Different Ways of Linking Behavioral and Neural Data Via Computational Cognitive Models

Supplement

Introduction

This supplement to the article "*Different ways of linking behavioral and neural data via computational cognitive models*" provides an extended version of the literature review in that paper.

Examples of Qualitative Structural Linking

Qualitative Structural Linking in Models of Evidence Accumulation

Evidence accumulation models have been used to explain simple decision-making processes for more than fifty years (1-3). More recently, attempts have been made to link the models with neural data. The earliest attempts, such as seminal work by Usher & McClelland (4), defined qualitative structural links. These links were structural in the sense that the constraints were applied to the structure of the model, not to the model's predictions, and the links were qualitative in the sense that the constraints revolved around the inclusion/exclusion of model elements, not to the quantitative parametric values taken. For example, the leaky competing accumulator model (LCA) of Usher & McClelland (4) specifically included structural elements such as mutual inhibition between competing accumulators. This inclusion was motivated by neural data which demonstrate the prevalence of inhibitory connections between nearby neurons within the same cortical stratum. Similarly, the LCA included passive decay of accumulated evidence, to respect the neural observation that membrane potential decays back to baseline in the absence of input. Evidence in favor of these links was inferred by the observation that the resulting cognitive model provided a good fit to behavioral data.

Smith (5) showed that a plausible model of how neurons encode sensory information at very short time scales (a Poisson shot noise process), converges, with some assumptions, to a Ornstein-Uhlenbeck velocity process. The integrated version of this process is, in turn, indistinguishable from a Wiener process and the Wiener process is the diffusion process that is assumed to underlie the standard drift diffusion model (2; 6). Smith thus showed that the DDM is a suitable abstract model of the neural dynamics of decision-making at larger time scales.

Ratcliff and Frank (7) showed, in a similar vein, that simulated “behavioral data” produced by a prominent neurocomputational model of perceptual decision-making in corticostriatal circuits (8) are well-explained by the DDM. Importantly, parametric changes in projection strengths in the neurocomputational model (those between the subthalamic nucleus and globus pallidus, internal segment) were reflected by corresponding parametric increases in decision threshold and non-decision time during high-conflict decisions when the parameters of the DDM were estimated on the behavioral output of the neurocomputational model.

Qualitative Structural Linking in Models of Reinforcement Learning

The classic parallel distributed processing models provided cognitive descriptions of learning including structural constraints from neural data (PDP; 9; 10). The models assumed massive parallelism and distributed information representation, reflecting key findings in the emerging neural literature on cortical structure. The models also used learning rules such as back-propogation, which were inspired by neural findings such as Hebbian plasticity.

An important contribution of the PDP approach was its demonstration that very simple structures -- such as those found in cortex -- were sufficient to support quite complex computations, when endowed with the appropriate representation and learning assumptions. This contribution emerged directly out of the effort to draw structural links between neural and cognitive data.

Qualitative Structural Linking in Models of Symbolic Reasoning

The ACT-R production framework (11) is a domain-general model of human cognition. This model began as a cognitive model purely aimed at behavioral data, but has since been extended in great detail to jointly consider behavioral and neural data (12; 13). The earliest linking of the ACT-R model to neural data was qualitative structural linking, which identified links between different cognitive modules in ACT-R and different brain regions. These links respected findings about the localization of brain function that were emerging at the time from the then-new method of fMRI. For example, the “visual module” of ACT-R was linked with lower occipital brain regions, and the “motor module” with motor cortices in the parietal and temporal lobes. These links defined the structure of the model and allowed the investigation of hypotheses about deficits due to brain lesions, for example.

Examples of Qualitative Predictive Linking

One could argue that almost all work in cognitive neuroscience, and particularly cognitive neuroimaging, applies qualitative predictive links, because this research usually assumes at least some cognitive concepts like “working memory”, “attention”, or “semantic memory” (14), which, taken together, could be interpreted as a cognitive model that is qualitatively linked to neural data. In most cases, these *conceptual* models are then related to neural data by means of experimental paradigms modulating the cognitive concept of interest. However, we limit our scope to the linking of explicit, computational models of cognition. Such models can, unlike conceptual models, quantitatively predict behavior (15) which make them also more suited for tighter forms of linking to neural data that we will describe later.

Qualitative Predictive Linking in Models of Evidence Accumulation

Hanes and Schall (16) recorded single-cell activity in the frontal eye fields (FEF) in behaving macaques. The activity of “movement neurons” predicted the execution of saccades. Hanes and Schall (16) showed that the ramping activity of these neurons preceding a saccade always ended with the same firing rate, but the rate of increase of firing rate was variable. The authors related these qualitative patterns to evidence accumulation models. In certain evidence accumulation models, evidence builds up gradually before a response is made, with two key properties: the rate of build-up (the “drift rate”) differs from decision to decision, but the amount of accumulated activity just before a response is issued (the “threshold”) is always the same. The authors interpreted their findings as showing that variability in response time could be explained by variability in drift rate as opposed to variability in threshold of the decision-making process, a claim that is hard to test when using only behavioral data. More and more electrophysiological work has since been interpreted in the framework offered by evidence accumulation models; for reviews, see (17) and (18).

Churchland, Kiani, and Shadlen (19) noted that some evidence accumulation models also predict increased decision thresholds when the number of choice alternatives increases. This change in threshold is required in some models to counter the effect of “statistical facilitation”; the tendency for the fastest-finishing of a set of accumulators to become even faster as the set grows larger (e.g., 20). Increased thresholds counter both the decreased RT as well as the increased error rates associated with statistical facilitation, and bring model predictions into line with behavioral data, such as Hick's Law. From a neural perspective, increased decision thresholds can be implemented in one of two ways: either by increasing the firing rate required to trigger a behavioral response, or by decreasing the baseline firing rate before a decision. Churchland, Kiani, and Shadlen investigated these two possibilities, by examination of neurons in the lateral intraparietal area (LIP) which behaved like evidence accumulators. Those neurons showed reduced baseline activity when more response choices were added. The subtle differences between a reduced baseline or an increased threshold are something that can be difficult to distinguish using behavioral data only (21).

Qualitative links between neural data and evidence accumulation models have also been drawn using fMRI methods. For example, Ho *et al.* (22) hypothesized that areas that implement evidence accumulation during a perceptual decision-making task should show a delayed, temporally extended hemodynamic response function (HRF) during difficult trials, as compared to easy trials. They identified areas where the shape of the HRF differed substantially between conditions, by testing for interactions between task difficulty and BOLD activity at a set of multiple time points throughout the trial. The raw signal in these areas was averaged over trials, and this indeed showed the predicted qualitative pattern of delayed and longer hemodynamic responses for trials from a hard condition, as compared to the easy condition.

Qualitative Predictive Linking in Models of Reinforcement Learning

The field of reinforcement learning and value-based decision-making has a long history of computational cognitive modeling (23). These computational models made it possible to design experiments that manipulated model parameters across conditions and compare the corresponding neural and behavioral data qualitatively (e.g., 24; 25).

An example is given by Nieuwenhuis *et al.* (26), who showed that modulating a single parameter in Holroyd & Coles' neurocomputational model of reinforcement learning could explain differences between age groups, by assuming plausible neuroanatomical differences. Nieuwenhuis *et al.* observed that the Holroyd and Coles' model could mirror the impaired performance of older adults in a probabilistic learning task, as well as the accompanying reduced error-related negativity (ERN) measured by EEG. It could do so by varying only one parameter in the model that represents to the efficiency of dopaminergic connections to the anterior cingulate cortex (ACC).

Since the early 2000s, the estimated parameters of reinforcement learning models have been quantitatively compared to differences in the neural signal, leading to a small revolution in the field, which will be discussed in the next section on quantitative predictive linking.

Qualitative Predictive Linking in Models of Symbolic Reasoning

The ACT-R model assumes distinct cognitive modules that perform different parts of cognitive tasks (27). For example, the cognitive steps necessary for performing some symbolic logic operation might be modeled as involving the visual module (to perceive the stimulus), the procedural and declarative memory modules (to remember the logical rules), and the motor module (to produce the desired behavioral response). From these assumptions, ACT-R can make predictions about differences in reaction time and accuracy between conditions. Many neuroimaging studies have related cognitive ACT-R models to fMRI data to localize the cognitive modules within the brain. Such localization assumptions are linking hypotheses, and subsequent studies have used qualitative predictive approaches to test those. For example, Borst *et al.* (28) constructed an ACT-R model of a task where both a subtraction operation, as well as a text entry had to be performed at the same time. The model made priori predictions about which modules (e.g., “Problem State”, “Declarative Memory”, “Manual”, and “Visual”) would be more activated during different combinations of easy/hard versions of the two tasks. These predictions about cognitive modules can be manifested as predictions about the behavioral data (via ACT-R's regular framework) and also as corresponding predictions for fMRI data, via the linking hypotheses which localize modules. These predictions were confirmed by (28), using a traditional region-of-interest fMRI study with different task contrasts.

The ACT-R community has come up with a “standard atlas”, predicting where in the brain BOLD activity should be expected, given hypothetical activity of cognitive modules (29; 30). The latest version of ACT-R software can even, given a model of a cognitive task, predict qualitative differences in BOLD-activity across different versions of a task. In recent work, these predictions have been made quantitative and can be empirically tested in fMRI data. This will be discussed in the next section.

Examples of Quantitative Predictive Linking

Computational cognitive models can also be linked to neural data in a quantitative way. This approach involves two models which are combined: a cognitive model that predicts behavior, as well as a neural model that predicts neural signals. Most often, one of these two models is very much simplified, usually reflecting the expertise, training, or focus of the modelers. For example, the cognitive model may be as basic as simple signal detection theory, or the neural model may be as atheoretical as the mean signal between two response-locked timepoints.

Quantitative Predictive Linking in models of Reinforcement Learning

Using quantitative outputs of a computational model of cognition to predict neural activity has been a successful strategy in the study of value-based decision-making and neuroeconomics (31). Especially prominent has been the *single-trial regression approach*, in which parameters of a reinforcement learning model are estimated from choice behavior during tasks involving the learning of reward values associated with different choices. These subject-specific parameter estimates can be used to calculate estimates of the subjective values of the different choice options to the subject, for every individual trial during the experiment (note that these values are dynamic due to learning). These subjective, trial-by-trial values can then be used as a hypothetical cognitive signal that tracks, for example, the difference between the expected reward after a choice and the reward that was actually delivered (the so-called “prediction error” or “delta” signal). To investigate a linking hypothesis, the researcher then hypothesizes that this cognitive signal is represented in the brain, at the *bridge locus*. Neural signals from the bridge locus should correspond to the phenomelogical concept under study, and to the hypothetical cognitive signal in this case (32). For example, the bridge locus of the prediction error signal might be some area in the brain where the neural signal consistently tracks the difference between the expected reward and the actual reward in a reinforcement learning task. At a practical level, the hypothetical cognitive signal, as estimated by the reinforcement learning model, can be transformed to a hypothetical BOLD fMRI-signal, by convolution with a hemodynamic response function (33). This creates a hypothetical fMRI signal corresponding to the prediction error signal, which can be used as a regressor in a general linear model (GLM), with additional regressors for other task-related activity (for example stimulus presentation). The parameters of this GLM are then estimated for all voxels in the brain. This yields a statistical parametric mapping of the brain that shows for which areas of the brain, BOLD activity correlates with the hypothetical neural signal representing stimulus value, and offers candidates for the bridge locus of interest.

The first studies in cognitive model-based fMRI studies that used the regression approach were in the field of reinforcement-learning. For example, the approach was used to show that the BOLD activity in orbitofrontral cortex (OFC) and in ventral striatum was correlated with the temporal difference error signal as estimated by a reinforcement learning model (34). Many studies using this approach were then published in the domain of reinforcement learning. It has, for example, been shown that different parts of the striatum code for different kinds of value representations (representing the value of states versus the value of actions; 35) and that uncertainty about the stability of the payoff structure of the environment is computed and learning rate accordingly modulated by the anterior cingulate cortex (36). Rodriguez, Aron, and Poldrack (37) applied the approach to the Rescorla-Wagner model during a categorization task, without explicit reward or punishment. Also more neurocomputational models, predicting how the different cortical inputs to striatum are integrated have been tested (38).

A recent study by den Ouden, Friston, Daw, McIntosh (39) extended the single-trial regression approach to link parameter estimates of a reinforcement learning model to trial-to-trial variability in not only the *magnitude of activation* in brain areas, but also to the *functional connectivity* between brain areas. The cognitive model was used to predict on which trials activity between two areas was more correlated. The authors applied the Rescorla-Wagner model, a simple reinforcement learning model, to a paradigm where auditory stimuli -- unrelated to the task at hand -- were predictive of visual stimuli that were also unrelated to the task. Because this relation was implicit, no behavior was available to fit the RL model to, but it was shown that functional connectivity between auditory and visual cortex was modulated by the prediction error. The authors used the single-trial prediction error, estimated by the RL model, as input for a dynamic causal modeling analysis (DCM; 40). Dynamic causal modeling can be used to study the causal influence of multiple brain areas on each other's BOLD signal. A DCM that assumed that the influence of the BOLD signal in auditory cortex on the BOLD signal in visual cortex was modulated by the prediction error of a trial was more likely than other models. This finding suggests that effective connectivity between sensory areas is increased when the degree of surprise in the environment is larger, with the direction of causality corresponding to the direction of information flow.

A follow-up study (41) used a perceptual decision-making task (faces vs. houses) with cues that predicted upcoming stimuli. DCM showed that BOLD activity in the ventral striatum, related to prediction error (as quantified by a reinforcement learning model) modulated the functional connectivity between visual areas and motor areas. This result provides evidence that the detection of uncertainty in the environment increases the neural influence of perceptual areas on motor areas.

Quantitative Predictive Linking In Models of Symbolic Reasoning

The single-trial regression approach has also gained a firm hold in the field of symbolic reasoning models. Recent versions of the ACT-R architecture predict quantitative differences in activation in different cognitive modules during a task. These predictions can be convolved with a canonical hemodynamic response function, generating quantitative hypotheses about which areas of the brain modulate their activity in correspondence with the activity of the proposed cognitive modules in the model. The first studies that applied such an approach fitted the hypothetical time course to signals from regions-of-interest that were chosen a priori (42; 43). Later, such time courses were also fitted to all the voxels throughout the entire brain, as had been done in the reinforcement learning before. For example, Borst *et al.* used a complicated multitasking paradigm, where either a subtraction or a text entry task had to be performed, while at the same time performing a listening comprehension task (44). The ACT-R model could predict, for every trial, the relative activity of “Problem State”, “Declarative Memory”, “Vision”, and “Manual” modules. Single-trial regression analysis showed correlated activity in corresponding cortical areas, largely overlapping with areas that had been related to these modules before by Anderson (27).

Borst & Anderson (45) extended the results of Borst *et al.* (44), by applying the same approach to five previously acquired datasets and combining the results into a meta-analysis. Declarative memory retrieval, as predicted by the ACT-R models, correlated with activity in the IFG and the ACC, whereas updating of working memory was also related to activity in the inferior parietal lobule (IFL). Model-based imaging helped to elucidate the roles of the different nodes in the frontoparietal network. This network includes both the IFG, ACC, as well as IFL and dissociating the individual roles of these nodes has turned out to be challenging when using the conventional subtraction-of-conditions approach, because the activity of the nodes in this network are highly correlated.

Quantitative Predictive Linking in Models of Evidence Accumulation

Linking evidence accumulation models of speeded decision-making to neuroimaging data represents a more difficult challenge, in many ways, than for the models of reinforcement learning and symbolic reasoning reviewed above. The challenge is more difficult because models of speeded decision-making have to explain variability in reaction times, and the processes underlying decision-making are therefore assumed to be stochastic. Across trials, there is variability in the amount of evidence that is necessary to make a decision, as well as variability in the average speed of evidence accumulation (as in the LBA; 46), possibly amongst even more variability (6). This means that, unlike in most reinforcement learning models, there is no one-to-one correspondence between data and the parameters of the model at the level of a single trial. This precludes the very popular single trial regression approach, at least for "out-of-the-box" evidence accumulation models, although different alternatives to resolve this issue have been proposed and performed.

One alternative is to change the unit of analysis from single trials to single subjects, focusing on the covariance of differences between subjects in neural and behavioral parameter estimates. In an fMRI study of decision-making, Forstmann *et al.* (47) instructed subjects to stress either the speed or accuracy of their decisions. The difference in BOLD-activity between accuracy- and speed-stressed trials in the striatum and the presupplementary motor area (pre-SMA) was correlated across subjects with the difference in model parameters related to response caution, estimated from behavioral data via the LBA model. In other words, participants who made large changes in their cognitive settings (for speed vs. caution) also showed large changes in fMRI responses, and vice versa. This provides some evidence for a role of these brain areas in setting a response threshold before a decision is made. In a follow-up study, Forstmann *et al.* (48) extended the earlier finding to structural connectivity measures. They identified a correlation between individual differences in response caution between speed- and accuracy-stressed trials and the strength of white matter connections between pre-SMA and striatum as measured by diffusion weighted imaging (DWI). The study showed evidence that the stronger these connections are, the more flexible subjects are in adjusting their response threshold to the current task demands.

Using a similar across-subjects approach, Mulder, Wagenmakers, Ratcliff, Boekel & Forstmann (49) used probabilistic payoffs to shift the decision biases of participants. As usual, these shifts were explained in a perceptual decision making model (the drift diffusion model) as a shift in the starting point parameter -- responses favored by bias were represented as having starting points for evidence accumulation that were closer to the response threshold. Mulder *et al.* showed that estimates of the start point, taken from behavioral data, were correlated with the difference in fMRI activity between biased and unbiased trials in frontoparietal regions involved in action preparation.

An alternative to the between-subjects approach is to link within-subject variability from neural and behavioral data by splitting the data on a neural measure and fitting a cognitive model to the subsets of behavioral data. Ratcliff, Philiastides, and Sajda (50) studied a perceptual decision-making task (houses vs. faces) and identified EEG components that classified trials as hard or as easy. Ratcliff *et al.* took trials from each single stimulus difficulty condition (in which nominal stimulus difficulty was constant) and applied a median split based on the amplitude of the EEG-component. Even though nominal stimulus difficulty was identical, estimated drift rates were lower in the trials with lower amplitude than trials with a higher EEG amplitude. Also, the estimated across-trial variability in difficulty was smaller for the median-split groups than when estimated from all trials at once. This pattern is consistent with the hypothesis that the EEG component indexes single-trial difficulty, and that restricting difficulty (estimated by EEG) reduces estimated variability in decision difficulty (measured by the model applied to the behavioral data).

In a different approach, van Maanen *et al.* (51) took not the neural, but the behavioral data as a starting point. In an attempt to leverage the power of the single-trial regressor approach, they developed an extension of the standard linear ballistic accumulator model (the LBA; 46), the “single-trial linear ballistic accumulator model'” (STLBA). The STLBA provided maximum-likelihood (ML) parameter estimates for the start point of the winning accumulator, for every trial. For many trials this value is identical, but the ML starting points of more extreme trials offer variability to explain corresponding variability in (neural) data. Van Maanen *et al.* used this model to analyze data from subjects performing a random-dot motion task, cued to stress either speed or accuracy. Van Maanen and colleagues did not use the standard single-trial regression approach, where convolved model-parameters are inserted into the GLM. Instead, the authors first estimated the height of the BOLD-response for the individual trials, assuming a canonical HRF (see 52). These single-trial estimates were then correlated with the amount of response caution of individual trials as estimated by the STLBA. The authors showed that BOLD activity in the pre-SMA and the dorsal ACC correlated significantly across trials with response caution during a speed-stressed response regime, but not during an accuracy-stressed response regime.

Boehm, van Maanen, Forstmann, & van Rijn (53) applied the STLBA of van Maanen *et al.* (51) to EEG data. The authors correlated single-trial estimates of response caution against the so-called contingent negative variation (CNV). The CNV is a buildup of negative EEG potential that occurs when subjects are shown a cue predicting an upcoming response stimulus. Its size was quantified by taking the mean potential between 100 ms and 200 ms before stimulus presentation at FCz, an EEG electrode close to pre-SMA. It turned out that the size of the CNV was linearly related to trial-to-trial variability in response caution within the same condition.

More recent approaches to linking evidence accumulation models to neural data start with the neural signal, and use this as input to an extended evidence accumulation model. Cavanagh *et al.* (54) estimated, separately for each trial in a decision-making experiment, the power in the theta frequency band from recorded EEG signals. These single-trial estimates of theta power were then used to inform parameter estimates in an extended version of the drift diffusion model (HDDM; 55). This model allowed different estimates of the threshold parameter on different trials, and a covariate model to assess the association of single-trial theta power with single-trial threshold estimates. When the model was fitted to data, its parameter estimates suggested that the coefficient of the covariate was probably larger than zero, which provides evidence that response caution (measured by the threshold parameter) is related to fluctuations in theta-power in medial prefrontal cortex.

Using the same model architecture, Frank *et al.* (56) extended the findings of Cavanagh *et al*. (54) to a reinforcement learning paradigm, where subjects had to learn to choose between two stimuli with different reward probabilities, while measuring both fMRI and EEG. A simple reinforcement learning model was fit to the responses of the subjects, as well as a drift diffusion model, using the approach of Wiecki *et al.* (55) to yield single-trial estimates of decision threshold. The parameter estimates were subsequently compared with multiple neural measures, as well as the output of the reinforcement learning model. Model comparison techniques showed that the fits of the DDM were improved by using the value-difference between the presented choices, as estimated by the RL model, as a covariate for the drift rate of the DDM. In addition, the DDM showed a better fit when the threshold parameter was included both the EEG and fMRI measures as covariates. This showed that the threshold parameter in the DDM model was reliably related to both cortical theta fluctuations as measured by EEG, as well as subcortical activations as measured by fMRI. For the fMRI data, single trial measures of BOLD activation were estimated using a slightly modified version of the canonical hemodynamic response function and a general linear model, applied to timecourses of predefined regions-of-interest. This was the first neuroscience study that related both EEG and fMRI to a cognitive computational model at the same time.

A similar approach to that of Cavanagh *et al.* and Frank *et al.* was developed in parallel by Turner *et al.* (57). Also in this “joint modeling approach'”, neural measures were used in addition to behavioral measures as input to an extended cognitive model. Turner *et al.*'s approach took the covariate-based analysis further, allowing for a general covariance matrix to link parameters of a behavioral model (the LBA model of decision-making) with the parameters of a neural model (a GLM). This approach supports more exploratory analyses, allowing the identification of different mappings from cognitive parameters to neural measures by studying the covariance matrix of the joint normal distribution; if a cognitive parameter is related to some neural measure, the covariance parameter that links them will be non-zero. Turner *et al.* (57) showed, using the data of Forstmann *et al.* (48), that this approach can find robust correlations between white-matter strength between pre-SMA and striatum, measured by diffusion-weighted magnetic resonance imaging (dMRI) and response caution, as quantified by the DDM.

In a more recent paper, Turner, van Maanen, and Forstmann (58) extended the joint modeling approach of (57) to relate single-trial BOLD activation estimates to single-trial parameter estimates of a drift diffusion model: the "neural drift diffusion model". The statistical approach was the same as used by Turner *et al.* (57): a Bayesian generative model assumed that both single-trial BOLD activity, as well as single-trial drift rates were sampled from a common multivariate normal distribution. Markov chain Monte-Carlo sampling (MCMC; 59; 60) offers the ability to sample from the marginal posterior distributions of the model parameters, after observing cognitive and neural data, and if neural activity is related to cognitive parameters, the corresponding covariance estimates of the joint normal distribution will be reliably different from zero. Turner *et al.* (58) showed that when activity in the "default mode network" was large at the onset of a trial, drift rates for that trial were lower. This could be explained as a lack of attention, or the presence of "task-unrelated thoughts", which decrease task performance.

Single Model Approach

The next step in the model\ling of neural and behavioral data is the development of models that can generate and predict both neural and behavioral data at the same time. Perhaps surprisingly, many of the approaches discussed above could in principle do that: if a researcher knows the size of a linear relationship between drift rate and BOLD amplitude, she can give the most-likely HRFs for a range of drift rates. However, in the applications of quantitative predictive modeling described above, the assumptions underlying the link are implicit and quite simplistic; usually, the neural model is nothing more than a descriptive measurement tool. Examples of such measurement models are the mean EEG amplitude in a fixed time window, or the linearly estimated height of the hemodynamic response in an fMRI signal, assuming a fixed shape across all subjects, areas and conditions. The assumed link between the cognitive parameter and neural measure is usually a strict linear one, with the added assumption of Gaussian noise.

Single Model Approaches in Evidence Accumulation

In some work in neurophysiology, the link between neural data and cognitive model is more explicit. The most complex models can take as input neural data from one source, and then predict neural data from another source, as well as behavior. Purcell *et al.* (61) identified and recorded from different clusters of cells in the frontal eye fields in awake macaque monkeys during a visual search task. Some neurons in the FEF only respond to specific visual inputs (“visual” neurons), while other neurons respond only just before a saccade (“motor” neurons), and some neurons respond to both (“visuomotor” neurons). Considered from the perspective of an evidence accumulation model of decision-making, the visual neurons might be interpreted as providing a continuous, noisy, representation of decision evidence, and the motor neurons might be interpreted as the accumulators which process that evidence. Purcell *et al.* did exactly this, and used the spike trains recorded from visual and visuomotor neurons as input to the accumulators of an evidence accumulation model. The authors showed that the model could use these inputs to reliably predict the behavioral data of the monkeys (response proportions and reaction time distributions). In this approach, the linking is made very explicit, and comparisons are made at the level of *distributions* over neural and behavioral data; no single trial estimates were made. The authors went further than just predicting behavior, and also compared the predictive performance of the model on neural data. For this, they used nine different architectures for evidence accumulation. These architectures differed in things like the presence or absence of leakage in the accumulation process, or mutual inhibition between accumulators. Interestingly, the response proportions and response time distributions were well explained by many of the different model architectures, even though those architectures made very different assumptions about neural structure. Purcell *et al.* suggested that one way to distinguish between their nine models was to make quantitative predictions about the properties of neurons that accumulate evidence, the “motor neurons”, and then probe which model most resembles the actual neural characteristics. Here, the linking of cognitive model and neural signal becomes very explicit and tight: it is assumed that the trajectory of evidence accumulation in the model maps identically on to spiking rates in motor neurons in FEF. The authors show that the onset of firing in motor neurons shows a correlation with RT, growth rate shows only a very minor, negative correlation, and baseline firing rate and ending firing rate do not correlate with RT at all. They then show that only one class of evidence accumulation architectures showed this pattern of correlation: the “gated” models, which assume that accumulation does not start until the amount of incoming evidence reaches a certain threshold.

Single Model Approaches in Symbolic Reasoning

Simple symbolic reasoning models have been combined with functional neuroimaging data in a single model using Hidden Semi-Markov Models (HSMMs). Such models assume that, in order to perform a task, subjects move through a discrete set of cognitive steps, or “states”, until they finished the trial (usually by giving a response). A HSMM can be fit to both behavioral and neuroimaging data, where it is assumed that both are dependent measurements of the same sequence of states.

Anderson, Betts, Ferris, and Fincham (62; 63) first applied this approach to a dataset where students solved linear algebra problems in an MRI scanner, where every step in solving the problem was made explicit using the task interface. The authors showed that, given both reaction times and functional neuroimaging data, it was possible to reliably predict in which state of solving the linear algebra problem the subject currently is. Extending this finding, Anderson and Fincham (64) showed that, in a paradigm where the different sub-steps of solving an algebra problem were not made explicit in the task interface, the number of cognitive states a subject goes through in a task can also be inferred by fitting different nested models. The authors provide evidence that these states correspond cognitive states like “Problem encoding”, “Planning”, “Solving” and “Responding”. One way that they do so is inspecting the prototypical brain activation patterns corresponding to the different cognitive states; for example, the “respond state” corresponds to high activation in left motor cortex. Another crucial argument is that the duration of these substates was specifically modulated for different experimental conditions. For example, response time differences across difficulty levels could almost exclusively be explained by differences in the length of the “solving” state. Such an effect would be very hard to investigate using only behavioral data.

An important disadvantage of fMRI is the temporally dispersed signal, providing a low temporal resolution. For experimental paradigms that (unlike algebra problems) take place on a sub-second timescale, EEG might be a more appropriate neuroimaging method to use in the HSMM framework, especially when one is interested in the possible number of substates that a subject goes through solving a task. Borst and Anderson (65) combined EEG with the HSMM framework in an associative recognition task. The authors showed that a familiarity process, an associative retrieval process, and a decision process play a role. The study therefore speaks in favor of a dual-process model where both processes of recollection and familiarity are present, and against a global matching model. Its results dissociate two main opposing theories in the literature on associative recognition and therefore provides an excellent example of how neuroimaging, when tightly linked to cognitive models, can help distinguish between different cognitive theories in a way behavioral data alone cannot.

Supplemental References

1. Stone M (1960): Models for choice-reaction time. *Psychometrika*.

2. Ratcliff R (1978): A theory of memory retrieval. *Psychological Review*.

3. Luce RD (1986): *Response times*. Oxford University Press.

4. Usher M, McClelland JL (2001): The time course of perceptual choice: the leaky, competing accumulator model. *Psychological Review*. 108: 550–592.

5. Smith PL (2010): Journal of Mathematical Psychology. *Journal of Mathematical Psychology*. 54: 266–283.

6. Ratcliff R, McKoon G (2008): The diffusion decision model: theory and data for two-choice decision tasks. *Neural Comput*. 20: 873–922.

7. Ratcliff R, Frank MJ (2012): Reinforcement-based decision making in corticostriatal circuits: mutual constraints by neurocomputational and diffusion models. *Neural Comput*. 24: 1186–1229.

8. Frank MJ (2006): Hold your horses: A dynamic computational role for the subthalamic nucleus in decision making. *Neural Networks*. 19: 1120–1136.

9. Rumelhart DE, Hinton GE, McClelland JL (1986): A General Framework for Parallel Distributed Processing. In: Rumelhart DE, McClelland JL, the PDP Research Group, editors. *Parallel Distributed Processing: Explorations in the Microstructure of Cognition (Vol 1)*. Cambridge, MA: MIT Press, pp 45–76.

10. Rumelhart DE, Hinton GE, Williams RJ (1986): Learning Internal Representations by Error Propagation. In: Rumelhart DE, McClelland JL, the PDP Research Group, editors. *Parallel Distributed Processing: Explorations in the Microstructure of Cognition (Vol 1)*. Cambridge, MA: MIT Press, pp 318–362.

11. Anderson JR (1992): Automaticity and the ACT* Theory. *American Journal of Psychology*. 105.

12. Sohn MH, Goode A, Stenger VA, Carter CS, Anderson JR (2003): Competition and representation during memory retrieval: Roles of the prefrontal cortex and the posterior parietal cortex. *Proceedings of the National Academy of Sciences*. 100: 7412–7417.

13. Qin Y, Sohn MH, Anderson JR, Stenger VA, Fissell K, Goode A, Carter CS (2003): Predicting the practice effects on the blood oxygenation level-dependent (BOLD) function of fMRI in a symbolic manipulation task. *Proceedings of the National Academy of Sciences*. 100: 4951–4956.

14. Gazzaniga MS, Ivry RB, Mangun GR (2007): *Cognitive neuroscience*. (M. S. Gazzaniga, R. B. Ivry, & G. R. Mangun, editors). Cambridge, Massachutes: MIT Press.

15. Lewandowsky S, Farrell S (2010): *Computational modeling in cognition: Principles and practice*. Sage.

16. Hanes DP, Schall JD (1996): Neural control of voluntary movement initiation. *Science*. 274: 427–430.

17. Gold JI, Shadlen MN (2001): Neural computations that underlie decisions about sensory stimuli. *Trends in Cognitive Sciences*. 5: 10–16.

18. Sequential sampling models in cognitive neuroscience: Advantages, applications, and extensions. (n.d.): Sequential sampling models in cognitive neuroscience: Advantages, applications, and extensions.

19. Churchland AK, Kiani R, Shadlen MN (2008): Decision-making with multiple alternatives. *Nat Neurosci*. 11: 693–702.

20. Usher M, Olami Z, McClelland JL (2002): Hick's Law in a Stochastic Race Model with Speed–Accuracy Tradeoff. *Journal of Mathematical Psychology*. 46: 704–715.

21. Smith PL, Ratcliff R (2004): Psychology and neurobiology of simple decisions. *Trends in Neurosciences*. 27: 161–168.

22. Ho TC, Brown S, Serences JT (2009): Domain General Mechanisms of Perceptual Decision Making in Human Cortex. *Journal of Neuroscience*. 29: 8675–8687.

23. Sutton RS, Barto AG (1998): *Introduction to Reinforcement Learning*, 1st ed. Cambridge, MA, USA: MIT Press.

24. Berns GS, McClure SM, Pagnoni G, Montague PR (2001): Predictability modulates human brain response to reward. *Journal of Neuroscience*. 21: 2793–2798.

25. Knutson B, Adams CM, Fong GW, Hommer D (2001): Anticipation of increasing monetary reward selectively recruits nucleus accumbens. *Journal of Neuroscience*. 21: RC159.

26. Nieuwenhuis S, Ridderinkhof KR, Talsma D, Coles MGH, Holroyd CB, Kok A, van der Molen MW (2002): A computational account of altered error processing in older age: dopamine and the error-related negativity. *Cognitive, Affective, & Behavioral Neuroscience*. 2: 19–36.

27. Anderson JR (2007): *How can the human mind occur in the physical universe?* New York, NY, USA: Oxford University Press.

28. Borst JP, Taatgen NA, Stocco A, van Rijn H (2010): The Neural Correlates of Problem States: Testing fMRI Predictions of a Computational Model of Multitasking. (B. J. Harrison, editor) *PLoS ONE*. 5: e12966.

29. Anderson JR, Fincham JM, Qin Y, Stocco A (2008): A central circuit of the mind. *Trends in Cognitive Sciences*. 12: 136–143.

30. Borst JP, Nijboer M, Taatgen NA, van Rijn H, Anderson JR (2015): Using Data-Driven Model-Brain Mappings to Constrain Formal Models of Cognition. (D. Margulies, editor) *PLoS ONE*. 10: e0119673.

31. Corrado GS, Sugrue LP, Brown JR, Newsome WT (2009): The Trouble with Choice: Studying DecisionVariables in the Brain. In:. *Neuroeconomics*. pp 1–19.

32. Teller DY (1984): Linking propositions. *Vision research*. 24: 1233–1246.

33. Glover GH (1999): Deconvolution of impulse response in event-related BOLD fMRI. *NeuroImage*. 9: 416–429.

34. O’Doherty JP, Dayan P, Friston KJ, Critchley H, Dolan RJ (2003): Temporal difference models and reward-related learning in the human brain. *Neuron*. 38: 329–337.

35. O’Doherty JP, Dayan P, Schultz J, Deichmann R, Friston KJ, Dolan RJ (2004): Dissociable roles of ventral and dorsal striatum in instrumental conditioning. *Science*. 304: 452–454.

36. Behrens TEJ, Woolrich MW, Walton ME, Rushworth MFS (2007): Learning the value of information in an uncertain world. *Nat Neurosci*. 10: 1214–1221.

37. Rodriguez PF, Aron AR, Poldrack RA (2006): Ventral–striatal/nucleus–accumbens sensitivity to prediction errors during classification learning. *Hum Brain Mapp*. 27: 306–313.

38. Haruno M, Kawato M (2006): Heterarchical reinforcement-learning model for integration of multiple cortico-striatal loops: fMRI examination in stimulus-action-reward association learning. *Neural Networks*. 19: 1242–1254.

39. Ouden den HEM, Friston KJ, Daw ND, McIntosh AR, Stephan KE (2009): A Dual Role for Prediction Error in Associative Learning. *Cerebral Cortex*. 19: 1175–1185.

40. Friston KJ, Harrison L, Penny W (2003): Dynamic causal modelling. *NeuroImage*. 19: 1273–1302.

41. Ouden den HEM, Daunizeau J, Roiser J, Friston KJ, Stephan KE (2010): Striatal Prediction Error Modulates Cortical Coupling. *Journal of Neuroscience*. 30: 3210–

3219.

42. Anderson JR, Qin Y, Sohn M-H, Stenger VA, Carter CS (2003): An information-processing model of the BOLD response in symbol manipulation tasks. *Psychon Bull Rev*. 10: 241–261.

43. Anderson JR (2005): Human symbol manipulation within an integrated cognitive architecture. *Cognitive Sc: A Multidisciplinary J*. 29: 313–341.

44. Borst JP, Taatgen NA, van Rijn H (2011): Using a symbolic process model as input for model-based fMRI analysis: Locating the neural correlates of problem state replacements. *NeuroImage*. 58: 137–147.

45. Borst JP, Anderson JR (2013): Using model-based functional MRI to locate working memory updates and declarative memory retrievals in the fronto-parietal network. *Proceedings of the National Academy of Sciences*. 110: 1628–1633.

46. Brown SD, Heathcote A (2008): The simplest complete model of choice response time: Linear ballistic accumulation. *Cognitive Psychology*. 57: 153–178.

47. Forstmann BU, Dutilh G, Brown S, Neumann J, Cramon Von DY, Ridderinkhof KR, Wagenmakers E-J (2008): Striatum and pre-SMA facilitate decision-making under time pressure. *Proceedings of the National Academy of Sciences*. 105: 17538–17542.

48. Forstmann BU, Anwander A, Schäfer A, Neumann J, Brown S, Wagenmakers E-J, *et al.* (2010): Cortico-striatal connections predict control over speed and accuracy in perceptual decision making. *Proceedings of the National Academy of Sciences*. 107: 15916–15920.

49. Mulder MJ, Wagenmakers E-J, Ratcliff R, Boekel W, Forstmann BU (2012): Bias in the Brain: A Diffusion Model Analysis of Prior Probability and Potential Payoff. *Journal of Neuroscience*. 32: 2335–2343.

50. Ratcliff R, Philiastides MG, Sajda P (2009): Quality of evidence for perceptual decision making is indexed by trial-to-trial variability of the EEG. *Proc Natl Acad Sci USA*. 106: 6539–6544.

51. van Maanen L, Brown SD, Eichele T, Wagenmakers E-J, Ho T, Serences J, Forstmann BU (2011): Neural Correlates of Trial-to-Trial Fluctuations in Response Caution. *Journal of Neuroscience*. 31: 17488–17495.

52. Mumford JA, Turner BO, Ashby FG, Poldrack RA (2012): Deconvolving BOLD activation in event-related designs for multivoxel pattern classification analyses. *NeuroImage*. 59: 2636–2643.

53. 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.

54. Cavanagh JF, Wiecki TV, Cohen MX, Figueroa CM, Samanta J, Sherman SJ, Frank MJ (2011): Subthalamic nucleus stimulation reverses mediofrontal influence over decision threshold. *Nature Publishing Group*. 14: 1462–1467.

55. Wiecki TV, Frank MJ (2013): A computational model of inhibitory control in frontal cortex and basal ganglia. *Psychological Review*. 120: 329–355.

56. Frank MJ, Gagne C, Nyhus E, Masters S, Wiecki TV, Cavanagh JF, Badre D (2015): fMRI and EEG Predictors of Dynamic Decision Parameters during Human Reinforcement Learning. *Journal of Neuroscience*. 35: 485–494.

57. Turner BM, Forstmann BU, Wagenmakers E-J, Brown SD, Sederberg PB, Steyvers M (2013): A Bayesian framework for simultaneously modeling neural and behavioral data. *NeuroImage*. 72: 193–206.

58. Turner BM, van Maanen L, Forstmann BU (2015): Informing cognitive abstractions through neuroimaging: the neural drift diffusion model. *Psychological Review*. 122: 312–336.

59. Kruschke JK (2011): *Doing Bayesian Data Analysis: A Tutorial with R and BUGS*. Barlington: Academic Press.

60. Turner BM, Sederberg PB, Brown SD, Steyvers M (2013): A method for efficiently sampling from distributions with correlated dimensions. *Psychol Methods*. 18: 368–384.

61. Purcell BA, Heitz RP, Cohen JY, Schall JD, Logan GD, Palmeri TJ (2010): Neurally constrained modeling of perceptual decision making. *Psychological Review*. 117: 1113–1143.

62. Anderson JR, Betts S, Ferris JL, Fincham JM (2010): Neural imaging to track mental states while using an intelligent tutoring system. *Proc Natl Acad Sci USA*. 107: 7018–7023.

63. Anderson JR (2012): Tracking problem solving by multivariate pattern analysis and Hidden Markov Model algorithms. *Neuropsychologia*. 50: 487–498.

64. Anderson JR, Fincham JM (2014): Discovering the sequential structure of thought. *Cognitive Science*. 38: 322–352.

65. Borst JP, Anderson JR (2015): The discovery of processing stages: Analyzing EEG data with hidden semi-Markov models. *NeuroImage*. 108: 60–73.