

1 **Title:** Adaptive design optimization as a promising tool for reliable and efficient
2 computational fingerprinting
3
4
5

6
7 **Authors:** Mina Kwon¹, Sang Ho Lee¹, Woo-Young Ahn^{1,2}
8
9

10
11
12 ¹Department of Psychology, Seoul National University, Seoul, Korea 08826
13

14 ²Department of Brain and Cognitive Sciences, Seoul National University, Seoul, Korea
15
16 08826
17
18
19
20

21 **Corresponding author:**
22

23
24 Woo-Young Ahn, Ph.D.
25

26 Department of Psychology
27

28 Seoul National University
29

30 Seoul, Korea 08826
31

32 Tel: +82-2-880-2538, Fax: +82-2-877-6428. E-mail: wahn55@snu.ac.kr
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

Abstract

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

A key challenge in understanding mental (dys)functions is their etiological and functional heterogeneity, and several multi-dimensional 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 the 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-art hierarchical modeling approach, reliable estimation of model parameters still requires a large number of trials. Recent works suggest that Bayesian adaptive design optimization (ADO) is a promising way to address the 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 works demonstrating that ADO increases reliability and efficiency of latent neurocognitive measures. We conclude by discussing the challenges and future directions of ADO and propose we develop ADO-based computational fingerprints to reliably and efficiently characterize the heterogeneous profiles of psychiatric disorders.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

Etiological and functional heterogeneity is a key problem in classifying and understanding mental dysfunctions. Research suggests that multi-dimensional, multi-level, and multi-domain assessment is needed to capture the heterogeneous characteristics of a specific mental disorder or dysfunction (1–3). To foster such new approaches, the Research Domain Criteria (RDoC) framework was proposed, which aims to characterize various multi-level measures in the major domains of human neurocognitive functions (1). Related projects include the Addictions Neuroclinical Assessment (ANA) for substance use disorders (SUDs) (4) and the Measurement and Treatment Research to Improve Cognition in Schizophrenia (MATRICS) (5) and the Cognitive Neuroscience Treatment Research to Improve Cognition in Schizophrenia (CNTRICS) initiatives for schizophrenia (6). A major challenge in the implementation of multi-dimensional assessments is that they involve lengthy assessment batteries, which may require up to 10 hours of administration (e.g., (7)). 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 (8,9). While the low replicability of many behavioral and neuroimaging findings (10–12) is well-known and raised public awareness in related fields, until recently researchers have not commonly assessed the test-retest reliability of laboratory task and its importance has been underestimated. Hedge et al. (13) reported that even the most well-established measures from neurocognitive tasks, such as 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, compared with self-report surveys for measuring

1 self-regulation (14,15). These inconsistent findings have prompted researchers to question the
2 existence of key psychological constructs central to psychiatric conditions, such as
3 impulsivity in SUDs (16). Critically, this reliability paradox suggests that many behavioral
4 and functional magnetic resonance imaging (fMRI)-based biomarkers that use such
5 behavioral tasks may also be unreliable (17,18). The lack of reliable and efficient
6 neurobehavioral measures is therefore one of the most formidable challenges in
7 characterizing individual differences and translating research findings into clinical practice.
8
9

10
11
12
13
14
15
16
17 Recent studies suggest that computational modeling provides more reliable measures
18 of cognitive functions and characterizes individual differences better than traditionally used
19 summary statistics (e.g., mean response time) (19,20). Computational models use
20 mathematical equations and parameters that reflect the cognitive properties of interest to
21 describe the latent neurocognitive processes that underlie observed behaviors. Furthermore,
22 computational modeling can be combined with neuroimaging studies to identify the brain
23 regions responsible for the cognitive processes (e.g., model-based fMRI (21) or model-based
24 electroencephalography (22)). This approach is also useful for computational psychiatry
25 because it provides computational explanations of complex psychiatric conditions; the model
26 parameters and latent variables delineate the aberrant neurocognitive processes in psychiatric
27 patients (23–35). For example, a growing number of clinical studies have used computational
28 models, such as reinforcement learning models (21,36) or models of decision-making (37–
29 39), and suggest that computational modeling provides novel computational markers that can
30 characterize various psychiatric disorders including SUDs (40–45), schizophrenia (40–42),
31 mood disorders (43,44,46,47), attention-deficit/hyperactivity disorder (48), and obsessive-
32 compulsive disorder (OCD) (49). Recent studies further suggest that such computational
33 markers may predict real-world clinical outcomes (e.g., (50)).
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33

Although computational modeling improves the reliability of neurocognitive measures, model parameter estimates might 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 (50,51) or non-Bayesian (52,53), can improve the test-retest reliability of model parameter estimates (20,54,55). In the hierarchical modeling framework, group-level information (e.g., 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) (56) of estimates often improves the accuracy of parameter estimation (57,58). Rouder and Haaf (20) 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.

34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

Even with hierarchical modeling, most behavioral tasks require too many trials for reliably estimating their model parameters. When the number of trials is insufficient, model parameters are overfitted or not well estimated (59,60). 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. 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 (60). To reduce the task duration without compromising the 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

1 experiment; however, many studies are designed on the basis of a prior study or are designed
2 using heuristic rules of experimental design (e.g., number of trials, reward/punishment
3 magnitudes, and frequency on a gambling task). Moreover, an optimal experimental design
4 presumably varies among individuals because of individual differences (61). Therefore,
5 ideally, we need to capitalize on individual differences observed in an experiment on the fly
6 to present customized designs for each participant (62). Several adjustment procedures (e.g.,
7 staircase method (39)) are widely used for this purpose; however, such procedures or rules
8 often lack theoretical justification.
9
10
11
12
13
14
15
16
17
18
19
20
21

22 **Adaptive design optimization (ADO)**

23
24 Adaptive design optimization (ADO), which originates from optimal experimental
25 design (OED) (63,64), is a promising way to address the aforementioned issues. OED (60,61)
26 in statistics offers useful tools for identifying the experimental designs that are likely to
27 provide the most informative data. Concretely, an adaptive experiment in an OED framework
28 selects a design (i.e., numerical values of design variables) that is optimal in an information-
29 theoretic sense (i.e., maximally informative) by analyzing observations from preceding trials
30 in real time. The selected optimal design is then used to make a new observation, which in
31 turn is used to identify the optimal design for the next trial. This adaptive process of design
32 optimization repeats trial after trial to make the data most informative within the smallest
33 number of trials.
34
35
36
37
38
39
40
41
42
43
44
45
46
47

48 ADO is a model-based method as it uses a parametric model that predicts task
49 performance to select optimal designs in an adaptive manner. To explain how ADO works
50 mathematically, a model is defined by its *likelihood function*, $p(y|\theta, d)$, which represents the
51 likelihood (i.e., probability) of observing an experimental outcome (y) given certain model
52 parameter values (θ) and experimental design (d). Based on the model, ADO selects the
53
54
55
56
57
58
59
60
61
62
63
64
65

optimal design d^* that maximizes the “global” utility function $U(d)$ (65), which quantifies the usefulness of design d . This function is defined as:

$$U(d) = \iint u(d, \theta, y) p(y|\theta, d) p(\theta) dy d\theta, \quad (1)$$

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 . The equation above indicates that the global utility is the mean of the local utility across all possible 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; (66)) between the parameter random variable Θ and the outcome random variable $Y(d)$, as follows:

$$U(d) = H(\Theta) - H(\Theta | Y(d)), \quad (2)$$

where the marginal entropy $H(\Theta)$ represents overall uncertainty about the parameter values, and the conditional entropy $H(\Theta | Y(d))$ represents reduced uncertainty about the parameter values given the outcome $Y(d)$.

ADO identifies the optimal design d^* on each trial, 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 three 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.

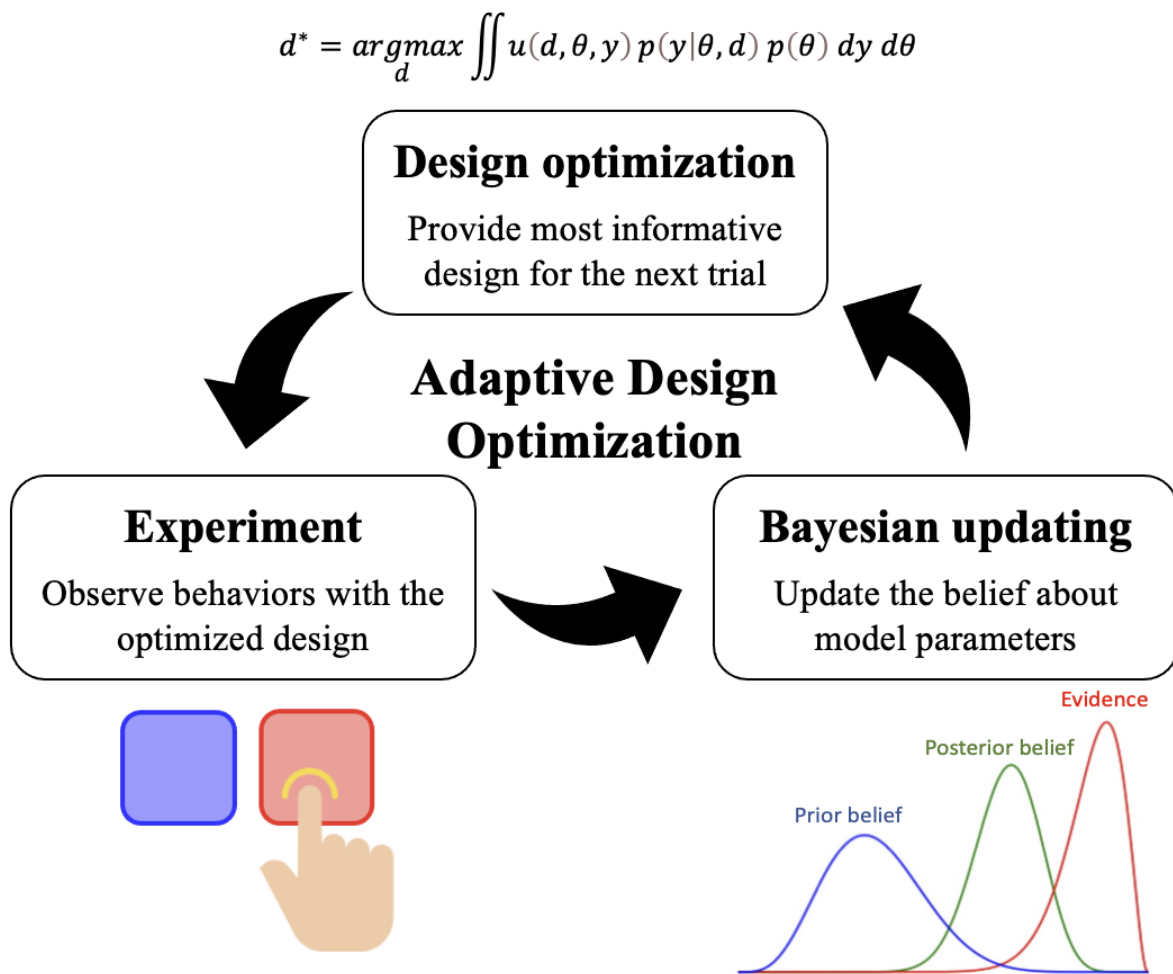


Figure 1. Schematic illustration of adaptive design optimization (ADO).

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

There is strong evidence that ADO improves task efficiency and the reliability of model parameter estimates (66). It 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. (67) applied ADO to a letter identification task in which participants viewed letter stimuli and reported 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 500–1,000 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 (68), memory retention (69), recognition memory (70), risk attitude (69), associative learning (71), and vision (72–74). These results provide further evidence that ADO improves the precision of parameter estimates, as well as task efficiency.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

Improving task efficiency using ADO is particularly beneficial for neuroimaging studies because ADO may substantially reduce the scan time and cost (75–77). Recent studies have further advanced ADO by incorporating neural and behavioral data for design optimization during neuroimaging experiments (78–80). In a proof-of-concept study, Bahg et al. (80) proposed an fMRI-based ADO where 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. (81) showed that an ADO-based delay discounting task (82) provides 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 10–20 trials, whereas a non-ADO (staircase) method (39) resulted in lower reliability even with a greater number of trials (**Figure 2**). Of note, 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 3–5 times more precise, 3–8 times more efficient and captured 10% more variance in test-retest reliability (please see (82) for definitions of precision and efficiency).

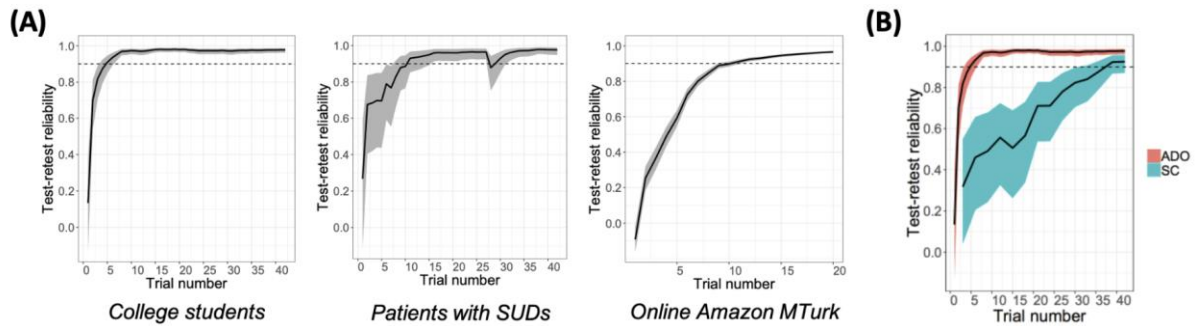


Figure 2. (A) Test-retest reliability of ADO-based discounting rates with college students, patients with SUDs, and online Amazon MTurk workers when assessed cumulatively in every trial. (B) Comparison of ADO and staircase (SC) test-retest reliabilities of temporal discounting rates with college students (adapted with permission from (82)).

Another study used ADO to investigate risk aversion in patients with OCD and hoarding disorder (HD) (83). In a novel decision-making task, participants indicated a preference between two 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 compared to healthy controls. The quantified risk attitude also showed significant correlations with several clinical indices, including 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 unclear whether ADO would provide more disorder-sensitive measures than a standard procedure.

Challenges and future directions of 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 (62,84). 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 fine-tuned based on the inadequate

1 model are likely to provide misleading parameter estimates. Model misspecification is a well-
2 acknowledged common problem in practice, given that no single model can account for all
3 possible variations of behaviors within and across individuals. Even when an ADO-based
4 task works well for most participants, behaviors of some participants might considerably
5 deviate from what the model predicts. This problem might get even worse when testing
6 clinical populations because patients with psychiatric conditions are well-known to be
7 heterogeneous and their behaviors might not be adequately explained by models developed
8 based on the data from healthy individuals.
9

10
11
12 In addressing the model misspecification problem when using ADO, recently a
13 model-free design optimization algorithm, dubbed Gaussian process with active learning
14 (GPAL; (85)) has been developed. While the design optimization procedure of GPAL is
15 much similar to ADO, an important and crucial difference between the two is that GPAL
16 does not make *a priori* assumptions about the functional form of the model underlying
17 behaviors. Instead, GPAL *learns* the function using a nonparametric modeling tool known as
18 Gaussian processes, while at the same time optimizing the design to facilitate model
19 inference. More specifically, GPAL selects a design on each trial to reduce the uncertainty
20 about the unknown model function in the fewest possible number of trials. This data-driven
21 method can make model development in a large design space feasible, even for participants
22 with short attention span (e.g., children), by minimizing the number of trials required for
23 model inference (86).
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47

48 It is worth mentioning that even with a correctly specified model, ADO can
49 compromise the reliability of parameter estimates on some occasions. Since by construct
50 ADO is sensitive to each and every response, a careless response or a simple mistake,
51 especially in the early stage of an experiment, can bias the design selection and thus the
52 parameter estimation. As another limiting factor of ADO, the algorithm tends to repeatedly
53
54
55
56
57
58
59
60
61
62
63
64
65

1 select the most informative stimuli; these are often difficult options that can cause fatigue or
2 negative feelings that influence task performance (e.g., increased random responses) (87). To
3 address this concern, we can either insert easy “filler” trials at certain times or stop data
4 collection when certain criteria are met (e.g., ADO constantly presents the same stimulus).
5 Relatedly, we developed and tested an algorithm that penalizes selection of recently chosen
6 stimulus and found that the approach led to similar performance to a conventional ADO
7 approach.

8
9 Besides the methodological challenges, another challenge is limited accessibility to
10 ADO methodology due to requirements of technical expertise in Bayesian statistics,
11 computational modeling, and programming skills. To address this issue, Yang et al. (88)
12 developed an open-source Python package called “ADOPY,” which provides a researcher-
13 friendly toolbox for ADO. This package enables researchers to readily run already
14 implemented ADO-based tasks and models with some minimal modifications, or to develop
15 their own tasks and models.

16
17 It is imperative to develop reliable and easy-to-use computational assays for clinical
18 populations. Relatedly, researchers have proposed that we develop “computational
19 fingerprint” to identify heterogeneous profiles of psychiatric disorders (89,90). Such a
20 computational fingerprint provides a holistic profile of each individual’s decision-making
21 process by combining multi-dimensional model parameters. We believe a battery of ADO-
22 based assays would be highly beneficial for the development of such multi-dimensional
23 clinical markers, because applying ADO will substantially reduce the time for neurocognitive
24 task administration without compromising reliability and precision of parameter estimates.
25 Furthermore, we could facilitate dissemination of such battery by combining it with web-
26 based platforms and digital technology (e.g., mobile apps), which will greatly enhance the
27 accessibility of computational assays.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
Practically, it is quite challenging to develop and validate an ADO-based battery and web/mobile platforms (**Figure 3**). We first need to carefully select multiple laboratory tasks suitable for inclusion in the ADO battery. The selection should be (i) informed by relevant theories (91), (ii) based on relevant criteria such as a valid measure of RDoC or other relevant domains, (iii) widely used in psychiatric research and prioritized by the research community or the National Institutes of Health, and (iv) 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., (81)). It is also necessary to assess the predictive utility (i.e., validity) of the newly developed ADO tasks for a psychiatric disorder compared to 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.

41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
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 multi-domain assessment have been critical barriers in computational psychiatry. An ADO-based battery and mobile technology might lead to reliable and efficient computational fingerprinting, while aiding in the development of individualized prevention and treatment of psychiatric conditions.

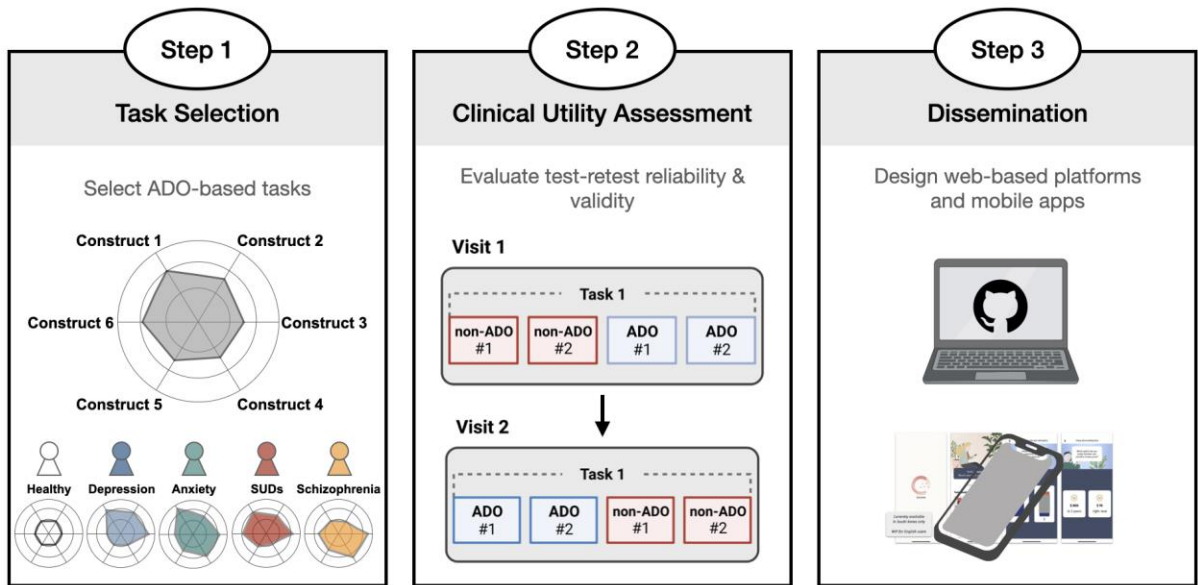


Figure 3. Future directions for an ADO-based battery and its implementation in web-based platforms and mobile apps for its dissemination.

Acknowledgements

This work was supported by the Basic Science Research Program through the National Research Foundation (NRF) of Korea funded by the Ministry of Science, ICT and Future Planning (MSIT) [NRF-2018R1C1B3007313]; the NRF grant funded by MSIT the Korea government [No. 2021M3E5D2A0102249311]; the BK21 FOUR Program [5199990314123] to W.-Y.A.; and the Creative-Pioneering Researchers Program through Seoul National University to W.-Y.A. The authors 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.

Financial disclosures/conflicts

The authors declare no competing interests.

References

1. Cuthbert BN, Insel TR (2013): Toward the future of psychiatric diagnosis: the seven pillars of RDoC. *Bmc Med* 11: 126–126.
2. Paulus MP, Huys QJM, Maia TV (2016): A Roadmap for the Development of Applied Computational Psychiatry. *Biological Psychiatry Cognitive Neurosci Neuroimaging* 1: 386–392.
3. García-Gutiérrez 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. *Frontiers 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 Psychiat* 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. *Schizophrenia Bull* 33: 1131–1137.
7. Kwako LE, Momenan R, Litten RZ, Koob GF, Goldman D (2016): Addictions Neuroclinical Assessment: A Neuroscience-Based Framework for Addictive Disorders. *Biol Psychiat* 80: 179–189.
8. 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. <https://doi.org/10.31234/osf.io/xr7y3>
9. 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.
10. Collaboration OS (2015): Estimating the reproducibility of psychological science. 349: aac4716–aac4716.
11. Kelly RE, Hoptman MJ (2022): Replicability in Brain Imaging. *Brain Sci* 12: 397.
12. Pashler H, Wagenmakers E (2012): Editors’ Introduction to the Special Section on Replicability in Psychological Science. *Perspect Psychol Sci* 7: 528–530.
13. Hedge C, Powell G, Sumner P (2017): The reliability paradox: Why robust cognitive tasks do not produce reliable individual differences. *Behavior Research Methods* 103: 1–21.

- 1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
14. 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 National Acad Sci* 116: 5472–5477.
 15. Eisenberg IW, Bissett PG, 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.
 16. Strickland JC, Johnson MW (2021): Rejecting Impulsivity as a Psychological Construct: A Theoretical, Empirical, and Sociocultural Argument. *Psychol Rev* 128: 336–361.
 17. 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.
 18. 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.
 19. 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. <https://doi.org/10.31234/osf.io/xr7y3>
 20. Rouder JN, Haaf JM (2019): A psychometrics of individual differences in experimental tasks. *Psychon B Rev* 26: 452–467.
 21. O’Doherty JP, Hampton A, Kim H (2007): Model- Based fMRI and Its Application to Reward Learning and Decision Making. *Ann Ny Acad Sci* 1104: 35–53.
 22. 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 National Acad Sci* 106: 6539–6544.
 23. Ahn WY, Dai J, Vassileva J, Busemeyer JR, Stout JC (2016): Chapter 3 Computational modeling for addiction medicine From cognitive models to clinical applications. *Prog Brain Res* 224: 53–65.
 24. Mollick JA, Kober H (2020): Computational Models of Drug Use and Addiction: A Review. *J Abnorm Psychol* 129: 544–555.
 25. Szalisznyó K, Silverstein DN (2021): Computational Predictions for OCD Pathophysiology and Treatment: A Review. *Frontiers Psychiatry* 12: 687062.
 26. Valton V, Romaniuk L, Steele JD, Lawrie S, Seriès P (2017): Comprehensive review: Computational modelling of schizophrenia. *Neurosci Biobehav Rev* 83: 631–646.
 27. Adams RA, Huys QJM, Roiser JP (2016): Computational Psychiatry: towards a mathematically informed understanding of mental illness. *J Neurology Neurosurg Psychiatry* 87: 53–63.

- 1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
28. Bishop SJ, Gagne C (2018): Anxiety, Depression, and Decision Making: A Computational Perspective. *Annu Rev Neurosci* 41: 1–18.
 29. Robinson OJ, Chase HW (2017): Learning and Choice in Mood Disorders: Searching for the Computational Parameters of Anhedonia. *Comput Psychiatry* 1: 208–233.
 30. Sharp PB, Eldar E (2019): Computational Models of Anxiety: Nascent Efforts and Future Directions. *Curr Dir Psychol Sci* 28: 170–176.
 31. 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.
 32. Raymond JG, Steele JD, Seriès P (2017): Modeling Trait Anxiety: From Computational Processes to Personality. *Frontiers Psychiatry* 8: 1.
 33. 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.
 34. 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.
 35. Loosen AM, Hauser TU (2020): Towards a computational psychiatry of juvenile obsessive-compulsive disorder. *Neurosci Biobehav Rev* 118: 631–642.
 36. Schultz W (2016): Dopamine reward prediction error coding. *Dialogues Clin Neurosci* 18: 23–32.
 37. Sokol-Hessner P, Hsu M, Curley NG, Delgado MR, Camerer CF, Phelps EA (2009): Thinking like a trader selectively reduces individuals’ loss aversion. *P Natl Acad Sci Usa* 106: 5035–40.
 38. Tversky A, Fox CR (1995): Weighing Risk and Uncertainty. *Psychol Rev* 102: 269–283.
 39. 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. Lawrence Erlbaum Associates, Inc., pp 55–73.
 40. Katthagen T, Kaminski J, Heinz A, Buchert R, Schlagenhauf F (2020): Striatal Dopamine and Reward Prediction Error Signaling in Unmedicated Schizophrenia Patients. *Schizophrenia 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. *Biological Psychiatry Cognitive Neurosci Neuroimaging* 1: 474–483.

- 1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
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. *Cognitive Affect Behav Neurosci* 18: 1338–1351.
 43. Macoveanu J, Stougaard ME, Kjørstad 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 Disorders* 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. Romeu RJ, Haines N, Ahn W-Y, Busemeyer JR, Vassileva J (2020): A computational model of the Cambridge gambling task with applications to substance use disorders. *Drug Alcohol Depen* 206: 107711.
 46. 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. *Neuropsychopharmacol* 43: 1581–1588.
 47. 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.
 48. 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 Psychiat* 71: 1165–1173.
 49. 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.
 50. Rouder JN, Lu J (2005): An introduction to Bayesian hierarchical models with an application in the theory of signal detection. *Psychon B Rev* 12: 573–604.
 51. Lee MD, Webb MR (2005): Modeling individual differences in cognition. *Psychon B Rev* 12: 605–621.
 52. Huys QJM, 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.
 53. Woltman H, Feldstain A, MacKay JC, Rocchi M (2012): An introduction to hierarchical linear modeling. *Tutorials Quantitative Methods Psychology* 8: 52–69.
 54. Brown VM, Chen J, Gillan CM, Price RB (2020): Improving the Reliability of Computational Analyses: Model-Based Planning and Its Relationship With Compulsivity. *Biological Psychiatry Cognitive Neurosci Neuroimaging* 5: 601–609.

- 1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
55. Waltmann M, Schlagenhaut F, Deserno L (2022): Sufficient reliability of the behavioral and computational readouts of a probabilistic reversal learning task. *Behav Res Methods* 1–22.
 56. Efron B, Morris C (1977): Stein’s Paradox in Statistics. *Sci Am* 236: 119–127.
 57. Ahn W-Y, Krawitz A, Kim W, Busemeyer JR, Brown JW (2011): A Model-Based fMRI Analysis With Hierarchical Bayesian Parameter Estimation. *J Neurosci Psychology Econ* 4: 95–110.
 58. Lehmann EL, Casella G (2006): *Theory of Point Estimation*. Springer Science & Business Media.
 59. Farrell S, Lewandowsky S (2018): *Computational Models of Cognition and Behavior*. Sage Publications.
 60. Busemeyer JR, Diederich A (2010): *Cognitive Modeling*. Sage Publications.
 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 Statistics* 27: 986–1005.
 64. Atkinson A, Donev A (1992): *Optimum Experimental Designs*. Oxford University Press.
 65. Chaloner K, Verdinelli I (1995): Bayesian Experimental Design: A Review. *Stat Sci* 10: 273–304.
 66. Cover TM, Thomas JA (2018): Elements of Information Theory. *Wiley Ser Telecommun*. <https://doi.org/10.1002/0471200611>
 67. Hou F, Lesmes LA, Kim W, Gu H, Pitt MA, Myung JI, Lu Z-L (2016): Evaluating the performance of the quick CSF method in detecting contrast sensitivity function changes. *J Vision* 16: 18.
 68. Cavagnaro DR, Aranovich GJ, McClure SM, Pitt MA, Myung JI (2016): On the functional form of temporal discounting: An optimized adaptive test. *J Risk Uncertainty* 52: 233–254.
 69. Cavagnaro DR, Pitt MA, Gonzalez R, Myung JI (2013): Discriminating among probability weighting functions using adaptive design optimization. *J Risk Uncertainty* 47: 255–289.
 70. Villarreal M, Stark CEL, Lee MD (2022): Adaptive design optimization for a Mnemonic Similarity Task. *J Math Psychol* 108: 102665.

- 1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
71. Melinscak F, Bach DR (2020): Computational optimization of associative learning experiments. *Plos Comput Biol* 16: e1007593.
 72. Gu H, Kim W, Hou F, Lesmes LA, Pitt MA, Lu Z-L, Myung JI (2016): A hierarchical Bayesian approach to adaptive vision testing: A case study with the contrast sensitivity function. *J Vision* 16: 15.
 73. Xu P, Lesmes LA, Yu D, Lu Z-L (2019): A novel Bayesian adaptive method for mapping the visual field. *J Vision* 19: 16.
 74. Zhao Y, Lesmes L, Lu Z-L (2019): Efficient assessment of the time course of perceptual sensitivity change. *Vision Res* 154: 21–43.
 75. Toubia O, Johnson E, Evgeniou T, Delquié P (2013): Dynamic Experiments for Estimating Preferences: An Adaptive Method of Eliciting Time and Risk Parameters. *Manage Sci* 59: 613–640.
 76. Soutschek A, Gvozdanovic G, Kozak R, Duvvuri S, Martinis N de, Harel B, *et al.* (2020): Dopaminergic D1 Receptor Stimulation Affects Effort and Risk Preferences. *Biol Psychiat* 87: 678–685.
 77. Soutschek A, Burke CJ, 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.
 78. 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.
 79. Sanchez G, Lecaigard 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.
 80. 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.
 81. Ahn W-Y, 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-uk* 10: 12091.
 82. Rachlin H, Raineri A, Cross D (1991): Subjective probability and delay. *J Exp Anal Behav* 55: 233–244.
 83. 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.
 84. Pitt MA, Myung JI (2019): Robust Modeling Through Design Optimization. *Comput Brain Behav* 2: 200–201.

- 1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
85. Chang J, Kim J, Zhang B-T, Pitt MA, Myung JI (2021): Data-driven experimental design and model development using Gaussian process with active learning. *Cognitive Psychol* 125: 101360.
 86. 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 B Rev* 29: 971–984.
 87. Ahn W-Y, Busemeyer JR (2016): Challenges and promises for translating computational tools into clinical practice. *Curr Opin Behav Sci* 11: 1–7.
 88. Yang J, Pitt MA, Ahn W-Y, Myung JI (2021): ADOpy: a python package for adaptive design optimization. *Behav Res Methods* 53: 874–897.
 89. Gueguen MC, 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.
 90. 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.
 91. Maia TV, Huys QJM, Frank MJ (2017): Theory-Based Computational Psychiatry. *Biol Psychiat* 82: 382–384.