Sequential sampling models in computational psychiatry:

Bayesian parameter estimation, model selection and classification

Thomas V. Wiecki

February 7, 2014

Abstract

Current psychiatric research is in crisis. In this review I will describe the causes of this crisis and highlight recent efforts to overcome current challenges. One particularly promising approach is the emerging field of computational psychiatry. By using methods and insights from computational cognitive neuroscience, computational psychiatry might enable us to move from a symptom-based description of mental illness to descriptors based on objective computational multidimensional functional variables. To exemplify this I will survey recent efforts towards this goal. I will then describe a set of methods that together form a toolbox of cognitive models to aid this research program. At the core of this toolbox are sequential sampling models which have been used to explain diverse cognitive neuroscience phenomena but have so far seen little adoption in psychiatric research. I will then describe how these models can be fitted to subject data and highlight how hierarchical Bayesian estimation provides a rich framework with many desirable properties and benefits compared to traditional optimization-based approaches. Finally, non-parametric Bayesian methods provide general solutions to the problem of classifying mental illness within this framework.
# Contents

1 Part I: Motivation  
  1.1 Current challenges in psychiatry  
  1.1.1 Diagnostic and Statistical Manual of Mental Disorders  

2 Part II: Potential Solutions  
  2.1 Research Domain Criteria Project  
  2.2 Neurocognitive phenotyping  
  2.3 Computational psychiatry  
  2.3.1 Case studies in the domain of decision making  

3 Part III: Quantitative Methods  
  3.1 Sequential Sampling models  
  3.1.1 Drift Diffusion Model  
  3.1.2 Linear Ballistic Accumulator  
  3.1.3 Relationship to cognitive neuroscience  
  3.1.4 Application to computational psychiatry  
  3.2 Parameter estimation  
  3.2.1 Maximum Likelihood and $\chi^2$  
  3.2.2 Hierarchical Bayesian models  
  3.2.3 Likelihood-free methods  
  3.3 Model selection  
  3.4 Mixture Models  
  3.4.1 Gaussian Mixture Models  
  3.4.2 Dirichlet Process Gaussian Mixture Models  

4 Conclusions  

5 Appendix  
  5.1 Drift-Diffusion Model  
  5.2 Bayesian Inference
5.2.1 Empirical Bayesian Approximation ........................................... 35
5.2.2 Markov-Chain Monte-Carlo ....................................................... 36
5.3 Likelihood free methods ............................................................ 38
5.4 Model Comparison ................................................................. 40
  5.4.1 Deviance Information Criterion ............................................. 40
  5.4.2 BIC .................................................................................. 40
  5.4.3 Bayes Factor ....................................................................... 41
5.5 Mixture Models ........................................................................ 42
  5.5.1 Gaussian Mixture Models ..................................................... 42
  5.5.2 Dirichlet Process Gaussian Mixture Models ......................... 43
1 Part I: Motivation

Imagine going to a doctor because of chest-pain that has been bothering you for a couple of weeks. The doctor would sit down with you, listen carefully to your description of symptoms and prescribe medication to lower blood pressure in case you have a heart condition. After a couple of weeks your pain has not subsided. The doctor now prescribes medication against reflux which finally seems to help. In this scenario not a single medical analysis (e.g. EKG, blood work or a gastroscopy) was performed and medication with potentially severe side-effects prescribed on a trial-and-error basis. While highly unlikely to occur if you walked into a primary care unit with these symptoms today, this scenario is actually a realistic description if you had mental problems and had seen a psychiatrist instead.

There are several reasons for this discrepancy in sophistication between psychiatry and other fields of medicine. The main cause is that by definition, mental illness affects the brain – the most complex biological system yet encountered. Compared to the level of scientific understanding achieved on other organs of the human body such as the heart, our understanding of the normally functioning brain is still, arguably, in its infancy.

Despite this complexity concerted efforts in the brain sciences have lead to an explosion of knowledge and understanding about the healthy and diseased brain in the last decades. The discovery of highly effective psychoactive drugs in the 50s and 60s raised expectations that psychiatry would progress in a similar fashion. Unfortunately, in retrospect it appears that these discoveries were serendipitous in nature as little progress has been made since then (e.g. Insel et al. 2010; Hyman 2012). This lack of progress also caused many major pharmaceuticals companies like AstraZeneca and GlaxoSmithKline to withdraw from psychiatric drug development and close large research centers (Nutt and Goodwin 2011; Cressey 2011). In sum, psychiatry is a field in crisis (Poland et al. 1994; Insel et al. 2010; Hyman 2012; Sahakian et al. 2010). As outlined in more detail below, the main reason for this crisis is a lack of measurable quantitative descriptors of mental illness. This lack results from an explanatory gap of how basic neurobiological aberrations result in complex disorders of the mind (Montague et al. 2011; Hyman 2012).
In part II I will review current challenges in psychiatry and recent efforts to overcome them. Several examples from the domain of decision making show the promise of moving away from symptom-based description of mental illness and instead formulate objective, quantifiable computational biomarkers as a basis for further psychiatric research. Part III introduces a computational cognitive toolbox that is suited to construct these computational biomarkers. Sequential sampling models serve as a case study for how computational models, when fit to behavior, have successfully been used to identify and quantify latent neurocognitive processes in healthy humans. Bayesian methods provide a resourceful framework to fit these models to behavior and establish individualized descriptors of neurocognitive function. After establishing the validity of these models to provide neurocognitive descriptors of individuals, I will review how clustering techniques can be used to construct a map of individual differences based on these neurocognitive descriptors. In sum, the objective of this review is to outline a research program to map the domain of neuropsychiatric disease.

In order to maintain a clear focus on this objective there are certain relevant issues not addressed here-within. While of critical importance to psychiatric patients, I will not discuss the clinical, pharmaceutical, environmental, social or developmental aspects of mental illness or rehabilitation programs. Moreover, I will treat mental illness as a disease of the brain with a focus on dysfunctional neurocircuitry (Insel, 2010).

1.1 Current challenges in psychiatry

While the current crisis in psychiatry has complex causes that are deeply rooted in existing classification systems, one of the core problems is what has been identified by Montague et al. (2011) as the explanatory gap. This gap refers to our lack of understanding of the causal processes linking genes, molecular, cellular, neurocircuitry and cognition to psychiatric symptoms. This explanatory gap, coupled with the “(almost) unreasonable effectiveness of psychotropic medication” (Montague et al., 2011) gave many a false premise to expect progress without understanding.

This explanatory gap suggests a new approach to research of mental disorders which aims to link cognitive and pure neuroscience to clinical symptoms without the restrictions of prior classification schemes (Poland and Von Eckardt, 2013; Cuthbert and Insel, 2010; Robbins et al., 2012). I will next specify problems current classification systems introduce followed by recent efforts to address them.
1.1.1 Diagnostic and Statistical Manual of Mental Disorders

For decades the Diagnostic and Statistical Manual of Mental Disorders (DSM) has been the basis of clinical diagnosis, treatment and research of mental illness. At its core, the DSM defines distinct disease categories like schizophrenia (SZ) and depression. These categories are mainly derived from translating subjective experience to objective symptomatology (Nordgaard et al., 2012) assuming unspecified biological, psychological, or behavioral dysfunctions (Poland et al., 1994).

While of certain value to clinicians the DSM is specifically designed to also serve as a classification system for scientific research with the goal of more easily translating results directly into clinical practice. While this translational research goal is commendable, decisions regarding systematic classification are more often based on perceptions of clinical utility rather than scientific merit (Poland and Von Eckardt, 2013). As a consequence, DSM-based research programs failed to deliver consistent, replicable and specific results (Kendell and Jablensky, 2003; Andreasen, 2007; Regier and Narrow, 2009; Kendler et al., 2009; Cuthbert and Insel, 2010; Hyman, 2010).

Heterogeneity and Comorbidity  One major problem of contemporary psychiatry classification is the heterogeneous symptomatology of patients receiving identical diagnoses. One striking example of this is SZ where one must show at least 2 out of 5 symptoms to receive a diagnosis (Heinrichs, 2001). It is thus possible to have patients with completely different symptomatology being diagnosed as schizophrenic.

Comorbidity is defined as the co-occurrence of multiple diseases in one individual. Importantly, we must differentiate between two relevant types of comorbidity: (i) True comorbidity is a result of independent disorders co-occurring; (ii) artificial comorbidity on the other hand is a result of separately classifying disorders that have a similar pathogenic cascade. It is now widely documented (Markon, 2010; Krueger and Markon, 2006) that “comorbidity between mental disorders is the rule rather than the exception, invading nearly all canonical diagnostic boundaries.” (Buckholtz and Meyer-Lindenberg, 2012). The authors further note that “It is important to understand that comorbidity in psychiatry does not imply the presence of multiple diseases or dysfunctions but rather reflects our current inability
[to formulate] a single diagnosis to account for all symptoms.” Together, these issues belie the assumption that DSM-disorders represent distinct and independent categories with a unique pathological cascade. Together, these issues belie the assumption that DSM-disorders represent distinct and independent categories with a unique pathological cascade [Krueger and Markon 2006; Hyman 2010; Cuthbert and Insel 2010; Andreasen 2007].

2 Part II: Potential Solutions

As outlined above, the shortcomings of the current DSM manual are well documented and a need for improvement has been recognized. In the following I will outline current efforts to address these challenges.

2.1 Research Domain Criteria Project

The Research Domain Criteria Project (RDoC) is an initiative by the National Institute for Mental Health (NIMH) [Insel et al. 2010]. RDoC improves on previous research efforts based on the DSM in the following ways. First, as the name implies it is conceptualized as a research framework only and is thus clearly separated from clinical practice. Second, RDoC is completely agnostic about DSM categories. Instead of a top-down approach which aims at identifying neural correlates of psychiatric disease, RDoC suggests a bottom-up approach that builds on the current understanding of neurobiological underpinnings of different cognitive processes and link those to clinical phenomena. Third, the RDoC research program integrates different levels of analysis like imaging, behavior and self-reports.

At its core, RDoC is structured into a matrix with columns representing different “units of analysis” and rows for research domains. The units of analysis include genes, molecules, cells, circuits, physiology, behavior, and self-reports. Research domains are clustered into negative and positive valence systems, cognitive systems, systems for social processes and arousal/regulatory systems. Each of these domains is further subdivided into distinct processes; for example, cognitive systems include attention, perception, working memory, declarative memory, language behavior and executive control.

Despite clear improvements over previous DSM-based research programs, the RDoC initiative currently lacks consideration of computational descriptors [Poland and Von Eckardt 2013]. As I will

1See Borsboom et al. (2011) for an alternative explanation based on causal relationships between symptoms.
outline below, computational methods show great promise to help link different levels of analysis, elucidate clinical symptoms and identify sub-groups of healthy and patient populations.

2.2 Neurocognitive phenotyping

In a recent review article, Robbins et al. (2012) suggest the use of neurocognitive endophenotypes to study psychiatric disease: “Neurocognitive endophenotypes would furnish more quantitative measures of deficits by avoiding the exclusive use of clinical rating scales, and thereby provide more accurate descriptions of phenotypes for psychiatric genetics or for assessing the efficacy of novel treatments. The use of such measures would likely also facilitate and improve the use of informative animal models in psychiatry by focusing on cognitive and neural processes that can often be investigated in parallel across species. Defining such endophenotypes might cut across traditional psychiatric classification, and hence begin to explain the puzzle of apparent comorbidities.”

Of particular interest are three studies that use such neurocognitive endophenotypes by constructing dimensional functional profiles (MFPs) from summary statistics of a battery of various neuropsychological tasks to identify subtypes of ADHD (Durston et al., 2008; Sonuga-Barke, 2005; Fair et al., 2012).

Durston et al. (2008) argues that there are distinct pathogenic cascades along which abnormalities in at least three different brain circuits can lead to similar symptomatology. Specifically, abnormalities in dorsal fronto-striatal, orbito-fronto-striatal, or fronto-cerebellar circuits can lead to impairments of cognitive control, reward processing and timing, respectively. Core deficits in one or multiple of these brain networks can thus result in a clinical diagnosis of ADHD (see figure 1) and provides a compelling explanation for the heterogeneity of the ADHD patient population. Preliminary evidence for this hypothesis is provided by Sonuga-Barke (2005) who used principal component analysis (PCA) on multi-dimensional functional profiles (based on a neuropsychological task battery) of ADHD patients and identified 3 distinct sub-types co-varying on timing, cognitive control, and reward.

A similar approach of identifying clusters in the ADHD population using MFPs was taken by Fair et al. (2012). While similar in spirit to Sonuga-Barke (2005) and Durston et al. (2008), Fair et al. (2012) do not only look at differences in the patient population but in both, healthy controls (HC)
and ADHD patients. The clustering is achieved by the application of graph theory. Interestingly, the authors find that HC and ADHD is not the predominant dimension along which clusters form. Instead, the authors uncover different functional profiles that apply to both, HC and ADHD patients. Critically, a classifier trained to predict HC and ADHD subjects inside of individual profiles achieved better performance than a classifier trained on the aggregated data. In addition, the abnormalities in cognitive functions of ADHD patients were different across different clusters (e.g. one cluster might show differences in response inhibition while another one shows differences in RT variability; see figure 2b). In other words, this implies that the overall population clusters into different cognitive profiles. ADHD affects individuals differently based on which cognitive profile they exhibit. Importantly, this study suggests that the source of heterogeneity may not only be distinct pathogenic cascades being labeled as the same disorder but may actually be a result of the inherent heterogeneity present in the overall population – healthy and diseased.

The above mentioned studies all exemplify the danger of lumping subjects at the level of disease and treating them as one homogeneous category with a single, identifiable pathological cascade. Instead, these studies use MFPs to find an alternative characterization of subjects independent of their DSM
classification that is (i) quantitatively measurable and (ii) a closer approximation to the underlying neurocircuitry (Robbins et al., 2012).

While a clear improvement on previous research efforts that use the DSM diagnosis as the sole descriptor this approach still has problems. First, although there is less reliance on DSM categories, these studies still use the diagnostic label for recruiting subjects. It could be imagined, for example, that patients with similar impulse control disorders like OCD or Tourette’s have abnormalities in similar brain circuits; thus, if only OCD patients are recruited a critical part of the picture might be missed. Second, the cognitive task battery only covers certain aspects of cognitive function. Other tasks that for example measure working memory or reinforcement learning would be a useful addition. Finally, performance on each individual task is assessed by an aggregate performance score. Recent behavioral and neuropsychological findings, however, suggest that executive control in a single task may instead be more accurately characterized as a collection of related but separable abilities (Baddeley, 1966; Collette et al., 2005), a pattern referred to as the unity and diversity of executive functions (Duncan et al., 1997; Miyake et al., 2000). Most cognitive tasks rely on a concerted and often intricate interaction of various neural networks and cognitive processes (see e.g. Collins and Frank, 2012). This task

Figure 2: Profile differences between ADHD and healthy controls (TDC) on based on functional descriptors. Lower left corner describes which each point along the x-axis represents. Upper left corner represents overall profile-differences. Right side shows differences between ADHD and HC functional profiles inside various clusters (i.e., profiles). As can be seen, group-differences vary over different profiles. Reproduced from Fair et al. (2012).
impurity problem (Burgess 1997; Phillips 1997) complicates identification of separate brain circuits based solely on MFPs.

In sum, while cognitive phenotypes provide a useful framework for measuring brain function there is still ambiguity when using behavioral scores that present an aggregate measure of various brain networks. This issue is also discussed by Buckholtz and Meyer-Lindenberg (2012) in relation to comorbidity: “The fact that a brain circuit can be involved in multiple cognitive domains helps explain why diverse psychiatric disorders can exhibit common deficits and symptoms (comorbidity).” This “common symptom, common circuit” model of psychopathology is illustrated by figure 3. Disentangling these transdiagnostic patterns of psychiatric symptoms thus requires identification and measurement of underlying brain circuits. While the authors propose the use of functional imaging studies and genetic analysis I will discuss how computational modeling can contribute to disambiguate the multiple pathways leading to behavioral features.

2.3 Computational psychiatry

How have other brain sciences dealt with one-to-many mapping problem trying to dissociate behavior on a cognitive task with brain circuits? Computational models at different levels of abstraction have had tremendous impact on the field of cognitive neuroscience. The aim is to construct a model based on integrated evidence from neuroscience and psychology to explain neural activity as well as cognitive behavior. While more detailed biologically inspired models such as biophysical and neural network models are generally more constrained by neurobiology they often have many parameters which make it very difficult to fit them directly to human behavior. More abstract, process based models on the other hand often have fewer parameters that allow them to be fit directly to data at the cost of being less detailed about the neurobiology. Critically, all of these models allow for increased specificity in the identification of different neuronal and psychological processes that are often lumped together when analyzing task behavior based on summary statistics.

Using computational models to infer dysfunctional latent processes in the brain is the premise of the newly emerging field of computational psychiatry. In their groundbreaking review, Montague et al. (2011) define the goal for computational psychiatry of “extracting computational principles around
Figure 3: Pathogenic cascade from Brain Circuit (BC) over multiple related cognitive processes (C1-C3) to symptoms (Sa-Si). “Some of these symptoms will constitute diagnostic criteria for categorical disorder A but not disorder B (yellow shading), and some symptoms will be relatively selective for disorder B but not disorder A (red shading). However, the plurality of symptoms will overlap the two diagnostic categories (transdiagnostic symptoms, orange shading). This highlights the idea that connectivity circuits convey cognitive and symptom domain-specific, but disorder-general, genetic risk for mental illness.” Reproduced from Buckholtz and Meyer-Lindenberg [2012].
which human cognition and its supporting biological apparatus is organized. Achieving this goal will require new types of phenotyping approaches, in which computational parameters are estimated (neurally and behaviorally) from human subjects and used to inform the models. This type of large-scale computational phenotyping of human behavior does not yet exist.” Also, see figure 4 for a visual depiction.

Based on this premise, Maia and Frank (2011) identify computational models as a “valuable tool in taming [the complex pathological cascades of mental illness] as they foster a mechanistic understanding that can span multiple levels of analysis and can explain how changes to one component of the system (for example, increases in striatal D2 receptor density) can produce systems-level changes that translate to changes in behavior”. Moreover, the authors define three concrete strategies for how computational models can be used to study brain dysfunction (see also figure 5) given a model of normal function:

- **Deductive approach**: Established neuronal models can be tested for how pathophysiologically plausible alterations in connectivity or neurotransmitter levels (e.g. dopamine is known to be
reduced in Parkinson’s disease) affect system level activations and behavior. This is essentially a bottom-up approach as it involves the study of how known or hypothesized neuronal changes affect higher-level functioning.

- **Abductive approach:** Computational models can be used to infer neurobiological causes from known behavioral differences. In essence, this is a top-down approach which tries to link behavioral consequences back to underlying latent causes.

- **Quantitative abductive approach:** Parameters of a computational model are fit to a subject’s behavior on a suitable task or task battery. Different parameter values point to differences in underlying neurocircuitry of the associated subject or subject group. These parameters can either be used comparatively to study group differences (e.g., healthy and diseased) or as a regressor with e.g., symptom severity. This approach is more common with abstract models than with neural network models as the former have often fewer parameters and thus can be more easily fit to data.

### 2.3.1 Case studies in the domain of decision making

One key area in which computational models have had tremendous success in elucidating how the different cognitive and neurobiological gears work together is the domain of decision making. In addition, many mental illnesses can be characterized by aberrant decision making of one sort or another (Maia and Frank, 2011; Wiecki and Frank, 2010; Montague et al., 2011). In the following I will review recent cases where computational models of decision making have been used to better understand brain disorders.

**Computational models of reinforcement learning**

**Parkinson’s Disease**  Our first case study concerns Parkinson’s disease (PD). Its most visible symptoms affect the motor system as manifest as hypokinesia, bradykinesia, akinesia, rigidity, tremor and progressive motor degeneration. However, recently, cognitive symptoms have received increased attention (e.g., Cools, 2005; Frank, 2005; Moustafa et al., 2008; Cunha et al., 2009). PD is an intriguing neuropsychiatric disorder because its pathogenic cascade is well identified to be the cell death of...
Figure 5: Different approaches computational models can inform mental health research. Given a computational model of normal function (a), research can provide a mechanistic bridge from neural abnormalities to explain or compare behavioral differences in a deductive approach (b). Contrary, in an abductive approach (c) behavioral differences are used to infer underlying neuronal abnormalities. Similarly, computational models can be quantitatively fit to behavior to infer underlying causes (i.e. quantitative abductive approach; d). See text for more information. Reproduced from Maia and Frank (2011).
midbrain dopaminergic neurons in the substantia nigra pars compacta (SNc), part of the basal ganglia (BG) (Kish et al., 1988). Neural network models of the BG (Frank, 2005, 2006) interpret this brain network as an adaptive action selection device that conditionally gates internal or external actions based on their previous reward history. DA is critically involved in learning from rewards and punishments which actions to facilitate and which actions to suppress in the future (Ljungberg et al., 1992; Montague et al., 1997; Schultz, 1998; Waelti et al., 2001; Pan et al., 2005; Bayer et al., 2007; Roesch et al., 2007; Sutton and Barto, 1990; Barto, 1995; Montague et al., 1997; Schultz et al., 1997). Behavioral reinforcement learning tasks show that the chronic low levels of DA in PD patients result in a bias towards learning from punishment at the cost of learning from rewards (Frank et al., 2004; Cohen and Frank, 2009). In extension, we have argued that PD is not a motor disorder per se but rather an action selection disorder in which the progressive decline of motor and cognitive function can be interpreted in terms of aberrant learning not to select actions (Wiecki and Frank, 2010; Wiecki et al., 2009).

In this case study, an existing biological model of normative brain function was paired with a known and well localized neuronal dysfunction to extend our understanding of the symptomatology of a brain disorder. Note, however, that the model was not fit to data quantitatively. In the terminology established by Maia and Frank (2011), this is an example of the deductive approach in which the model provides a mechanistic bridge that explains how abnormal behavior can result from neurocircuit dysfunctions.

**Depression** Our second case study involves how depression may be understood as an action planning disorder. Planning a series of future actions is an exponentially complex problem as each individual action can have different outcomes which themselves enable new actions essentially forming a decision tree. One way to deal with this complexity is to prune the decision tree and not consider certain actions (see figure 6). Recently, Huys et al., (2012) proposed a model of how humans perform this approximation. Briefly, the authors suggest that actions that would lead to comparatively bad outcomes will not be further considered and pruned from the decision tree (note that this is provably not optimal as early bad outcomes could be more than accounted for later on). To test this theory the authors test healthy subjects on a novel behavioral task that requires execution of a sequence of actions each associated with winning or losing a certain number of points. Critically, in one task condition,
Figure 6: Depiction of decision trees with 2 possible actions (pressing of 'U' key (green) or 'I' key (yellow). Each action (starting at the top node) leads to winning or losing points and subsequent action choice. After 3 subsequent actions points are aggregated. Critically, subjects experience rewards and have to learn the reward structure (which remains stable across a number of trials) to maximize rewards. Complete enumeration of a decision tree (left) has exponential complexity and is thus computationally infeasible. Thus, the decision tree must be pruned at a certain point. One possible strategy is to prune once a large negative reward is encountered (as depicted). Reproduced from Huys et al. (2012).

Subjects had two action paths available to them: (i) one path in which a very large loss occurred early that, however, was more than accounted for later on; and (ii) another less favorable one action path without an early loss but lower cumulative rewards overall. Participants overwhelmingly chose the non-optimal action sequence that did not result in an early large loss suggesting that humans prune the decision tree once actions are encountered leading to unfavorable outcomes, even in scenarios where this strategy does not result in optimal performance.

The authors formulate, fit and compare various computational algorithmic models with different pruning strategies. The model that provided the best fit (measured with an integrated BIC score that penalizes model complexity; see below) had three parameters: specific pruning, general pruning and reward sensitivity. Specific pruning represents the probability of a participant to stop evaluating sub-trees that are associated with large losses while general pruning represents the probability to stop considering sub-trees irrespective of rewards. Reward sensitivity captured the tendency to evaluate a loss of e.g. -140 to be of larger absolute magnitude than a win of 140 points although they cancel each other (i.e. loss aversion).

Participants were also given questionnaires to assess their (sub-clinical) levels of depression (BDI). Intriguingly, this depression rating correlated with the specific pruning parameter of the model (see...
Figure 7: Correlation between specific pruning parameter \( \lambda_S \) and residual depression rating (BDI). Each circle represents one subject. Diameter of the circle corresponds to estimation uncertainty. See text for further model details. Reproduced from Huys et al. (2012).

This effect was specific to this parameter as there was no such correlation with the general pruning parameter. In other words, subjects with higher depressive ratings were quicker to discard plans that lead to bad outcomes early on – sometimes missing large rewards available to them later on. Finally, the authors speculate that the specific pruning depends on the serotonergic transmitter system.

In sum, Huys et al. (2012) proposed an algorithmic model of normative cognitive computation. By quantitatively fitting the model to behavioral data on a novel task and regressing an independent clinical variable (rating on the depression scale) with the fitted parameter values the authors are able pinpoint the cognitive computation underlying a clinically significant symptom. In terms of the Maia and Frank (2011) terminology, Huys et al. (2012) used a quantitative abductive approach.

Schizophrenia Despite SZ being the focus of intense research over the last decades, no single theory of its underlying neural causes has been able to explain the diverse set of symptoms that can lead to a SZ diagnosis. The symptomatology is structured in terms of positive symptoms like psychosis, negative symptoms like anhedonia which refers to the inability to experience pleasure from activities usually found enjoyable such as social interaction.

Recent progress has been made by the application of RL models to understand individual symptoms or a single symptom category (e.g. negative symptoms) rather than SZ as a whole (Waltz et al. 2011; Gold et al. 2008; 2012; Strauss et al. 2011).
In a recent behavioral study using a RL task, Waltz et al. (2007) found that SZ patients show reduced performance in selecting previously rewarded stimuli compared with HCs. Moreover, this performance deficit is most pronounced in patients with severe negative symptoms. Notably, SZ and HC did not differ in their ability to avoid actions leading to negative outcomes. However, this behavioral analysis did not allow to differentiate whether SZ patients were impaired at learning from positive outcomes or from a failure in representation of the prospective reward values during decision making.

This dichotomy in learning vs representation is also present in two types of RL models – actor-critic and Q-learning models (Sutton and Barto, 1998). An actor-critic model consists of two modules: an actor and a critic. The critic learns the expected rewards of states and trains the actor to perform actions that lead to better-than-expected outcomes. Q-learning models on the other hand have an explicit representation of the outcomes that are associated with each action. Thus, while a Q-learning model chooses actions based on their absolute reward values, an actor-critic chooses actions based on whether they lead to better-than-expected outcomes.

In a follow-up study, Gold et al. (2012) administered a new task that paired a neutral stimulus in one context with a positive and in another context with a negative stimulus. While the neutral stimulus has the same value of zero in both contexts, it is known that DA signals reward prediction errors (RPE) that drive learning in the BG are coding outcomes relative to the expected reward (Sutton and Barto, 1990; Barto, 1995; Montague et al., 1997; Schultz et al., 1997). Thus, in the negative context, receiving nothing is better than expected and will result in a positive RPE, driving learning in the BG to select this action in the future (Frank, 2005). In a test period in which no rewards were presented, participants had to choose between an action that had been rewarding and one that had simply avoided a loss. Both actions should have been associated with better-than-expected outcomes. An actor-critic model should thus show a tendency to select the neutral stimulus while a Q-learning model with explicit representation of the reward contingencies should mainly select the one with a higher reward. Intriguingly, when both of these models were fit to participant data, the actor-critic model produced a better fit for SZ patients with high degree of negative symptoms while HC and SZ with low negative symptoms were better fit by a Q-learning model. In other words, the high negative symptoms group largely based decisions on learned stimulus-response associations instead of expected reward values. Notably, HC and the low negative symptom group did not differ significantly in their
RL behavior. Moreover, this also rules out possibly confounding effects of antipsychotic drugs, as both patient groups were similarly medicated.

In a related line of work, Strauss et al. (2011) tested HC and SZ patients on a reinforcement learning task that allowed subjects to either adopt a safe strategy and exploit the rewards of actions with previously experienced rewards, or, to explore new actions with perhaps even higher payoffs. Frank et al. (2009) develop a computational model that can be fit to subjects’ behavior and recover how individual subjects balance this exploration-exploitation trade-off. Intriguingly, applying this model to SZ patients, Strauss et al. (2011) found that patients with high anhedonia ratings where less willing to explore their environment and uncover potentially better actions. This result suggests a reinterpretation of the computational cognitive process underlying anhedonia. For example, one might assume that the lack of engagement of social activities of anhedonistic patients results from an inability to experience pleasure and as a consequence, a failure to learn the positive value of social interaction. Instead, this study suggests that anhedonia is a result of an inability to consider the prospective benefit of doing something that might lead to better outcomes.

In sum, Gold et al. (2012) and Strauss et al. (2011) used a quantitative abductive approach to infer aberrant computational cognitive processes in RL in a subgroup of SZ patients. By grouping subjects according to symptom severity instead of diagnosis the authors addressed the problem of heterogeneity. Moreover, linking the results to prior neural network modeling efforts (Frank and Claus, 2006), as described above, also points to the OFC as a promising target for further investigation as a neural source of negative symptoms in SZ patients.

**Computational models of response inhibition** Besides RL, response inhibition is another widely studied phenomenon in cognitive neuroscience of relevance to mental illness. Response inhibition is required when actions in the planning or execution stage are no longer appropriate and must be suppressed. The antisaccade task is one such task that is often used in a psychiatric setting (e.g. Aichert et al. 2012; Fukumoto-Motoshita et al. 2009). It requires subjects to inhibit a prepotent response to a salient stimulus and instead saccade to the opposite side (Hallett 1979). A wealth of literature has demonstrated reduced performance of psychiatric patients with disorders including
attention-deficit/hyperactivity disorder (ADHD) (Nigg, 2001; Oosterlaan et al., 1998; Schachar and Logan, 1990), obsessive compulsive disorder (OCD) (Chamberlain et al., 2006; Menzies et al., 2007; Penadés et al., 2007; Morein-Zamir et al., 2009), schizophrenia (SZ) (Huddy et al., 2009; Bellgrove et al., 2006; Badcock et al., 2002), Parkinson’s disease (PD) (van Koningsbruggen et al., 2009) and substance abuse disorders (Monterosso et al., 2005; Nigg et al., 2006). However, as demonstrated by Wiecki and Frank (2013), even a supposedly simple behavioral task such as the antisaccade task requires a finely orchestrated interplay between various brain regions including frontal cortex and basal ganglia. It thus can not be said that decreased accuracy in this task is evidence of response inhibition deficits per se as the source of this performance impairment can be manifold.

In follow-up work, we have formulated a psychological process model which summarizes the higher-level computations of the neural network and has fewer parameters. In an attempt to bridge these two levels of abstraction, we fit this process model to the outputs of this neural network model for which the biological modulations can be tightly controlled (Wiecki and Frank, 2010). Interestingly, by modulating different biological parameters in the neural network model and recovering which parameter of the process model was affected by this modulation, we were able to associate high-level computational processes with their neural correlates. The hope is that these associations, once validated, allow us to infer specific neural aberrations from behavioral performance.

In sum, computational models like the DDM allow mapping of behavior to psychological processes and could thus be categorized as the computational abductive approach. However, ambiguity of how psychological processes relate to the underlying neurocircuitry still have to be disambiguated. By combining different levels of modeling these ambiguities can be better identified and studied. Ultimately, this might allow development of tasks that use specific conditions (e.g. speed-accuracy trade-off, reward modulations and conflict) to disambiguate the mapping of psychological processes to their neurocircuitry. Using biological process models to test different hypotheses about the behavioral and cognitive effects of neurocircuit modulations would correspond to the deductive approach. In other words, by combining the research approaches outlined by Maia and Frank (2011) we can use our understanding of the different levels of processing to inform and validate how these levels interact in the healthy and dysfunctional brain.
In sum, there are a few example studies which applied established computational models to identify model parameters (which aim to describe a single cognitive function) and relate it to the severity of a specific clinical symptom. In the following, I will review sequential sampling models and Bayesian methods and show how they are well suited to solve many of the issues encountered in computational psychiatry.

3 Part III: Quantitative Methods

Part I described the current issues in psychiatry. Several examples presented in part II highlighted computational psychiatry as an area with a lot of potential to solve these problems. In part III I will review several quantitative methods that can solve the problems associated with quantifying cognitive function. Specifically, sequential sampling models present a versatile tool to model cognitive function. Fitting these models to data – especially with the small number of trials often found in clinical experiments – is a challenge well addressed by hierarchical Bayesian models that share statistical power by assuming similarity between subjects. However, there are two limitations with this basic approach. (i) Traditional inference methods require a likelihood function which is often intractable for more nuanced formulations of sequential sampling models. Likelihood free methods solve this problem as they only require a generative process from which the likelihood is estimated. (ii) While the similarity assumption made by hierarchical Bayesian modeling is reasonable we can not know the exact form of this similarity ahead of time. Bayesian mixture models address this problem by inferring clusters from the data. While traditional methods like Gaussian Mixture Models require specification of the number of clusters to be found in the data ahead of time, Bayesian non-parametrics relax this restriction and infer the number of clusters from the data. In the following I will focus on how each quantitative method helps to solve the above mentioned issues. Mathematical details can be found in the appendix. Finally, while these methods are described with the motivation of estimating sequential sampling models, the Bayesian methods are applicable more broadly (e.g. to the estimation of RL models).
3.1 Sequential Sampling models

Cognition spans many mental processes that include attention, social cognition, memory, emotion, and reasoning, to name a few. As outlined above, RL models have already proven to be a valuable tool in explaining neuropsychological disorders and their symptoms. A computational psychiatric framework that aims to explain the multi-faceted domain of mental illness must thus include computational cognitive neuroscience models that cover a broad range of cognitive processes. As such a review is outside of our scope (but see O’Reilly and Munakata (e.g. 2000)) I will focus on sequential sampling models as an illustrative example for how these models work, how they have been applied to study normal and aberrant neurocognitive phenomena, how they can be fit to data using Bayesian estimation and how subgroups of subjects can be inferred using mixture models.

Sequential sampling models (SSMs) (e.g. Townsend and Ashby, 1983) have established themselves as the de facto standard for modeling data from simple decision making tasks (e.g. Smith and Ratcliff, 2004). Each decision is modeled as a sequential extraction and accumulation of information from the environment and/or internal representations. Once the accumulated evidence crosses a threshold, a corresponding response is executed. This simple assumption about the underlying psychological process has the intriguing property of reproducing reaction time distributions and choice probability in simple two-choice decision making tasks.

SSMs generally fall into one of two classes: (i) diffusion models which assume that relative evidence is accumulated over time and (ii) race models which assume independent evidence accumulation and response commitment once the first accumulator crossed a boundary (e.g. LaBerge, 1962; Vickers, 1970). While there are many variants of these models they are often closely related on a computational level and sometimes mathematically equivalent under certain assumptions (Bogacz et al., 2006). As such, I will restrict discussion to two exemplar models from each class widely used in the literature: the drift diffusion model (DDM) (Ratcliff and Rouder, 1998; Ratcliff and McKoon, 2008) belonging to the class of diffusion models and the linear ballistic accumulator (LBA) (Brown and Heathcote, 2008) belonging to the class of race models.
3.1.1 Drift Diffusion Model

The DDM models decision making in two-choice tasks. Each choice is represented as an upper and lower boundary. A drift-process accumulates evidence over time until it crosses one of the two boundaries and initiates the corresponding response (Ratcliff and Rouder, 1998; Smith and Ratcliff, 2004). The speed with which the accumulation process approaches one of the two boundaries is called the drift rate and represents the relative evidence for or against a particular response. Because there is noise in the drift process, the time of the boundary crossing and the selected response will vary between trials. The distance between the two boundaries (i.e. threshold) influences how much evidence must be accumulated until a response is executed. A lower threshold makes responding faster in general but increases the influence of noise on decision making while a higher threshold leads to more cautious responding. Reaction time, however, is not solely comprised of the decision making process – perception, movement initiation and execution all take time and are summarized into one variable called non-decision time. The starting point of the drift process relative to the two boundaries can influence if one response has a prepotent bias. This pattern gives rise to the reaction time distributions of both choices (see figure 8; mathematical details can be found in the appendix).

Later on, the DDM was extended to include inter-trial variability in the drift-rate, the non-decision time and the starting point in order to account for two phenomena observed in decision making tasks – early and late errors. Models that take this into account are referred to as the full DDM (Ratcliff and Rouder, 1998).

3.1.2 Linear Ballistic Accumulator

The Linear Ballistic Accumulator (LBA) model belongs to the class of race models (Brown and Heathcote, 2008). Instead of one drift process and two boundaries, the LBA contains one drift process for each possible response with a single boundary each. Thus, the LBA can model decision making when more than two responses are possible. Moreover, unlike the DDM, the LBA drift process has no intra-trial variance. RT variability is obtained by including inter-trial variability in the drift-rate and the starting point distribution (see figure 9). Note that the simplifying assumption of a noiseless drift-process simplifies the math significantly leading to a computationally faster likelihood function.
Figure 8: Trajectories of multiple drift-processes (blue and red lines, middle panel). Evidence is accumulated over time (x-axis) with drift-rate $v$ until one of two boundaries (separated by threshold $a$) is crossed and a response is initiated. Upper (blue) and lower (red) panels contain histograms over boundary-crossing-times for two possible responses. The histogram shapes match closely to that observed in reaction time measurements of research participants.
Figure 9: Two linear ballistic accumulators (left and right) with different noiseless drifts (arrows) sampled from a normal distribution initiated at different starting points sampled from uniform distribution. In this case, accumulator for response alternative 1 reaches criterion first and gets executed. Because of this race between two accumulators towards a common threshold these model are called race-models. Reproduced from Donkin et al. (2011).

In a simulation study it was shown that the LBA and DDM lead to similar results as to which parameters are affected by certain manipulations (Donkin et al., 2011).

3.1.3 Relationship to cognitive neuroscience

SSMs were originally developed from a pure information processing point of view and primarily used in psychology as a high-level approximation of the decision process. More recent efforts in cognitive neuroscience have simultaneously (i) validated core assumptions of the model by showing that neurons indeed integrate evidence probabilistically during decision making (Smith and Ratcliff, 2004; Gold and Shadlen, 2007) and (ii) applied this model to understand and describe neural correlates of cognitive processes (e.g. Forstmann et al., 2010a; Cavanagh et al., 2011).

Multiple routes to decision threshold modulation have been identified. Decision threshold in the speed-accuracy trade-off is modulated by changes in the functional connectivity between pre-SMA and striatum (Forstmann et al., 2010a). Neural network modeling (Frank, 2006; Ratcliff and Frank, 2012) validated by studies of PD patients with a deep-brain-stimulator (DBS) in their subthalamic nucleus.
(STN) (Frank et al., 2007a) suggest that this node is implicated in raising the decision threshold when there is conflict between two options associated with similar rewards. This result was further corroborated by Cavanagh et al. (2011) who found that frontal theta power (as measured by electroencephalography and thought to correspond to conflict (Cavanagh et al., 2012)) is correlated with decision threshold increase on a trial-by-trial basis. As predicted, this relationship was broken in PD patients with DBS turned on (but, critically, not when DBS was turned off thus showing the effect is not a result of the disease). In other words, by interfering with STN function through stimulation we were able to show that this brain area is causally involved in decision threshold modulation despite intact experience of conflict (as measured by theta power). Interestingly, these results provide a computational cognitive explanation for the clinical symptom of impulsivity observed in PD patients receiving DBS (Frank et al., 2007a).

3.1.4 Application to computational psychiatry

Despite its long history, the DDM has only recently been applied to the study of psychopathology. For example, threat/no-threat categorization tasks (e.g. “Is this word threatening or not?”) are used in anxiety research to explore biases to threat responses. Interestingly, participants with high anxiety are more likely to classify a word as threatening than low anxiety participants. One hypothesis assumes that this behavior results from an increased response bias towards threatening words in anxious people (Becker and Rinck, 2004; Manguno-Mire et al., 2005; Windmann and Krüger, 1998). Using DDM analysis, White (2009) showed that instead of a response bias (or a shifted starting-point in DDM terminology), anxious people actually showed a perceptual bias towards classifying threatening words as indicated by an increased DDM drift-rate.

In a recent review article, White et al. (2010) use this case-study to highlight the potential of the DDM to elucidate research into mental disease. Note that in this study the authors did not attempt to hypothesize about the underlying neural cause of this threat-bias. While there is some evidence that bias in decision making is correlated with activity in the parietal network (Forstmann et al., 2010b) this was not tested in respect to threatening words. Ultimately, I suggest that this research strategy should be applied to infer neural correlates of psychological DDM decision making parameters using functional methods like fMRI to the study neuropsychopathology (as outlined above).
The DDM has also been successfully used to show that ADHD subjects were less able to raise their decision threshold when accuracy demands were high [Mulder et al., 2010]. Interestingly, the amount by which ADHD subjects failed to modulate their decision threshold correlated strongly with patients’ impulsivity/hyperactivity rating. Moreover, this correlation was specific to impulsivity as no correlation was found between decision threshold modulation and inattentiveness.

In sum, SSMs show great promise as a tool for computational psychiatry. However, their applicability depend on the ability to accurately estimate them to construct individual CMDFs. In the following I will review different parameter estimation techniques. Special focus will be given to Bayesian methods. Finally, once SSMs can be fit accurately the question arises how to construct a map of mental illness based on these CMDF. Towards this goal I will review clustering methods that can be expressed in the Bayesian framework.

3.2 Parameter estimation

To identify computational biomarkers in a variable clinical population with the DDM it is critical to have robust and sensitive estimation methods. In the following I will describe traditional parameter estimation methods and their pitfalls. I will then argue how Bayesian estimation provides a complete framework that avoids these pitfalls.

3.2.1 Maximum Likelihood and $\chi^2$

Traditionally, fitting of computational models is treated as an optimization problem in which an objective function is minimized. Different objective functions for the DDM have been proposed. Most common is the quantile method which calculates the quantiles of the DDM likelihood and uses the $\chi^2$ statistic as the objective to compare it to the quantiles observed in the data [Ratcliff and Tuerlinckx 2002]. While this approach requires a fair amount of data to get stable quantile estimation it naturally deals with outlier RTs as the quantiles average extreme values out. In comparison, maximum likelihood estimation (MLE) directly uses the DDM likelihood function derived from the DDM generative process [Navarro and Fuss 2009].
Psychological experiments often test multiple subjects on the same behavioral task. Models are then either fit to individual subjects or to the aggregated group data. Both approaches are not ideal. When models are fit to individual subjects we neglect any similarity the parameters are likely to have. While we do not necessarily have to make use of this property to make useful inference if we have lots of data, the ability to infer subject parameters based on the estimation of other subjects generally leads to more accurate parameter recovery [Sofer et al.] in cases where little data is available as is often the case in clinical and neurocognitive experiments. One alternative is to aggregate all subject data into one meta-subject and estimate one set of parameters for the whole group. While useful in some settings, this approach is unsuited for the setting of computational psychiatry as individual differences play a huge role.

3.2.2 Hierarchical Bayesian models

Statistics and machine learning have developed efficient and versatile Bayesian methods to solve various inference problems [Poirier 2006]. More recently, they have seen wider adoption in applied fields such as genetics [Stephens and Balding 2009] and psychology (e.g. Clemens et al. 2011). One reason for this Bayesian revolution is the ability to quantify the certainty one has in a particular estimation. Moreover, hierarchical Bayesian models provide an elegant solution to the problem of estimating parameters of individual subjects outlined above. Under the assumption that participants within each group are similar to each other, but not identical, a hierarchical model can be constructed where individual parameter estimates are constrained by group-level distributions [Nilsson et al. 2011; Shiffrin et al. 2008].

Bayesian methods require specification of a generative process in form of a likelihood function that produced the observed data $x$ given some parameters $\theta$. By specifying our prior belief we can use Bayes formula to invert the generative model and make inference on the probability of parameters $\theta$:

$$P(\theta|x) = \frac{P(x|\theta) * P(\theta)}{P(x)}$$ (1)

Where $P(x|\theta)$ is the likelihood and $P(\theta)$ is the prior probability. Computation of the marginal likelihood $P(x)$ requires integration (or summation in the discrete case) over the complete parameter space $\Theta$: 
\[ P(x) = \int_{\Theta} P(x|\theta) \, d\theta \]  \hspace{1cm} (2)

Note that in most scenarios this integral is analytically intractable. Sampling methods like Markov-Chain Monte Carlo (MCMC) \cite{Gamerman2006} circumvent this problem by providing a way to produce samples from the posterior distribution. These methods have been used with great success in many different scenarios \cite{Gelman2003} and will be discussed in more detail below.

Another nice property of the Bayesian method is that it lends itself naturally to a hierarchical design. In such a design, parameters for one distribution can themselves come from a different distribution which allows chaining together of distributions of arbitrary complexity and map the structure of the data onto the model.

This hierarchical property has a particular benefit to cognitive modeling where data is often scarce. We can construct a hierarchical model to more adequately capture the likely similarity structure of our data. As above, observed data points of each subject \( x_{i,j} \) (where \( i = 1, \ldots, S \) data points per subject and \( j = 1, \ldots, N \) for \( N \) subjects) are distributed according to some likelihood function \( f|\theta \).

We now assume that individual subject parameters \( \theta_j \) are normal distributed around a group mean with a specific group variance (\( \lambda = (\mu, \sigma) \) with hyperprior \( G_0 \)) resulting in the following generative description:

\[
\mu, \sigma \sim G_0() \hspace{1cm} (3) \\
\theta_j \sim N(\mu, \sigma^2) \hspace{1cm} (4) \\
x_{i,j} \sim f(\theta_j) \hspace{1cm} (5)
\]

See figure 10 for the corresponding graphical model description.

Another way to look at this hierarchical model is to consider that our fixed prior on \( \theta \) from formula (1) is actually a random variable (in our case a normal distribution) parameterized by \( \lambda \) which leads to the following posterior formulation:

\[
P(\theta, \lambda|x) = \frac{P(x|\theta) * P(\theta|\lambda) * P(\lambda)}{P(x)} \hspace{1cm} (6)
\]

30
Note that we can factorize $P(x|\theta)$ and $P(\theta|\lambda)$ due to their conditional independence. This formulation also makes apparent that the posterior contains estimation of the individual subject parameters $\theta_j$ and group parameters $\lambda$.

Several inference methods to estimate the posterior distribution have been developed. For details on commonly used sampling algorithms, see the appendix.

In sum, hierarchical Bayesian estimation leverages similarity between individual subject to share statistical power and increase sensitivity in our parameter estimation. However, note that in our computational psychiatry application the homogeneity assumption that all subjects come from the same normal distribution is almost certainly violated (see above). To deal with the heterogeneous data often encountered in psychiatry I will discuss mixture models further down below. Next, I will describe algorithms to estimate this posterior distribution.

### 3.2.3 Likelihood-free methods

Most models in cognitive neuroscience and mathematical psychology like the DDM are described by a latent generative process (see e.g. \footnote{7}). However, this generative description is usually ill suited
for inference as it does not directly provide us with a closed-form likelihood function \( p(x|\theta) \) of how observed data (e.g. the Wiener first passage time; see above) arise from this generative process.

While the DDM is used partly because it has a tractable likelihood functions, many interesting considerations require models for which a generative process but no tractable likelihood can be specified. Recent examples of these efforts include changes of mind as new evidence is processed (Resulaj et al., 2009), the influence of attention (Krajbich and Rangel, 2011), and reward-based decision making given conflict in values of alternative actions (Cavanagh et al., 2011; Ratcliff and Frank, 2012). In these cases, a likelihood function must instead be approximated by simulation from the generative process. This type of inference is commonly called likelihood free.

It can be expected that the full spectrum of cognitive function will be relevant in computational psychiatry. Being able to formulate (and estimate) drift-diffusion models very flexibly potentially allows its application to new domains.

For a detailed explanation of these methods, see the appendix.

### 3.3 Model selection

Computational models often allow formulation of several plausible accounts of cognitive behavior. We anticipate that this problem will also occur in computational psychiatry where multiple theories of cognitive dysfunction must be tested. One way to differentiate between these various plausible hypothesis as expressed by alternative models is model comparison. In the following I will review various methods and metrics to compare hierarchical models. The most critical property for model comparison is that model complexity gets penalized because more complex models have greater explanatory power by design. Several model comparison measures have been devised. I refer the reader to the appendix section for mathematical details.

### 3.4 Mixture Models

In this section I will review different mixture models that allow estimation of clusters in data in a Bayesian framework. These are relevant to our objective as they (i) deal with the heterogeneity
encountered in computational psychiatry and (ii) have the potential to bootstrap a new psychiatric classification system based on measurable, quantitative, computational endophenotypes. Because we are describing a toolbox using hierarchical Bayesian estimation techniques I will focus this section on mixture models as they are easily integrated into this framework. Where possible, I will highlight connections to more traditional clustering methods like k-means.

### 3.4.1 Gaussian Mixture Models

GMMs assume parameters to be distributed according to one of several Gaussian distributions (i.e. clusters). Specifically, given the number of clusters $k$, each cluster mean and variance gets estimated from the data. This type of model is capable of solving our above identified problem of assuming heterogeneous subjects to be normally distributed: by allowing individual subject parameters to be assigned to different clusters we allow estimation of different sub-groups in our patient and healthy population. Note, however, that the number $k$ of how many clusters should be estimated must be specified a-priori in a GMM and remain fixed for the course of the estimation. This is problematic as we do not necessarily know how many sub-groups to expect in advance. As we will see below, Bayesian non-parametrics solve this issue by inferring the number of clusters from data.

### 3.4.2 Dirichlet Process Gaussian Mixture Models

Dirichlet processes Gaussian mixture models (DPGMMs) belong to the class of Bayesian non-parametrics (Antoniak, 1974). They can be viewed as a variant of GMMs with the critical difference that they assume an infinite number of potential mixture components (see Gershman and Blei (2012) for a review).

Mixture models can infer sub-groups when the data is heterogeneous as is generally the case in patient populations. While the mindset describing these methods was their application towards the SSM their applicability is much more general than that. For example, the case-studies described above which used, among others, RL models to identify differences between HC and psychiatric patients could easily be embedded into this hierarchical Bayesian mixture model framework I outlined here. There are multiple benefits to such an approach. First, computational models fitted via hierarchical Bayesian estimation provide a tool to accurately describe the neurocognitive functional profile of individuals. Second, the mixture model approach is ideally suited to deal with the heterogeneity in patients but
also healthy controls (Fair et al., 2012). Third, by testing psychiatric patients with a range of diagnoses (as opposed to most previous research studies that only compare patients with a single diagnosis, e.g. SZ, to controls) we might be able to identify shared pathogenic cascades as suggested by Buckholtz and Meyer-Lindenberg (2012).

4 Conclusions

As outlined above, computational psychiatry is an emerging field that shows great promise to understand aberrant biological processes in mental disease and address current challenges encountered in mental health research. By fitting computational models to behavioral data we can construct computational multi-dimensional features to replace symptom-based classification as implemented by the DSM. Decision making appears to provide a good framework for studying psychiatric disease as many disorders show abnormalities in core decision making processes. Sequential sampling models have a good track record in describing individual differences and can be linked to neuronal processes. Hierarchical Bayesian estimation provides a compelling toolbox to fit these models directly to data as it (i) provides an uncertainty measure; (ii) allows estimation of individual and group-level parameters simultaneously; (iii) allows for direct model comparison; (iv) can be used in scenarios where a likelihood cannot be easily formulated; and (v) enable deconstruction of symptoms by identifying latent clusters. For example, impulsivity is a core symptom of impulse control disorders like ADHD, OCD, Tourette syndrome, substance abuse and eating disorder (Robbins et al., 2012). Computational cognitive models have already started to deconstruct this broadly defined behavioral symptom and identified separate pathways that can all lead to alterations in impulse control (Dalley et al., 2011) including reduced motor inhibition (Chamberlain et al., 2006, 2008) early temporal discounting of future rewards, insensitivity towards negative relative to positive outcomes (Frank et al., 2007b; Cockburn and Holroyd, 2010), or an inability to adjust the decision threshold appropriately (Mulder et al., 2010; Cavanagh et al., 2011; Frank et al., 2007a). Ultimately, the hope is to find a way to describe and diagnose psychiatric disease based on objective computational neurocognitive markers rather than the current subjective symptom-based approach. I believe that this combination of computational tools described here-within is powerful enough to lead the charge towards a new level of understanding of
mental illness based on identifiable and reproducible neurocognitive computational multi-dimensional features.

5 Appendix

The following serves as a reference for the mathematical details of the methods motivated above.

5.1 Drift-Diffusion Model

Mathematically, the DDM is defined by a stochastic differential equation called the Wiener process with drift:

$$dW \sim \mathcal{N}(v, \sigma^2)$$  \hspace{1cm} (7)

where \(v\) represents the drift-rate and \(\sigma\) the variance. As we often only observe the response times of subjects we are interested in the wiener first passage time (wfpt) – the time it takes \(W\) to cross one of two boundaries. Assuming two absorbing boundaries of this process and through some fairly sophisticated math (see e.g. Smith, 2000) it is possible to analytically derive the time this process will first pass one of the two boundaries (i.e. the wiener first passage time; wfpt). This probability distribution\(^2\) then serves as the likelihood function for the DDM.

5.2 Bayesian Inference

5.2.1 Empirical Bayesian Approximation

Empirical Bayes can be regarded as an approximation of equation (6). To derive this approximation consider \(P(\theta|x)\) which we can calculate by integrating over \(P(\lambda)\):

$$P(\theta|x) = \frac{P(x|\theta)}{P(x)} \int P(\theta|\lambda)P(\lambda) \, d\lambda$$  \hspace{1cm} (8)

Now, if the true distribution \(P(\theta|\lambda)\) is sharply peaked, the integral can be replaced with the point estimate of its peak \(\lambda^*\):

\(^2\)the wfpt will not be a distribution rather than a single value because of the stochasticity of the wiener process
\[ P(\theta|x) \approx \frac{P(x|\theta)P(\theta|\lambda^*)}{P(x|\lambda^*)} \quad (9) \]

Note, however, that \( \lambda^* \) depends itself on \( P(\theta|x) \). One algorithm to solve this interdependence is Expectation Maximization (EM) (Dempster et al., 1977). EM is an iterative algorithm that alternates between computing the expectation of \( P(\theta|x) \) (this can be easily done by Laplace Approximation (Azevedo-filho and Shachter, 1994)) and then maximizing the prior point estimate \( \lambda^* \) based on the current values obtained by the expectation step. This updated point estimate is then used in turn to recompute the expectation. The algorithm is run until convergence or some other criterion in reached. This approach is used for example by Huys et al. (2012) to fit their reinforcement learning models.

5.2.2 Markov-Chain Monte-Carlo

As mentioned above, the posterior is often intractable to compute analytically. While Empirical Bayes provides a useful approximation, an alternative approach is to estimate the full posterior by drawing samples from it. One way to achieve this is to construct a Markov-Chain that has the same equilibrium distribution as the posterior (Gamerman and Lopes, 2006). Algorithms of this class are called Markov-Chain Monte Carlo (MCMC) samplers.

One common and widely applicable algorithm is Metropolis-Hastings (Chib and Greenberg, 1995; Andrieu et al., 2003). Assume we wanted to generate samples \( \theta \) from the posterior \( p(\theta|x) \). In general, we can not sample from \( p(\theta|x) \) directly. Metropolis-Hastings instead generates samples \( \theta^t \) from a proposal distribution \( q(\theta^t|\theta^{t-1}) \) where the next position \( \theta^t \) only depends on the previous position at \( \theta^{t-1} \) (i.e. the Markov-property). For simplicity we will assume that this proposal distribution is symmetrical; i.e. \( q(\theta^t|\theta^{t-1}) = q(\theta^{t-1}|\theta^t) \). A common choice for the proposal distribution is the Normal distribution, formally:

\[ \theta^t \sim \mathcal{N}(\theta^{t-1}, \sigma^2) \quad (10) \]

The proposed jump to \( \theta^t \) is then accepted with probability \( \alpha \):

\[ \alpha = \min(1, \frac{p(\theta^t|x)}{p(\theta^{t-1}|x)}) \quad (11) \]
In other words, the probability of accepting a jump depends on the probability ratio of the proposed jump position $\theta^t$ to the previous position $\theta^{t-1}$. Critically, in this probability ratio, the intractable integral in the denominator (i.e. $p(x) = \int p(x|\theta) \, d\theta$) cancels out. This can be seen by applying Bayes formula (1):

$$
p(\theta^t | x) \frac{p(\theta^t)}{p(\theta^{t-1})} = \frac{p(x|\theta^t)p(\theta^t)}{p(x|\theta^{t-1})p(\theta^{t-1})}
$$

(12)

Thus, to calculate the probability of accepting a jump we only have to evaluate the likelihood and prior, not the intractable posterior.

Note that $\theta^0$ has to be initialized at some position and can not directly be sampled from the posterior. From this initial position, the Markov chain will explore other parts of the parameter space and only gradually approach the posterior region. The first samples generated are thus not from the true posterior and are often discarded as “burn-in”. Note moreover that once the algorithm reaches a region of high probability it will continue to explore lower probability regions in the posterior, albeit with lower frequency. This random-walk behavior is due to the probability ratio $\alpha$ which allows Metropolis-Hastings to also sometimes accept jumps from a high probability position to a low probability position.

Another common algorithm is Gibbs sampling that iteratively updates each individual random variable conditional on the other random variables set to their last sampled value (e.g. Frey and Jojic). Starting at some configuration $\theta^0$, the algorithm makes $T$ iterations over each random variable $\theta_i$. At each iteration $t$ each random variable is sampled conditional on the current $(t-1)$ value of all other random variables that it depends on:

$$
\theta_i^t \sim p(\theta_i^t | \theta_{i\neq j}^{(t-1)})
$$

(13)

Critically, $\theta_{i\neq j}^{(t-1)}$ are treated as constant. The sampled value of $\theta_i^{(t)}$ will then be treated as fixed while sampling the other random variables.

Note that while Gibbs sampling never rejects a sample (which often leads to faster convergence and better mixing), in contrast to Metropolis-Hastings, it does require sampling from the conditional distribution which is not always tractable.
Several likelihood-free methods have emerged in the past (for a review, see Turner and Van Zandt 2012). Instead of an analytical solution of the likelihood function, these methods require a sampling process that can simulate a set of data points from a generative model for each $\theta$. We will call the simulated data $y$ and the observed data $x$. Approximate Bayesian Computation (ABC) relies on a distance measure $\rho(x, y)$ that compares how similar the simulated data $y$ is to the observed data $x$ (commonly, this distance measure relies on summary statistics). We can then use the Metropolis-Hastings algorithm introduced before and change the acceptance ratio $\alpha$ (11) to use $\rho(x, y)$ instead of a likelihood function.

$$\alpha = \begin{cases} \min(1, \frac{p(\theta^t)}{p(\theta^{t-1})}) & \text{if } \rho(x, y) \leq \epsilon_0 \\ 0 & \text{if } \rho(x, y) \geq \epsilon_0 \end{cases}$$

where $\epsilon_0$ is an acceptance threshold. Large $\epsilon_0$ will result in higher proposal acceptance probability but a worse estimation of the posterior while small $\epsilon_0$ will lead to better posterior estimation but slower convergence.

An alternative approach to ABC is to construct a synthetic likelihood function based on summary statistics (Wood 2010). Specifically, we sample $N_r$ multiple data sets $y_1, \ldots, y_r$ from the generative process. We then compute summary statistics $s_1, \ldots, s_r$ for each simulated data set. Based on these summary statistics we then construct the synthetic likelihood function to evaluate $\theta$ (see figure 11 for an illustration):

$$p(x|\theta) \simeq \mathcal{N}(S(x); \mu_\theta, \Sigma_\theta)$$

This synthetic likelihood function based on summary statistics can then be used as a drop-in replacement for e.g. the Metropolis-Hastings algorithm outlined above.

*The summary statistics must (i) be sufficient and (ii) normally distributed*
Figure 11: Construction of a synthetic likelihood. To evaluate parameter vector $\theta$, $N_r$ data sets $y_1, \ldots, y_{N_r}$ are sampled from the generative model. On each sampled data set summary statistics $s_1, \ldots, s_{N_r}$ are computed. Based on these summary statistics a multivariate normal distribution is approximated with mean $\mu_\theta$ and covariance matrix $\Sigma_\theta$. The likelihood is approximated by evaluating summary statistics of the actual data on the log normal distribution with the estimated $\mu_\theta$ and $\Sigma_\theta$. Reproduced from [Wood 2010].
5.4 Model Comparison

5.4.1 Deviance Information Criterion

The Deviance Information Criterion (DIC) is a measure which trades off model complexity and model fit (Spiegelhalter et al. 2002). Several similar measures exist such as Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC). However, both these measures use the number of parameters as a proxy for model complexity. While a reasonable approximation to the complexity of non-hierarchical models, the relationship between model parameters (some of which are latent) and complexity in hierarchical models is more intricate. The DIC measure instead infers the number of parameters from the posterior. The DIC is computed as follows:

\[
DIC = \bar{D} + pD
\]

(16)

where

\[
pD = \bar{D} - \hat{D}
\]

(17)

\(\bar{D}\) is the posterior mean of the deviance (i.e. \(-2 \times \log(likelihood)\)) and \(\hat{D}\) is a point estimate of the deviance obtained by substituting in the posterior means. Loosely, \(\bar{D}\) represents how well the model fits the data on average while \(\hat{D}\) captures the deviance at the best fitting parameter combination. \(pD\) then acts as a measure related to the posterior variability and used as a proxy for the effective number of parameters. Complex models with many parameters will tend to have higher posterior variability and thus result in increased \(pD\) penalization.

Note that the only parameters that affect \(\hat{D}\) directly in our hierarchical model (equation 6) are the subject parameters \(\theta_i\). Thus, DIC estimates model fit based on how well individual subjects explain the observed data.

5.4.2 BIC

The Bayesian Information Criterion (BIC) is defined as follows:
BIC = \(-2 \log p(x|\hat{\theta}^{ML}) + k \log(n)\)  

(18)

where \(k\) is the number of free parameters, \(n\) is the number of data points, \(x\) is the observed data and \(\log p(x|k)\) is the likelihood of the parameters given the data (Schwarz, 1978).

While BIC can not directly be applied to hierarchical models (as outlined above), it is possible to integrate out individual subject parameters (e.g. Huys et al., 2012):  

\[
\log p(x|\hat{\theta}^{ML}) = \sum_i \log \int p(x_i|h)p(h|\hat{\theta}^{ML}) \, dh
\]

(19)

where \(x_i\) is the data belonging to the \(i\)th subject. The resulting score is called integrated BIC.

Since the subject parameters are integrated out, integrated BIC estimates how well the group parameters are able to explain the observed data.

5.4.3 Bayes Factor

Another measure to compare two models is the Bayes Factor (BF) (Kass and Raftery, 1995). It is defined as the ratio between the marginal model probabilities of the two models:

\[
BF = \frac{p(x|M_1)}{p(x|M_2)} = \frac{\int p(\theta_1|M_1)p(x|\theta_1, M_1) \, d\theta_1}{\int p(\theta_2|M_2)p(x|\theta_2, M_2) \, d\theta_2}
\]

(20)

The magnitude of this ratio informs the degree one should believe in one model compared to the other.

As BF integrates out subject and group parameters this model comparison measure should be used when different classes of models are to be compared in their capacity to explain observed data.
5.5 Mixture Models

5.5.1 Gaussian Mixture Models

Mixture models infer \( k \) number of clusters in a data set. The assumption of normal distributed clusters leads to a Gaussian Mixture Model (GMM) with a probability density function as follows:

\[
p(x|\pi, \mu_1, \ldots, \mu_k, \sigma_1^2, \ldots, \sigma_k^2) = \sum_{k=1}^{K} \pi_k N(x_i|\mu_k, \sigma_k^2)
\]  

(21)

Each observed data point \( x_i \) can be created by drawing a sample from the normal distribution selected by the unobserved indicator variable \( z_i \) which itself is distributed according to a multinomial distribution \( \pi \):

\[
\mu_k, \sigma_k \sim G_0()
\]  

(22)

\[
z_i \sim \pi
\]  

(23)

\[
x_i \sim N(\mu_{z_i}, \sigma_{z_i}^2)
\]  

(24)

where the base measure \( G_0 \) defines the prior for \( \mu_k \) and \( \sigma_k \). To simplify the inference it is often advisable to use a conjugate prior for these parameters. For example, the normal distribution is the conjugate prior for a normal distribution with known variance:

\[
\mu_k \sim N(\mu_0, \sigma_0)
\]  

(25)

In a similar fashion, we can assign the mixture weights a symmetric Dirichlet prior:

\[
\pi \sim Dir(\frac{\alpha}{K}, \ldots, \frac{\alpha}{K})
\]  

(26)

Note that the GMM assumes a mixture distribution on the level of the observed data \( x_i \). However, in our relevant case of a multi-level hierarchical model we need to place the mixture at the level of the latent subject parameters instead of the observed data. As before, we use the subject index
\(j = 1, \ldots, N.\)

\[
\begin{align*}
\mu_k, \sigma_k & \sim G_0() \quad (27) \\
\pi & \sim \text{Dir}(\alpha) \quad (28) \\
z_j & \sim \text{Categorical}(\pi) \quad (29) \\
\theta_j & \sim \mathcal{N}(\mu_{z_j}, \sigma_{z_j}^2) \quad (30) \\
x_{i,j} & \sim f(\theta_j) \quad (31)
\end{align*}
\]

Where \(f\) denotes the likelihood function.

Interestingly, the famous K-Means clustering algorithm is identical to a Gaussian Mixture Model (GMM) in the limit \(\sigma^2 \to 0\) [Kulis et al., 2012]. K-Means is an expectation maximization (EM) algorithm that alternates between an expectation step during which data points are assigned to their nearest cluster centroids and a maximization step during which new cluster centroids are estimated. This algorithm is repeated until convergence is reached (i.e. no points are reassigned to new clusters).

### 5.5.2 Dirichlet Process Gaussian Mixture Models

\[
p(x|\pi, \mu_1, \ldots, \mu_\infty, \sigma_1, \ldots, \sigma_\infty) = \sum_{k=1}^{\infty} \pi_k \mathcal{N}(x_i|\mu_k, \sigma_k^2) \quad (32)
\]

As above, we specify our generative mixture model:

\[
\begin{align*}
\mu_k, \sigma_k & \sim G_0() \quad (33) \\
z_i & \sim \text{Categorical}(\pi) \quad (34) \\
x_i & \sim \mathcal{N}(\mu_{z_i}, \sigma_{z_i}^2) \quad (35)
\end{align*}
\]

with the critical difference of replacing the hyperprior \(\pi\) with the *stick breaking process* [Sethuraman](#).
The stick-breaking process is a realization of a Dirichlet process (DP). Specifically, $\pi = \{\pi_k\}_{k=1}^{\infty}$ is an infinite sequence of mixture weights derived from the following process:

$$\beta_k \sim \text{Beta}(1, \alpha)$$

$$\pi_k \sim \beta_k \cdot \prod_{i=1}^{k-1} (1 - \beta_i)$$

with $\alpha > 0$. See figure 12 for a visual explanation.

The Chinese Restaurant Process (CRP) – named after the apparent infinite seating capacity in Chinese restaurants – allows for a more succinct model formulation. Consider that customers $z_i$ are coming into the restaurant and are seated at table $k$ with probability:

$$\pi \sim \text{StickBreaking}(\alpha)$$

(36)

The stick-breaking process is a realization of a Dirichlet process (DP). Specifically, $\pi = \{\pi_k\}_{k=1}^{\infty}$ is an infinite sequence of mixture weights derived from the following process:

$$\beta_k \sim \text{Beta}(1, \alpha)$$

$$\pi_k \sim \beta_k \cdot \prod_{i=1}^{k-1} (1 - \beta_i)$$

with $\alpha > 0$. See figure 12 for a visual explanation.

The Chinese Restaurant Process (CRP) – named after the apparent infinite seating capacity in Chinese restaurants – allows for a more succinct model formulation. Consider that customers $z_i$ are coming into the restaurant and are seated at table $k$ with probability:
\[ p(z_i = k|z_1,\ldots,n-1,\alpha,K) = \frac{n_k + \alpha/K}{n - 1 + \alpha} \]

where \( k = 1\ldots K \) is the table and \( n_k \) is the number of customers already sitting at table \( k \) (see figure 13 for an illustration). It can be seen that in the limit as \( K \to \infty \) this expression becomes:

\[ p(z_i = k|z_1,\ldots,n-1,\alpha) = \frac{n_k}{n - 1 + \alpha} \]

Thus, as customers are social, the probability of seating customer \( z_i \) to table \( k \) is proportional the number of customers already sitting at that table. This desirable clustering property is also known as the “rich get richer”.

Note that for an individual empty table \( k \) at which no customer has been seated (i.e. \( n_k = 0 \)) the probability of seating a new customer to that table goes to 0 in the limit as \( K \to \infty \). However, at the same time the number of empty tables approaches infinity. Consider that we have so far seated \( L \) customers to tables and the set \( Q \) contains all empty tables such that there are \( |Q| = K - L \) empty tables in the restaurant. The probability of seating a customer \( z_i \) at an empty table becomes:

\[ p(z_i \in Q|z_1,\ldots,n-1,\alpha) = \frac{\alpha}{n - 1 + \alpha} \]

As can be seen, the probability of starting a new table is proportional to the concentration parameter \( \alpha \). Intuitively, large values of the dispersion parameter \( \alpha \) lead to more clusters being used.

Thus, while the Stick-Breaking process sampled mixture weights from which we had to infer cluster assignments, the CRP allows for direct sampling of cluster assignments. The resulting model can then be written as:

\[ \mu_k, \sigma_k \sim G_0() \quad (39) \]
\[ z_1,\ldots,N \sim \text{CRP}(\alpha) \quad (40) \]
\[ x_i \sim \mathcal{N}(\mu_{z_i},\sigma_{z_i}^2) \quad (41) \]
Figure 13: Illustration of the Chinese Restaurant Process. Customers are seated at tables with parameters $\theta$. The more customers are already seated at a table, the higher the probability that future customers are seated at the same table (i.e. clustering property). Taken from Gershman and Blei (2012).

Finally, in a hierarchical group model we would need to place the infinite mixture on the subject level rather than the observed data level:

\[
\begin{align*}
\mu_k, \sigma_k & \sim G_0() \quad (42) \\
z_j & \sim \text{CRP}(\alpha) \quad (43) \\
\theta_j & \sim \mathcal{N}(\mu_{z_j}, \sigma_{z_j}^2) \quad (44) \\
x_{i,j} & \sim F(\theta_j) \quad (45)
\end{align*}
\]

See figure 14 for a graphical model description.

Note that while the potential number of clusters is infinite, any realization of this process will always lead to a finite number of clusters as we always have finite amounts of data. However, this method allows the addition (or subtraction) of new clusters as new data becomes available.

References

Désirée S Aichert, Nicola M Wöstmann, Anna Costa, Christine Macare, Johanna R Wenig, Hans-Jürgen Möller, Katya Rubia, and Ulrich Ettinger. Associations between trait impulsivity and prepotent response inhibition. *Journal of clinical and experimental neuropsychology, 00*(00), August 2012. ISSN 1744-411X. doi: 10.1080/13803395.2012.706261. URL [http://www.ncbi.nlm.nih.gov/pubmed/22888795](http://www.ncbi.nlm.nih.gov/pubmed/22888795).
Figure 14: Graphical model representation of the hierarchical Dirichlet process mixture model. Group parameters $\lambda_k = (\mu_k, \sigma_k)$. See text for details.

NC Andreasen. DSM and the death of phenomenology in America: an example of unintended consequences. *Schizophrenia Bulletin*, 2007. URL [http://schizophreniabulletin.oxfordjournals.org/content/33/1/108.short](http://schizophreniabulletin.oxfordjournals.org/content/33/1/108.short).

C Andrieu, N De Freitas, A Doucet, and MI Jordan. An introduction to MCMC for machine learning. *Machine learning*, 2003. URL [http://www.springerlink.com/index/xh62794161k70540.pdf](http://www.springerlink.com/index/xh62794161k70540.pdf).

CE Antoniak. Mixtures of Dirichlet processes with applications to Bayesian nonparametric problems. *The annals of statistics*, 1974. URL [http://www.jstor.org/stable/10.2307/2958336](http://www.jstor.org/stable/10.2307/2958336).

Adriano Azevedo-filho and Ross D Shachter. Laplace’s Method Approximations for Probabilistic Inference in Belief Networks with Continuous Variables. pages 28–36, 1994.

J. C. Badcock, P. T. Michie, L. Johnson, and J. Combrinck. Acts of control in schizophrenia: dissociating the components of inhibition. *Psychological medicine*, 32(2):287–297, February 2002. ISSN 0033-2917. URL [http://view.ncbi.nlm.nih.gov/pubmed/11866323](http://view.ncbi.nlm.nih.gov/pubmed/11866323).

A D Baddeley. The capacity for generating information by randomization. *The Quarterly journal of*
A. G. Barto. Adaptive critics and the basal ganglia. In J. C. Houk, J. L. Davis, and D. G. Beiser, editors, *Models of Information Processing in the Basal Ganglia*, pages 215–232. MIT Press, Cambridge, MA, December 1995.

Hannah M Bayer, Brian Lau, and Paul W Glimcher. Statistics of midbrain dopamine neuron spike trains in the awake primate. *J Neurophysiol*, 98(3):1428–1439, Sep 2007. URL http://dx.doi.org/10.1152/jn.01140.2006.

E Becker and M Rinck. Sensitivity and response bias in fear of spiders. *Cognition and Emotion*, 2004.

M. A. Bellgrove, C. D. Chambers, A. Vance, N. Hall, M. Karamitsios, and J. L. Bradshaw. Lateralized deficit of response inhibition in early-onset schizophrenia. *Psychological medicine*, 36(4):495–505, April 2006. ISSN 0033-2917. doi: 10.1017/S0033291705006409. URL http://dx.doi.org/10.1017/S0033291705006409.

Rafal Bogacz, Eric Brown, Jeff Moehlis, Philip Holmes, and Jonathan D Cohen. 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, 10 2006. URL http://www.ncbi.nlm.nih.gov/pubmed/17014301.

Denny Borsboom, Angélique O J Cramer, Verena D Schmittmann, Sacha Epskamp, and Lourens J Waldorp. The small world of psychopathology. *PloS one*, 6(11):e27407, January 2011. ISSN 1932-6203. doi: 10.1371/journal.pone.0027407. URL http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3219664&tool=pmcentrez&rendertype=abstract.

Scott D Brown and Andrew Heathcote. The simplest complete model of choice response time: linear ballistic accumulation. *Cognitive psychology*, 57(3):153–78, November 2008. ISSN 1095-5623. doi: 10.1016/j.cogpsych.2007.12.002. URL http://www.ncbi.nlm.nih.gov/pubmed/18243170.

Joshua W. Buckholtz and Andreas Meyer-Lindenberg. Psychopathology and the Human Connectome: Toward a Transdiagnostic Model of Risk For Mental Illness. *Neuron*, 74(6):990–1004, June 2012.
PW Burgess. Theory and methodology in executive function research. Methodology of frontal and executive function, 1997. URL http://books.google.com/books?hl=en&lr=&id=L-q3q0Xo0eEC&oi=fnd&pg=PA79&dq=Theory+and+methodology+in+executive+function+research&ots=x16CN9B8wY&sig=JYOOHw0IaSkBLhdrcMhsnVVFejU

James F Cavanagh, Thomas V Wiecki, Michael X Cohen, Christina M Figueroa, Johan Samanta, Scott J Sherman, and Michael J Frank. Subthalamic nucleus stimulation reverses mediofrontal influence over decision threshold. Nature neuroscience, 14:1462–1467, Sep 2011. URL http://www.ncbi.nlm.nih.gov/pubmed/21946325

James F Cavanagh, Laura Zambrano-Vazquez, and John J B Allen. Theta lingua franca: a common mid-frontal substrate for action monitoring processes. Psychophysiology, 49(2):220–38, February 2012. ISSN 1540-5958. doi: 10.1111/j.1469-8986.2011.01293.x. URL http://www.ncbi.nlm.nih.gov/pubmed/22091878

Samuel R. Chamberlain, Naomi A. Fineberg, Andrew D. Blackwell, Trevor W. Robbins, and Barbara J. Sahakian. Motor inhibition and cognitive flexibility in obsessive-compulsive disorder and trichotillomania. Am J Psychiatry, 163(7):1282–1284, July 2006. doi: 10.1176/appi.ajp.163.7.1282. URL http://dx.doi.org/10.1176/appi.ajp.163.7.1282

Samuel R Chamberlain, Lara Menzies, Adam Hampshire, John Suckling, Naomi A Fineberg, Natalia del Campo, Mike Aitken, Kevin Craig, Adrian M Owen, Edward T Bullmore, Trevor W Robbins, and Barbara J Sahakian. Orbitofrontal dysfunction in patients with obsessive-compulsive disorder and their unaffected relatives. Science, 321, Jul 2008. URL http://www.ncbi.nlm.nih.gov/pubmed/18635808

S Chib and E Greenberg. Understanding the metropolis-hastings algorithm. The American Statistician, 1995. URL http://amstat.tandfonline.com doi/abs/10.1080/00031305.1995.10476177

Ivar A H Clemens, Maaike De Vrijer, Luc P J Selen, Jan A M Van Gisbergen, and W Pieter Medendorp. Multisensory processing in spatial orientation: an inverse probabilistic approach. The Journal of
neuroscience : the official journal of the Society for Neuroscience, 31(14):5365–77, April 2011. ISSN 1529-2401. doi: 10.1523/JNEUROSCI.6472-10.2011. URL http://www.jneurosci.org/content/31/14/5365.short

Jeffrey Cockburn and Clay B Holroyd. Focus on the positive: computational simulations implicate asymmetrical reward prediction error signals in childhood attention-deficit/hyperactivity disorder. Brain research, 1365:18–34, December 2010. ISSN 1872-6240. doi: 10.1016/j.brainres.2010.09.065. URL http://www.ncbi.nlm.nih.gov/pubmed/20875804.

Michael X. Cohen and Michael J. Frank. Neurocomputational models of basal ganglia function in learning, memory and choice. Behavioural Brain Research, 199(1):141–156, Apr 2009. URL http://www.ncbi.nlm.nih.gov/pubmed/18950662.

Fabienne Collette, Martial Van der Linden, Steven Laureys, Guy Delfiore, Christian Degueldre, Andre Luxen, and Eric Salmon. Exploring the unity and diversity of the neural substrates of executive functioning. Human brain mapping, 25(4):409–23, August 2005. ISSN 1065-9471. doi: 10.1002/hbm.20118. URL http://www.ncbi.nlm.nih.gov/pubmed/15852470.

Anne G E Collins and Michael J Frank. How much of reinforcement learning is working memory, not reinforcement learning? A behavioral, computational, and neurogenetic analysis. The European journal of neuroscience, 35(7):1024–35, April 2012. ISSN 1460-9568. doi: 10.1111/j.1460-9568.2011.07980.x. URL http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3390186&tool=pmcentrez&rendertype=abstract.

Roshan Cools. Dopaminergic modulation of cognitive function-implications for l-dopa treatment in parkinson’s disease. Neuroscience and biobehavioral reviews, 30(1):1–23, 12 2005. URL http://www.ncbi.nlm.nih.gov/pubmed/15935475.

Daniel Cressey. Psychopharmacology in crisis. Nature, June 2011. ISSN 1476-4687. doi: 10.1038/news.2011.367. URL http://www.nature.com/news/2011/110614/full/news.2011.367.html.

Claudio Da Cunha, Evellyn Claudia Wietzkoski, Patricia Dombrowski, Mariza Bortolanza, Lucília Mendes Santos, Suelen Lucio Boschen, and Edmar Miyoshi. Learning processing in the basal ganglia: a mosaic of broken mirrors. Behavioural brain research, 199(1):157–170, Apr 2009. URL http://www.ncbi.nlm.nih.gov/pubmed/18977393.
BN Cuthbert and TR Insel. Toward new approaches to psychotic disorders: the NIMH Research Domain Criteria project. *Schizophrenia bulletin*, 2010. URL http://schizophreniabulletin.oxfordjournals.org/content/36/6/1061.short

Jeffrey W Dalley, Barry J Everitt, and Trevor W Robbins. Impulsivity, compulsivity, and top-down cognitive control. *Neuron*, 69(4):680–94, February 2011. ISSN 1097-4199. doi: 10.1016/j.neuron.2011.01.020. URL http://www.ncbi.nlm.nih.gov/pubmed/21338879

AP Dempster, NM Laird, and DB Rubin. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society*. . . ., 39(1):1–38, 1977. URL http://www.jstor.org/stable/10.2307/2984875

Chris Donkin, Scott Brown, Andrew Heathcote, and Eric-Jan Wagenmakers. Diffusion versus linear ballistic accumulation: different models but the same conclusions about psychological processes? *Psychonomic bulletin & review*, 18(1):61–9, February 2011. ISSN 1531-5320. doi: 10.3758/s13423-010-0022-4. URL http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3042112&tool=pmcentrez&rendertype=abstract

John Duncan, Roger Johnson, and Charles Freer. Frontal love deficits after head injury: Unity and diversity of function. *Cognitive Neuropsychology*, 14:713, January 1997.

Sarah Durston, John A Fossella, Martijn J Mulder, B J Casey, Tim B Ziemans, M Nathalie Vessaz, and Herman Van Engeland. Dopamine transporter genotype conveys familial risk of attention-deficit/hyperactivity disorder through striatal activation. *Journal of the American Academy of Child and Adolescent Psychiatry*, 47(1):61, Jan 2008. URL http://www.ncbi.nlm.nih.gov/pubmed/18174826

D. a. Fair, D. Bathula, M. a. Nikolas, and J. T. Nigg. Distinct neuropsychological subgroups in typically developing youth inform heterogeneity in children with ADHD. *Proceedings of the National Academy of Sciences*, April 2012. ISSN 0027-8424. doi: 10.1073/pnas.1115365109. URL http://www.pnas.org/cgi/doi/10.1073/pnas.1115365109

Birte U Forstmann, Alfred Anwander, Andreas Schfer, Jane Neumann, Scott Brown, Eric-Jan Wagenmakers, Rafal Bogacz, and Robert Turner. Cortico-striatal connections predict control over speed
and accuracy in perceptual decision making. *Proceedings of the National Academy of Sciences of the United States of America*, pages 1–5 online, Aug 2010a. URL http://www.ncbi.nlm.nih.gov/pubmed/20733082.

Birte U. Forstmann, Scott Brown, Gilles Dutilh, Jane Neumann, and Eric-Jan Wagenmakers. The neural substrate of prior information in perceptual decision making: a model-based analysis. *Frontiers in human neuroscience*, 4:1–12 (online), 2010b. URL http://www.ncbi.nlm.nih.gov/pubmed/20577592.

M. J. Frank. Dynamic dopamine modulation in the basal ganglia: A neurocomputational account of cognitive deficits in medicated and non-medicated Parkinsonism. *Journal of Cognitive Neuroscience*, 17:51–72, January 2005.

M. J. Frank, L. C. Seeberger, and R. C. O’Reilly. By carrot or by stick: Cognitive reinforcement learning in Parkinsonism. *Science*, 306(5703):1940–1943, January 2004.

M. J. Frank, B.B Doll, J. Oas-Terpstra, and F. Moreno. Prefrontal and striatal dopaminergic genes predict individual differences in exploration and exploitation. *Nature Neuroscience*, 12:1062–1068, 2009.

Michael J Frank. Hold your horses: a dynamic computational role for the subthalamic nucleus in decision making. *Neural networks : the official journal of the International Neural Network Society*, 19:1120–1136, 10 2006. URL http://www.ncbi.nlm.nih.gov/pubmed/16945502.

Michael J. Frank and Eric D. Claus. Anatomy of a decision: Striato-orbitofrontal interactions in reinforcement learning, decision making, and reversal. *Psychological Review*, 113(2):300–326, 04 2006. URL http://www.ncbi.nlm.nih.gov/pubmed/16637763.

Michael J Frank, Johan Samanta, Ahmed A Moustafa, and Scott J Sherman. Hold your horses: impulsivity, deep brain stimulation, and medication in parkinsonism. *Science (New York, N.Y.)*, 318:1309–1312, 11 2007a. URL http://www.ncbi.nlm.nih.gov/pubmed/17962524.

Michael J Frank, Amy Santamaria, Randall C O’Reilly, and Erik Willcutt. Testing computational models of dopamine and noradrenaline dysfunction in attention deficit/hyperactivity disorder. *Neu-
BJ Frey and N Jojic. A comparison of algorithms for inference and learning in probabilistic graphical models. Pattern Analysis and Machine Intelligence, ..., 2005. URL http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=1471706

M. Fukumoto-Motoshita, M. Matsuura, T. Ohkubo, H. Ohkubo, N. Kanaka, E. Matsushima, M. Taira, T. Kojima, and T. Matsuda. Hyperfrontality in patients with schizophrenia during saccade and antisaccade tasks: a study with fmri. Psychiatry and clinical neurosciences, 63(2):209–217, April 2009. ISSN 1440-1819. doi: 10.1111/j.1440-1819.2009.01941.x. URL http://dx.doi.org/10.1111/j.1440-1819.2009.01941.x

D Gamerman and HF Lopes. Markov chain Monte Carlo: stochastic simulation for Bayesian inference. 2006. URL http://books.google.com/books?hl=en&lr=&id=yPvECi_L3bwC&oi=fnd&pg=PR13&dq=gamerman+bayesian&ots=NisYwWffgg&sig=W9fgGzxMiklkMGA2fnwQACTz8BY

A Gelman, JB Carlin, HS Stern, and DB Rubin. Bayesian data analysis. 2003. URL http://books.google.com/books?hl=en&lr=&id=TNYhnkXQSjAC&oi=fnd&pg=PP1&dq=Gelman+Carlin+Stern+04&ots=5H4S8DAwH3&sig=W9fgGzxMiklkMGA2fnwQACTz8BY

Samuel J. Gershman and David M. Blei. A tutorial on Bayesian nonparametric models. Journal of Mathematical Psychology, 56(1):1–12, February 2012. ISSN 00222496. doi: 10.1016/j.jmp.2011.08.004. URL http://linkinghub.elsevier.com/retrieve/pii/S002224961100071X

James M Gold, James A Waltz, Kristen J Prentice, Sarah E Morris, and Erin A Heerey. Reward processing in schizophrenia: a deficit in the representation of value. Schizophrenia bulletin, 34:835–847, 08 2008. URL http://www.ncbi.nlm.nih.gov/pubmed/18591195

James M Gold, James A Waltz, Tatyana M Matveeva, Zuzana Kasanova, Gregory P Strauss, Ellen S Herbener, Anne G E Collins, and Michael J Frank. Negative Symptoms and the Failure to Represent the Expected Reward Value of Actions. 69(2):129–138, 2012.

Joshua I Gold and Michael N Shadlen. The neural basis of decision making. Annual review of neuroscience, 30, 2007. URL http://www.ncbi.nlm.nih.gov/pubmed/17600525
P E Hallett. Primary and secondary saccades to goals defined by instructions. *Vision research*, 18:1270–1296, 02 1979. URL http://www.ncbi.nlm.nih.gov/pubmed/726270

RW Heinrichs. *In search of madness: Schizophrenia and neuroscience*. 2001. URL http://books.google.com/books?hl=en&lr=&id=pIrY28CApGQC&oi=fnd&pg=PA3&dq=In+Search+of+Madness&ots=BFzEK9gERs&sig=Q400C7kQLo7ETJCjb02iY1g75os

V. C. Huddy, A. R. Aron, M. Harrison, T. R. E. Barnes, T. W. Robbins, and E. M. Joyce. Impaired conscious and preserved unconscious inhibitory processing in recent onset schizophrenia. *Psychological Medicine*, 39(06):907–916, 2009. doi: 10.1017/S0033291708004340. URL http://dx.doi.org/10.1017/S0033291708004340

Quentin J. M. Huys, Neir Eshel, Elizabeth O’Nions, Luke Sheridan, Peter Dayan, and Jonathan P. Roiser. Bonsai Trees in Your Head: How the Pavlovian System Sculpts Goal-Directed Choices by Pruning Decision Trees. *PLoS Computational Biology*, 8(3):e1002410, March 2012. ISSN 1553-7358. doi: 10.1371/journal.pcbi.1002410. URL http://dx.plos.org/10.1371/journal.pcbi.1002410

SE Hyman. The diagnosis of mental disorders: the problem of reification. *Annual Review of Clinical Psychology*, 2010. URL http://www.annualreviews.org/doi/abs/10.1146/annurev.clinpsy.3.022806.091532

Steven E Hyman. Revolution Stalled. 4(155):1–5, 2012.

T. Insel, B. Cuthbert, M. Garvey, R. Heinssen, D. S. Pine, K. Quinn, C. A. Sanislow, and P. W. Wang. Research domain criteria (RDoC): toward a new classification framework for research on mental disorders. *American Journal of Psychiatry*, 2010. URL http://works.bepress.com/charles_sanislow/2/

Thomas R Insel. Rethinking schizophrenia. *Nature*, 468(7321):187–93, November 2010. ISSN 1476-4687. doi: 10.1038/nature09552. URL http://dx.doi.org/10.1038/nature09552

Robert E. Kass and Adrian E. Raftery. Bayes Factors. *Journal of the American Statistical Association*, 90(430):773–795, June 1995. ISSN 0162-1459. doi: 10.1080/01621459.1995.10476572. URL http://www.tandfonline.com/doi/abs/10.1080/01621459.1995.10476572

54
R Kendell and A Jablensky. Distinguishing between the validity and utility of psychiatric diagnoses. *American journal of psychiatry*, 2003. URL `http://journals.psychiatryonline.org/article.aspx?articleid=175948`

K Kendler, D Kupfer, W Narrow, K Phillips, and J Fawcett. Guidelines for making changes to DSM-V. … manuscript, Washington, DC: …, 2009. URL `http://scholar.google.com/scholar?hl=en&q=Guidelines+for++making+changes+to+DSM-V.&btnG=&as_sdt=1,40&as_sdtp=#0`

S. J. Kish, K. Shannak, and O. Hornykiewicz. Uneven pattern of dopamine loss in the striatum of patients with idiopathic parkinson’s disease. *New England Journal of Medicine*, 318:876–880, January 1988.

I. Krajbich and A. Rangel. Multialternative drift-diffusion model predicts the relationship between visual fixations and choice in value-based decisions. *Proceedings of the National Academy of Sciences*, 108(33), August 2011. ISSN 0027-8424. doi: 10.1073/pnas.1101328108. URL `http://www.pnas.org/cgi/doi/10.1073/pnas.1101328108`

Robert F. Krueger and Kristian E. Markon. Understanding Psychopathology: Melding Behavior Genetics, Personality, and Quantitative Psychology to Develop an Empirically Based Model. *Current Directions in Psychological Science*, 15(3):113–117, June 2006. ISSN 0963-7214. doi: 10.1111/j.0963-7214.2006.00418.x. URL `http://cdp.sagepub.com/content/15/3/113.full`

Brian Kulis, Michael I Jordan, Jordan Eecs, and Berkeley Edu. Revisiting k-means: New Algorithms via Bayesian Nonparametrics. 2012.

David LaBerge. A recruitment theory of simple behavior. *Psychometrika*, 27(4):375–396, December 1962. ISSN 0033-3123. doi: 10.1007/BF02289645. URL `http://www.springerlink.com/index/10.1007/BF02289645`

T Ljungberg, P Apicella, and W Schultz. Responses of monkey dopamine neurons during learning of behavioral reactions. *Journal of Neurophysiology*, 67:145–163, 04 1992. URL `http://www.ncbi.nlm.nih.gov/pubmed/1552316`

Tiago V Maia and Michael J Frank. From reinforcement learning models to psychiatric and neurological
disorders. *Nature neuroscience*, 14(2):154–62, February 2011. ISSN 1546-1726. doi: 10.1038/nn.2723. URL [http://www.ncbi.nlm.nih.gov/pubmed/21270784](http://www.ncbi.nlm.nih.gov/pubmed/21270784).

GM Manguno-Mire, JI Constans, and JH Geer. Anxiety-related differences in affective categorizations of lexical stimuli. *Behaviour research and therapy*, 2005. URL [https://www.sciencedirect.com/science/article/pii/S000579670400049X](https://www.sciencedirect.com/science/article/pii/S000579670400049X).

KE Markon. Modeling psychopathology structure: a symptom-level analysis of Axis I and II disorders. *Psychological medicine*, 2010. URL [http://journals.cambridge.org/production/action/cjoGetFulltext?fulltextid=5830732](http://journals.cambridge.org/production/action/cjoGetFulltext?fulltextid=5830732).

L. Menzies, S. Achard, S. R. Chamberlain, N. Fineberg, C. H. Chen, N. del Campo, B. J. Sahakian, T. W. Robbins, and E. Bullmore. Neurocognitive endophenotypes of obsessive-compulsive disorder. *Brain: a journal of neurology*, 130(Pt 12):3223–3236, December 2007. ISSN 1460-2156. doi: 10.1093/brain/awm205. URL [http://dx.doi.org/10.1093/brain/awm205](http://dx.doi.org/10.1093/brain/awm205).

A Miyake, N P Friedman, M J Emerson, A H Witzki, A Howarter, and T D Wager. The unity and diversity of executive functions and their contributions to complex "frontal lobe” tasks: a latent variable analysis. *Cognitive psychology*, 41:49–100, 09 2000. URL [http://www.ncbi.nlm.nih.gov/pubmed/10945922](http://www.ncbi.nlm.nih.gov/pubmed/10945922).

P R Montague, P Dayan, and T J Sejnowski. A framework for mesencephalic dopamine systems based on predictive hebbian learning. *The Journal of Neuroscience*, 16:1936–1947, 01 1997. URL [http://www.ncbi.nlm.nih.gov/pubmed/8774460](http://www.ncbi.nlm.nih.gov/pubmed/8774460).

PR Read Montague, RJ Raymond J. Dolan, KJ Karl J. Friston, and Peter Dayan. Computational psychiatry. *Trends in Cognitive Sciences*, 16(1):1–9, December 2011. ISSN 13646613. doi: 10.1016/j.tics.2011.11.018. URL [http://linkinghub.elsevier.com/retrieve/pii/S1364661311002518](http://linkinghub.elsevier.com/retrieve/pii/S1364661311002518).

John R Monterosso, Adam R Aron, Xochitl Cordova, Jiansong Xu, and Edythe D London. Deficits in response inhibition associated with chronic methamphetamine abuse. *Drug and alcohol dependence*, 79(2):273–7, August 2005. ISSN 0376-8716. doi: 10.1016/j.drugalcdep.2005.02.002. URL [http://www.ncbi.nlm.nih.gov/pubmed/15967595](http://www.ncbi.nlm.nih.gov/pubmed/15967595).
S. Morein-Zamir, N. A. Fineberg, T. W. Robbins, and B. J. Sahakian. Inhibition of thoughts and actions in obsessive-compulsive disorder: extending the endophenotype? *Psychological medicine*, pages 1–10, July 2009. ISSN 1469-8978. doi: 10.1017/S003329170999033X. URL [http://dx.doi.org/10.1017/S003329170999033X](http://dx.doi.org/10.1017/S003329170999033X).

A. A. Moustafa, S. J. Sherman, and M. J. Frank. A dopaminergic basis for working memory, learning, and attentional shifting in parkinson’s disease. *Neuropsychologia*, 46:3144–3156, January 2008.

Martijn J Mulder, Dienke Bos, Juliette M H Weusten, Janna van Belle, Sarai C van Dijk, Patrick Simen, Herman van Engeland, and Sarah Durston. Basic impairments in regulating the speed-accuracy tradeoff predict symptoms of attention-deficit/hyperactivity disorder. *Biological psychiatry*, 68(12):1114–9, December 2010. ISSN 1873-2402. doi: 10.1016/j.biopsych.2010.07.031. URL [http://www.ncbi.nlm.nih.gov/pubmed/20926067](http://www.ncbi.nlm.nih.gov/pubmed/20926067).

D.J. Daniel J. Navarro and I.G. Ian G. Fuss. Fast and accurate calculations for first-passage times in Wiener diffusion models. *Journal of Mathematical Psychology*, 53(4):222–230, August 2009. ISSN 00222496. doi: 10.1016/j.jmp.2009.02.003. URL [http://linkinghub.elsevier.com/retrieve/pii/S0022249609000200](http://linkinghub.elsevier.com/retrieve/pii/S0022249609000200).

J. T. Nigg. Is adhd a disinhibitory disorder? *Psychological bulletin*, 127(5):571–598, September 2001. ISSN 0033-2909. URL [http://view.ncbi.nlm.nih.gov/pubmed/11548968](http://view.ncbi.nlm.nih.gov/pubmed/11548968).

Joel T Nigg, Maria M Wong, Michelle M Martel, Jennifer M Jester, Leon I Puttler, Jennifer M Glass, Kenneth M Adams, Hiram E Fitzgerald, and Robert A Zucker. Poor response inhibition as a predictor of problem drinking and illicit drug use in adolescents at risk for alcoholism and other substance use disorders. *Journal of the American Academy of Child and Adolescent Psychiatry*, 45(4):468–75, April 2006. ISSN 0890-8567. doi: 10.1097/01.chi.0000199028.76452.a9. URL [http://www.ncbi.nlm.nih.gov/pubmed/16601652](http://www.ncbi.nlm.nih.gov/pubmed/16601652).

Hå kan Nilsson, Jörg Rieskamp, and Eric-Jan Wagenmakers. Hierarchical Bayesian parameter estimation for cumulative prospect theory. *Journal of Mathematical Psychology*, 55(1):84–93, February 2011. ISSN 00222496. doi: 10.1016/j.jmp.2010.08.006. URL [http://linkinghub.elsevier.com/retrieve/pii/S0022249610001070](http://linkinghub.elsevier.com/retrieve/pii/S0022249610001070).

57
Julie Nordgaard, Louis A Sass, and Josef Parnas. The psychiatric interview: validity, structure, and subjectivity. *European archives of psychiatry and clinical neuroscience*, September 2012. ISSN 1433-8491. doi: 10.1007/s00406-012-0366-z. URL http://www.ncbi.nlm.nih.gov/pubmed/23001456.

David Nutt and Guy Goodwin. ECNP Summit on the future of CNS drug research in Europe 2011: report prepared for ECNP by David Nutt and Guy Goodwin. *European neuropsychopharmacology: the journal of the European College of Neuropsychopharmacology*, 21(7):495–9, July 2011. ISSN 1873-7862. doi: 10.1016/j.euroneuro.2011.05.004. URL http://www.ncbi.nlm.nih.gov/pubmed/21684455.

J Oosterlaan, G Logan, and J Sergeant. Response inhibition in AD/HD, CD, comorbid AD/HD+ CD, anxious, and control children: A metaanalysis of studies with the stop task. *Journal of Child...*, 1998. URL http://onlinelibrary.wiley.com/doi/10.1111/1469-7610.00336/abstract.

Randall C. O'Reilly and Yuko Munakata. *Computational Explorations in Cognitive Neuroscience: Understanding the Mind by Simulating the Brain*. The MIT Press, Cambridge, MA, January 2000.

Wei-Xing Pan, Robert Schmidt, Jeffery R Wickens, and Brian I Hyland. Dopamine cells respond to predicted events during classical conditioning: Evidence for eligibility traces in the reward-learning network. *Journal of Neuroscience*, 25(26):6235–6242, 06 2005. URL http://www.ncbi.nlm.nih.gov/pubmed/15987953.

R. Penadés, R. Catalán, K. Rubia, S. Andrés, M. Salamero, and C. Gastó. Impaired response inhibition in obsessive compulsive disorder. *European psychiatry: the journal of the Association of European Psychiatrists*, 22(6):404–410, September 2007. ISSN 0924-9338. doi: 10.1016/j.eurpsy.2006.05.001. URL http://dx.doi.org/10.1016/j.eurpsy.2006.05.001.

LH Phillips. Do frontal tests measure executive function? Issues of assessment and evidence from fluency tests. *Methodology of frontal and executive function*, 1997. URL http://books.google.com/books?hl=en&lr=&id=L-q3qOXo0eEC&oi=fnd&pg=PA185&dq=measure+executive+function?+Issues+of+assessment+and+evidence+from+fluency+tests&ots=x16CN9B7z_&sig=TEJBBby8J8R6mr89f2DoaGo8NvI8.

Dale J. Poirier. The growth of Bayesian methods in statistics and economics since 1970. *Bayesian Anal-
J Poland, B Von Eckardt, and W Spaulding. Problems with the DSM approach to classifying psychopathology. 1994. URL http://psycnet.apa.org/psycinfo/1995-97231-011

Jeffrey Poland and Barbara Von Eckardt. Mapping the Domain of Mental Illness. In The Oxford Handbook of Philosophy of Psychiatry. 2013.

Roger Ratcliff and Michael J. Frank. Reinforcement-based decision making in corticostriatral circuits: mutual constraints by neurocomputational and diffusion models. Neural Computation, 24(5):1186–1229, May 2012. URL http://www.ncbi.nlm.nih.gov/pubmed/22295983

Roger Ratcliff and Gail McKoon. The diffusion decision model: theory and data for two-choice decision tasks. Neural computation, 20:873–922, Apr 2008. URL http://www.ncbi.nlm.nih.gov/pubmed/18085991

Roger Ratcliff and Jeffrey N. Rouder. Modeling response times for two-choice decisions. Psychological Science, 9:347, January 1998.

Roger Ratcliff and Francis Tuerlinckx. Estimating parameters of the diffusion model: approaches to dealing with contaminant reaction times and parameter variability. Psychonomic bulletin & review, 9(3):438–81, September 2002. ISSN 1069-9384. URL http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2474747&tool=pmcentrez&rendertype=abstract

D Regier and W Narrow. The conceptual development of DSM-V. American Journal of . . ., 2009. URL http://journals.psychiatryonline.org/article.aspx?Volume=166&page=645&journalID=13

Arbora Resulaj, Roozbeh Kiani, Daniel M Wolpert, and Michael N Shadlen. Changes of mind in decision-making. Nature, 461(7261):263–266, 2009. ISSN 0028-0836. doi: 10.1038/nature08275. URL http://dx.doi.org/10.1038/nature08275

Trevor W Robbins, Claire M Gillan, Dana G Smith, Sanne de Wit, and Karen D Ersche. Neurocognitive endophenotypes of impulsivity and compulsivity: towards dimensional psychiatry. Trends in cognitive sciences, 16(1):81–91, January 2012. ISSN 1879-307X. doi: 10.1016/j.tics.2011.11.009. URL http://www.ncbi.nlm.nih.gov/pubmed/22155014
Matthew R Roesch, Donna J Calu, and Geoffrey Schoenbaum. Dopamine neurons encode the better option in rats deciding between differently delayed or sized rewards. *Nature Neuroscience*, 10(12): 1615–1624, 11 2007. URL [http://www.ncbi.nlm.nih.gov/pubmed/18026098](http://www.ncbi.nlm.nih.gov/pubmed/18026098)

Barbara J Sahakian, Gavin Malloch, and Christopher Kennard. A UK strategy for mental health and wellbeing. *Lancet*, 375(9729):1854–5, May 2010. ISSN 1474-547X. doi: 10.1016/S0140-6736(10)60817-3. URL [http://www.thelancet.com/journals/a/article/PIIS0140-6736(10)60817-3/fulltext](http://www.thelancet.com/journals/a/article/PIIS0140-6736(10)60817-3/fulltext)

Russell Schachar and Gordon D. Logan. Impulsivity and inhibitory control in normal development and childhood psychopathology. *Developmental Psychology*, 26(5):710–720, 1990. ISSN 0012-1649. doi: 10.1037/0012-1649.26.5.710. URL [http://dx.doi.org/10.1037/0012-1649.26.5.710](http://dx.doi.org/10.1037/0012-1649.26.5.710)

W Schultz, P Dayan, and P R Montague. A neural substrate of prediction and reward. *Science*, 275: 1593, 03 1997. URL [http://www.ncbi.nlm.nih.gov/pubmed/9054347](http://www.ncbi.nlm.nih.gov/pubmed/9054347)

Wolfram Schultz. Predictive reward signal of dopamine neurons. *Journal of neurophysiology*, 80:1–27, 08 1998. URL [http://www.ncbi.nlm.nih.gov/pubmed/9658025](http://www.ncbi.nlm.nih.gov/pubmed/9658025)

Gideon Schwarz. Estimating the Dimension of a Model. *The Annals of Statistics*, 6(2):461–464, March 1978. ISSN 2168-8966. URL [http://projecteuclid.org/euclid.aos/1176344136](http://projecteuclid.org/euclid.aos/1176344136)

J Sethuraman. A constructive definition of Dirichlet priors. 1991. URL [http://oai.dtic.mil/oai/oai?verb=getRecord&metadataPrefix=html&identifier=ADA238689](http://oai.dtic.mil/oai/oai?verb=getRecord&metadataPrefix=html&identifier=ADA238689)

RM Shiffrin, MD Lee, and W Kim. A survey of model evaluation approaches with a tutorial on hierarchical Bayesian methods. *Cognitive ...*, 2008. URL [http://onlinelibrary.wiley.com/doi/10.1080/03640210802414826/full](http://onlinelibrary.wiley.com/doi/10.1080/03640210802414826/full)

Philip L Smith and Roger Ratcliff. Psychology and neurobiology of simple decisions. *Trends in neurosciences*, 27(3):161–8, March 2004. ISSN 0166-2236. doi: 10.1016/j.tins.2004.01.006. URL [http://www.ncbi.nlm.nih.gov/pubmed/15036882](http://www.ncbi.nlm.nih.gov/pubmed/15036882)

Pl Smith. Stochastic Dynamic Models of Response Time and Accuracy: A Foundational Primer. *Journal of mathematical psychology*, 44(3):408–463, September 2000. ISSN 0022-2496. doi: 10.1006/jmps.1999.1260. URL [http://www.ncbi.nlm.nih.gov/pubmed/10973778](http://www.ncbi.nlm.nih.gov/pubmed/10973778)
Imri Sofer, Thomas V. Wiecki, and Michael J. Frank. Hierarchical Bayesian estimation of the drift diffusion model in practice.

Edmund J S Sonuga-Barke. Causal models of attention-deficit/hyperactivity disorder: from common simple deficits to multiple developmental pathways. *Biological psychiatry*, 57(11):1231–1238, 06 2005. URL http://www.ncbi.nlm.nih.gov/pubmed/15949993

DJ Spiegelhalter, NG Best, and Bradley P. Carlin. Bayesian measures of model complexity and fit. *Journal of the Royal ..., 2002*. URL http://onlinelibrary.wiley.com/doi/10.1111/1467-9868.00353/full

Matthew Stephens and David J Balding. Bayesian statistical methods for genetic association studies. *Nature reviews. Genetics*, 10(10):681–90, October 2009. ISSN 1471-0064. doi: 10.1038/nrg2615. URL http://dx.doi.org/10.1038/nrg2615

Gregory P. Strauss, Michael J. Frank, James a. Waltz, Zuzana Kasanova, Ellen S. Herbener, and James M. Gold. Deficits in Positive Reinforcement Learning and Uncertainty-Driven Exploration Are Associated with Distinct Aspects of Negative Symptoms in Schizophrenia. *Biological Psychiatry*, 69(5):424–431, March 2011. ISSN 00063223. doi: 10.1016/j.biopsych.2010.10.015. URL http://linkinghub.elsevier.com/retrieve/pii/S0006322310011066

R. S. Sutton and A. G. Barto. Time-derivative models of pavlovian reinforcement. In J. W. Moore and M. Gabriel, editors, *Learning and Computational Neuroscience*, pages 497–537. MIT Press, Cambridge, MA, January 1990.

R. S. Sutton and A. G. Barto. *Reinforcement Learning: An Introduction*. MIT Press, Cambridge, MA, January 1998.

James T. Townsend and F. Gregory Ashby. *The stochastic modeling of elementary psychological processes*. Cambridge University Press, 1983. ISBN 0521241812. doi: 102-212-801.

Brandon M. Turner and Trisha Van Zandt. A tutorial on approximate Bayesian computation. *Journal of Mathematical Psychology*, 56(2):69–85, April 2012. ISSN 00222496. doi: 10.1016/j.jmp.2012.02.005. URL http://linkinghub.elsevier.com/retrieve/pii/S0022249612000272
Martijn G. van Koningsbruggen, Tom Pender, Liana Machado, and Robert D. Rafal. Impaired control of the oculomotor reflexes in Parkinson’s disease. *Neuropsychologia*, June 2009. ISSN 1873-3514. doi: 10.1016/j.neuropsychologia.2009.06.018. URL http://dx.doi.org/10.1016/j.neuropsychologia.2009.06.018

D Vickers. Evidence for an accumulator model of psychophysical discrimination. *Ergonomics*, 13(1): 37–58, January 1970. ISSN 0014-0139. doi: 10.1080/00140137008931117. URL http://dx.doi.org/10.1080/00140137008931117

P Waelti, A Dickinson, and W Schultz. Dopamine responses comply with basic assumptions of formal learning theory. *Nature*, 412:43–48, 07 2001. URL http://www.ncbi.nlm.nih.gov/pubmed/11452299

James A Waltz, Michael J Frank, Benjamin M Robinson, and James M Gold. Selective reinforcement learning deficits in schizophrenia support predictions from computational models of striatal-cortical dysfunction. *Biological psychiatry*, 62:756–764, 09 2007. URL http://www.ncbi.nlm.nih.gov/pubmed/17300757

James A. Waltz, Michael J. Frank, Thomas V. Wiecki, and James M. Gold. Altered probabilistic learning and response biases in schizophrenia: behavioral evidence and neurocomputational modeling. *Neuropsychology*, 25(1):86–97, Jan 2011. URL http://www.ncbi.nlm.nih.gov/pubmed/21090899

C White. A Model-Based Analysis of Anxiety and Biased Processing of Threatening Information. 2009. URL http://kb.osu.edu/dspace/handle/1811/44499

CN White, R Ratcliff, MW Vasey, and G McKoon. Using diffusion models to understand clinical disorders. *Journal of Mathematical ...,*, 2010. URL http://www.sciencedirect.com/science/article/pii/S0022249610000155

Thomas V. Wiecki and Michael J. Frank. A computational model of inhibitory control in frontal cortex and basal ganglia. *Psychological Review.*, 2013.

Thomas V. Wiecki, Katrin Riedinger, Andreas von Ameln-Mayerhofer, Werner J. Schmidt, and Michael J. Frank. A neurocomputational account of catalepsy sensitization induced by D2 receptor
blockade in rats: context dependency, extinction, and renewal. *Psychopharmacology (Berl)*, 204(2): 265–277, Jun 2009. URL http://www.ncbi.nlm.nih.gov/pubmed/19169674.

T.V. Wiecki and M.J. Frank. Neurocomputational models of motor and cognitive deficits in Parkinson’s disease. In Anders Bjorklund and M. Angela Cenci, editors, *Progress in Brain Research: Recent Advances in Parkinson’s Disease - Part I: Basic Research*, volume 183, chapter 14, pages 275–297. Elsevier, 2010. URL http://www.ncbi.nlm.nih.gov/pubmed/20696325.

S Windmann and T Krüger. Subconscious detection of threat as reflected by an enhanced response bias. *Consciousness and Cognition*, 1998. URL https://www.sciencedirect.com/science/article/pii/S1053810098903373.

Simon N Wood. Statistical inference for noisy nonlinear ecological dynamic systems. *Nature*, 466 (7310):1102–4, August 2010. ISSN 1476-4687. doi: 10.1038/nature09319. URL http://www.ncbi.nlm.nih.gov/pubmed/20703226.