Reliability of decision-making and reinforcement learning computational parameters

Anahit Mkrtchian*[1], Vincent Valton[1], Jonathan P. Roiser[1]

[1]Neuroscience and Mental Health Group, Institute of Cognitive Neuroscience, University College London, London, UK.

***Correspondence**

Anahit Mkrtchian

Institute of Cognitive Neuroscience, University College London

Alexandra House, 17-19 Queen Square

London, WC1N 3AZ

United Kingdom

Email: a.mkrtchian@ucl.ac.uk

Running Title: Reliability of computational parameters

Keywords: Reliability, Computational psychiatry, Decision-making, Reinforcement learning, Prospect theory, Computational modelling

Abstract

Background: Computational models can offer mechanistic insight into cognition and therefore have the potential to transform our understanding of psychiatric disorders and their treatment. For translational efforts to be successful, it is imperative that computational measures capture individual characteristics reliably. To date, this issue has received little consideration. **Methods:** Here we examine the reliability of canonical reinforcement learning and economic models derived from two commonly used tasks. Healthy individuals (N=50) completed a restless four-armed bandit and a calibrated gambling task twice, two weeks apart. **Results:** Reward and punishment processing parameters from the reinforcement learning model showed fair-to-good reliability, while risk/loss aversion parameters from a prospect theory model exhibited good-to-excellent reliability. Both models were further able to predict future behaviour above chance within individuals. **Conclusions:** These results suggest that reinforcement learning, and particularly prospect theory measures, represent relatively reliable decision-making mechanisms, which are also unique across individuals, indicating the translational potential of clinically-relevant computational parameters for precision psychiatry.

Introduction

Cognitive and neural processes are increasingly conceptualized in computational terms(1). Generative computational models offer the advantage of examining behaviourally unobservable, but important, latent processes that drive behaviour and can be closely linked to neurobiology(2, 3). As such, they provide a mathematically precise framework for specifying hypotheses about the cognitive processes that generate behaviour. These features make modelling a powerful tool to provide mechanistic accounts into (a)typical behaviours, including those associated with psychiatric symptoms.

The increasing adoption of computational approaches in cognitive neuroscience inspired the emerging discipline of computational psychiatry, which aims to better understand mental illness through computational methods with the ultimate goal of transforming such knowledge into new personalised treatment strategies (2, 4-16). For such translational endeavours to be successful, however, it is vital that computational measures capture individual characteristics reliably(5, 12).

One area of computational modelling that has received particular attention is the reward and punishment processes underlying decision-making(17). This progress has fostered research into the computational mechanisms that may drive psychiatric symptoms, as disrupted decision-making and reward processing are important features of many disorders(2, 18-26). In particular, two classes of computational models, respectively originating from computer science and behavioural economics, have been influential in characterizing the cognitive

mechanisms underlying decision-making: reinforcement learning (RL) and prospect theory(17, 27-32).

RL models describe how agents learn from rewards and punishments through trial-and-error(31). Within the field of computational psychiatry, this set of models has perhaps been the most influential. For example, reward and punishment sensitivity (reflecting subjective valuation of the outcomes) and learning rate (reflecting how quickly individuals learn from better- or worse-than-expected outcomes) have been associated with distinct symptomatology and neural signals(9, 23, 29, 33). A commonly-used RL task is the multi-armed “bandit”, in which individuals choose between multiple slot machines with fluctuating, unknown probabilities of reward and punishment, with the goal of maximizing earnings(34-38). Participants must decide on each trial whether to persist with the previously sampled slot machine or explore others which may yield better outcomes. The cognitive processes thought to underlie these decisions can be captured effectively by RL models(20, 35).

Prospect theory models, on the other hand, describe the cognitive processes driving decision-making biases under known risks, and have been extremely influential in understanding economic decision-making(28, 30, 32, 39, 40). These processes are often examined by asking participants to choose between a guaranteed outcome (e.g., £0 gain) and a 50% gamble with two possible outcomes (e.g., £30 gain or £10 loss). It is commonly observed that humans tend to prefer a sure payment over a risky payment with equivalent or higher expected value. For example, you may prefer an investment with a fixed return over one with a potentially higher

but uncertain return. Prospect theory proposes that these observations can be accounted for by two different processes: 1) risk aversion – the preference for certain over uncertain gains, and 2) loss aversion – weighting losses more heavily than gains. Risk and loss aversion vary across individuals, and these differences have been associated with psychiatric diagnoses and affective states(41-50). Importantly, computational modelling has allowed researchers to dissociate risk and loss aversion and their contribution to symptoms(43).

Parameters from RL and prospect theory models show promise in generating insights into the mechanisms underlying psychiatric symptoms. However, as of yet, the psychometric properties of computational parameters have received limited attention(5, 12, 51, 52). Model-agnostic measures derived from reward processing tasks (e.g., percent correct) often have modest or poor reliability(53-55), and the handful of studies that have examined the reliability of computational parameters have reported markedly varied results. For example, risk aversion parameters have been reported to be fairly stable in a decision-making task in both healthy and depressed individuals(56). Similarly, parameters from a prospect theory model based on a gambling task showed significant correlations over time(57, 58). In contrast, reliability assessment of parameters derived from a go/no-go RL task exhibited poor reliability(59), as did those from a two-step decision-making task assessing model-based versus model-free RL(60). Interestingly, the reliability of parameters derived from the latter task was substantially improved through hierarchical estimation procedures, suggesting that the estimation procedure matters(61). However, only one of the above studies(61) quantified test-retest reliability using the standard intraclass correlation coefficient (ICC), which allows assessing both

the relative rank and absolute agreement between measurements(62). Additionally, no studies to date have reported the reliability of decision-making processes derived from a bandit task with fluctuating ("restless") reward and punishment probabilities(35) or an individually calibrated gambling task(43).

A complementary perspective to understanding the reliability of computational cognitive models can be obtained through prediction. Generative models offer a substantial advantage in that they can both explain and predict behaviour. Unlike model-agnostic measures, computational parameters fit to one dataset should be able to predict future behaviour in the same individual. In other words, computational models can additionally be assessed by their ability to forecast future behaviour, equivalent to out-of-sample validation. This type of validation assesses model generalizability and is often referred to as predictive accuracy(57, 58, 63), but it has rarely been used as a metric of reliability. Therefore, in the current study we assessed the reliability of computational parameters derived from two widely-used decision-making tasks (a restless four-armed bandit and a calibrated gambling task) using both approaches: standard measures of stability and reliability (respectively, practice effects and ICCs); and out-of-sample predictive accuracy.

Methods and Materials

Participants

Fifty-four healthy participants were recruited from the UCL Institute of Cognitive Neuroscience Subject Database. Four participants were excluded for failing to complete the second session (final N=50: 32 females [64%]; age range=19-38; mean age=25.16, SD±5.48 years; mean education=17.38, SD=±3.24 years). Participants reported no current or past psychiatric or neurological disorder, cannabis use in the past 31 days, alcohol consumption in the past 24 hours, or any other recreational drug use in the week prior to participation. Participants provided written informed consent and were compensated at the end of their second session with a flat rate of £30 and a bonus of up to £20 based on task winnings. The study was approved by the UCL Psychology and Language Sciences Research Ethics Committee (Project ID Number: fMRI/2013/005).

Sample size was determined by an *a priori* power analysis in G*Power(64). The power analysis was based on the smallest effect size of interest, r=0.4, since reliability below this threshold is conventionally considered poor(65). Detecting an effect size of this magnitude, at the one-tailed 0.05 alpha level with 90% power, requires 47 participants.

Study procedure and Tasks

Participants completed a battery of computerized tasks, including a restless four-armed bandit(35, 36) and an individually calibrated gambling task(43) over two sessions (mean test-

retest interval = 13.96 days, SD=0.20). On each trial in the bandit task, participants chose one out of four bandits and received one out of four possible outcomes: reward, punishment, neither reward nor punishment or both reward and punishment. Win and loss probabilities fluctuated independently over time (Figure 1a; Supplemental Information). In the gambling task, participants chose between a 50-50 gamble and a sure option without receiving feedback. An initial training phase was used to create individually calibrated offers in a second phase. Trials were classified as either mixed (50% chance to win or lose money gamble, or a sure option of 0 points) or gain-only (50% chance to win or receive nothing gamble, or a variable sure gain; Figure 1b; Supplemental Information). Calibration failed for one participant resulting in N=49 participants for this task.

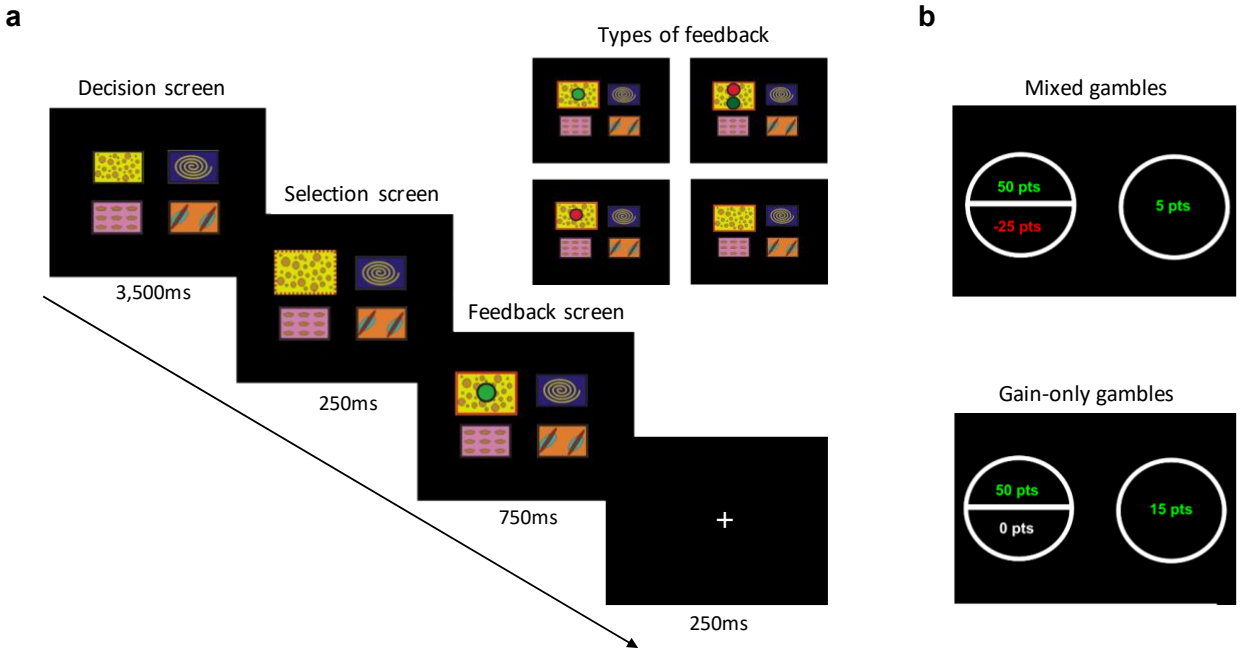

Figure 1. Four-armed bandit and gambling task. Example trial of the four-armed bandit task (a). On each trial, participants chose one out of four bandits and received one out of four possible outcomes: reward (green token), punishment (red token), neither reward nor punishment (empty box) or both reward and punishment (red and green token). On each gambling task trial, participants chose between a 50-50 gamble and a sure (guaranteed amount of points) option (b). Trials were either mixed gambles (50-50 chance of winning or losing points or sure option of 0 points) or gain-only trials (50-50 chance of winning or receiving nothing or sure gain).

Data analysis

Data were processed in Matlab (R2019b) and analysed in SPSS (v25, IBM Corp, Armonk, NY). Computational modelling was performed with the hBayesDM package for R (v. 3.6.0; https://github.com/CCS-Lab/hBayesDM)(66), which uses hierarchical Bayesian modelling in Stan (v.2.21.2). For all analyses, $p<0.05$ (two-tailed) was considered statistically significant. Cohen's d_z (within-subject) effect sizes are reported for practice effects (67). Details on model-agnostic task measures are in Supplemental Information.

Computational modelling

The bandit task data was fit with seven different RL models previously described in detail(20)(Supplemental Information). Three prospect theory models were fit to the gambling task(28, 43, 66, 68). Modelling was conducted on the second phase of the gambling task (i.e., on individually calibrated trials). The models were fit for each session separately, using separate hierarchical priors (group-level parameters), as this has shown to provide more accurate fits(69), and we wished to avoid artificially inflating reliability estimates. We also estimated model fits under a single hierarchical prior (session 1 and session 2 data together) as a

sensitivity analysis (Supplemental Information). Model comparison was performed with leave-one-out information criterion (LOOIC) where the winning model was the one with the lowest LOOIC. Several model validation checks were completed for the winning models(70-72)(Supplemental Information).

Reliability analysis

Test-retest reliability was assessed with ICCs (ratios of intra-individual to inter-individual variability (62, 73)), with values of <0.40 interpreted as poor, 0.4-0.6 as fair, 0.6-0.75 as good, and >0.75 as excellent reliability(65). A two-way mixed effects model based on single measures ICC was used (fixed effect: testing time-interval, random effect: subject). We report consistency ICCs here (assesses relative ranking of participants over time) but note that there is little difference compared with absolute ICCs (additionally assess changes in scores over time; Supplemental Information).

Posterior predictive performance

To assess to what extent an individual's future behaviour can be predicted using a generative model fit to their own task performance two weeks earlier, we calculated the probability of participants' choices on each trial (i.e., the softmax output), given their session 2 data and model parameter estimates from session 1. Probabilities were averaged across trials for each individual.

Since hierarchical parameter estimation produces ‘shrinkage’, effectively pulling parameter estimates from different individuals closer to each other (which improves estimation accuracy), it is possible that future performance may also be predicted above-chance using other participants’ parameter estimates from session 1 (e.g., participant A’s parameter estimates from session 1 predicting participant B’s session 2 choices). We therefore assessed whether using an individual’s model parameter estimates from session 1 predicted the same individual’s choices on session 2 better than using all other subjects’ model parameter estimates. To construct the latter measure, for each subject, we predicted trial-by-trial choices on session 2 based on parameter estimates from every other participant’s session 1 model, and averaged the probabilities across all participants.

Results

Four-armed bandit task: Model-agnostic results

Basic behaviour and practice effects

As expected, there was a main effect of outcome type on behaviour ($F_{(2,98)}$=117.39, p<0.001, η_p^2=0.71; Figure 2a). The probability to repeat a choice was significantly greater after wins compared with both losses and outcomes on which neither wins nor losses occurred, and greater after neither compared with losses (all p<0.001). There was no significant main effect of testing session ($F_{(1, 49)}$=0.01, p=0.91, η_p^2<0.001), but there was a significant outcome-by-session interaction ($F_{(2,98)}$=3.12, p=0.049, η_p^2=0.06), reflecting slightly increased repeated choices after wins and decreased repeated choices after losses. However, the difference in the tendency to repeat a choice between session 1 and session 2 did not reach significance following any of the outcome types (loss: $t_{(49)}$=1.45, p=0.15, d_z=0.21; win: $t_{(49)}$=0.87, p=0.39, d_z=0.12; neither: $t_{(49)}$=0.54, p=0.59, d_z=0.08), and therefore we do not interpret this result further.

Test-retest reliability

The model-agnostic outcome measures of the bandit task exhibited good reliability (Figure 2b).

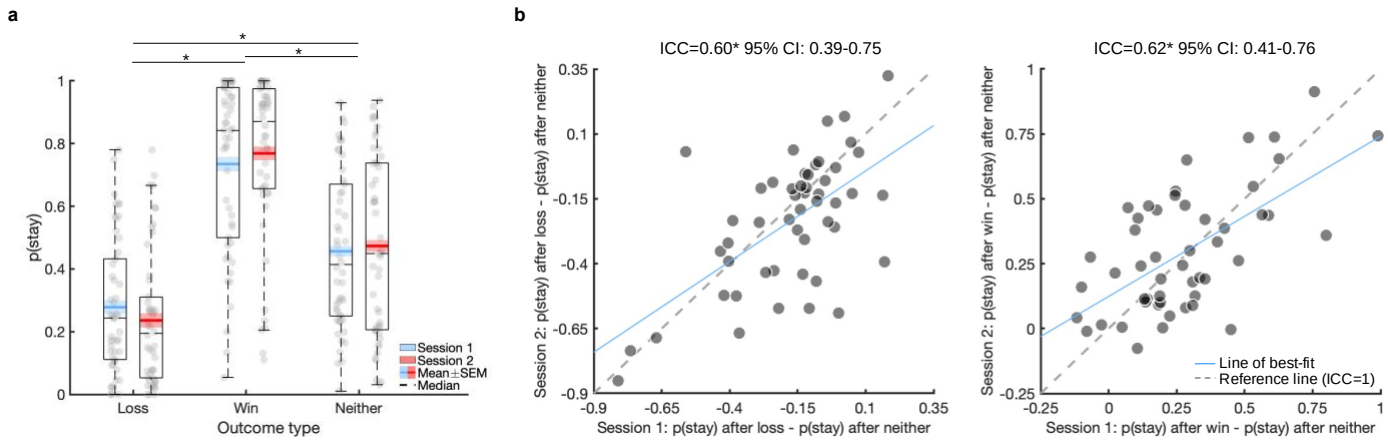

Figure 2: Basic behaviour, practice effects, and test-retest reliability of model-agnostic measures on the four-armed bandit task. Boxplots of the four-armed bandit task showing probability to stay after a certain outcome in session 1 and 2 (a). The probability to stay was significantly different after each outcome type (Loss<Neither<Win) but no clear practice effect was evident. Scatter plots and reliability of the four-armed bandit task model-agnostic measures comparing behaviour on two testing sessions approximately 2 weeks apart (b). Consistency (assesses relative ranking over time) intraclass correlation coefficients (ICC) are presented. Lightly shaded regions in Figure 2a represent within-subjects standard error of the mean (SEM). CI: confidence interval. *p<0.001

Four-armed bandit task: Modelling results

Model comparison indicated that the winning (most parsimonious) model was the five-parameter 'Bandit4arm_lapse' (nomenclature from the hBayesDM package) model, with reward and punishment learning rate parameters, reward and punishment sensitivity parameters and a lapse parameter (final parameter captures random responding; Table S1), consistent with previous reports(20). Several model diagnostics were conducted to check the model performance (Figure S1-5). Three individuals were excluded due to difficulties in obtaining mean parameter estimates, as multiple peaks were evident in the posterior distribution of at least one parameter (Figure S2). The Bandit4arm_lapse model was therefore

re-fit without these participants. Excluding these participants did not affect test-retest reliability inference (Table S2). All parameters other than the lapse parameter showed high recoverability (Figure S3), and synthetic data from the winning model accurately recapitulated real data (Figure S4).

Practice effects

On session 2 there were significant increases in the reward sensitivity ($t_{(46)}$=3.00, p=0.004, d_z=0.44) and lapse parameters ($t_{(46)}$=8.88, p<0.001, d_z=1.29), but not on any of the other parameters (reward learning rate: $t_{(46)}$=1.28, p=0.21, d_z=0.19; punishment learning rate: $t_{(46)}$=1.74, p=0.09, d_z=0.25; punishment sensitivity: $t_{(46)}$=1.28, p=0.21, d_z=0.19; Figure 3a). However, there were no significant practice effects when the data was fit under a single hierarchical prior (Table S3).

Test-retest reliability

All estimated Bandit4arm_lapse model parameters, except the lapse parameter, demonstrated fair-to-good reliability (Figure 3b), which did not substantially change when parameters were estimated under a single hierarchical prior (Table S3).

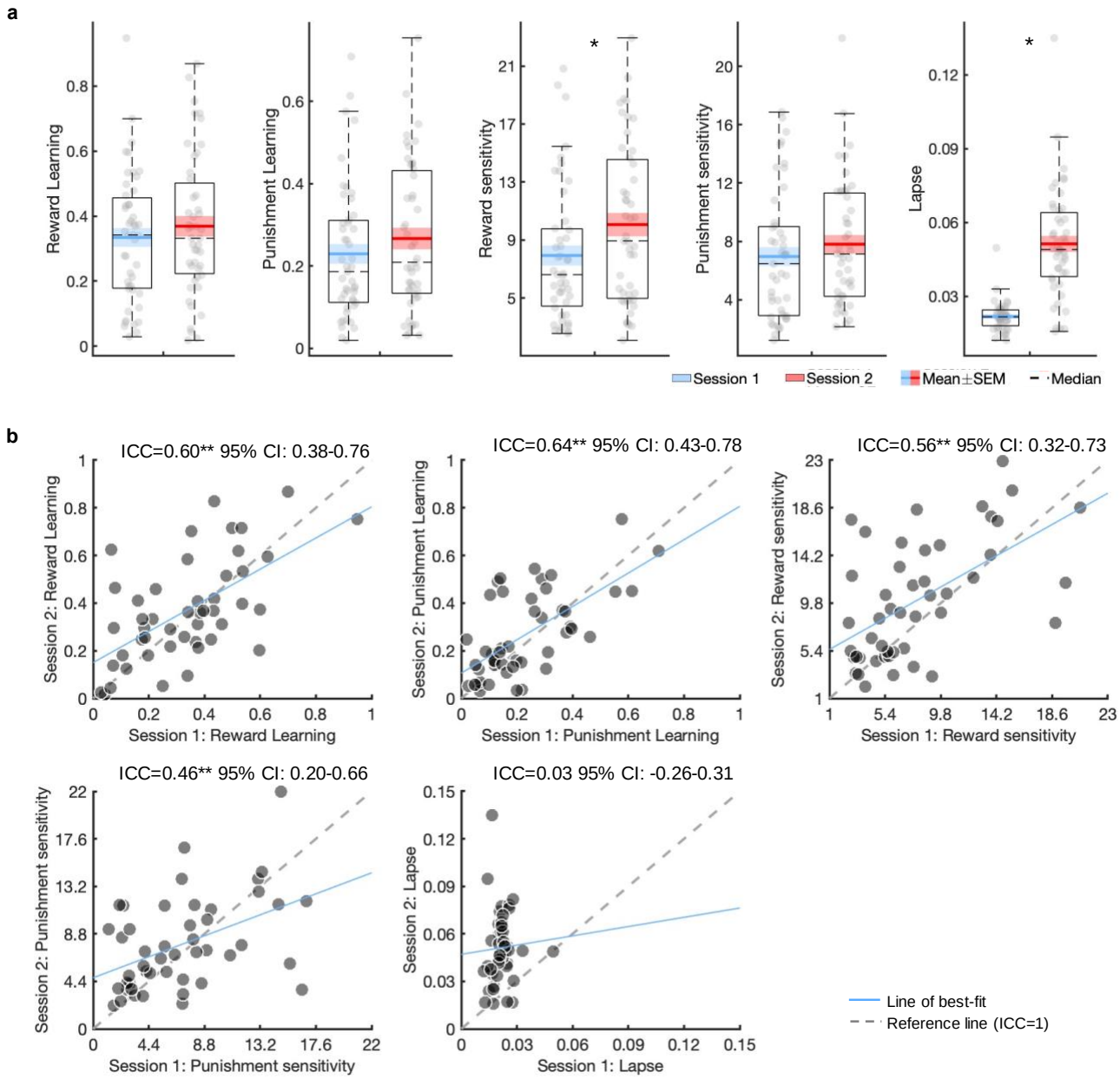

Figure 3: Practice effects and test-retest reliability of the winning reinforcement learning model parameters derived from the four-armed bandit task. Boxplots show point estimates of the Bandit4arm_lapse model parameters in session 1 and 2, fit under separate priors (a). Scatter plots and reliability of the Bandit4arm_lapse model parameters over session 1 and 2 are presented (b). Consistency (assesses relative ranking score over time) intraclass correlation coefficients (ICC) are presented. SEM: standard error of the mean; CI: confidence interval; S1: session 1; S2: session 2. *p<0.05, **p<0.001.

Posterior predictive performance

Parameter estimates from session 1 predicted task performance on session 2 substantially better than chance (mean = 42%, chance = 25% accuracy; $t_{(46)}$=9.10, p<0.001; Figure 4a), indicating that the model could predict future choices by using a generative model fit to the same participants' data two weeks earlier. Using an individual's parameter estimates to predict their own future choices was significantly better than when that prediction was based on the average of the other participants' session 1 estimates ($t_{(46)}$=3.20, p=0.003; Figure 4b), signifying good individual-level model generalizability.

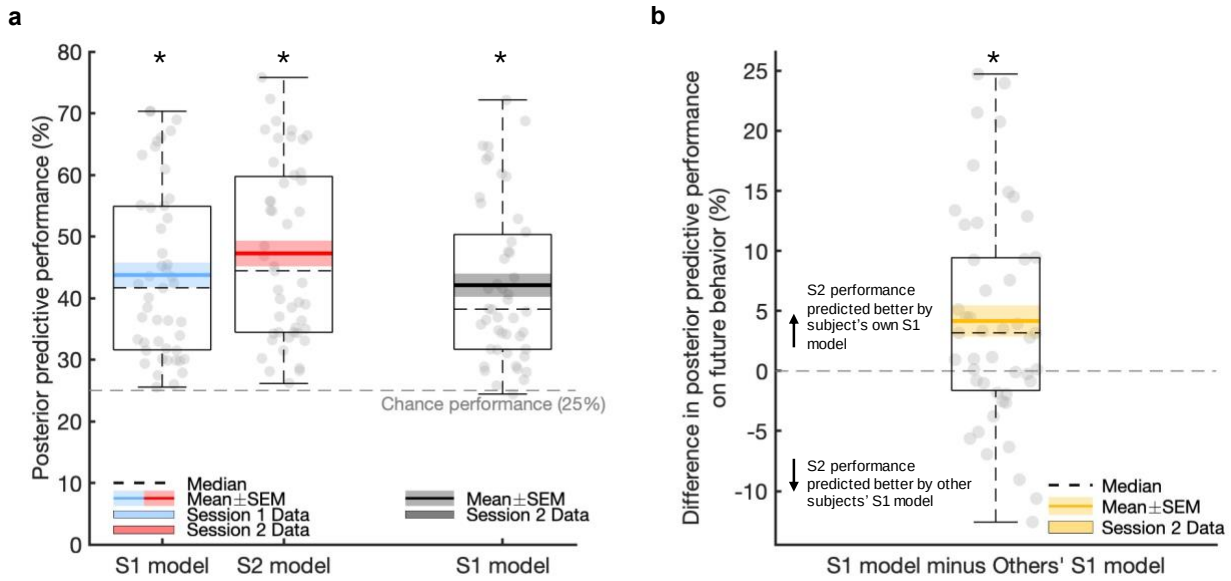

Figure 4: Posterior predictive performance of the winning reinforcement learning model derived from the four-armed bandit task. Boxplots depicting accuracy of bandit4arm_lapse model in predicting choices (a). Model estimates from session 1 (S1) predicted future session 2 (S2) behaviour above chance (black boxplot). Both S1 and S2 model estimates also predicted behaviour on the same session significantly above chance (blue and red

boxplots). Predicting future performance (session 2 data) using a participant's own model parameter estimates was significantly better than using other participants' S1 model parameter estimates (b – zero indicates mean accuracy from other participants' parameter estimates). SEM: standard error of the mean. *p<0.01.

Gambling task: Model-agnostic results

Basic behaviour and practice effects

As expected, propensity to gamble was significantly higher on mixed gambles ($F_{(1, 48)}$=13.71, p=0.001, η_p^2=0.22). There were no significant main ($F_{(1, 48)}$=0.76, p=0.40, η_p^2=0.02) or interaction ($F_{(1, 48)}$=1.07, p=0.31, η_p^2=0.02) effects of session on the propensity to gamble (session differences: probability to gamble on mixed trials $t_{(48)}$=0.23, p=0.82, d_z=0.03; probability to gamble on gain-only trials $t_{(48)}$=1.51, p=0.14, d_z=0.22; Figure 5a).

Test-retest reliability

Model-agnostic outcome measures on the gambling task exhibited good reliability (Figure 5b).

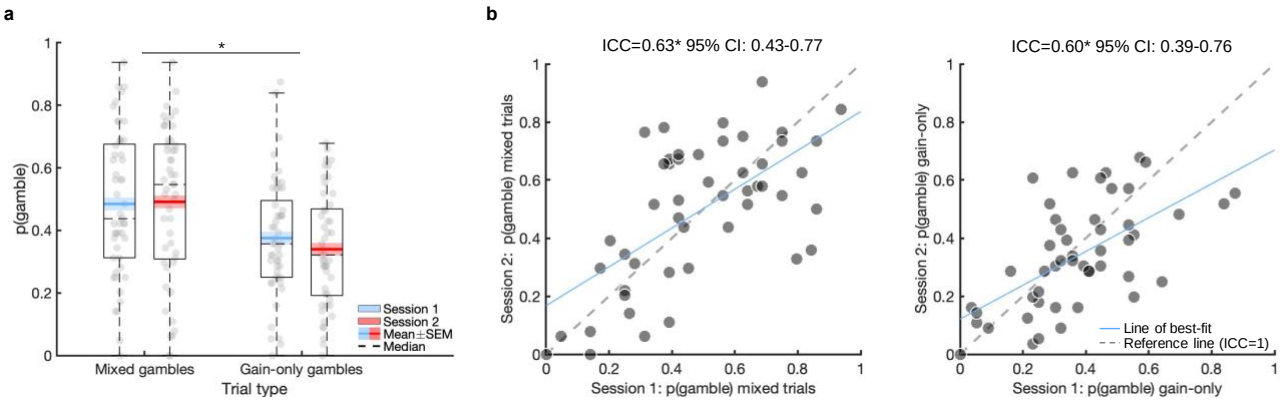

Figure 5: Basic behaviour, practice effects, and test-retest reliability of model-agnostic measures on the gambling task. Boxplots show the probability to gamble based on the trial type in session 1 and 2, with no significant session effects (a). Scatter plots and reliability of the gambling task model-agnostic measures over session 1 and 2 (b). Consistency (assesses relative score ranking over time) intraclass correlation coefficients (ICC) are presented. Lightly shaded regions in Figure 5a represent within-subjects standard error of the mean (SEM). CI: confidence interval. *p<0.001.

Gambling task: Modelling results

The winning model was the prospect theory model ('ra_prospect' in the hBayesDM package) with loss aversion, risk aversion and inverse temperature parameters (last parameter represents choice consistency; Table S4), consistent with previous reports(43). A loss aversion parameter above 1 represents overweighting of losses to gains, while a risk aversion parameter less than 1 indicates aversion to risk. Model diagnostics indicated that this model performed well (Figure S6). All parameters showed high recoverability (Figure S7), and synthetic data from the winning model accurately recapitulated real data (Figure S8). Neither test-retest nor practice effects were substantially altered when the model was fit under a single hierarchical prior (Table S5).

Practice effects

There were significant session effects on all prospect theory model parameters (on session 2: decreased loss aversion: $t_{(48)}=2.17$, $p=0.04$, $d_z=0.31$; decreased risk aversion: $t_{(48)}=4.04$, $p<0.001$, $d_z=0.58$; increased inverse temperature: $t_{(48)}=3.07$, $p=0.004$, $d_z=0.44$; Figure 6a).

Test-retest reliability

All estimated parameters demonstrated good-to-excellent reliability (Figure 6b).

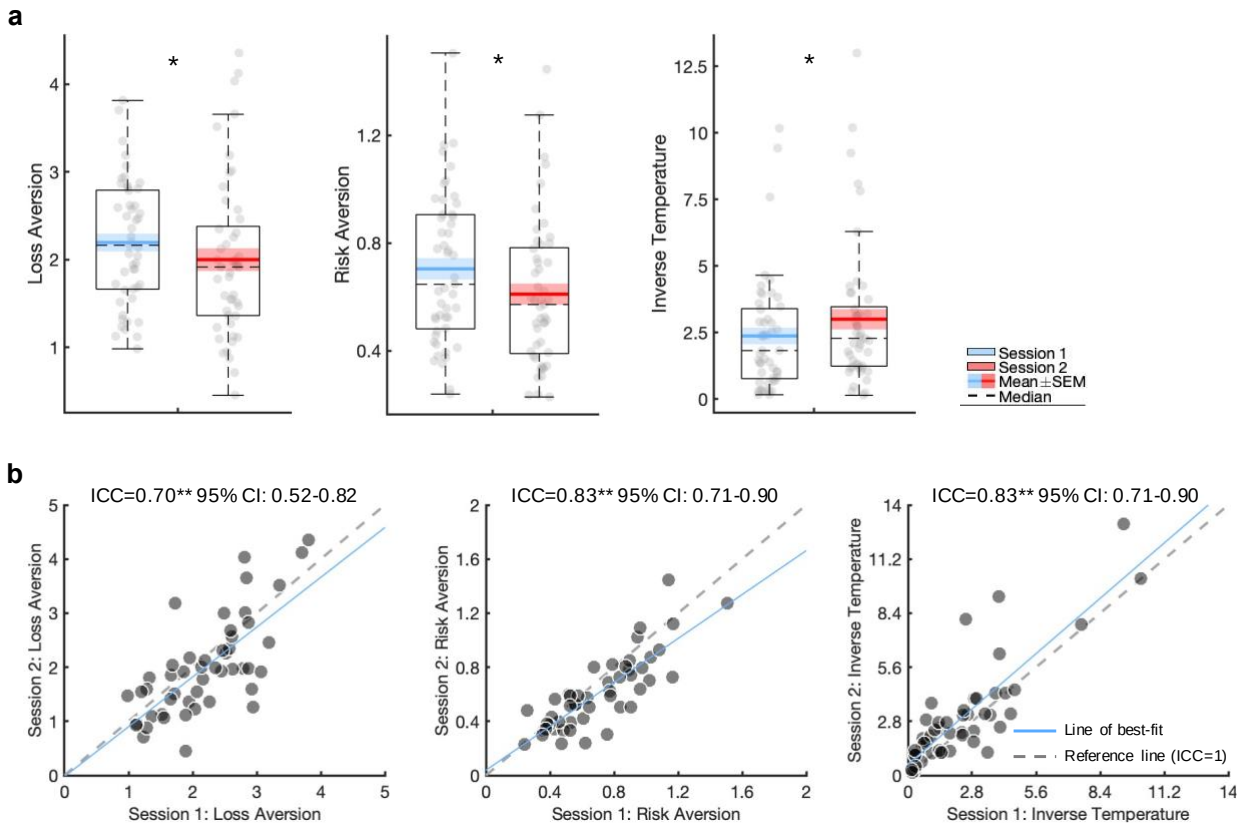

Figure 6: Practice effects and test-retest reliability of the prospect theory model derived from the gambling task. Boxplots show point estimates of the prospect theory model parameters in session 1 and 2, fit under separate

priors (a). Scatter plots and reliability of the prospect theory model parameters over session 1 and 2 are presented (b). Consistency (assesses relative score ranking over time) intraclass correlation coefficients (ICC) are presented. SEM: standard error of the mean; CI: confidence interval; S1: session 1; S2: session 2. *p<0.05, **p<0.001.

Posterior predictive performance

Prospect theory model parameters from session 1 predicted future choices at session 2 significantly above chance (mean = 68%, chance = 50% accuracy; $t_{(48)}$= 12.08, p<0.001; Figure 7a). Predicting future performance at session 2 was significantly higher when based on participants' own parameter estimates from session 1 compared with model parameter estimates of other participants from session 1 ($t_{(48)}$=8.38, p<0.001; Figure 7b).

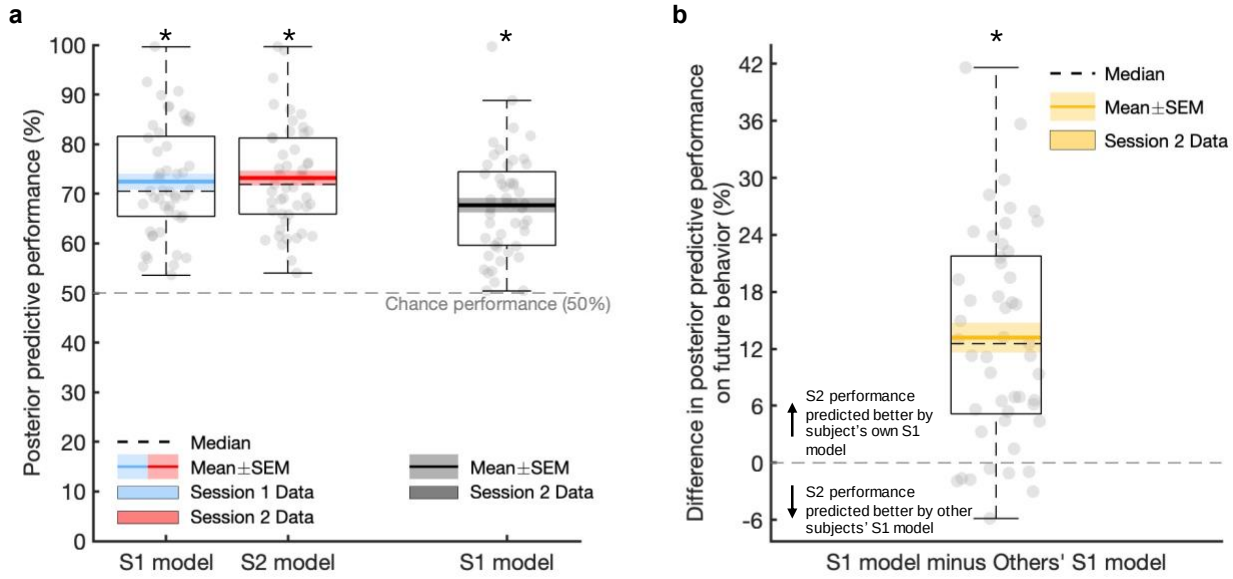

Figure 7: Posterior predictive performance of the prospect theory model derived from the gambling task. Boxplots depicting accuracy of prospect theory model in predicting choices (a). Session 1 (S1) model estimates predicted S1 behaviour significantly above chance (blue boxplot), as did session 2 (S2) model estimates on S2 data (red boxplot). Importantly, model parameter estimates from S1 predicted task performance from S2 above chance

(black boxplot). Predicting future S2 performance using a participant's own S1 model parameter estimates was significantly better than using other participants' S1 model parameter estimates (b – zero indicates mean accuracy from other participants' parameter estimates). SEM: standard error of the mean. *p<0.001.

Discussion

Reliability has garnered increased attention in recent years, with worryingly low reliability across conventional measures from cognitive tasks and functional neuroimaging(53, 74-78). However, few attempts have been made to examine the reliability of computational cognitive measures. Here we assessed the psychometric properties of computational models derived from a restless four-armed bandit and a calibrated gambling task. Overall, most parameters reflecting RL and decision-making processes exhibited adequate reliability and predicted future performance well. These results provide promise for their use in clinical settings. However, this conclusion depends on the specific parameters assessed in each task, highlighting the complexities of translating tasks to the clinic.

Reward and punishment learning rates from the bandit task demonstrated good reliability while reward and punishment sensitivity showed fair reliability, suggesting that this task may be more suitable for assessing learning rates than sensitivity. Elevated punishment learning rates (faster learning in the face of negative outcomes) and lapse values (random responding or exploration) have been associated with greater mood and anxiety symptoms, representing potential measurable mechanistic treatment targets (20). However, the present study suggests that the lapse parameter, which exhibited poor reliability as assessed by the bandit task, may not be suitable as a potential target. This parameter measures responding not captured by the model (including goal-directed and random exploration), and the sources of this 'noise' might differ across sessions. It is therefore perhaps unsurprising that this parameter was unreliable.

Crucially, the lapse parameter showed poor recoverability, which places an upper limit on its potential reliability. Some of this poor recoverability may be explained by limited lapse variation, especially in session 1 (Figure S3). The distribution of the group-level standard deviation lapse parameter was biased towards smaller values here such that the lapse parameter did not vary greatly between individuals (Figure S5). This suggests that the lapse parameter could be replaced with a constant and inference on this parameter is not advised.

Although no prior studies have specifically investigated ICC properties of the current RL model, one previous study found similarly poor reliability of the lapse parameter across six months in a go/no-go RL model in adolescents(59). In contrast to our results, this study also reported poor reliability of both reward and punishment learning rates. These differences may arise for a multitude of reason. Importantly, however, we provide evidence that it is possible to achieve at least moderate reliability for some canonical RL parameters.

Interestingly, the model-agnostic outcome measures of the bandit task exhibited similar reliability to the computational measures. Model-agnostic measures of cognitive tasks have often been reported to exhibit poor-to-moderate reliability(53, 55, 78, 79). It has been argued that this may be due to their inability to capture the generative process underlying task performance(23, 80). Our results suggest that it should not be assumed that computational parameters will always provide greater reliability than non-computational ones. However, the model-agnostic outcome measures are only a proxy of the processes the bandit task assesses, as it is difficult to compute model-agnostic equivalents of some parameters, such as

reward/punishment sensitivity. Indeed, models make explicit and falsifiable predictions of the components driving behaviour, which can be refined and used to simulate artificial data to generate new predictions. Thus, computational modelling is a more rigorous and preferable method for assessing behaviour than model-agnostic measures, which unlike computational methods, lack the mechanistic insights into the underlying processes generating behaviour.

While reliability of parameters ranged from poor-to-good in the bandit task, all parameters from the gambling task showed good-to-excellent test-retest reliability. These were also substantially higher than the reliability of the model-agnostic measures, suggesting that computational models may offer advantages in psychometric properties here, in addition to their advantage in specifying mechanisms. In particular, the risk aversion parameter, which has previously been associated with anxiety(43), exhibited excellent reliability (ICC>0.8), providing promise for use in clinical research. These results are consistent with previous studies, but we generally found considerably higher reliability than previously reported (previous reliability reports: loss aversion r≈0.25-0.61, risk aversion r≈0.50-0.60, inverse temperature r≈0.30-0.60(56-58)). These studies all used different estimation procedures, including hierarchical Bayesian, and employed both longer and shorter testing time-windows than the current study, suggesting that these factors may not fully explain the differences. It is possible that our results instead stem from different prospect model specifications(56, 58), as well as different task designs.

The mean loss and risk aversion parameters observed are consistent with previous literature indicating that losses are weighed about twice as much as gains, and that people are on average risk averse(28, 30, 32, 40). Indeed, prospect theory models have shown to be highly replicable across different contexts(39). This implementation of the gambling task may therefore represent a case where group-average results are highly reproducible but the task is also suitable to assess individual differences. It has been observed that this is rarely the case, since tasks that are designed to produce reliable average effects do so by minimizing between-subject variance, whereas what makes tasks reliable and suitable for inter-individual assessments are large between-subject but small within-subject variances(79). This may have occurred because of the calibrated design of the present gambling task. A similar approach of dynamically updating parameter values to each individual during task performance has previously been suggested as a solution to unreliable cognitive tasks(81).

In addition to examining reliability using the traditional ICC method, we also examined how well the models predicted future task performance. This approach provides complementary perspective on reliability, which is unique to computationally-informed measures. Generative models were consistently able to predict participants' future behaviour above chance. Notably, on average, each participant's own parameter estimates from the first session were better at predicting their future performance compared with using parameter estimates from all other participants. This indicates that individuals may also reliably differ in the cognitive mechanisms underlying their decisions, and offers reassurance that hierarchical estimation procedures are suitable for inter-individual inferences(58, 61, 70). In other words, individuals show relatively

unique computational decision-making profiles. This is consistent with two previous studies using a different prospect theory model and gambling task(57, 58).

Some of the reliable parameters, however, showed small-to-medium session effects. Practice effects cans either obscure a true effect or lead to false treatment claims if appropriate controls are not employed. Quantifying session effects, however, allows such changes to be accounted for. In the RL model, reward sensitivity increased in the second session, while all other reliable RL processes were fairly stable. In contrast, all prospect theory parameters showed significant session effects. However, the sensitivity analysis showed no substantial practice effects of RL parameters when data were fit under a single prior, while the practice effects of prospect theory parameters remained. Thus, the apparent practice effects on the RL task should be interpreted with caution, as these may be overestimated under the two prior estimation approach(69). Importantly, the ICC estimates were hardly influenced by the estimation approach. Although practice effects were evident in the computational parameters, no clear changes were observed in the model-agnostic measures, suggesting further that model-agnostic measures lack the precision to quantify these decision-making processes.

A potential limitation of our study is the sample tested, as the reliability of tasks in healthy individuals may differ from that in clinical groups. Similarly, our results only speak to reliability over two weeks. Thus, it is possible that longer time periods may produce lower reliability, which should be assessed in future studies.

The RL parameters showed relatively modest reliability, suggesting that these processes are more vulnerable to state influences than prospect theory parameters. A consequence of this is that larger sample sizes may be required to examine effects, as effect sizes would be expected to be lower, relative to prospect theory tasks(65). Interpreting the marked difference in reliability between the prospect theory and RL models is not straightforward, as these models measure complementary aspects of cognition. The bandit task is a learning paradigm that requires constant updating of optimal choices. It is possible that in the first session individuals had not yet stabilized on a cognitive strategy and were still learning the task structure, as indicated by lower evidence of the winning model in session 1 compared with session 2 (Table S1). It would be interesting to explore if an initial baseline session, familiarizing participants with the task, would improve test-retest reliability. It is also possible that estimating test-retest reliability within a generative model, may provide a more accurate theoretical representation of reliability(61, 82); for this reason we may have slightly underestimated reliability.

In summary, we show that commonly-used computational parameters derived from an RL 'restless' bandit task and a calibrated gambling task exhibit fair-to-excellent reliability. These models can further be used to predict future behaviour in the same individuals, indicating that the RL and decision-making processes assessed in these tasks represent relatively consistent and unique characteristics of an individual. These findings take us one step closer to translating computational measures of behaviour into clinical application.

Acknowledgements

This work was supported by a Wellcome Trust - NIH PhD studentship (200934/Z/16/Z) to A.M.; a National Institute for Health Research (NIHR) Biomedical Research Center (BRC) fellowship to V.V.; and by a Wellcome Investigator Award (101798/Z/13/Z) to J.P.R. The funders had no role in the study design, data collection and analysis, decision to publish or preparation of the manuscript. The authors thank Eoin Travers and Oliver J. Robinson for valuable input on an earlier version of the manuscript.

Disclosures

The authors declare no competing interests.

References

1. Palminteri S, Wyart V, Koechlin E (2017): The Importance of Falsification in Computational Cognitive Modeling. *Trends Cogn Sci*. 21:425-433.
2. Huys QJ, Maia TV, Frank MJ (2016): Computational psychiatry as a bridge from neuroscience to clinical applications. *Nat Neurosci*. 19:404-413.
3. Montague PR, Hyman SE, Cohen JD (2004): Computational roles for dopamine in behavioural control. *Nature*. 431:760-767.
4. Adams RA, Huys QJ, Roiser JP (2016): Computational Psychiatry: towards a mathematically informed understanding of mental illness. *J Neurol Neurosurg Psychiatry*. 87:53-63.
5. Browning M, Carter CS, Chatham C, Den Ouden H, Gillan CM, Baker JT, et al. (2020): Realizing the Clinical Potential of Computational Psychiatry: Report From the Banbury Center Meeting, February 2019. *Biol Psychiatry*. 88:e5-e10.
6. Friston KJ, Redish AD, Gordon JA (2017): Computational Nosology and Precision Psychiatry. *Comput Psychiatr*. 1:2-23.
7. Huys QJ, Moutoussis M, Williams J (2011): Are computational models of any use to psychiatry? *Neural Netw*. 24:544-551.
8. Huys QJM (2018): Advancing Clinical Improvements for Patients Using the Theory-Driven and Data-Driven Branches of Computational Psychiatry. *JAMA Psychiatry*. 75:225-226.
9. Maia TV, Frank MJ (2011): From reinforcement learning models to psychiatric and neurological disorders. *Nat Neurosci*. 14:154-162.

10. Montague PR, Dolan RJ, Friston KJ, Dayan P (2012): Computational psychiatry. *Trends Cogn Sci*. 16:72-80.

11. Patzelt EH, Hartley CA, Gershman SJ (2018): Computational Phenotyping: Using Models to Understand Individual Differences in Personality, Development, and Mental Illness. *Personal Neurosci*. 1:e18.

12. Paulus MP, Huys QJ, Maia TV (2016): A Roadmap for the Development of Applied Computational Psychiatry. *Biol Psychiatry Cogn Neurosci Neuroimaging*. 1:386-392.

13. Paulus MP, Thompson WK (2019): Computational approaches and machine learning for individual-level treatment predictions. *Psychopharmacology (Berl)*.

14. Teufel C, Fletcher PC (2016): The promises and pitfalls of applying computational models to neurological and psychiatric disorders. *Brain*. 139:2600-2608.

15. Wang XJ, Krystal JH (2014): Computational psychiatry. *Neuron*. 84:638-654.

16. Wiecki TV, Poland J, Frank MJ (2015): Model-Based Cognitive Neuroscience Approaches to Computational Psychiatry. *Clinical Psychological Science*. 3:378-399.

17. Maia TV (2009): Reinforcement learning, conditioning, and the brain: Successes and challenges. *Cogn Affect Behav Neurosci*. 9:343-364.

18. (2013): *Diagnostic and statistical manual of mental disorders: DSM-5*. Fifth ed. Arlington, VA: American Psychiatric Association.

19. Eshel N, Roiser JP (2010): Reward and punishment processing in depression. *Biol Psychiatry*. 68:118-124.

20. Aylward J, Valton V, Ahn WY, Bond RL, Dayan P, Roiser JP, et al. (2019): Altered learning under uncertainty in unmedicated mood and anxiety disorders. *Nat Hum Behav*. 3:1116–1123.

21. Browning M, Behrens TE, Jocham G, O'Reilly JX, Bishop SJ (2015): Anxious individuals have difficulty learning the causal statistics of aversive environments. *Nat Neurosci*. 18:590-596.

22. Gillan CM, Kosinski M, Whelan R, Phelps EA, Daw ND (2016): Characterizing a psychiatric symptom dimension related to deficits in goal-directed control. *Elife*. 5.

23. Huys QJM, Browning M, Paulus MP, Frank MJ (2021): Advances in the computational understanding of mental illness. *Neuropsychopharmacology*. 46:3-19.

24. Mkrtchian A, Aylward J, Dayan P, Roiser JP, Robinson OJ (2017): Modeling Avoidance in Mood and Anxiety Disorders Using Reinforcement Learning. *Biol Psychiatry*. 82:532-539.

25. Mkrtchian A, Roiser JP, Robinson OJ (2017): Threat of shock and aversive inhibition: Induced anxiety modulates Pavlovian-instrumental interactions. *J Exp Psychol Gen*. 146:1694-1704.

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. Dayan P, Niv Y (2008): Reinforcement learning: the good, the bad and the ugly. *Curr Opin Neurobiol*. 18:185-196.

28. Kahneman D, Tversky A (1979): Prospect Theory: An Analysis of Decision under Risk. *Econometrica*. 47:263-291.

29. Niv N (2009): Reinforcement learning in the brain. *Journal of Mathematical Psychology*. 53:139-154.

30. Schonberg T, Fox CR, Poldrack RA (2011): Mind the gap: bridging economic and naturalistic risk-taking with cognitive neuroscience. *Trends Cogn Sci*. 15:11-19.

31. Sutton RS, Barto AG (2018): *Reinforcement learning: An introduction*. Second ed. Cambridge, MA: MIT Press.

32. Tversky A, Kahneman D (1992): Advances in Prospect Theory: Cumulative Representation of Uncertainty. *Journal of Risk and Uncertainty*. 5:297-323.

33. Daw ND, Doya K (2006): The computational neurobiology of learning and reward. *Curr Opin Neurobiol*. 16:199-204.

34. Cohen JD, McClure SM, Yu AJ (2007): Should I stay or should I go? How the human brain manages the trade-off between exploitation and exploration. *Philos Trans R Soc Lond B Biol Sci*. 362:933-942.

35. Daw ND, O'Doherty JP, Dayan P, Seymour B, Dolan RJ (2006): Cortical substrates for exploratory decisions in humans. *Nature*. 441:876-879.

36. Seymour B, Daw ND, Roiser JP, Dayan P, Dolan R (2012): Serotonin selectively modulates reward value in human decision-making. *J Neurosci*. 32:5833-5842.

37. Speekenbrink M, Konstantinidis E (2015): Uncertainty and exploration in a restless bandit problem. *Top Cogn Sci*. 7:351-367.

38. Yi MSK, Steyvers M, Lee M (2009): Modeling Human Performance in Restless Bandits with Particle Filters. *The Journal of Problem Solving*. 2.

39. Ruggeri K, Ali S, Berge ML, Bertoldo G, Bjorndal LD, Cortijos-Bernabeu A, et al. (2020): Replicating patterns of prospect theory for decision under risk. *Nat Hum Behav*. 4:622-633.

40. Sokol-Hessner P, Rutledge RB (2019): The Psychological and Neural Basis of Loss Aversion. *Current Directions in Psychological Science*. 28:20-27.

41. Baek K, Kwon J, Chae JH, Chung YA, Kralik JD, Min JA, et al. (2017): Heightened aversion to risk and loss in depressed patients with a suicide attempt history. *Sci Rep*. 7:11228.

42. Brown JK, Waltz JA, Strauss GP, McMahon RP, Frank MJ, Gold JM (2013): Hypothetical decision making in schizophrenia: the role of expected value computation and "irrational" biases. *Psychiatry Res*. 209:142-149.

43. Charpentier CJ, Aylward J, Roiser JP, Robinson OJ (2017): Enhanced Risk Aversion, But Not Loss Aversion, in Unmedicated Pathological Anxiety. *Biol Psychiatry*. 81:1014-1022.

44. Charpentier CJ, De Martino B, Sim AL, Sharot T, Roiser JP (2016): Emotion-induced loss aversion and striatal-amygdala coupling in low-anxious individuals. *Soc Cogn Affect Neurosci*. 11:569-579.

45. Charpentier CJ, De Neve JE, Li X, Roiser JP, Sharot T (2016): Models of Affective Decision Making: How Do Feelings Predict Choice? *Psychol Sci*. 27:763-775.

46. Hadlaczky G, Hokby S, Mkrtchian A, Wasserman D, Balazs J, Machin N, et al. (2018): Decision-Making in Suicidal Behavior: The Protective Role of Loss Aversion. *Front Psychiatry*. 9:116.

47. Hartley CA, Phelps EA (2012): Anxiety and decision-making. *Biol Psychiatry*. 72:113-118.

48. Klaus F, Chumbley JR, Seifritz E, Kaiser S, Hartmann-Riemer M (2020): Loss Aversion and Risk Aversion in Non-Clinical Negative Symptoms and Hypomania. *Front Psychiatry*. 11:574131.

49. Sip KE, Gonzalez R, Taylor SF, Stern ER (2017): Increased Loss Aversion in Unmedicated Patients with Obsessive-Compulsive Disorder. *Front Psychiatry*. 8:309.

50. Tremeau F, Brady M, Saccente E, Moreno A, Epstein H, Citrome L, et al. (2008): Loss aversion in schizophrenia. *Schizophr Res*. 103:121-128.

51. Ahn WY, Busemeyer JR (2016): Challenges and promises for translating computational tools into clinical practice. *Curr Opin Behav Sci*. 11:1-7.

52. Nair A, Rutledge RB, Mason L (2020): Under the Hood: Using Computational Psychiatry to Make Psychological Therapies More Mechanism-Focused. *Front Psychiatry*. 11:140.

53. Enkavi AZ, Eisenberg IW, Bissett PG, Mazza GL, MacKinnon DP, Marsch LA, et al. (2019): Large-scale analysis of test-retest reliabilities of self-regulation measures. *Proc Natl Acad Sci U S A*. 116:5472-5477.

54. Bland AR, Roiser JP, Mehta MA, Schei T, Boland H, Campbell-Meiklejohn DK, et al. (2016): EMOTICOM: A Neuropsychological Test Battery to Evaluate Emotion, Motivation, Impulsivity, and Social Cognition. *Front Behav Neurosci*. 10:25.

55. Plichta MM, Schwarz AJ, Grimm O, Morgen K, Mier D, Haddad L, et al. (2012): Test-retest reliability of evoked BOLD signals from a cognitive-emotive fMRI test battery. *Neuroimage*. 60:1746-1758.

56. Chung D, Kadlec K, Aimone JA, McCurry K, King-Casas B, Chiu PH (2017): Valuation in major depression is intact and stable in a non-learning environment. *Sci Rep*. 7:44374.

57. Glockner A, Pachur T (2012): Cognitive models of risky choice: parameter stability and predictive accuracy of prospect theory. *Cognition*. 123:21-32.

58. Scheibehenne B, Pachur T (2015): Using Bayesian hierarchical parameter estimation to assess the generalizability of cognitive models of choice. *Psychon Bull Rev*. 22:391-407.

59. Moutoussis M, Bullmore ET, Goodyer IM, Fonagy P, Jones PB, Dolan RJ, et al. (2018): Change, stability, and instability in the Pavlovian guidance of behaviour from adolescence to young adulthood. *PLoS Comput Biol*. 14:e1006679.

60. Shahar N, Hauser TU, Moutoussis M, Moran R, Keramati M, consortium N, et al. (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.

61. 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.

62. Koo TK, Li MY (2016): A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *J Chiropr Med*. 15:155-163.

63. Busemeyer JR, Wang YM (2000): Model Comparisons and Model Selections Based on Generalization Criterion Methodology. *J Math Psychol*. 44:171-189.

64. Faul F, Erdfelder E, Lang AG, Buchner A (2007): G*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behav Res Methods*. 39:175-191.

65. Fleiss JL (2011): Reliability of Measurement. *The Design and Analysis of Clinical Experiments*, pp 1-32.

66. Ahn WY, Haines N, Zhang L (2017): Revealing Neurocomputational Mechanisms of Reinforcement Learning and Decision-Making With the hBayesDM Package. *Comput Psychiatr*. 1:24-57.

67. Lakens D (2013): Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and ANOVAs. *Frontiers in Psychology*. 4:863.

68. 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 U S A*. 106:5035-5040.

69. Valton V, Wise T, Robinson OJ (2020): Recommendations for Bayesian hierarchical model specifications for case-control studies in mental health. *Machine Learning for Health (ML4H) at NeurIPS 2020, 34th Conference on Neural Information Processing Systems arXiv:201101725 [csCY]*

70. Daw ND (2011): Trial-by-trial data analysis using computational models. *Decision Making, Affect, and Learning: Attention and Performance XXIII*: Oxford University Press.

71. Kruschke JK (2015): *Doing Bayesian Data Analysis: A tutorial with R, JAGS, and Stan*. Second ed. Amsterdam: Academic Press.

72. Wilson RC, Collins AG (2019): Ten simple rules for the computational modeling of behavioral data. *Elife*. 8.

73. McGraw KO, Wong SP (1996): Forming inferences about some intraclass correlation coefficients. *Psychological Methods*. 1:30-46.

74. 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.

75. Frey R, Pedroni A, Mata R, Rieskamp J, Hertwig R (2017): Risk preference shares the psychometric structure of major psychological traits. *Sci Adv*. 3:e1701381.

76. Noble S, Scheinost D, Constable RT (2019): A decade of test-retest reliability of functional connectivity: A systematic review and meta-analysis. *Neuroimage*. 203:116157.

77. Nord CL, Gray A, Charpentier CJ, Robinson OJ, Roiser JP (2017): Unreliability of putative fMRI biomarkers during emotional face processing. *Neuroimage*. 156:119-127.

78. Rodebaugh TL, Scullin RB, Langer JK, Dixon DJ, Huppert JD, Bernstein A, et al. (2016): Unreliability as a threat to understanding psychopathology: The cautionary tale of attentional bias. *J Abnorm Psychol*. 125:840-851.

79. 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.

80. Price RB, Brown V, Siegle GJ (2019): Computational Modeling Applied to the Dot-Probe Task Yields Improved Reliability and Mechanistic Insights. *Biol Psychiatry*. 85:606-612.

81. Palminteri S, Chevallier C (2018): Can We Infer Inter-Individual Differences in Risk-Taking From Behavioral Tasks? *Front Psychol*. 9:2307.

82. Haines N, Kvam P, Irving L, Smith C, Beauchaine TP, Pitt MA, et al. (2020): Learning from the Reliability Paradox: How Theoretically Informed Generative Models Can Advance the Social, Behavioral, and Brain Sciences. *PsyArXiv*.