Computational cognitive methods for precision psychiatry

Quentin J.M. Huys[1,2]

1 Division of Psychiatry and Max Planck UCL Centre for Computational Psychiatry and Ageing Research, University College London, London, UK **2** Camden and Islington NHS Foundation Trust, London, UK

Corresponding Author
Quentin JM Huys
Division of Psychiatry and
Max Planck Centre for Computational Psychiatry and Ageing Research
University College London
10-12 Russell Square, London WC1B 5EH, UK
Email: qhuys@cantab.net

Word count
Abstract: 149
Total: 3460
Figures: 3
Box: 1
References:

Abstract

When an organ is unable to meet the demands placed on it, illness can arise. As the main functions of the brain are to compute and learn, an understanding of mental illnesses will benefit from an understanding of the computational and learning functions the brain performs, and how these are affected in states of ill-health. Tasks allow highly specific computational and learning processes to be probed. After a brief introduction into computational psychiatry more broadly, this chapter explores how tasks can be used for precision psychiatry. It is argued that identification of disease mechanisms via tasks can facilitate the development of targeted interventions and the targeted administration of therapies. However, the clinical use of tasks currently face issues around reliability and robustness. This chapter describes and discusses four reasons for this: time, strategy, noise and research setting, and describes how these are at least partially amenable to computational techniques.

1 INTRODUCTION

The burden of mental illnesses is large (WHO, 2017) and the treatment gap remains wide. While many different treatment approaches may exist, it is often hard to predict who will respond to which treatment, and the illness course is relapsing or chronic in a substantial fraction of those affected (e.g. Angst et al. 2003). In this setting, it would be clearly useful if clinicians were able to target treatments more precisely. Treatment precision has two facets. On the one hand, precise treatments remove the cause of the illness with minimal other effects. On the other hand, individuals differ and hence will react to the same treatment differently even if suffering from the putatively same causative agent. Hence, precision treatment aims also to be personalized and take constitutional or illness factors of individuals and their individual disease processes into account when refining the choice of treatment (Collins and Varmus, 2015). Although this is a standard feature of many aspects of medicine, advances in molecular techniques have dramatically increased the scope of personalization and precision. This has arguably been most successful in the field of oncology, where genomic signatures of individual cancers are used to guide treatment.

In psychiatry, parallel efforts have led to the use of genetic information about liver metabolic enzymes affecting the pharmacokinetic properties of medications as the ideal dosing will depend on the metabolism of the medication, though these tests have yet to enter standard clinical practice (Peterson et al., 2017). On a different level, personalization has of course long been woven tightly into the treatments of mental illnesses. Cognitive Behaviour Therapy, for instance, aims to identify and then modify the individual's underlying core beliefs, and a similar focus on the details of an individual's illness are prominent features of many psychotherapies. The understanding that mental illnesses are usually worsened by stress is reflected in the broad acceptance of the importance of holistic care, e.g. in Engel (1980)'s biopsychosocial model, and this in turn has facilitated specific social, financial, housing and other support interventions, which certainly are 'personalized'. Precision psychiatry, as understood in this book, however aims to leverage our recent advances in the understanding of the brain and the explosion in technological capacities to further improve how treatments are developed and targeted (Fernandes et al., 2017; Perna et al., 2018).

This chapter will first briefly outline the motivation for a computational approach to mental illness, before focusing on the use of tasks to probe specific computational processes. A major hindrance to the translation of tasks into clinical practice is their apparently low reliability, and the chapter will hence examine possible causes of this in some detail.

2 COMPUTATIONAL PSYCHIATRY

When considering illnesses arising from a particular organ, it is critically important to keep the main function of that organ in mind. For instance, many features of heart failure are only understandable with respect to the fact that it has a pumping function and is key to maintaining appropriate pressure gradients across the cardiovascular system. The same is true for the brain, the key function of which is to compute - storing information, deriving succinct summaries of it, and predicting the future. What sets the brain apart from all other organs is that its ability to process information changes as a function of the information it has already processed and stored - it learns. While other organs adapt, like muscles, the changes in the brain's ability at a higher level are not just quantitative (such as more strength), but qualitative: through learning, the brain can come to perform novel computations; and through its ability to compute, it can change how and what it learns. Just as the inability to meet pumping demands is the defining feature of heart disease, the inability to meet computing and storage demands shapes the signs and symptoms of brain disease (see Box 1 for an example). While some computational demands have obvious consequences in terms of movement or sensory deficits, others are more subtle, affecting a brain's ability to solve abstract or complex social problems, i.e. to perform higher cognitive functions. An understanding of these dysfunctions is likely to require an understanding of the functions affected, and if these functions are primarily computational, then a computational approach might well be necessary.

Computational psychiatry is a young field at the intersection of psychiatry, psychology, neuroscience, mathematics, statistics and machine learning. It attempts to harness advances in theoretical and computational insights to address clinical issues in the realm of psychiatric illnesses (Huys et al., 2011; Montague et al., 2012; Stephan and Mathys, 2014; Wang and Krystal, 2014; Huys et al., 2016b; Stephan et al., 2016b,a, 2017; Rutledge et al., 2019). The motivation for using mathematical and computational methods to approach subjective phenomena as mood, paranoia and trauma is broadly-speaking two-fold, reflecting the above theoretical considerations about the nature of mental illnesses, but also data-analytic ones (Huys et al., 2016b; Huys, 2018; Bennett et al., 2019).

BOX 1: Computational components of depressive symptoms.

Computational processes are likely involved in both the aetiology and the treatment of major depressive disorder, which is a syndrome with low mood, anhedonia and low energy at its core (American Psychiatric Association, 2013; World Health Organization, 1990; Mitchell et al., 2009). The odds of a first episode being increased by a factor of nearly 10 after a major life event (Kendler et al., 2000). The causal link from experiences such as live events, which are external to the brain, to symptoms such as mood and energy must flow via some form of interpretation of the events with long-term consequences, i.e. it must involve a computational process and learning. Indeed, biases in information processing have long been established as risk factors for depression (Alloy et al., 1999). A similar argument can be made for psychotherapeutic treatments such as behavioural activation (Jacobson et al., 1996; Dimidjian et al., 2006). Although specific implementations of this therapy vary, they broadly involve teaching individuals to act from 'outside in', i.e. to act according to their goals rather than their current emotional state and support them in formulating specific, achievable, measurable, realistic and temporally defined (SMART) goals. The aim is to increase the rate at which activities with positive consequences are performed, leading to an overall increased rate of experienced rewards and thereby a reduction in negative mood. While behavioural activation is a particularly clear example, the causal link of any psychotherapeutic intervention must at some point involve computational and learning processes as there are no direct influences on the brain systems determining the symptoms. Interestingly, the original behavioural activation study found that response to the behavioural component appeared to relate to a change in cognitive features (attributional style) engendered by the intervention (Jacobson et al., 1996).

The data-analytic side itself has multiple aspects (Woo et al., 2017; Stephan et al., 2017; Bzdok and Meyer-Lindenberg, 2018). First, testing computational theories about brain functions and their involvement in mental illness requires the use of advanced analytic methods. Theories of learning, for instance, are most thoroughly tested by building generative computational models and examining how well they explain data (Wetzels et al., 2010; Piray et al., 2019). Second, data of increasing richness, complexity and volume are now being gathered routinely thanks to advances in neuroimaging, mobile devices, data storage, online and computer technology (e.g. Gillan and Daw 2016). Researchers and clinicians are hence faced with such large datasets with increasing frequency. Deriving insights from them and correctly interpreting them requires familiarity with computational methods ranging from programming to complex machine-learning. For instance, large datasets raise fundamental issues around the stability and validity of inference: some inference problems such as regression become ill-posed when the dimensionality of the data (e.g. the number of data points per subject) is too high, and require sophisticated methods such as regularization, dimensionality reduction, Bayesian model evidence estimation, cross-validation or the training of deep neural networks. An important contribution of data-analytic approaches is a renewed emphasis on cross-validation (Stone, 1974). The term prediction has often been used to describe associations in correlational analyses, e.g. regressions, but such associations often do not generalize to novel datasets, and hence represent instances of overfitting (Huys et al., 2016b). These can be addressed by validating them on a separate dataset, and techniques such as cross-validation provide estimates of the likely generalizability of any findings. More generally, machine-learning approaches allow for the pragmatic discovery of potential signatures of illness or treatment response. This is becoming more attractive as the cost of acquiring vast datasets is reduced (Rutledge et al., 2019).

These arguments suggest why computational methods are likely to play an important role in the development of novel targets and treatments for mental illnesses (Wang and Krystal, 2014; Maia et al., 2017). They also suggest that researchers in the area of mental health are likely to be faced with computational challenges, and hence that ensuring literacy with computational methods through teaching programmes might be useful.

3 Tasks to measure computational functions

The ways in which computational tools can come to support precision psychiatry in everyday clinical practice is hence manifold. We have previously described a broad procedure for bringing computational tools into the clinic that is modelled on the drug developmental pipeline, with a large number of tools being examined preclinically, and the most promising ones being optimized for robustness prior to being put through randomized clinical trials (Paulus et al., 2016). Clearly, this is not something to be fulfilled by a single lab, and there is an urgent need for large-scale collaborations (Browning et al., 2019).

In this chapter, I will focus on one important role computational tools will likely assume in daily clinical practice: that of measurement using tasks. Measurements of relevant computational processes could have many different applications, such as for diagnosis, treatment allocation, treatment monitoring and risk assessment. Tasks are likely to play a key role in measurement because they represent the most direct approach to measuring specific learning and computational functions. There is now a large and rapidly growing wealth of tasks that activate, and thereby measure, increasingly complex and well-defined computational functions. (Frank et al., 2004; Mathews and MacLeod, 2005; Pizzagalli et al., 2005; Daw et al., 2011; Huys et al., 2013, 2012; Browning et al., 2015; Mkrtchian et al., 2017; Huys and Renz, 2017; Rutledge et al., 2017; Aylward et al., 2019; Berwian et al., 2019). The recent increase in mobile devices and computers, coupled with the development of toolboxes for efficient task deployment in browsers or apps, has profoundly reduced barriers to deploying tasks as probes in clinical settings (Gillan and Daw, 2016; Rutledge et al., 2019).

The two-step task (Fig 1), for instance, is a widely used task which attempts to capture an individual's tendency to learn and make decisions via one of two strategies - model-free or model-based (Daw et al., 2011). Model-free learning has been related to habitual decisions, and model-based learning to goal-directed behaviour (Friedel et al., 2014). In model-free learning, individuals learn by summing up prediction errors over multiple repetitions. Model-free decisions relying on these values hence change slowly over time. By contrast, model-based decisions require on-the-fly inference. While this is computationally demanding, it is also able to more rapidly adjust to any new information. Patients with a variety of compulsive disorders including obsessive-compulsive disorder, binge eating and methamphetamine addiction show a characteristic pattern on this and related tasks, with a bias towards model-free and away from model-based reasoning in large samples across multiple different settings (Voon et al. 2015; Gillan et al. 2016, 2019; Patzelt et al. 2019; though not in alcohol addiction Huys et al. 2016a; Nebe et al. 2018).

Another example task is the affective go-nogo task (Guitart-Masip et al. 2012; Fig 2). Tasks measuring Pavlovian influences on instrumental choice have shown robust sensitivity to alcohol dependence (Garbusow et al., 2019), anxiety (Mkrtchian et al., 2017), trauma (Ousdal et al., 2018) and suicidality (Millner et al., 2019). The latter finding in particular is noteworthy. The authors modified the task so that individuals could learn to avoid or escape unpleasant sounds either through active (go) or passive (nogo) behaviour. Patients with lifetime non-fatal suicidal thoughts and behaviours showed a selective increase in the tendency to actively escape from the aversive noise. Tasks such as these have great potential as structured probes to directly measure, with high precision and at low cost, high-level processes not accessible to techniques such as self-report, observation, biochemical or neuroimaging assays (Barch et al., 2008).

For the ideal clinical scenario, such tasks would result in task-derived measures of specific computational or learning processes (TDMs) that a) are mechanistically involved in causing illness and b) are amenable to interventions. The impact of a particular intervention might then be mediated by its impact on the TDMs (Figure 3A). In this situation, measuring the TDMs would have substantial value for precision psychiatry. The presence of raised or reduced TDM would indicate the presence of the particular aetiological process. This in turn would allow for differential treatment allocation to those interventions known to impact this particular TDM. In the absence of any abnormalities in any TDMs on the other hand futile exposure to likely unhelpful treatment could be avoided.

The key steps towards these goals hinge on the notion of discovering and engaging mechanisms relevant to mental illnesses. First, tasks need to be designed that yield reliable, robust and clinically deployable individual differences in a computational or learning process (Figure 3B). This is currently an area in great need for development, and we will focs on it below. Once such robust TDMs have been established, their relevance to particular symptoms or illnesses needs to be established, for instance via traditional case-control or correlational dimensional studies. Here, the advent of online methods promises to greatly accelerate the examination of new probes (e.g. Gillan et al. 2016; Gillan and Daw 2016; Rouault et al. 2018). It is however worth pointing to the value of longitudinal studies here. Although still correlational, examining how symptoms covary with a TDM within individuals over time avoids at least some of the more

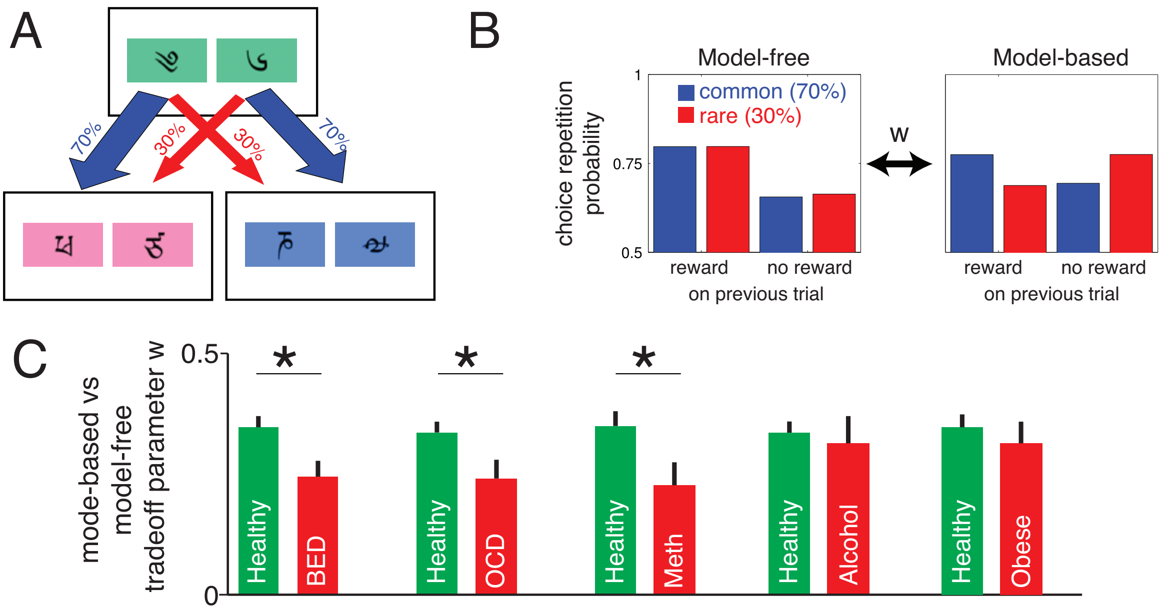

FIGURE 1: Two-step task. **A**: Participants have to first choose amongst two of the green stimuli. Each of the stimuli probabilistically leads to one of the second-stage stimulus sets with high probability, and to the other set with low probability. Participants then choose one of the two resulting second-stage stimuli and obtain a reward or not. **B**: A model-free strategy here corresponds to repeating the first-stage (green) choice if the second-stage choice was rewarded, irrespective of the frequency of the transition observed. A model-based strategy takes the transition probability into account: after a rare transition, a reward leads to a switch at the first stage. Consider choosing the left green choice, but transitioning to the blue second stage and then obtaining a reward. In order to gain another reward from the same blue stimulus, the best strategy takes the transition probability into account and leads to a switch of the unchosen first-stage stimulus. Individuals typically use a mixtures of these two strategies which can be measured by the parameter w. **C**: Patients with binge eating disorder (BED), OCD and methamphetamine depencence, but not with obesity or alcohol dependence show a reduction in the parameter w that trades off between these strategies, i.e. they show a shift towards mode-free decision-making. Panels A and B adapted from Daw et al. (2011). Panel C adapted and redrawn from Voon et al. (2015).

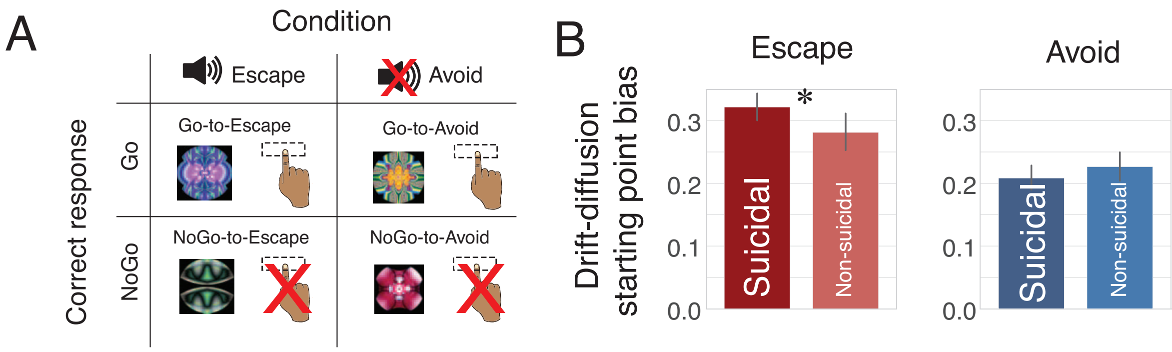

FIGURE 2: Affective go-nogo task. **A**: Individuals were taught to choose whether to go or nogo (respond on a button) for different stimuli. With some stimuli (fractals), an unpleasant tone could be escaped or entirely avoided by a go response (top row). With other stimuli, the unpleasant tone could be escaped or avoided by performing a nogo. **B**: A computational model fitted to the data extract a key parameter on which groups differed. Participants with a lifetime history of suicidal ideation showed a selective bias towards actively escaping, but did not show a bias in the avoid condition. The bias here was the starting point of a drift-diffusion model (Ratcliff and Smith, 2004). Figures adapted from Millner et al. (2019).

fundamental problems inherent in cross-sectional designs (Borsboom et al., 2009; Molenaar and Campbell, 2009). It then has to be examined whether the task-derived measures can be engaged by interventions, and whether a change in the task-derived measures must mediate the improvement in symptoms due to targeted interventions. Clearly, this is a very high bar. Indeed, there are few tasks which have been examined in all these scenarios. One exception is the two-step task, which has not shown changes with clinical state after psychotherapy (Wheaton et al., 2019). Though negative, this may relate to the poor reliability of the task (see below), and such research is critical for the development of precision tools.

4 LIMITATIONS OF CURRENT TASK-BASED MEASUREMENTS

Computational probes involving tasks must deliver reliable measurements if they are to be used clinically (Barch et al., 2008; Savitz et al., 2013). It has recently become clear that the reliability of many task-derived measures is still below the level of reliability deemed necessary for potential clinical utility (Barch et al., 2008; Savitz et al., 2013). Strikingly, this is even true for classical tasks that have stood the test of time. For instance, Hedge et al. (2018) examined tasks such as the Stroop, Eriksen Flanker and stop-signal tasks, and found that although the group effects were reliable, the effects at the individual level were not. Even the Stroop reaction time cost, i.e. the difference in reaction times to congruent and incongruent stimuli, only showed a test-retest Intraclass Correlation Coefficient of 0.6. A meta-analysis of published test-retest reliability measurements and a large-scale online study suggest a reliability somewhere between 0.3 and 0.6 across a wide variety of tasks (Enkavi et al., 2019). Importantly, this was substantially lower than the reliability of self-report surveys around 0.6 - 0.7.

The causes for low reliability fall into four categories: time, strategy, noise and research setting.

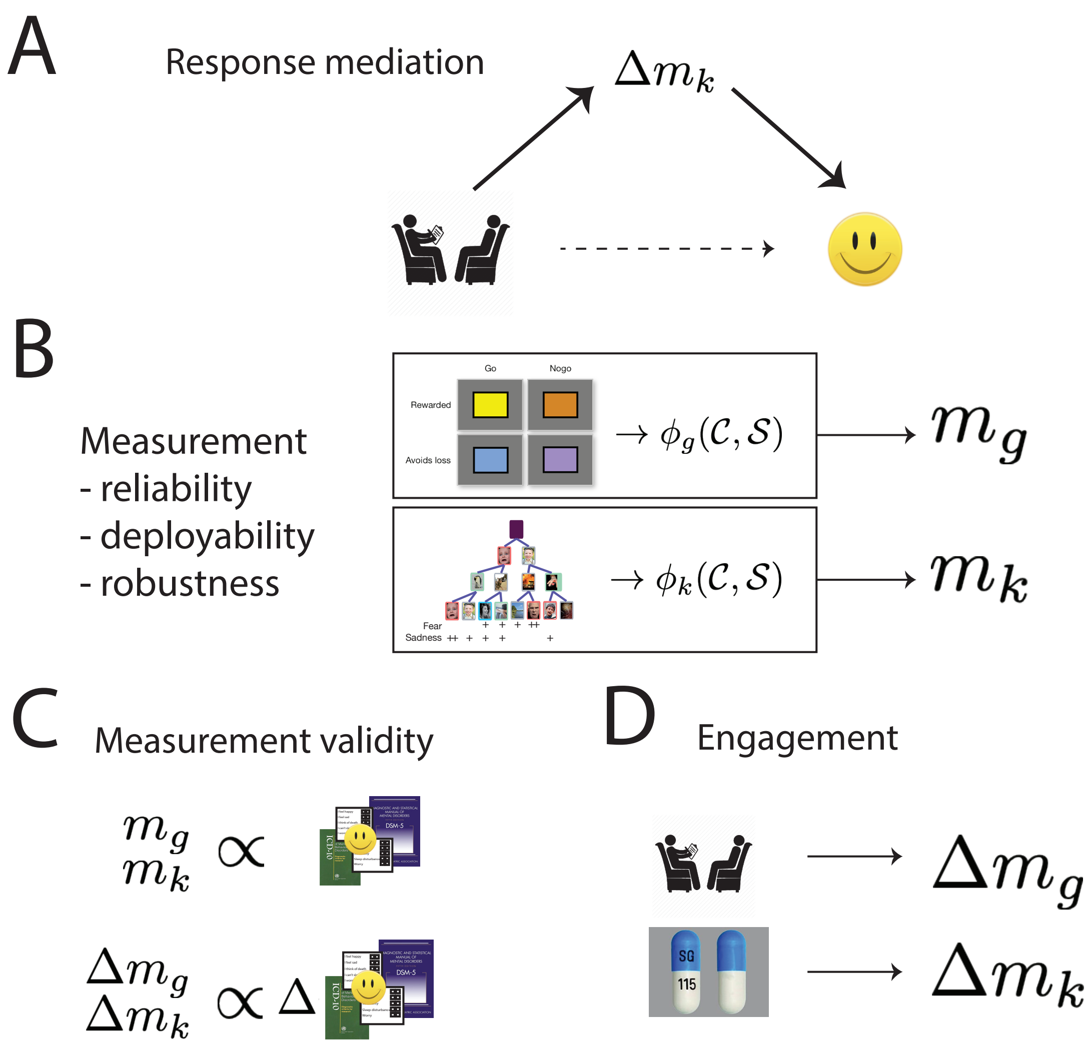

FIGURE 3: Measuring mechanisms for the clinic. **A**: For precision targeting, the measures derived from computational probes should mediate the effect of interventions. The measure m_k can be used to decide whether to apply intervention k if intervention k reduces measure m_k, and this measure m_k relates in a mechanistic or causal way to the illness. For instance, if antibiotics reduce certain bacterial cell counts, and these bacterial cell counts cause symptoms such as fever, then applying this antibiotic is likely to lead to an improvement in symptoms via its impact on the bacteria. **B**: For this to be feasible, computational probes, which might involve the results of a task being analyzed with some computational model and producing a measure m_k or m_g must be reliable at the individual level. The probes must also be deployable in clinical settings, and be robust to typical clinical situations. **C**: Measurements derived from computational probes must be valid, i.e. changes in these measurements should covary with changes in other measures of illness within individuals over time, and between individuals. **D**: Treatments, be they novel or established, should impact the measurement.

First, if the underlying cognitive mechanism changes with **time**, the measure will appear less reliable. A variety of task-derived measures do show excellent split-half reliabilities or reliability when the task is repeated immediately, but this reliability drops over weeks to months (Garbusow et al., 2014; Pooseh et al., 2018; Ahn et al., 2019; Shahar et al., 2019). For instance, Ahn et al. (2019) report reliabilities above 0.9 for immediate test-retest, but these fall to 0.8 within a month. A number of reports on low test-retest reliability of tasks are based on examination of reliability over delays of months over even years (Hedge et al., 2018; Enkavi et al., 2019). While this may be the right timescale to examine cognitive processes relating to stable personality factors, cognitive processes of relevance to the treatment of mental illnesses are expected to change on the same timescale as psychopathological states change, and should be amenable to therapeutic interventions. As such, the aim should be the establishment of TDMs that are highly reliable over short periods, but sensitive to relevant psychopathological changes (Duff, 2012) over weeks.

Second, tasks may appear unreliable because the **strategy** an individual employs to solve the task changes. This will also reduce internal consistency (Hajcak et al., 2017). For instance, through practice, individuals might discover shortcuts to solving a task, or engage demanding processes less due to fatigue. The *presence* of such inconsistencies can be examined using measures such as split-half reliability. Their *nature* can be identified through computational modelling, allowing particular strategies to be formulated as generative models. These can then be run on the task and provide quantitative measures as to how well a particular strategy explains the behaviour on a task (Guitart-Masip et al., 2012; Huys et al., 2012; Schlagenhauf et al., 2014; Huys et al., 2015; Berwian et al., 2019). Indeed, a change in strategy is likely to account for some of the changes seen with time, which have traditionally been ascribed to a change in the underlying cognitive mechanism.

Third, **noise** in a task-derived measure will reduce the estimated reliability. This noise can sometimes be reduced by making tasks longer (Rouder and Haaf, 2019), though this does not improve all task-derived measures equally (Enkavi et al., 2019): in learning tasks, for instance, early trials are informative about the learning process, but once a strategy has been selected and reached stability, additional trials no longer provide information about the learning process. The best strategy to reduce noise in the estimation of a particular process hence depends on the specific process. Techniques such as active learning or adaptive design optimization provide principled ways of maximising the amount of information acquired per trial (MacKay, 1992; Chaloner and Verdinelli, 1995; Paninski, 2005; Myung et al., 2013). As an example, consider delay discounting (Pooseh et al., 2018; Ahn et al., 2019): here, individuals have to repeatedly choose between receiving a small monetary amount sooner (e.g. £5 now), or a larger amount later (e.g. £10 in 2 months). Traditional approaches ask a fixed set of questions. However, if the subject has already accepted to wait 2 months for £10, then they are very likely to also be willing to wait 2 months for £30, and hence this item will not add much information. An alternative is to choose an option where the evidence gathered so far suggests equal probabilities for choosing the early and late option–an example of uncertainty sampling (Settles, 2012; Pooseh et al., 2018; Schulz et al., 2018). Indeed, these ideas underlie adaptive testing in item response theory approaches (Embretson and Reise, 2000), but have only rarely been exploited in the setting of tasks for mental health (Aranovich et al., 2017).

Strong guarantees exist about the usefulness of information-guided adaptive optimization if the underlying process is static and does not interact with the process (Paninski, 2005). However, human task strategies may change, and in particular subject may respond to the presence of changes in the task with shifts in strategy. More global optimizations problems are computationally challenging (Krause et al., 2008). But promising approaches include a combination of dynamic programming with adaptive design (Kim et al., 2017), and there is a dearth of optimization work that has explicitly attempted to avoid inducing changes in the underlying cognitive process, or indeed measured this. Another important aspect of noise in TDMs is the reduction of estimated correlation with other processes of interest (Spearman, 1904). This may lead to processes being deemed irrelevant when this is not the case (Rouder and Haaf, 2019). One approach is to take uncertainty into account, either via computational models (Huys et al., 2012, 2013; Shahar et al., 2019; Price et al., 2019; Yang et al., 2019) or via hierarchical estimation procedures (Gelman et al., 2013; Wetzels et al., 2010; Huys et al., 2012, 2013; Rouder and Haaf, 2019). Accounting for noise in the context of temporal reliability allows for a theoretical upper limit on reliability to be estimated. Tasks with higher theoretical reliability can in principle be lengthened or optimized to achieve this reliability, and hence may have clinical value. Taking uncertainty into account when examining covariation with other variables prevents disregarding potentially important processes.

Fourth, what exactly is viewed as 'the same result' varies in different **research settings** (Cronbach, 1957; Borsboom et al., 2009). In a traditional experimental psychology setting, a task is viewed as reliable if it produces an effect at the group level when repeated in a different sample. Variation between individuals hurts this notion of reliability: if individuals vary substantially, then the effect at the group level, usually measured by dividing the mean by the standard deviation, will necessarily be lower. On the other hand, in the context of the individual differences literature, a reliable measure is one that ranks individuals in the same order on repeated administrations. Here, variation between individuals, measured by correlation coefficients, generally increases reliability as the between-individual variability appears in the numerator (Hedge et al., 2018). Unlike questionnaires, tasks have generally been designed in the experimental psychology tradition, and hence are typically geared at maximising group-level reliability. As such, the component of variability attributable to differences between individuals is generally lower (Enkavi et al., 2019; Hedge et al., 2018). This is one major reason for the comparatively low reliability of tasks compared to questionnaires, and is a major hurdle for translation of tasks into a clinical setting.

5 CONCLUSION

In summary, computational approaches to mental health are motivated by the computational and learning functions of the brain, and by the complexity and quantity of data being acquired. Tasks are likely to be important for precision psychiatry as they allow specific learning and computational functions to be probed. However, tasks need to be further developed to achieve the reliability and robustness necessary for clinical deployment. As described here, computational models are likely to play an important role in this in that they can account for noise and strategy changes, and also facilitate adaptive sampling techniques. Once tasks have been designed that are both valid and reliable, researchers will need to shift attention towards studies that ask whether the processes can be engaged by therapies, and whether they mediate therapeutic improvement. It might even be advantageous to consider such longitudinal studies (particularly in the setting of treatments) early on.

Although the focus here has mainly been on tasks, similar arguments can be made for other techniques including in particular neuroimaging. Here, too, very reliable tasks have not reached reliabilities necessary for clinical deployment (Braun et al., 2012; Plichta et al., 2012; Savitz et al., 2013), and at least as far as they are to be used as measurements, similar arguments as for tasks can be made.

5.1 Acknowledgements and Disclosures

Support by the Max Planck UCL Centre for Computational Psychiatry is acknowledged. The author reports no biomedical financial interests or potential conflicts of interest.

References

Ahn, W.-Y., Gu, H., Shen, Y., Haines, N., Hahn, H. A., Teater, J. E., Myung, J. I., and Pitt, M. A. (2019). Rapid, precise, and reliable phenotyping of delay discounting using a bayesian learning algorithm. *bioRxiv*, 567412.

Alloy, L. B., Abramson, L. Y., Whitehouse, W. G., Hogan, M. E., Tashman, N. A., Steinberg, D. L., Rose, D. T., and Donovan, P. (1999). Depressogenic cognitive styles: predictive validity, information processing and personality characteristics, and developmental origins. *Behav Res Ther*, 37(6):503–531.

American Psychiatric Association (2013). *Diagnostic and Statistical Manual of Mental Disorders (DSM-5®)*. American Psychiatric Pub.

Angst, J., Gamma, A., Sellaro, R., Lavori, P. W., and Zhang, H. (2003). Recurrence of bipolar disorders and major depression. a life-long perspective. *Eur Arch Psychiatry Clin Neurosci*, 253(5):236–240.

Aranovich, G. J., Cavagnaro, D. R., Pitt, M. A., Myung, J. I., and Mathews, C. A. (2017). A model-based analysis of decision making under risk in obsessive-compulsive and hoarding disorders. *J Psychiatric Research*, 90:126–132.

Aylward, J., Valton, V., Ahn, W.-Y., Bond, R. L., Dayan, P., Roiser, J. P., and Robinson, O. J. (2019). Altered learning under uncertainty in unmedicated mood and anxiety disorders. *Nat. Hum. Behaviour*.

Barch, D. M., Carter, C. S., and Committee, C. E. (2008). Measurement issues in the use of cognitive neuroscience tasks in drug development for impaired cognition in schizophrenia: a report of the second consensus building conference of the cntrics initiative. *Schiz. Bull.*, 34:613–618.

Bennett, D., Silverstein, S. M., and Niv, Y. (2019). The two cultures of computational psychiatry. *JAMA Psychiatry*.

Berwian, I. M., Wenzel, J., Collins, A. G., Seifritz, E., Stephan, K. E., Walter, H., and Huys, Q. J. M. (2019). Computational mechanisms of effort and reward decisions in depression and their relationship to relapse after antidepressant discontinuation. *Revision. under review, JAMA Psychiatry*.

Borsboom, D., Kievit, R. A., Cervone, D., and Hood, S. B. (2009). The two disciplines of scientific psychology, or: The disunity of psychology as a working hypothesis. In Valsiner, J., Molenaar, P. C. M., Lyra, M. C. D. P., and Chaudhary, N., editors, *Dynamic process methodology in the social and developmental sciences*. Springer Science + Business Media.

Braun, U., Plichta, M. M., Esslinger, C., Sauer, C., Haddad, L., Grimm, O., Mier, D., Mohnke, S., Heinz, A., Erk, S., Walter, H., Seiferth, N., Kirsch, P., and Meyer-Lindenberg, A. (2012).

Test-retest reliability of resting-state connectivity network characteristics using fmri and graph theoretical measures. *Neuroimage*, 59(2):1404–1412.

Browning, M., Behrens, T. E., Jocham, G., O'Reilly, J. X., and Bishop, S. J. (2015). Anxious individuals have difficulty learning the causal statistics of aversive environments. *Nat Neurosci*, 18(4):590–596.

Browning, M., Carter, C. S., Chatham, C., Ouden, H. D., Gillan, C. M., Baker, J. T., Chekroud, A. M., Cools, R., Dayan, P., Gold, J., Goldstein, R. Z., Hartley, C. A., Kepecs, A., Lawson, R. P., Mourao-Miranda, J., Phillips, M. L., Pizzagalli, D. A., Powers, A., Rindskopf, D., Roiser, J. P., Schmack, K., Schiller, D., Sebold, M., Stephan, K. E., Frank, M. J., Huys, Q. J. M., and Paulus, M. (2019). Realizing the clinical potential of computational psychiatry: Report from the banbury center meeting, february 2019. *psyArxiv*, 10.31234/osf.io/5qbxp.

Bzdok, D. and Meyer-Lindenberg, A. (2018). Machine learning for precision psychiatry: Opportunities and challenges. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, 3(3):223–230.

Chaloner, K. and Verdinelli, I. (1995). Bayesian experimental design: A review. *Statistical Science*.

Collins, F. S. and Varmus, H. (2015). A new initiative on precision medicine. *New England Journal of Medicine*, 372(9):793–795.

Cronbach, L. J. (1957). The two disciplines of scientific psychology. *American Psychologist*, 12:671–684.

Daw, N. D., Gershman, S. J., Seymour, B., Dayan, P., and Dolan, R. J. (2011). Model-based influences on humans' choices and striatal prediction errors. *Neuron*, 69(6):1204–1215.

Dimidjian, S., Hollon, S. D., Dobson, K. S., Schmaling, K. B., Kohlenberg, R. J., Addis, M. E., Gallop, R., McGlinchey, J. B., Markley, D. K., Gollan, J. K., Atkins, D. C., Dunner, D. L., and Jacobson, N. S. (2006). Randomized trial of behavioral activation, cognitive therapy, and antidepressant medication in the acute treatment of adults with major depression. *J Consult Clin Psychol*, 74(4):658–670.

Duff, K. (2012). Evidence-based indicators of neuropsychological change in the individual patient: relevant concepts and methods. *Arch. Clinical Neuropsychology*, 27:248–261.

Embretson, S. E. and Reise, S. P. (2000). *Item Response Theory for Psychologists*. Lawrence Erlbaum Associates, Inc.

Engel, G. L. (1980). The clinical application of the biopsychosocial model. *The American journal of psychiatry*, 137:535–544.

Enkavi, A. Z., Eisenberg, I. W., Bissett, P. G., Mazza, G. L., MacKinnon, D. P., Marsch, L. A., and Poldrack, R. A. (2019). Large-scale analysis of test-retest reliabilities of self-regulation measures. *Proc. Nat. Acad. Sci. USA*, 116:5472–5477.

Fernandes, B. S., Williams, L. M., Steiner, J., Leboyer, M., Carvalho, A. F., and Berk, M. (2017). The new field of 'precision psychiatry'. *BMC medicine*, 15:80.

Frank, M. J., Seeberger, L. C., and O'Reilly, R. C. (2004). By carrot or by stick: cognitive reinforcement learning in Parkinsonism. *Science*, 306(5703):1940–3.

Friedel, E., Schlagenhauf, F., Beck, A., Dolan, R. J., Huys, Q. J. M., Rapp, M. A., and Heinz, A. (2014). The effects of life stress and neural learning signals on fluid intelligence. *Eur Arch Psychiatry Clin Neurosci*.

Garbusow, M., Nebe, S., Sommer, C., Kuitunen-Paul, S., Sebold, M., Schad, D. J., Friedel, E., Veer, I. M., Wittchen, H.-U., Rapp, M. A., Ripke, S., Walter, H., Huys, Q. J. M., Schlagenhauf, F., Smolka, M. N., and Heinz, A. (2019). Pavlovian-to-instrumental transfer and alcohol consumption in young male social drinkers: Behavioral, neural and polygenic correlates. *Journal of Clinical Medicine*, 8.

Garbusow, M., Schad, D. J., Sommer, C., Jünger, E., Sebold, M., Friedel, E., Wendt, J., Kathmann, N., Schlagenhauf, F., Zimmermann, U. S., Heinz, A., Huys, Q. J. M., and Rapp, M. A. (2014). Pavlovian-to-instrumental transfer in alcohol dependence: a pilot study. *Neuropsychobiology*, 70(2):111–121.

Gelman, A., Carlin, J., Stern, H., Dunson, D., Vehtari, A., and Rubin, D. (2013). *Bayesian Data Analysis*. Chapman and Hall/CRC Press, 3rd edition.

Gillan, C. M. and Daw, N. D. (2016). Taking psychiatry research online. *Neuron*, 91:19–23.

Gillan, C. M., Kalanthroff, E., Evans, M., Weingarden, H. M., Jacoby, R. J., Gershkovich, M., Snorrason, I., Campeas, R., Cervoni, C., Crimarco, N. C., Sokol, Y., Garnaat, S. L., McLaughlin, N. C. R., Phelps, E. A., Pinto, A., Boisseau, C. L., Wilhelm, S., Daw, N. D., and Simpson, H. B. (2019). Comparison of the association between goal-directed planning and self-reported compulsivity vs obsessive-compulsive disorder diagnosis. *JAMA psychiatry*, pages 1–10.

Gillan, C. M., Kosinski, M., Whelan, R., Phelps, E. A., and Daw, N. D. (2016). Characterizing a psychiatric symptom dimension related to deficits in goal-directed control. *Elife*, 5.

Guitart-Masip, M., Huys, Q. J. M., Fuentemilla, L., Dayan, P., Duzel, E., and Dolan, R. J. (2012). Go and no-go learning in reward and punishment: interactions between affect and effect. *Neuroimage*, 62(1):154–166.

Hajcak, G., Meyer, A., and Kotov, R. (2017). Psychometrics and the neuroscience of individual differences: Internal consistency limits between-subjects effects. *J Abnormal Psychology*, 126:823–834.

Hedge, C., Powell, G., and Sumner, P. (2018). The reliability paradox: Why robust cognitive tasks do not produce reliable individual differences. *Behavior Research Methods*, 50:1166–1186.

Huys, Q. J., Deserno, L., Obermayer, K., Schlagenhauf, F., and Heinz, A. (2016a). Model-free temporal-difference learning and dopamine in alcohol dependence: Examining concepts from theory and animals in human imaging. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, 1(5):401 – 410.

Huys, Q. J. M. (2018). Advancing clinical improvements for patients using the theory-driven and data-driven branches of computational psychiatry. *JAMA psychiatry*, 75:225–226.

Huys, Q. J. M., Dayan, P., and Daw (2015). Depression: A Decision-Theoretic Account. *Ann. Rev. Neurosci.*, 38:1–23.

Huys, Q. J. M., Eshel, N., O'Nions, E., Sheridan, L., Dayan, P., and Roiser, J. P. (2012). Bonsai trees in your head: how the Pavlovian system sculpts goal-directed choices by pruning decision trees. *PLoS Comput Biol*, 8(3):e1002410.

Huys, Q. J. M., Maia, T. V., and Frank, M. J. (2016b). Computational psychiatry as a bridge from neuroscience to clinical applications. *Nat Neurosci*, 19(3):404–413.

Huys, Q. J. M., Moutoussis, M., and Williams, J. (2011). Are computational models of any use to psychiatry? *Neural Netw*, 24(6):544–551.

Huys, Q. J. M., Pizzagalli, D. A., Bogdan, R., and Dayan, P. (2013). Mapping anhedonia onto reinforcement learning: A behavioural meta-analysis. *Biol Mood Anxiety Disord*, 3(1):12.

Huys, Q. J. M. and Renz, D. (2017). A formal valuation framework for emotions and their control. *Biological Psychiatry*, 82:413–420.

Jacobson, N. S., Dobson, K. S., Truax, P. A., Addis, M. E., Koerner, K., Gollan, J. K., Gortner, E., and Prince, S. E. (1996). A component analysis of cognitive-behavioral treatment for depression. *J Consult Clin Psychol*, 64(2):295–304.

Kendler, K. S., Thornton, L. M., and Gardner, C. O. (2000). Stressful life events and previous episodes in the etiology of major depression in women: An evaluation of the "kindling" hypothesis. *Am. J. Psychiatry*, 157:1243–51.

Kim, W., Pitt, M. A., Lu, Z.-L., and Myung, J. I. (2017). Planning beyond the next trial in adaptive experiments: A dynamic programming approach. *Cognitive Science*, 41:2234–2252.

Krause, A., Singh, A., and Guestrin, C. (2008). Near-optimal sensor placements in gaussian processes: Theory, efficient algorithms and empirical studies. *J. Machine Learning Research*, 9:235–284.

MacKay, D. J. C. (1992). Information-based objective functions for active data selection. *Neural Computation*, 4:590–604.

Maia, T. V., Huys, Q. J. M., and Frank, M. J. (2017). Theory-based computational psychiatry. *Biological psychiatry*, 82:382–384.

Mathews, A. and MacLeod, C. (2005). Cognitive vulnerability to emotional disorders. *Annu Rev Clin Psychol*, 1:167–195.

Millner, A. J., den Ouden, H. E. M., Gershman, S. J., Glenn, C. R., Kearns, J. C., Bornstein, A. M., Marx, B. P., Keane, T. M., and Nock, M. K. (2019). Suicidal thoughts and behaviors are associated with an increased decision-making bias for active responses to escape aversive states. *Journal of Abnormal Psychology*, 128:106–118.

Mitchell, A. J., McGlinchey, J. B., Young, D., Chelminski, I., and Zimmerman, M. (2009). Accuracy of specific symptoms in the diagnosis of major depressive disorder in psychiatric outpatients: data from the midas project. *Psychol Med*, 39(7):1107–1116.

Mkrtchian, A., Aylward, J., Dayan, P., Roiser, J. P., and Robinson, O. J. (2017). Modeling avoidance in mood and anxiety disorders using reinforcement learning. *Biol. Psychiatry*, 82:532–539.

Molenaar, P. C. and Campbell, C. G. (2009). The new person-specific paradigm in psychology. *Current Directions in Psychological Science*, 18(2):112–117.

Montague, P. R., Dolan, R. J., Friston, K. J., and Dayan, P. (2012). Computational psychiatry. *Trends Cogn Sci*, 16(1):72–80.

Myung, J. I., Cavagnaro, D. R., and Pitt, M. A. (2013). A tutorial on adaptive design optimization. *J Mathematical Psychology*, 57:53–67.

Nebe, S., Kroemer, N. B., Schad, D. J., Bernhardt, N., Sebold, M., Müller, D. K., Scholl, L., Kuitunen-Paul, S., Heinz, A., Rapp, M. A., Huys, Q. J. M., and Smolka, M. N. (2018). No association of goal-directed and habitual control with alcohol consumption in young adults. *Addiction Biology*, 23:379–393.

Ousdal, O. T., Huys, Q. J., Milde, A. M., Craven, A. R., Ersland, L., Endestad, T., Melinder, A., Hugdahl, K., and Dolan, R. J. (2018). The impact of traumatic stress on pavlovian biases. *Psychological medicine*, 48:327–336.

Paninski, L. (2005). Asymptotic theory of information-theoretic experimental design. *Neural Computation*, 17:1480–1507.

Patzelt, E. H., Kool, W., Millner, A. J., and Gershman, S. J. (2019). Incentives boost model-based control across a range of severity on several psychiatric constructs. *Biological Psychiatry*, 85:425–433.

Paulus, M. P., Huys, Q. J. M., and Maia, T. V. (2016). A roadmap for the development of applied computational psychiatry. *Biological psychiatry: CNNI*, 1:386–392.

Perna, G., Grassi, M., Caldirola, D., and Nemeroff, C. B. (2018). The revolution of personalized psychiatry: will technology make it happen sooner? *Psychological medicine*, 48:705–713.

Peterson, K., Dieperink, E., Anderson, J., Boundy, E., Ferguson, L., and Helfand, M. (2017). Rapid evidence review of the comparative effectiveness, harms, and cost-effectiveness of pharmacogenomics-guided antidepressant treatment versus usual care for major depressive disorder. *Psychopharmacology*, 234(11):1649–1661.

Piray, P., Dezfouli, A., Heskes, T., Frank, M. J., and Daw, N. D. (2019). Hierarchical bayesian inference for concurrent model fitting and comparison for group studies. *PLoS computational biology*, 15:e1007043.

Pizzagalli, D. A., Jahn, A. L., and O'Shea, J. P. (2005). Toward an objective characterization of an anhedonic phenotype: a signal-detection approach. *Biol Psychiatry*, 57(4):319–327.

Plichta, M. M., Schwarz, A. J., Grimm, O., Morgen, K., Mier, D., Haddad, L., Gerdes, A. B., Sauer, C., Tost, H., Esslinger, C., Colman, P., Wilson, F., Kirsch, P., and Meyer-Lindenberg, A. (2012). Test–retest reliability of evoked BOLD signals from a cognitive–emotive fMRI test battery. *NeuroImage*, 60(3):1746–1758.

Pooseh, S., Bernhardt, N., Guevara, A., Huys, Q. J. M., and Smolka, M. N. (2018). Value-based decision-making battery: A bayesian adaptive approach to assess impulsive and risky behavior. *Behavior Research Methods*, 50:236–249.

Price, R. B., Brown, V., and Siegle, G. J. (2019). Computational modeling applied to the dot-probe task yields improved reliability and mechanistic insights. *Biological Psychiatry*, 85:606–612.

Ratcliff, R. and Smith, P. L. (2004). A comparison of sequential sampling models for two-choice reaction time. *Psychological review*, 111:333–367.

Rouault, M., Seow, T., Gillan, C. M., and Fleming, S. M. (2018). Psychiatric symptom dimensions are associated with dissociable shifts in metacognition but not task performance. *Biological psychiatry*, 84:443–451.

Rouder, J. N. and Haaf, J. M. (2019). A psychometrics of individual differences in experimental tasks. *Psychonomic Bulletin & Review*, 26:452–467.

Rutledge, R. B., Chekroud, A. M., and Huys, Q. J. (2019). Machine learning and big data in psychiatry: toward clinical applications. *Current Opinion in Neurobiology*, 55:152–159.

Rutledge, R. B., Moutoussis, M., Smittenaar, P., Zeidman, P., Taylor, T., Hrynkiewicz, L., Lam, J., Skandali, N., Siegel, J. Z., Ousdal, O. T., Prabhu, G., Dayan, P., Fonagy, P., and Dolan, R. J. (2017). Association of neural and emotional impacts of reward prediction errors with major depression. *JAMA Psychiatry*, 74:790–797.

Savitz, J. B., Rauch, S. L., and Drevets, W. C. (2013). Clinical application of brain imaging for the diagnosis of mood disorders: the current state of play. *Mol Psychiatry*, 18(5):528–539.

Schlagenhauf, F., Huys, Q. J. M., Deserno, L., Rapp, M. A., Beck, A., Heinze, H.-J., Dolan, R., and Heinz, A. (2014). Striatal dysfunction during reversal learning in unmedicated schizophrenia patients. *Neuroimage*, 89:171–180.

Schulz, E., Wu, C. M., Huys, Q. J. M., Krause, A., and Speekenbrink, M. (2018). Generalization and search in risky environments. *Cognitive science*, 42:2592–2620.

Settles, B. (2012). *Active Learning*. Morgan & Claypool Publishers.

Shahar, N., Hauser, T. U., Moutoussis, M., Moran, R., Keramati, M., consortium, N., and Dolan, R. J. (2019). Improving the reliability of model-based decision-making estimates in the two-stage decision task with reaction-times and drift-diffusion modeling. *PLoS Comput. Biol.*, 15:e1006803.

Spearman, C. (1904). The proof and measurement of association between two things. *The American Journal of Psychology*, 15(1):72–101.

Stephan, K. E., Bach, D. R., Fletcher, P. C., Flint, J., Frank, M. J., Friston, K. J., Heinz, A., Huys, Q. J. M., Owen, M. J., Binder, E. B., Dayan, P., Johnstone, E. C., Meyer-Lindenberg, A., Montague, P. R., Schnyder, U., Wang, X.-J., and Breakspear, M. (2016a). Charting the landscape of priority problems in psychiatry, part 1: classification and diagnosis. *Lancet Psychiatry*, 3:77–83.

Stephan, K. E., Binder, E. B., Breakspear, M., Dayan, P., Johnstone, E. C., Meyer-Lindenberg, A., Schnyder, U., Wang, X.-J., Bach, D. R., Fletcher, P. C., Flint, J., Frank, M. J., Heinz, A., Huys,

Q. J. M., Montague, P. R., Owen, M. J., and Friston, K. J. (2016b). Charting the landscape of priority problems in psychiatry, part 2: pathogenesis and aetiology. *Lancet Psychiatry*, 3:84–90.

Stephan, K. E. and Mathys, C. (2014). Computational approaches to psychiatry. *Curr Opin Neurobiol*, 25:85–92.

Stephan, K. E., Schlagenhauf, F., Huys, Q. J. M., Raman, S., Aponte, E. A., Brodersen, K. H., Rigoux, L., Moran, R. J., Daunizeau, J., Dolan, R. J., Friston, K. J., and Heinz, A. (2017). Computational neuroimaging strategies for single patient predictions. *NeuroImage*, 145:180–199.

Stone, M. (1974). Cross-validation and multinomial prediction. *Biometrika*, 61(3):509–515.

Voon, V., Derbyshire, K., Rück, C., Irvine, M. A., Worbe, Y., Enander, J., Schreiber, L. R. N., Gillan, C., Fineberg, N. A., Sahakian, B. J., Robbins, T. W., Harrison, N. A., Wood, J., Daw, N. D., Dayan, P., Grant, J. E., and Bullmore, E. T. (2015). Disorders of compulsivity: a common bias towards learning habits. *Mol Psychiatry*, 20(3):345–352.

Wang, X.-J. and Krystal, J. H. (2014). Computational psychiatry. *Neuron*, 84(3):638–654.

Wetzels, R., Vandekerckhove, J., Tuerlinckx, F., and Wagenmakers, E.-J. (2010). Bayesian parameter estimation in the expectancy valence model of the iowa gambling task. *J. Math. Psychol.*, 54:14–27.

Wheaton, M. G., Gillan, C. M., and Simpson, H. B. (2019). Does cognitive-behavioral therapy affect goal-directed planning in obsessive-compulsive disorder? *Psychiatry research*, 273:94–99.

WHO (2017). Depression and Other Common Mental Disorders: Global Health Estimates. World Health Organization, Geneva, CH.

Woo, C.-W., Chang, L. J., Lindquist, M. A., and Wager, T. D. (2017). Building better biomarkers: brain models in translational neuroimaging. *Nature Neuroscience*, 20(3):365–377.

World Health Organization (1990). *International Classification of Diseases*. World Health Organization Press.

Yang, J., Pitt, M. A., Ahn, W.-Y., and Myung, J. I. (2019). ADOpy: A python package for adaptive design optimization. *PsyArXiv*, mdu23.