Review

Biological Psychiatry: CNNI

Adaptive Design Optimization as a Promising Tool for Reliable and Efficient Computational Fingerprinting

Mina Kwon, Sang Ho Lee, and Woo-Young Ahn

ABSTRACT

A key challenge in understanding mental (dys)functions is their etiological and functional heterogeneity, and several multidimensional assessments have been proposed for their comprehensive characterization. However, such assessments require lengthy testing, which may hinder reliable and efficient characterization of individual differences due to increased fatigue and distraction, especially in clinical populations. Computational modeling may address this challenge as it often provides more reliable measures of latent neurocognitive processes underlying observed behaviors and captures individual differences better than traditional assessments. However, even with a state-of-the-art hierarchical modeling approach, reliable estimation of model parameters still requires a large number of trials. Recent work suggests that Bayesian adaptive design optimization (ADO) is a promising way to address these challenges. With ADO, experimental design is optimized adaptively from trial to trial to extract the maximum amount of information about an individual's characteristics. In this review, we first describe the ADO methodology and then summarize recent work demonstrating that ADO increases the reliability and efficiency of latent neurocognitive measures. We conclude by discussing the challenges and future directions of ADO and proposing development of ADO-based computational fingerprints to reliably and efficiently characterize the heterogeneous profiles of psychiatric disorders.

https://doi.org/10.1016/j.bpsc.2022.12.003

Etiological and functional heterogeneity is a key problem in classifying and understanding mental dysfunctions. Research suggests that multidimensional, multilevel, and multidomain assessments are needed to capture the heterogeneous characteristics of a specific mental disorder or dysfunction (1–3). To foster such new approaches, the Research Domain Criteria framework was proposed, which aims to characterize various multilevel measures in the major domains of human neurocognitive functions (1). Related projects include the Addictions Neuroclinical Assessment for substance use disorders (SUDs) (4) and the Measurement and Treatment Research to Improve Cognition in Schizophrenia (5) and the Cognitive Neuroscience Treatment Research to Improve Cognition in Schizophrenia initiatives for schizophrenia (6). A major challenge in the implementation of multidimensional assessments is that they involve lengthy assessment batteries, which may require up to 10 hours of administration [e.g., (4)]. Such lengthy assessments are unsuitable for clinical practice because participants can easily become fatigued and distracted, which may lead to increased measurement error and hinder reliable characterization of individual differences.

An important statistical issue for reliably capturing individual differences is the "reliability paradox," which refers to when certain effects are consistently and robustly replicated at the group level, but individual-level measures show poor test-retest reliability (7,8). While the low replicability of many behavioral and neuroimaging findings (9–11) is well-known and has raised public awareness in related fields, until recently, researchers have not commonly assessed the test-retest reliability of laboratory tasks, and its importance has been underestimated. Hedge *et al.* (8) reported that even the most well-established measures from neurocognitive tasks, such as the Stroop, Eriksen Flanker, Stop-Signal, and Go/NoGo tasks have low test-retest reliability. Additionally, recent large-scale studies have shown that behavioral tasks have much lower test-retest reliability than self-report surveys for measuring self-regulation (12,13). These inconsistent findings have prompted researchers to question the existence of key psychological constructs central to psychiatric conditions, such as impulsivity in SUDs (14). Critically, this reliability paradox suggests that many behavioral and functional magnetic resonance imaging (fMRI)–based biomarkers that use such behavioral tasks may also be unreliable (15,16). The lack of reliable and efficient neurobehavioral measures is therefore one of the most formidable challenges in characterizing individual differences and translating research findings into clinical practice.

Recent studies suggest that computational modeling provides more reliable measures of cognitive functions and characterizes individual differences better than traditionally used summary statistics (e.g., mean response time) (7,17). Computational models use mathematical equations and parameters that reflect the cognitive properties of interest to describe the latent neurocognitive processes that underlie

ISSN: 2451-9022

observed behaviors. Furthermore, computational modeling can be combined with neuroimaging studies to identify the brain regions responsible for the cognitive processes [e.g., model-based fMRI (18) or model-based electroencephalography (19)]. This approach is also useful for computational psychiatry because it provides computational explanations of complex psychiatric conditions; the model parameters and latent variables delineate the aberrant neurocognitive processes in patients with psychiatric disorders (20–32). For example, a growing number of clinical studies have used computational models, such as reinforcement learning models (18,33) or models of decision making (34–36), and suggest that computational modeling provides novel computational markers that can characterize various psychiatric disorders including SUDs (21,37–39), schizophrenia (40–42), mood disorders (43–46), attention-deficit/hyperactivity disorder (47), and obsessive-compulsive disorder (OCD) (48). Recent studies further suggest that such computational markers may predict real-world clinical outcomes [e.g., (38)].

Although computational modeling improves the reliability of neurocognitive measures, model parameter estimates may be still unreliable because of large trial-by-trial variation (i.e., random noise) or insufficient data within individuals. Recent studies have shown that hierarchical modeling, whether Bayesian (49,50) or non-Bayesian (51,52), can improve the test-retest reliability of model parameter estimates (17,53,54). In the hierarchical modeling framework, group-level information (e.g., information related to all participants in an experiment) associated with the parameter values provides prior beliefs (e.g., prior distributions in Bayesian modeling) that constrain parameter estimation at an individual level. Therefore, hierarchical estimation regularizes individual parameter estimates that deviate from the group-level distribution. This regularization (or shrinkage) (55) of estimates often improves the accuracy of parameter estimation (56,57). Rouder and Haaf (17) suggested that hierarchical models improve test-retest reliability by reducing within-individual variability in parameter estimates (i.e., regularization), with the model-based estimates accurately reflecting across-individual variation alone.

Even with hierarchical modeling, most behavioral tasks require too many trials for reliable estimation of their model parameters. When the number of trials is insufficient, model parameters are overfitted or not well estimated (58,59). The number of trials needed for reliable parameter estimation depends on the model complexity or the number of parameters of the model for a behavioral task. Rouder *et al.* (60) showed that hierarchical modeling may not be sufficient to reveal individual differences and recover latent correlations among tasks measuring a similar psychological construct, especially when there are insufficient numbers of trials per person. Although commonly used approaches include an increased number of trials and thorough sampling of the experimental stimuli from all possible design spaces, prolonged experiments can induce fatigue and boredom in participants, thereby leading to careless responses (59). To reduce task duration without compromising estimation precision, we should selectively use the trials that are the most informative of parameter values, as some trials might be too easy or obvious and provide little information for parameter estimation. This problem should be carefully considered when designing an experiment; however, many studies are designed on the basis of a prior study or are designed using heuristic rules of experimental design (e.g., number of trials, reward/punishment magnitudes, and frequency on a gambling task). Moreover, an optimal experimental design presumably varies among individuals because of individual differences (61). Therefore, ideally, we need to capitalize on individual differences observed in an experiment on the fly to present customized designs for each participant (62). Several adjustment procedures [e.g., staircase method (36)] are widely used for this purpose; however, such procedures or rules often lack theoretical justification.

ADAPTIVE DESIGN OPTIMIZATION

Adaptive design optimization (ADO), which originates from optimal experimental design (63,64), is a promising way to address the aforementioned issues (59,60,65). Optimal experimemntal design in statistics offers useful tools for identifying the experimental designs that are likely to provide the most informative data. Concretely, an adaptive experiment in an optimal experimental design framework selects a design (i.e., numerical values of design variables) that is optimal in an information-theoretic sense (i.e., maximally informative) by analyzing observations from preceding trials in real time. The selected optimal design is then used to make a new observation, which in turn is used to identify the optimal design for the next trial. This adaptive process of design optimization repeats trial after trial to make the data most informative within the smallest number of trials.

ADO is a model-based method as it uses a parametric model that predicts task performance to select optimal designs in an adaptive manner. To explain how ADO works mathematically, a model is defined by its likelihood function, $p(y|\theta, d)$, which represents the likelihood (i.e., probability) of observing an experimental outcome (y) given certain model parameter values (θ) and experimental design (d). Based on the model, ADO selects the optimal design d^* that maximizes the global utility function $U(d)$ (66), which quantifies the usefulness of design d. This function is defined as:

$$U(d) = \int\int u(d,\theta,y)p(y|\theta,d)\,p(\theta)\,dy\,d\theta \tag{1}$$

where $p(\theta)$ is the Bayesian prior distribution of parameters, and $u(d,\theta,y)$ is the local utility function that measures the utility of a hypothetical experimental trial with design d when the model outputs y given the parameter values θ. Equation 1 indicates that the global utility is the mean of the local utility across all possible values of y and θ, weighted by the likelihood $p(y|\theta,d)$ and the prior distribution $p(\theta)$.

The local utility function is defined as $u(d,\theta,y) = log\frac{p(\theta|y,d)}{p(\theta)}$ to favor the design that reduces the uncertainty about the parameter values to the greatest extent. With this form of utility function, equation 1 reduces to the mutual information (in information theory) (67) between the parameter random variable Θ and the outcome random variable $Y(d)$ as follows:

$$U(d) = H(\Theta) - H(\Theta\,|Y(d)) \tag{2}$$

where the marginal entropy $H(\Theta)$ represents overall uncertainty about the parameter values, and the conditional entropy

$H(\Theta \mid Y(d))$ represents reduced uncertainty about the parameter values given the outcome $Y(d)$.

ADO identifies the optimal design d^* on each trial and then observes the outcome y_{obs} using the optimal design. The new observation is used to update the prior distribution $p(\theta)$ via Bayes' rule. The updated posterior distribution $p(\theta|y_{obs})$ becomes the new prior distribution for the design selection in the next trial. Figure 1 illustrates these 3 steps of ADO: design optimization, experiment, and Bayesian updating. By iterating this procedure throughout the course of an experiment, ADO estimates the model parameters as quickly and precisely as possible. This procedure is highly sensitive to individual differences because stimulus selection is fine-tuned based on each participant's prior responses and up-to-date model parameters.

There is strong evidence that ADO improves task efficiency and the reliability of model parameter estimates (65). This is because ADO searches for the most informative experimental design with respect to a specific objective such as parameter estimation or model comparison. Additionally, ADO reduces the influence of random variation within individuals (i.e., random noise) by minimizing inconsistent or careless responses. For example, if a participant produces two extremely different responses to a certain stimulus, ADO is likely to select the stimulus again because it is likely to maximally reduce the uncertainty about the participant's model parameters. Increased task efficiency with ADO is particularly useful for testing clinical populations and children who may find a standard laboratory task very demanding or lengthy. Therefore, ADO is a promising tool for providing reliable and efficient computational markers of various psychiatric dysfunctions.

ADO APPLICATION

Recent studies have empirically demonstrated that ADO enhances the reliability and accuracy of parameter estimates in healthy individuals. For example, Hou *et al.* (68) applied ADO to a letter identification task in which participants viewed letter stimuli and reported the identities of the letters. In each trial, ADO selected the contrast (e.g., intensity of stimuli) and size (e.g., spatial frequency) of the letter stimuli to optimally estimate the model parameters of the contrast sensitivity function. With the use of ADO, the task length was substantially reduced from 500 to 1000 trials to 50 trials without compromising the test-retest reliability (i.e., 0.974). Several other studies have also demonstrated the utility of ADO for optimizing cognitive tasks including delay discounting (69), memory retention (70), recognition memory (71), risk attitude (70), associative learning (72), and vision (73–75) tasks. These results provide further evidence that ADO improves the precision of parameter estimates, as well as task efficiency.

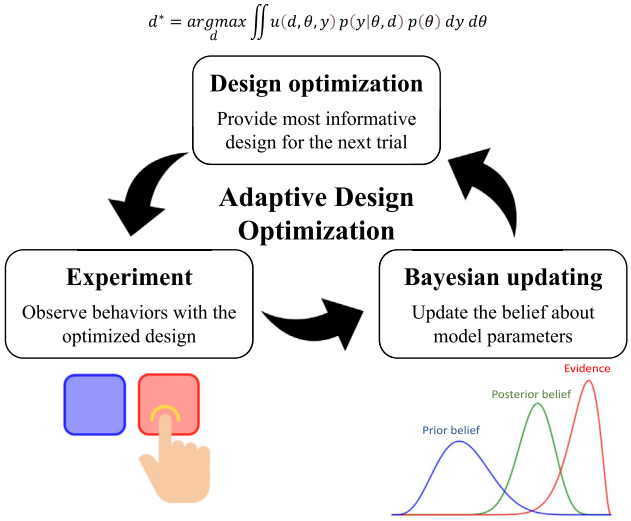

Figure 1. Schematic illustration of adaptive design optimization.

Improving task efficiency using ADO is particularly beneficial for neuroimaging studies because ADO may substantially reduce scan time and cost (76–78). Recent studies have further advanced ADO by incorporating neural and behavioral data for design optimization during neuroimaging experiments (79–81). In a proof-of-concept study, Bahg *et al.* (81) proposed an fMRI-based ADO wherein they used behavioral and fMRI data jointly to optimize the experimental design of a contrast discrimination task. When tested in simulated and real experiments, the precision of parameter estimates was better with the fMRI-based ADO than without it. These results suggest that ADO can incorporate both neural and behavioral data to further improve design optimization and enhance data quality.

Thus far, only a handful of studies have utilized ADO in clinical research despite its potential and promises in the field. Ahn *et al.* (82) showed that an ADO-based delay discounting task (83) provided highly reliable estimates of the delay discounting rate, which is a strong candidate endophenotype for addictive disorders and related conditions. The results showed that ADO led to 0.95 or higher test-retest reliability of the delay discounting rate within 10 to 20 trials, whereas a non-ADO (staircase) method (36) resulted in lower reliability even with a greater number of trials (Figure 2). Notably, ADO showed excellent reliability not just in healthy populations, but also in patients with SUDs and Amazon MTurk workers. Compared with the staircase method, ADO was 3 to 5 times more precise, 3 to 8 times more efficient, and captured 10% more variance in test-retest reliability [please see (82) for definitions of precision and efficiency]. Note that ADO and non-ADO led to similar delay discounting rates, with correlations ranging from 0.733 to 0.903. When ADO and non-ADO approaches provide slightly different parameter estimates, we prefer the estimates from ADO because of their higher consistency within and across multiple experimental conditions, which is supported by theoretically motivated estimation principles [see (82) for more details].

Another study used ADO to investigate risk aversion in patients with OCD and hoarding disorder (HD) (84). In a novel decision-making task, participants indicated a preference between 2 gamble options, and ADO optimized the probabilities of the gambles in each trial. The results suggest that individuals with OCD and HD exhibit lower-risk aversion than healthy control participants. The quantified risk attitude also showed significant correlations with several clinical indices, including measures of OCD and HD symptoms. Taken together, these findings suggest that risk aversion assessed by ADO might be related to OCD and HD symptoms. However, a non-ADO version was not tested in this study; thus, it is

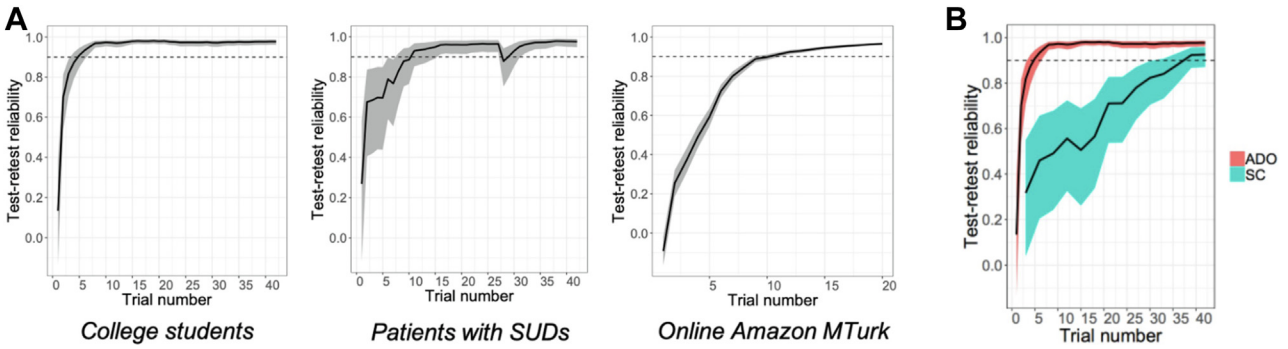

Figure 2. **(A)** Test-retest reliability of adaptive design optimization (ADO)–based discounting rates with college students, patients with substance use disorders (SUDs), and online Amazon MTurk workers when assessed cumulatively in every trial. **(B)** Comparison of ADO and staircase (SC) test-retest reliabilities of temporal discounting rates with college students [Adapted from Ahn *et al.* (82)].

unclear whether ADO would provide more disorder-sensitive measures than a standard procedure.

CHALLENGES AND FUTURE DIRECTIONS FOR ADO

While ADO is a promising approach for reliable and efficient evaluation of psychiatric conditions, there are remaining challenges and shortcomings of ADO. A fundamental limitation is that ADO may not be robust to model misspecification (62,85). As mentioned earlier, ADO is a model-based approach built upon the assumption that observed behaviors are generated from a particular model under consideration. If the model fails to adequately describe the behaviors, experimental designs that have been fine-tuned based on the inadequate model are likely to provide misleading parameter estimates. Model misspecification is a well-acknowledged common problem in practice, given that no single model can account for all possible variations of behaviors within and across individuals. This is a critical challenge, especially when we use a highly complex computational model. For example, models that account for multimodal data (e.g., choice, response time, and neurophysiological measures) sometimes have complex structures to explain specific associations between different measures, which might not generalize across participants. Even when an ADO-based task works well for most participants, behaviors of some participants might considerably deviate from what the model predicts. This problem may get even worse when testing clinical populations because patients with psychiatric conditions are well-known to be heterogeneous, and their behaviors might not be adequately explained by models developed based on data from healthy individuals.

In addressing the model misspecification problem when using ADO, a model-free design optimization algorithm called Gaussian process with active learning (GPAL) (86) has recently been developed. While the design optimization procedure of GPAL is very similar to ADO, a crucial difference between the two is that GPAL does not make a priori assumptions about the functional form of the model underlying behaviors. Instead, GPAL learns the function using a nonparametric modeling tool known as Gaussian processes, while at the same time optimizing the design to facilitate model inference. More specifically, GPAL selects a design on each trial to reduce the uncertainty about the unknown model function in the fewest possible number of trials. This data-driven method can make model development in a large design space feasible, even for participants with a short attention span (e.g., children), by minimizing the number of trials required for model inference (87).

It is worth mentioning that even with a correctly specified model, ADO can compromise the reliability of parameter estimates on some occasions. Because by construct ADO is sensitive to each and every response, a careless response or a simple mistake, especially in the early stage of an experiment, can bias the design selection and thus the parameter estimation. Another limiting factor of ADO is that the algorithm tends to repeatedly select the most informative stimuli; these are often difficult options that can cause fatigue or negative feelings that influence task performance (e.g., increased random responses) (88). To address this concern, we can either insert easy filler trials at certain times or stop data collection when certain criteria have been met (e.g., ADO constantly presents the same stimulus). Relatedly, we developed and tested an algorithm that penalizes selection of recently chosen stimuli and found that the approach led to performance similar to that seen when using a conventional ADO approach.

Besides the methodological challenges, another challenge is limited accessibility to ADO methodology due to requirements of technical expertise in Bayesian statistics, computational modeling, and programming skills. To address this issue, Yang *et al.* (89) developed an open-source Python package called ADOpy, which provides a researcher-friendly toolbox for ADO. This package enables researchers to readily run already implemented ADO-based tasks and models with some minor modifications or to develop their own tasks and models.

It is imperative to develop reliable and easy-to-use computational assays for clinical populations. Relatedly, researchers have proposed that we develop a "computational fingerprint" to identify heterogeneous profiles of psychiatric disorders (90,91). Such a computational fingerprint provides a holistic profile of each individual's decision-making process by combining multidimensional model parameters. We believe that a battery of ADO-based assays would be highly beneficial for the development of such multidimensional clinical markers because applying ADO will substantially reduce the time needed for neurocognitive task administration without compromising reliability and precision of parameter estimates. Furthermore, we could facilitate dissemination of such a

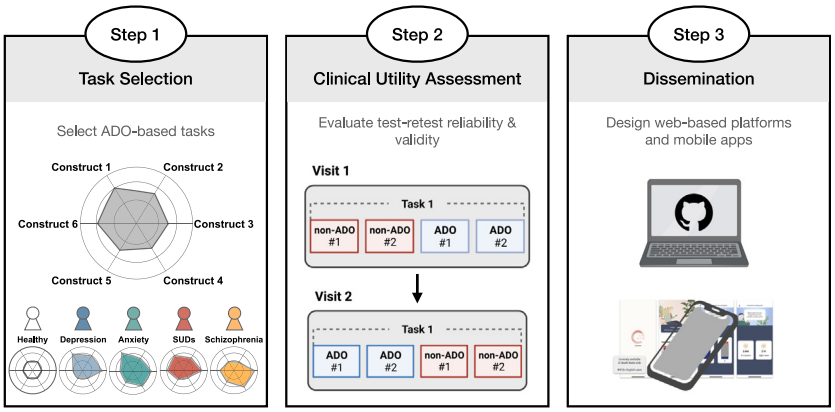

Figure 3. Future directions for an adaptive design optimization (ADO)–based battery and its implementation in web-based platforms and mobile apps for its dissemination. SUDs, substance use disorders.

battery by combining it with web-based platforms and digital technology (e.g., mobile apps), which will greatly enhance the accessibility of computational assays.

Practically, it is quite challenging to develop and validate an ADO-based battery and web/mobile platforms (Figure 3). First, we need to carefully select multiple laboratory tasks suitable for inclusion in the ADO battery. The selection should be 1) informed by relevant theories (92), 2) based on relevant criteria such as a valid measure of Research Domain Criteria or other relevant domains, 3) widely used in psychiatric research and prioritized by the research community or the National Institutes of Health, and 4) amenable to ADO (i.e., tasks with adequate computational models). When the tasks are selected, the reliability of the ADO-based tasks should be compared with the standard (non-ADO) versions of the tasks in neurotypical controls and in participants with psychiatric conditions [e.g., (82)]. It is also necessary to assess the predictive utility (i.e., validity) of the newly developed ADO tasks for a psychiatric disorder compared with non-ADO tasks. Finally, it would be highly beneficial for its dissemination if we develop web-based testing platforms and mobile apps. For example, we developed an app called Cocomo (available for iOS and Android) with ADO-based tasks and used it for predicting future cigarette use during and after a smoking cessation program in South Korea. Because of the substantially reduced number of trials required for ADO-based tasks, we found that many participants completed the ADO-based tasks daily on a mobile app, which would be much more challenging with a standard testing platform.

CONCLUSION

In conclusion, ADO is a promising approach that could provide reliable and efficient computational tools for assessing various psychiatric conditions. ADO optimizes and fine-tunes the experimental design for each individual using a Bayesian learning algorithm to maximize information gain on each trial. Heterogeneity of psychiatric disorders and lengthy testing required for multidomain assessment have been critical barriers in computational psychiatry. An ADO-based battery and mobile technology may lead to reliable and efficient computational fingerprinting, while aiding in the development of individualized prevention and treatment of psychiatric conditions.

ACKNOWLEDGMENTS AND DISCLOSURES

This work was supported by the Basic Science Research Program through the National Research Foundation of Korea funded by the Ministry of Science, Information and Communication Technologies and Future Planning (Grant No. NRF-2018R1C1B3007313 [to W-YA]); the National Research Foundation grant funded by the Ministry of Science, Information and Communication Technologies and Future Planning, the Korean Government (Grant No. 2021M3E5D2A0102249311 [to W-YA]); the BK21 FOUR Program (Grant No. 5199990314123 [to W-YA]); the Creative-Pioneering Researchers Program through Seoul National University (to W-YA); and the Creative Challenge Research Program through the National Research Foundation of Korea funded by the Ministry of Education (Grant No. 2022R1I1A1A01066530 [to SHL]).

We thank Jay I. Myung, Mark A. Pitt, and Jasmin Vassileva for their comments on an earlier draft of the manuscript as well as Jeung-Hyun Lee and Jooyeon J. Im for their help in making figures.

The authors report no biomedical financial interests or potential conflicts of interest.

ARTICLE INFORMATION

From the Department of Psychology, Seoul National University, Seoul, Korea (MK, SHL, W-YA); and the Department of Brain and Cognitive Sciences, Seoul National University, Seoul, Korea (SHL, W-YA).

Address correspondence to Woo-Young Ahn, Ph.D., at wahn55@snu.ac.kr.

Received Aug 31, 2022; revised Nov 21, 2022; accepted Dec 11, 2022.

REFERENCES

1. Cuthbert BN, Insel TR (2013): Toward the future of psychiatric diagnosis: The seven pillars of RDoC. BMC Med 11:126.
2. Paulus MP, Huys QJ, Maia TV (2016): A roadmap for the development of applied computational psychiatry. Biol Psychiatry Cogn Neurosci Neuroimaging 1:386–392.
3. García-Gutièrrez MS, Navarrete F, Sala F, Gasparyan A, Austrich-Olivares A, Manzanares J (2020): Biomarkers in psychiatry: Concept, definition, types and relevance to the clinical reality. Front Psychiatry 11:432.
4. Kwako LE, Momenan R, Litten RZ, Koob GF, Goldman D (2016): Addictions neuroclinical assessment: A neuroscience-based framework for addictive disorders. Biol Psychiatry 80:179–189.
5. Marder SR, Fenton W (2004): Measurement and Treatment Research to Improve Cognition in Schizophrenia: NIMH MATRICS initiative to support the development of agents for improving cognition in schizophrenia. Schizophr Res 72:5–9.

6. Carter CS, Barch DM (2007): Cognitive neuroscience-based approaches to measuring and improving treatment effects on cognition in schizophrenia: The CNTRICS initiative. Schizophr Bull 33:1131–1137.
7. Haines N, Kvam PD, Irving LH, Smith C, Beauchaine TP, Pitt MA, *et al.* (2020): Theoretically informed generative models can advance the psychological and brain sciences: Lessons from the reliability paradox. PsyArXiv. https://doi.org/10.31234/osf.io/xr7y3.
8. Hedge C, Powell G, Sumner P (2018): The reliability paradox: Why robust cognitive tasks do not produce reliable individual differences. Behav Res Methods 50:1166–1186.
9. Open Science Collaboration (2015): Estimating reproducibility of psychological science. Science 349:aac4716.
10. Kelly RE, Hoptman MJ (2022): Replicability in brain imaging. Brain Sci 12:397.
11. Pashler H, Wagenmakers EJ (2012): Editors' introduction to the special section on replicability in psychological science: A crisis of confidence? Perspect Psychol Sci 7:528–530.
12. Enkavi AZ, Eisenberg IW, Bissett PG, Mazza GL, MacKinnon DP, Marsch LA, Poldrack RA (2019): Large-scale analysis of test–retest reliabilities of self-regulation measures. Proc Natl Acad Sci USA 116:5472–5477.
13. Eisenberg IW, Bissett PG, Zeynep Enkavi AZ, Li J, MacKinnon DP, Marsch LA, Poldrack RA (2019): Uncovering the structure of self-regulation through data-driven ontology discovery. Nat Commun 10:2319.
14. Strickland JC, Johnson MW (2021): Rejecting impulsivity as a psychological construct: A theoretical, empirical, and sociocultural argument. Psychol Rev 128:336–361.
15. Elliott ML, Knodt AR, Ireland D, Morris ML, Poulton R, Ramrakha S, *et al.* (2020): What is the test-retest reliability of common task-functional MRI measures? New empirical evidence and a meta-analysis. Psychol Sci 31:792–806.
16. Kennedy JT, Harms MP, Korucuoglu O, Astafiev SV, Barch DM, Thompson WK, *et al.* (2022): Reliability and stability challenges in ABCD task fMRI data. Neuroimage 252:119046.
17. Rouder JN, Haaf JM (2019): A psychometrics of individual differences in experimental tasks. Psychon Bull Rev 26:452–467.
18. O'Doherty JP, Hampton A, Kim H (2007): Model-based fMRI and its application to reward learning and decision making. Ann N Y Acad Sci 1104:35–53.
19. Ratcliff R, Philiastides MG, Sajda P (2009): Quality of evidence for perceptual decision making is indexed by trial-to-trial variability of the EEG. Proc Natl Acad Sci USA 106:6539–6544.
20. Ahn WY, Dai J, Vassileva J, Busemeyer JR, Stout JC (2016): Computational modeling for addiction medicine from cognitive models to clinical applications. Prog Brain Res 224:53–65.
21. Mollick JA, Kober H (2020): Computational models of drug use and addiction: A review. J Abnorm Psychol 129:544–555.
22. Szalisznyó K, Silverstein DN (2021): Computational predictions for OCD pathophysiology and treatment: A review. Front Psychiatry 12:687062.
23. Valton V, Romaniuk L, Douglas Steele JD, Lawrie S, Seriès P (2017): Comprehensive review: Computational modelling of schizophrenia. Neurosci Biobehav Rev 83:631–646.
24. Adams RA, Huys QJ, Roiser JP (2016): Computational Psychiatry: Towards a mathematically informed understanding of mental illness. J Neurol Neurosurg Psychiatry 87:53–63.
25. Bishop SJ, Gagne C (2018): Anxiety, depression, and decision making: A computational perspective. Annu Rev Neurosci 41:371–388.
26. Robinson OJ, Chase HW (2017): Learning and choice in mood disorders: Searching for the computational parameters of anhedonia. Comput Psychiatr 1:208–233.
27. Sharp PB, Eldar E (2019): Computational models of anxiety: Nascent efforts and future directions. Curr Dir Psychol Sci 28:170–176.
28. Rusch T, Steixner-Kumar S, Doshi P, Spezio M, Gläscher J (2020): Theory of mind and decision science: Towards a typology of tasks and computational models. Neuropsychologia 146:107488.
29. Raymond JG, Steele JD, Seriès P (2017): Modeling trait anxiety: From computational processes to personality. Front Psychiatry 8:1.
30. Hauser TU, Fiore VG, Moutoussis M, Dolan RJ (2016): Computational psychiatry of ADHD: Neural gain impairments across Marrian levels of analysis. Trends Neurosci 39:63–73.
31. Fradkin I, Adams RA, Parr T, Roiser JP, Huppert JD (2020): Searching for an anchor in an unpredictable world: A computational model of obsessive compulsive disorder. Psychol Rev 127:672–699.
32. Loosen AM, Hauser TU (2020): Towards a computational psychiatry of juvenile obsessive-compulsive disorder. Neurosci Biobehav Rev 118:631–642.
33. Schultz W (2016): Dopamine reward prediction error coding. Dialogues Clin Neurosci 18:23–32.
34. Sokol-Hessner P, Hsu M, Curley NG, Delgado MR, Camerer CF, Phelps EA (2009): Thinking like a trader selectively reduces individuals' loss aversion. Proc Natl Acad Sci USA 106:5035–5040.
35. Tversky A, Fox CR (1995): Weighing risk and uncertainty. Psychol Rev 102:269–283.
36. Mazur JE (1987): An adjusting procedure for studying delayed reinforcement. In: Commons ML, Mazur JE, Nevin JA, Rachlin H, editors. The Effect of Delay and of Intervening Events on Reinforcement Value, vol. 5: London: Psychology Press, 55–73.
37. Romeu RJ, Haines N, Ahn WY, Busemeyer JR, Vassileva J (2020): A computational model of the Cambridge gambling task with applications to substance use disorders. Drug Alcohol Depend 206:107711.
38. Konova AB, Lopez-Guzman S, Urmanche A, Ross S, Louie K, Rotrosen J, Glimcher PW (2020): Computational markers of risky decision-making for identification of temporal windows of vulnerability to opioid use in a real-world clinical setting. JAMA Psychiatry 77:368–377.
39. Ahn WY, Dai J, Vassileva J, Busemeyer JR, Stout JC (2016): Computational modeling for addiction medicine: From cognitive models to clinical applications. Prog Brain Res 224:53–65.
40. Katthagen T, Kaminski J, Heinz A, Buchert R, Schlagenhauf F (2020): Striatal dopamine and reward prediction error signaling in unmedicated schizophrenia patients. Schizophr Bull 46:1535–1546.
41. Culbreth AJ, Westbrook A, Xu Z, Barch DM, Waltz JA (2016): Intact ventral striatal prediction error signaling in medicated schizophrenia patients. Biol Psychiatry Cogn Neurosci Neuroimaging 1:474–483.
42. Hernaus D, Xu Z, Brown EC, Ruiz R, Frank MJ, Gold JM, Waltz JA (2018): Motivational deficits in schizophrenia relate to abnormalities in cortical learning rate signals. Cogn Affect Behav Neurosci 18:1338–1351.
43. Macoveanu J, Stougaard ME, Kjærstad HL, Knudsen GM, Vinberg M, Kessing LV, Miskowiak KW (2022): Trajectory of aberrant reward processing in patients with bipolar disorder – A longitudinal fMRI study. J Affect Disord 312:235–244.
44. Bakker JM, Goossens L, Kumar P, Lange IMJ, Michielse S, Schruers K, *et al.* (2019): From laboratory to life: Associating brain reward processing with real-life motivated behaviour and symptoms of depression in non-help-seeking young adults. Psychol Med 49:2441–2451.
45. Kumar P, Goer F, Murray L, Dillon DG, Beltzer ML, Cohen AL, *et al.* (2018): Impaired reward prediction error encoding and striatal-midbrain connectivity in depression. Neuropsychopharmacology 43:1581–1588.
46. Chen C, Takahashi T, Nakagawa S, Inoue T, Kusumi I (2015): Reinforcement learning in depression: A review of computational research. Neurosci Biobehav Rev 55:247–267.
47. Hauser TU, Iannaccone R, Ball J, Mathys C, Brandeis D, Walitza S, Brem S (2014): Role of the medial prefrontal cortex in impaired decision making in juvenile attention-deficit/hyperactivity disorder. JAMA Psychiatry 71:1165–1173.
48. Hauser TU, Iannaccone R, Dolan RJ, Ball J, Hättenschwiler J, Drechsler R, *et al.* (2017): Increased fronto-striatal reward prediction errors moderate decision making in obsessive–compulsive disorder. Psychol Med 47:1246–1258.
49. Rouder JN, Lu J (2005): An introduction to Bayesian hierarchical models with an application in the theory of signal detection. Psychon Bull Rev 12:573–604.

50. Lee MD, Webb MR (2005): Modeling individual differences in cognition. Psychon Bull Rev 12:605–621.
51. Huys QJ, Cools R, Gölzer M, Friedel E, Heinz A, Dolan RJ, Dayan P (2011): Disentangling the roles of approach, activation and valence in instrumental and Pavlovian responding. PLoS Comput Biol 7: e1002028.
52. Woltman H, Feldstain A, MacKay JC, Rocchi M (2012): An introduction to hierarchical linear modeling. Tutor Quant Methods Psychol 8: 52–69.
53. Brown VM, Chen J, Gillan CM, Price RB (2020): Improving the reliability of computational analyses: Model-based planning and its relationship with compulsivity. Biol Psychiatry Cogn Neurosci Neuroimaging 5:601–609.
54. Waltmann M, Schlagenhauf F, Deserno L (2022): Sufficient reliability of the behavioral and computational readouts of a probabilistic reversal learning task. Behav Res 54:2993–3014.
55. Efron B, Morris C (1977): Stein's paradox in statistics. Sci Am 236:119–127.
56. Ahn WY, Krawitz A, Kim W, Busemeyer JR, Brown JW (2011): A model-based fMRI analysis with hierarchical Bayesian parameter estimation. J Neurosci Psychol Econ 4:95–110.
57. Lehmann EL, Casella G (2006): Theory of Point Estimation. New York, NY: Springer.
58. Farrell S, Lewandowsky S (2018): Computational Models of Cognition and Behavior. Cambridge, England: Cambridge University Press.
59. Busemeyer JR, Diederich A (2010): Cognitive Modeling. Thousand Oaks, CA: Sage Publications.
60. Rouder J, Kumar A, Haaf JM (2019): Why most studies of individual differences with inhibition tasks are bound to fail. PsyArXiv. https://doi.org/10.31234/osf.io/3cjr5.
61. Myung JI, Pitt MA (2009): Optimal experimental design for model discrimination. Psychol Rev 116:499–518.
62. Myung JI, Cavagnaro DR, Pitt MA (2013): A tutorial on adaptive design optimization. J Math Psychol 57:53–67.
63. Lindley DV (1956): On a measure of the information provided by an experiment. Ann Math Statist 27:986–1005.
64. Atkinson A, Donev A (1992): Optimum Experimental Designs. Oxford, England: Oxford University Press.
65. Cavagnaro DR, Myung JI, Pitt MA, Kujala JV (2010): Adaptive design optimization: A mutual information-based approach to model discrimination in cognitive science. Neural Comput 22:887–905.
66. Chaloner K, Verdinelli I (1995): Bayesian experimental design: A review. Statist Sci 10:273–304.
67. Cover TM, Thomas JA (2018): Elements of Information Theory. Hoboken, NJ: John Wiley & Sons Inc.
68. Hou F, Lesmes LA, Kim W, Gu H, Pitt MA, Myung JI, Lu ZL (2016): Evaluating the performance of the quick CSF method in detecting contrast sensitivity function changes. J Vis 16:18.
69. Cavagnaro DR, Aranovich GJ, McClure SM, Pitt MA, Myung JI (2016): On the functional form of temporal discounting: An optimized adaptive test. J Risk Uncertain 52:233–254.
70. Cavagnaro DR, Pitt MA, Gonzalez R, Myung JI (2013): Discriminating among probability weighting functions using adaptive design optimization. J Risk Uncertain 47:255–289.
71. Villarreal M, Stark CEL, Lee MD (2022): Adaptive design optimization for a Mnemonic similarity task. J Math Psychol 108:102665.
72. Melinscak F, Bach DR (2020): Computational optimization of associative learning experiments. PLoS Comput Biol 16:e1007593.
73. Gu H, Kim W, Hou F, Lesmes LA, Pitt MA, Lu ZL, Myung JI (2016): A hierarchical Bayesian approach to adaptive vision testing: A case study with the contrast sensitivity function. J Vis 16:15.
74. Xu P, Lesmes LA, Yu D, Lu ZL (2019): A novel Bayesian adaptive method for mapping the visual field. J Vis 19:16.
75. Zhao Y, Lesmes L, Lu ZL (2019): Efficient assessment of the time course of perceptual sensitivity change. Vision Res 154:21–43.
76. Toubia O, Johnson E, Evgeniou T, Delquié P (2013): Dynamic experiments for estimating preferences: An adaptive method of eliciting time and risk parameters. Manag Sci 59:613–640.
77. Soutschek A, Gvozdanovic G, Kozak R, Duvvuri S, de Martinis N, Harel B, *et al.* (2020): Dopaminergic D1 receptor stimulation affects effort and risk preferences. Biol Psychiatry 87:678–685.
78. Soutschek A, Burke CJ, Raja Beharelle AR, Schreiber R, Weber SC, Karipidis II, *et al.* (2017): The dopaminergic reward system underpins gender differences in social preferences. Nat Hum Behav 1:819–827.
79. Sanchez G, Daunizeau J, Maby E, Bertrand O, Bompas A, Mattout J (2014): Toward a new application of real-time electrophysiology: Online optimization of cognitive neurosciences hypothesis testing. Brain Sci 4:49–72.
80. Sanchez G, Lecaignard F, Otman A, Maby E, Mattout J (2016): Active SAmpling protocol (ASAP) to optimize individual neurocognitive hypothesis testing: A BCI-inspired dynamic experimental design. Front Hum Neurosci 10:347.
81. Bahg G, Sederberg PB, Myung JI, Li X, Pitt MA, Lu Z-L, Turner BM (2020): Real-time adaptive design optimization within functional MRI experiments. Comput Brain Behav 3:400–429.
82. Ahn WY, Gu H, Shen Y, Haines N, Hahn HA, Teater JE, *et al.* (2020): Rapid, precise, and reliable measurement of delay discounting using a Bayesian learning algorithm. Sci Rep 10:12091.
83. Rachlin H, Raineri A, Cross D (1991): Subjective probability and delay. J Exp Anal Behav 55:233–244.
84. Aranovich GJ, Cavagnaro DR, Pitt MA, Myung JI, Mathews CA (2017): A model-based analysis of decision making under risk in obsessive-compulsive and hoarding disorders. J Psychiatr Res 90:126–132.
85. Pitt MA, Myung JI (2019): Robust modeling through design optimization. Comput Brain Behav 2:200–201.
86. Chang J, Kim J, Zhang BT, Pitt MA, Myung JI (2021): Data-driven experimental design and model development using Gaussian process with active learning. Cogn Psychol 125:101360.
87. Lee SH, Kim D, Opfer JE, Pitt MA, Myung JI (2022): A number-line task with a Bayesian active learning algorithm provides insights into the development of non-symbolic number estimation. Psychon Bull Rev 29:971–984.
88. Ahn WY, Busemeyer JR (2016): Challenges and promises for translating computational tools into clinical practice. Curr Opin Behav Sci 11:1–7.
89. Yang J, Pitt MA, Ahn WY, Myung JI (2021): ADOpy: A python package for adaptive design optimization. Behav Res Methods 53:874–897.
90. Gueguen MCM, Schweitzer EM, Konova AB (2021): Computational theory-driven studies of reinforcement learning and decision-making in addiction: What have we learned? Curr Opin Behav Sci 38:40–48.
91. Yechiam E, Busemeyer JR, Stout JC, Bechara A (2005): Using cognitive models to map relations between neuropsychological disorders and human decision-making deficits. Psychol Sci 16:973–978.
92. Maia TV, Huys QJ, Frank MJ (2017): Theory-based computational psychiatry. Biol Psychiatry 82:382–384.