretical foundation, as it provides trial-by-trial estimates for distinct model components. In Gluth et al. (2012), the discrepancy between model predictions and observations (in terms of too early or too late predictions) was used to approximate fluctuations in the initial decision threshold. However, these very general model predictions are also affected by other parameters (i.e., the decrease of thresholds with time, the standard deviation of the drift rate). Our new approach allows isolating single model components (by holding the rest of the model constant), which also enables a comparison of the respective contributions of different parameters to variability in the data (one simply has to repeat the analysis outlined in Section 3 for other parameters). In van Maanen et al. (2011), maximum likelihood estimates for single-trial parameters (LBA's start points and drift rates) were derived analytically, which is desirable but often infeasible to derive. More importantly, van Maanen et al. did not implement priors for parameter values to constrain these maximum likelihood estimates. This step is not mandatory to get trial-by-trial predictions, but it helps avoid that some single-trial estimates are very extreme or equally likely (e.g., see the blue lines in Fig. 3B).

Two conceptually related approaches were proposed by Ratcliff, Philiastides, and Sajda (2009) and by Ide, Shenoy, Yu, and Li (2013), respectively. Using data from a perceptual decision making task, Ratcliff et al. linked variability in the drift rate of the drift diffusion model (DDM) to an EEG component for which single-trial estimates could be obtained. Trials were split on the basis of the amplitude of the EEG component, and the DDM was estimated for each half of the trials separately. The authors were able to show that the drift rate was higher for high-amplitude compared to low-amplitude trials even within single conditions of perceptual noise. The main difference to our approach is that the neural data was used to inform the behavioral modeling, whereas our approach relies on directly extracting variability via behavioral modeling and then informing the neuroimaging analysis. Ide et al.

used a dynamic belief updating model to estimate trial-by-trial changes in the expectancy of stop signals in a stop-signal detection task and to relate this to the neural encoding of expectancy and surprise signals in the anterior cingulate cortex (see also Harlé et al., 2015, who relates this model to substance abuse in adolescents). In contrast to our approach, the actual behavior of the subjects did not inform the estimates of belief updating, but the model simply assumed that all subjects updated their expectancy in the same manner.

Another different approach to relating trial-by-trial variability in cognitive and neural processes was proposed by Michael Frank and colleagues (Cavanagh et al., 2011; Cavanagh, Wiecki, Kochar, & Frank, 2014; Frank et al., 2015; Wiecki, Sofer, & Frank, 2013). Their technique is based on implementing trial-by-trial changes in neuroimaging data, such as variable peaks in the hemodynamic response function of fMRI data or trial-specific power estimates taken from EEG time-frequency analyses, in hierarchical Bayesian modeling of Ratcliff's drift diffusion model (Ratcliff, 1978; Wiecki et al., 2013). For example, Cavanagh et al. (2011) showed that the decision threshold can be affected by fluctuations in the EEG power of theta-band frequencies (4–8 Hz) at frontal electrodes, presumably reflecting an adaptive conflict-control mechanism mediated by interactions of the medial prefrontal cortex with the subthalamic nucleus. Practically, the supposedly relevant part of the neuroimaging data is first extracted on a trial-by-trial basis and then implemented in the hierarchical drift diffusion model using a regression approach. Posterior probability estimates of the regression coefficients (that relate the brain data with the model) and model comparisons determine whether adding the neural term to the model is warranted. The central difference from our approach is that the brain signal has to be extracted first to implement it in the model that predicts behavior (instead of first extracting trial-by-trial predictions from the behavioral model to predict neuroimaging data, as in our case). Hence, this method requires much more specific hypotheses about which brain signals could be related to trial-by-trial changes in cognitive processes: The researcher must have a good idea about which kind of signal (e.g.,

theta-band power) in which brain region (e.g., medial prefrontal cortex) could be linked to which component (e.g., decision threshold) of the cognitive model. Exploring the entire space of possible brain–behavior relationships with this technique would be far too time consuming. Furthermore, the software tool provided by the research group (Wiecki et al., 2013) is restricted to the drift diffusion model (and the LBA).

A similar but more advanced method of combining neuroimaging and behavioral data in a Bayesian modeling framework was introduced by Turner et al. (2013). Their idea was to model both behavioral and brain data sets simultaneously with hierarchical Bayesian modeling to restrict parameter estimation on both the “neural model” and the “behavioral model.” A central difference from Frank’s approach is that one does not have to specify which brain region might be related to which model parameter in advance. Instead, an entire set of areas (potentially covering the entire brain) can be used simultaneously, and their relationship to the model is computed for every parameter. In addition, the joint modeling of brain and behavioral data makes it statistically reciprocal; that is, the neural data can constrain parameter estimations of the behavioral model and vice versa. This method has recently been applied to improve predictions of Ratcliff’s drift diffusion model for a perceptual decision-making task by means of adding information from an fMRI data set (Turner, van Maanen, & Forstmann, 2015). The approach is as general as ours and has the advantage of statistical reciprocity, but its complexity might discourage many cognitive neuroscientists from using it for their models and data sets. In contrast, our method can be realized quite easily by simply restructuring formulas that are already part of the cognitive model. In other words, everyone who has managed to set up and fit a cognitive model should be able to go this one step further and apply our technique.

7. Potential findings and what they mean

In this section, we present three scenarios of how trial-by-trial variability in behavior might be explained by variability in model parameters and related to neuroimaging data. We also discuss implications for theoretical developments in cognitive neuroscience. We try to keep the discussion at a high level of generalizability but we complement it with real examples from the literature. All scenarios are based on the (methodologically simple but straightforward) idea that the trial-by-trial parameter variability, derived by applying our technique, is used as a predictor in a general linear model (GLM) of the neuroimaging data analysis. The GLM is particularly popular in fMRI; here, trial-by-trial parameter variability could be implemented as a parametric modulator: It predicts how the hemodynamic response of the fMRI blood-oxygen-level-dependent signal is modulated as a function of the parameter variability at specific time points in the experiment (for more information, see Büchel et al., 1998; Büchel, Wise, Mummery, Poline, & Friston, 1996; O’Doherty et al., 2007). These specific time points could, for instance, refer to when decisions are made or when subjects are instructed to emphasize speed or accuracy in the upcoming trial. There are, of course, many other possible ways of linking parameter variability to neuroimaging data (see also Turner, Forstmann, Love, Palmeri, & van Maanen, submitted), but an in-depth discussion is beyond the scope of this article.

Scenario 1: Converging evidence. Parameter π and activity in brain region x are correlated across subjects; applying the trial-by-trial analysis proposed in this paper reveals that parameter π is similarly related to brain region x on the single-trial level.

This first scenario is essentially what has been found for the association of response caution with activation in the pre-SMA and CN (Forstmann et al., 2008; Gluth et al., 2012; van Maanen et al., 2011) (see Section 2), and it can be regarded as most desired and least surprising. Variability across and within subjects in (activation of) x is reflected in variability

across and within subjects in (behavior-based estimation of) π . Importantly, the trial-by-trial analysis increases the evidence for the brain–behavior link quantitatively but also qualitatively: As discussed in Section 2, a relationship between parameter and brain activation across subjects does not necessarily imply the same relationship within subjects. The trial-by-trial analysis is therefore more than a mere replication of the interindividual correlation.

Scenario 2: New insights. Parameter π and activity in brain region x are correlated across subjects; applying the trial-by-trial analysis proposed in this paper reveals, however, that parameter π is not related to brain region x but to another brain region y on the single-trial level.

This scenario is definitely less likely—but still possible as shown above.⁴ One explanation could be that while region x sets the overall (i.e., subject- or condition-wise) level of parameter p , region y is responsible for the trial-by-trial fine-tuning. In this case, there should be a functional connection between x and y , which could be revealed using coupling analyses such as psychophysiological interaction (Friston et al., 1997) or dynamic causal modeling (Friston, Harrison, & Penny, 2003). A more in-depth analysis of the behavioral data may also help clarify this complex pattern of results. For instance, making an error often triggers adaptations on the next trial (Rabbitt & Rodgers, 1977), including changes in response caution (Dutilh et al., 2012). Region y could thus be a brain area associated with error processing and cognitive control, such as the anterior cingulate cortex (Holroyd & Coles, 2002; Shenhav, Botvinick, & Cohen, 2013), which exerts an additional influence on

⁴ Because this scenario is less likely, more empirical evidence would be needed to establish that π and y are indeed linked to each other. For instance, a small-volume correction approach to the multiple-comparison problem for fMRI (Poldrack, 2006) can be used to increase the sensitivity to detect a relationship of π with region x (because x has already been shown to be linked to π across subjects) but not with region y (unless there is another, strong theoretical prediction for why activity in region y should be associated with variability in π).

response caution under high conflict (Botvinick, Nystrom, Fissell, Carter, & Cohen, 1999; Cavanagh et al., 2011).

Scenario 3: Inconsistent evidence. Applying the trial-by-trial analysis proposed in this paper does not reveal any relationship of parameter π with the neuroimaging data.

At first glance, such a finding could be taken simply as a null finding without any theoretical implications at all. However, a lack of neural support for variability in parameter π might be because the model assumption of variability in π is not useful to describe the brain data (and thus might also not be useful to describe the behavioral data). As outlined in Section 4, behavior is always variable because the neurocognitive system is always affected by “random noise.” Applying our technique to a randomly fluctuating behavioral measure will attribute these fluctuations to the parameter of interest. Yet, it is possible that this purported parameter variability does not exist. To give an example, neurophysiological studies in monkeys suggest that the evidence accumulation process, reflected in the gradual increase of firing rates in cortical areas such as the lateral intraparietal cortex and the frontal eye fields, stops at a stereotyped level that is independent of response times and task difficulty (Hanes & Schall, 1996; Roitman & Shadlen, 2002). This has led to the proposal that adaptations in response caution are mediated by a change in the start point of accumulation rather than by an adjustment of the decision threshold itself (Bogacz, Wagenmakers, Forstmann, & Nieuwenhuis, 2010; but see Simen, 2012). Consequently, looking for a neural correlate of response caution variability at the time when the choice is made (i.e., when the decision variable crosses the threshold) might not reveal any significant results. Instead, the more promising time window to look at is when the decision process starts, or even earlier, when people are instructed or incentivized to behave more or less cautiously in the upcoming trial (Boehm et al., 2014; Forstmann et al., 2008; Gluth et al., 2012; Gluth, Rieskamp, & Büchel,

2013; Ivanoff, Branning, & Marois, 2008; van Veen, Krug, & Carter, 2008). Hence, not finding a relationship between trial-by-trial variability in cognitive model components and neuroimaging data might support or even promote theoretical developments in cognitive neuroscience.

Another reason for not finding an effect of trial-by-trial variability in the neuroimaging data could be the low signal-to-noise ratio of the latter. Both fMRI and EEG analyses require many repetitions of trials to separate the faint hemodynamic or electrical signal from all the differences sources of random fluctuations, including movements, physiological artifacts such as heart beat and respiration, or artifacts of the measurement tools (Huettel, Song, & McCarthy, 2014; Luck, 2014). This might raise skepticism about whether these techniques are in any form suited to capture the trial-by-trial changes which we are addressing here.

However, we would argue that there is no reason to believe that the low signal-to-noise ratio of neuroimaging tools should be more problematic for investigating parametric effects as compared to simpler comparisons of experimental conditions. In fact, it seems more likely that simple contrasts such as “task vs. rest” or “emotional vs. neutral” in an fMRI study are systematically confounded by heart rate or respiration. Having said this, we would recommend that researchers seeking to apply our methods should consider (and maybe even test) whether the targeted trial-wise variations in model parameters might be contaminated by (regular variations in) physiological artifacts.

8. Conclusion

We have introduced a general, comparatively simple technique to capture trial-by-trial variability in cognitive model parameters (and similar components) to explain trial-by-trial variability in behavior and link it to the analysis of neuroimaging data. Our approach relies on deriving probability estimates for different parameter values over all trials as well as for each individual trial and then combining these estimates by means of Bayes’s rule to obtain the

most likely posterior estimate in every trial. Crucially, the method can be applied to (almost) any type of mathematical model as well as to a broad range of neuroimaging data and analyses (but especially to event-related fMRI data and time-frequency EEG analyses). In addition, the approach can also be used in combination with physiological measures such as skin conductance, electromyography, heart rate, eye tracking, or pupil dilation. We hope to encourage other researchers in cognitive and systems neuroscience to employ our technique to tighten the links between their neuroimaging data and their favorite cognitive models to allow for theory advancement.

Acknowledgments

We thank Anita Todd for editing the manuscript. This research was supported by a grant from the Swiss National Science Foundation (SNSF grant 100014_153616) to both authors.

References

- Bartra, O., McGuire, J. T., & Kable, J. W. (2013). The valuation system: a coordinate-based meta-analysis of BOLD fMRI experiments examining neural correlates of subjective value. *NeuroImage*, 76, 412–427. <http://doi.org/10.1016/j.neuroimage.2013.02.063>
- Behrens, T. E. J., Woolrich, M. W., Walton, M. E., & Rushworth, M. F. S. (2007). Learning the value of information in an uncertain world. *Nature Neuroscience*, 10(9), 1214–1221. <http://doi.org/10.1038/nn1954>
- Boehm, U., van Maanen, L., Forstmann, B., & van Rijn, H. (2014). Trial-by-trial fluctuations in CNV amplitude reflect anticipatory adjustment of response caution. *NeuroImage*, 96, 95–105. <http://doi.org/10.1016/j.neuroimage.2014.03.063>
- Bogacz, R., Brown, E., Moehlis, J., Holmes, P., & Cohen, J. D. (2006). The physics of optimal decision making: a formal analysis of models of performance in two-alternative forced-choice tasks. *Psychological Review*, 113(4), 700–765. <http://doi.org/10.1037/0033-295X.113.4.700>
- Bogacz, R., Wagenmakers, E.-J., Forstmann, B. U., & Nieuwenhuis, S. (2010). The neural basis of the speed-accuracy tradeoff. *Trends in Neurosciences*, 33(1), 10–16. <http://doi.org/10.1016/j.tins.2009.09.002>
- Boll, S., Gamer, M., Gluth, S., Finsterbusch, J., & Büchel, C. (2013). Separate amygdala subregions signal surprise and predictiveness during associative fear learning in humans. *The European journal of neuroscience*, 37(5), 758–767. <http://doi.org/10.1111/ejn.12094>
- Botvinick, M., Nystrom, L. E., Fissell, K., Carter, C. S., & Cohen, J. D. (1999). Conflict monitoring versus selection-for-action in anterior cingulate cortex. *Nature*, 402(6758), 179–181. <http://doi.org/10.1038/46035>
- Brandstätter, E., Gigerenzer, G., & Hertwig, R. (2006). The priority heuristic: Making choices without trade-offs. *Psychological Review*, 113, 409–432.
- Brown, S. D., & Heathcote, A. (2008). The simplest complete model of choice response time: linear ballistic accumulation. *Cognitive psychology*, 57(3), 153–178. <http://doi.org/10.1016/j.cogpsych.2007.12.002>
- Büchel, C., Holmes, A. P., Rees, G., & Friston, K. J. (1998). Characterizing stimulus-response functions using nonlinear regressors in parametric fMRI experiments. *NeuroImage*, 8(2), 140–148. <http://doi.org/10.1006/nimg.1998.0351>
- Büchel, C., Wise, R. J., Mummery, C. J., Poline, J. B., & Friston, K. J. (1996). Nonlinear regression in parametric activation studies. *NeuroImage*, 4(1), 60–66. <http://doi.org/10.1006/nimg.1996.0029>
- Bussemeyer, J. R., & Townsend, J. T. (1993). Decision field theory: a dynamic-cognitive approach to decision making in an uncertain environment. *Psychological Review*, 100(3), 432–459.
- Cavanagh, J. F., Wiecki, T. V., Cohen, M. X., Figueroa, C. M., Samanta, J., Sherman, S. J., & Frank, M. J. (2011). Subthalamic nucleus stimulation reverses mediofrontal influence over decision threshold. *Nature Neuroscience*, 14(11), 1462–1467. <http://doi.org/10.1038/nn.2925>
- Cavanagh, J. F., Wiecki, T. V., Kochar, A., & Frank, M. J. (2014). Eye tracking and pupillometry are indicators of dissociable latent decision processes. *Journal of Experimental Psychology. General*,

143(4), 1476–1488. <http://doi.org/10.1037/a0035813>

Clithero, J. A., & Rangel, A. (2014). Informatic parcellation of the network involved in the computation of subjective value. *Social Cognitive and Affective Neuroscience*, 9(9), 1289–1302. <http://doi.org/10.1093/scan/nst106>

Craigmile, P. F., Peruggia, M., & Van Zandt, T. (2010). Hierarchical Bayes models for response time data. *Psychometrika*, 75(4), 613–632. <http://doi.org/10.1007/s11336-010-9172-6>

Daw, N. D., O'Doherty, J. P., Dayan, P., Seymour, B., & Dolan, R. J. (2006). Cortical substrates for exploratory decisions in humans. *Nature*, 441(7095), 876–879. <http://doi.org/10.1038/nature04766>

Dutilh, G., Vandekerckhove, J., Forstmann, B. U., Keuleers, E., Brysbaert, M., & Wagenmakers, E.-J. (2012). Testing theories of post-error slowing. *Attention, Perception, & Psychophysics*, 74(2), 454–465. <http://doi.org/10.3758/s13414-011-0243-2>

Eichele, T., Debener, S., Calhoun, V. D., Specht, K., Engel, A. K., Hugdahl, K., ... Ullsperger, M. (2008). Prediction of human errors by maladaptive changes in event-related brain networks. *Proceedings of the National Academy of Sciences of the United States of America*, 105(16), 6173–6178. <http://doi.org/10.1073/pnas.0708965105>

Farrell, S., & Ludwig, C. J. H. (2008). Bayesian and maximum likelihood estimation of hierarchical response time models. *Psychonomic Bulletin & Review*, 15(6), 1209–1217. <http://doi.org/10.3758/PBR.15.6.1209>

Forstmann, B. U., Dutilh, G., Brown, S., Neumann, J., von Cramon, D. Y., Ridderinkhof, K. R., & Wagenmakers, E.-J. (2008). Striatum and pre-SMA facilitate decision-making under time pressure. *Proceedings of the National Academy of Sciences, USA*, 105(45), 17538–17542. <http://doi.org/10.1073/pnas.0805903105>

Frank, M. J., Gagne, C., Nyhus, E., Masters, S., Wiecki, T. V., Cavanagh, J. F., & Badre, D. (2015). fMRI and EEG predictors of dynamic decision parameters during human reinforcement learning. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, 35(2), 485–494. <http://doi.org/10.1523/JNEUROSCI.2036-14.2015>

Friston, K. J., Buechel, C., Fink, G. R., Morris, J., Rolls, E., & Dolan, R. J. (1997). Psychophysiological and modulatory interactions in neuroimaging. *NeuroImage*, 6(3), 218–229. <http://doi.org/10.1006/nimg.1997.0291>

Friston, K. J., Harrison, L., & Penny, W. D. (2003). Dynamic causal modelling. *NeuroImage*, 19(4), 1273–1302.

Garrison, J., Erdeniz, B., & Done, J. (2013). Prediction error in reinforcement learning: a meta-analysis of neuroimaging studies. *Neuroscience and Biobehavioral Reviews*, 37(7), 1297–1310. <http://doi.org/10.1016/j.neubiorev.2013.03.023>

Gläscher, J., Hampton, A. N., & O'Doherty, J. P. (2009). Determining a role for ventromedial prefrontal cortex in encoding action-based value signals during reward-related decision making. *Cerebral Cortex*, 19(2), 483–495. <http://doi.org/10.1093/cercor/bhn098>

Glimcher, P. W. (2005). Indeterminacy in brain and behavior. *Annual Review of Psychology*, 56, 25–56. <http://doi.org/10.1146/annurev.psych.55.090902.141429>

- Gluth, S., Rieskamp, J., & Büchel, C. (2012). Deciding when to decide: time-variant sequential sampling models explain the emergence of value-based decisions in the human brain. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, 32(31), 10686–10698. <http://doi.org/10.1523/JNEUROSCI.0727-12.2012>
- Gluth, S., Rieskamp, J., & Büchel, C. (2013). Classic EEG motor potentials track the emergence of value-based decisions. *NeuroImage*, 79, 394–403. <http://doi.org/10.1016/j.neuroimage.2013.05.005>
- Gluth, S., Rieskamp, J., & Büchel, C. (2014). Neural evidence for adaptive strategy selection in value-based decision-making. *Cerebral Cortex (New York, N.Y.: 1991)*, 24(8), 2009–2021. <http://doi.org/10.1093/cercor/bht049>
- Gluth, S., Sommer, T., Rieskamp, J., & Büchel, C. (2015). Effective Connectivity between Hippocampus and Ventromedial Prefrontal Cortex Controls Preferential Choices from Memory. *Neuron*, 86(4), 1078–1090. <http://doi.org/10.1016/j.neuron.2015.04.023>
- Hanes, D. P., & Schall, J. D. (1996). Neural control of voluntary movement initiation. *Science (New York, N.Y.)*, 274(5286), 427–430.
- Harlé, K. M., Stewart, J. L., Zhang, S., Tapert, S. F., Yu, A. J., & Paulus, M. P. (2015). Bayesian neural adjustment of inhibitory control predicts emergence of problem stimulant use. *Brain*, 138(11), 3413–3426. <http://doi.org/10.1093/brain/awv246>
- Hey, J. D. (1995). Experimental investigations of errors in decision making under risk. *European Economic Review*, 39(3-4), 633–640. [http://doi.org/10.1016/0014-2921\(09\)40007-4](http://doi.org/10.1016/0014-2921(09)40007-4)
- Hindi Attar, C., Finckh, B., & Büchel, C. (2012). The influence of serotonin on fear learning. *PLoS ONE*, 7(8), e42397. <http://doi.org/10.1371/journal.pone.0042397>
- Holroyd, C. B., & Coles, M. G. H. (2002). The neural basis of human error processing: reinforcement learning, dopamine, and the error-related negativity. *Psychological Review*, 109(4), 679–709.
- Ho, T., Brown, S., van Maanen, L., Forstmann, B. U., Wagenmakers, E.-J., & Serences, J. T. (2012). The optimality of sensory processing during the speed-accuracy tradeoff. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, 32(23), 7992–8003. <http://doi.org/10.1523/JNEUROSCI.0340-12.2012>
- Huettel, S. A., Song, A. W., & McCarthy, G. (2014). *Functional magnetic resonance imaging* (Third edition). Sunderland, Massachusetts, U.S.A: Sinauer Associates, Inc., Publishers.
- Ide, J. S., Shenoy, P., Yu, A. J., & Li, C. R. (2013). Bayesian prediction and evaluation in the anterior cingulate cortex. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, 33(5), 2039–2047. <http://doi.org/10.1523/JNEUROSCI.2201-12.2013>
- Ivanoff, J., Branning, P., & Marois, R. (2008). fMRI evidence for a dual process account of the speed-accuracy tradeoff in decision-making. *PloS One*, 3(7), e2635. <http://doi.org/10.1371/journal.pone.0002635>
- Kahneman, D., & Tversky, A. (1979). Prospect theory: an analysis of decision under risk. *Econometrica*, 47(2), 263–292. <http://doi.org/10.2307/1914185>
- Kim, H. (2011). Neural activity that predicts subsequent memory and forgetting: a meta-analysis of 74 fMRI studies. *NeuroImage*, 54(3), 2446–2461. <http://doi.org/10.1016/j.neuroimage.2010.09.045>

- Krugel, L. K., Biele, G., Mohr, P. N. C., Li, S.-C., & Heekeren, H. R. (2009). Genetic variation in dopaminergic neuromodulation influences the ability to rapidly and flexibly adapt decisions. *Proceedings of the National Academy of Sciences of the United States of America*, 106(42), 17951–17956. <http://doi.org/10.1073/pnas.0905191106>
- Kruschke, J. K. (2010). What to believe: Bayesian methods for data analysis. *Trends in Cognitive Sciences*, 14(7), 293–300. <http://doi.org/10.1016/j.tics.2010.05.001>
- Lee, M. D., & Wagenmakers, E.-J. (2014). *Bayesian Cognitive Modeling A Practical Course*. Cambridge University Press.
- Lewandowsky, S., & Farrell, S. (2011). *Computational Modeling in Cognition: Principles and Practice*. Los Angeles, Calif.: Sage Publications.
- Loomes, G., & Sugden, R. (1995). Incorporating a stochastic element into decision theories. *European Economic Review*, 39(3-4), 641–648. [http://doi.org/10.1016/0014-2921\(94\)00071-7](http://doi.org/10.1016/0014-2921(94)00071-7)
- Luce, R. D. (1986). *Response Times*. New York; Oxford: Oxford University Press ; Clarendon Press.
- Luck, S. J. (2014). *An introduction to the event-related potential technique* (Second edition). Cambridge, Massachusetts: The MIT Press.
- McFadden, D. (2001). Economic choices. *The American Economic Review*, 91(3), 351–378.
- McGuire, J. T., Nassar, M. R., Gold, J. I., & Kable, J. W. (2014). Functionally dissociable influences on learning rate in a dynamic environment. *Neuron*, 84(4), 870–881. <http://doi.org/10.1016/j.neuron.2014.10.013>
- O’Doherty, J. P., Dayan, P., Schultz, J., Deichmann, R., Friston, K., & Dolan, R. J. (2004). Dissociable roles of ventral and dorsal striatum in instrumental conditioning. *Science*, 304(5669), 452–454. <http://doi.org/10.1126/science.1094285>
- O’Doherty, J. P., Hampton, A., & Kim, H. (2007). Model-based fMRI and its application to reward learning and decision making. *Annals of the New York Academy of Sciences*, 1104, 35–53. <http://doi.org/10.1196/annals.1390.022>
- Paller, K. A., & Wagner, A. D. (2002). Observing the transformation of experience into memory. *Trends in Cognitive Sciences*, 6(2), 93–102.
- Park, S. Q., Kahnt, T., Talmi, D., Rieskamp, J., Dolan, R. J., & Heekeren, H. R. (2012). Adaptive coding of reward prediction errors is gated by striatal coupling. *Proceedings of the National Academy of Sciences of the United States of America*, 109(11), 4285–4289. <http://doi.org/10.1073/pnas.1119969109>
- Payzan-LeNestour, E., Dunne, S., Bossaerts, P., & O’Doherty, J. P. (2013). The neural representation of unexpected uncertainty during value-based decision making. *Neuron*, 79(1), 191–201. <http://doi.org/10.1016/j.neuron.2013.04.037>
- Pessiglione, M., Petrovic, P., Daunizeau, J., Palminteri, S., Dolan, R. J., & Frith, C. D. (2008). Subliminal instrumental conditioning demonstrated in the human brain. *Neuron*, 59(4), 561–567. <http://doi.org/10.1016/j.neuron.2008.07.005>
- Poldrack, R. A. (2006). Region of interest analysis for fMRI. *Social Cognitive and Affective*

Neuroscience, 2(1), 67–70. <http://doi.org/10.1093/scan/nsm006>

Rabbitt, P., & Rodgers, B. (1977). What does a man do after he makes an error? an analysis of response programming. *Quarterly Journal of Experimental Psychology*, 29(4), 727–743. <http://doi.org/10.1080/14640747708400645>

Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review*, 85(2), 59–108. <http://doi.org/10.1037/0033-295X.85.2.59>

Ratcliff, R., & McKoon, G. (2008). The diffusion decision model: theory and data for two-choice decision tasks. *Neural Computation*, 20(4), 873–922. <http://doi.org/10.1162/neco.2008.12-06-420>

Ratcliff, R., Philastides, M. G., & Sajda, P. (2009). Quality of evidence for perceptual decision making is indexed by trial-to-trial variability of the EEG. *Proceedings of the National Academy of Sciences of the United States of America*, 106(16), 6539–6544. <http://doi.org/10.1073/pnas.0812589106>

Ratcliff, R., & Rouder, J. N. (1998). Modeling response times for two-choice decisions. *Psychological Science*, 9(5), 347–356. <http://doi.org/10.1111/1467-9280.00067>

Rieskamp, J. (2008). The probabilistic nature of preferential choice. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 34(6), 1446–1465. <http://doi.org/10.1037/a0013646>

Robinson, W. S. (1950). Ecological correlations and the behavior of individuals. *American Sociological Review*, 15(3), 351–357.

Roitman, J. D., & Shadlen, M. N. (2002). Response of neurons in the lateral intraparietal area during a combined visual discrimination reaction time task. *Journal of Neuroscience*, 22(21), 9475–9489.

Shenhav, A., Botvinick, M. M., & Cohen, J. D. (2013). The expected value of control: an integrative theory of anterior cingulate cortex function. *Neuron*, 79(2), 217–240. <http://doi.org/10.1016/j.neuron.2013.07.007>

Simen, P. (2012). Evidence Accumulator or Decision Threshold - Which Cortical Mechanism are We Observing? *Frontiers in Psychology*, 3, 183. <http://doi.org/10.3389/fpsyg.2012.00183>

Singer, T., Seymour, B., O’Doherty, J., Kaube, H., Dolan, R. J., & Frith, C. D. (2004). Empathy for pain involves the affective but not sensory components of pain. *Science (New York, N.Y.)*, 303(5661), 1157–1162. <http://doi.org/10.1126/science.1093535>

Smith, P. L., & Ratcliff, R. (2004). Psychology and neurobiology of simple decisions. *Trends in Neurosciences*, 27(3), 161–168. <http://doi.org/10.1016/j.tins.2004.01.006>

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., & van der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 64(4), 583–639. <http://doi.org/10.1111/1467-9868.00353>

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., & van der Linde, A. (2014). The deviance information criterion: 12 years on. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 76(3), 485–493. <http://doi.org/10.1111/rssb.12062>

Turner, B. M., Forstmann, B. U., Love, B., Palmeri, T. J., & van Maanen, L. (submitted). Structures of analysis in model-based cognitive neuroscience.

Turner, B. M., Forstmann, B. U., Wagenmakers, E.-J., Brown, S. D., Sederberg, P. B., & Steyvers, M. (2013). A Bayesian framework for simultaneously modeling neural and behavioral data. *NeuroImage*, 72, 193–206. <http://doi.org/10.1016/j.neuroimage.2013.01.048>

Turner, B. M., van Maanen, L., & Forstmann, B. U. (2015). Informing cognitive abstractions through neuroimaging: the neural drift diffusion model. *Psychological Review*, 122(2), 312–336. <http://doi.org/10.1037/a0038894>

van Maanen, L., Brown, S. D., Eichele, T., Wagenmakers, E.-J., Ho, T., Serences, J., & Forstmann, B. U. (2011). Neural correlates of trial-to-trial fluctuations in response caution. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, 31(48), 17488–17495. <http://doi.org/10.1523/JNEUROSCI.2924-11.2011>

van Ravenzwaaij, D., & Oberauer, K. (2009). How to use the diffusion model: Parameter recovery of three methods: EZ, fast-dm, and DMAT. *Journal of Mathematical Psychology*, 53(6), 463–473. <http://doi.org/10.1016/j.jmp.2009.09.004>

van Veen, V., Krug, M. K., & Carter, C. S. (2008). The neural and computational basis of controlled speed-accuracy tradeoff during task performance. *Journal of Cognitive Neuroscience*, 20(11), 1952–1965. <http://doi.org/10.1162/jocn.2008.20146>

Von Neumann, J., & Morgenstern, O. (1947). *Theory of Games and Economic Behavior* (2. Aufl.). Princeton, N.J.; Woodstock: Princeton University Press.

Wiecki, T. V., Sofer, I., & Frank, M. J. (2013). HDDM: Hierarchical Bayesian estimation of the Drift-Diffusion Model in Python. *Frontiers in Neuroinformatics*, 7, 14. <http://doi.org/10.3389/fninf.2013.00014>

Zandbelt, B., Purcell, B. A., Palmeri, T. J., Logan, G. D., & Schall, J. D. (2014). Response times from ensembles of accumulators. *Proceedings of the National Academy of Sciences of the United States of America*, 111(7), 2848–2853. <http://doi.org/10.1073/pnas.1310577111>