JAMA Psychiatry | Original Investigation

Reinforcement Learning in Patients With Mood and Anxiety Disorders vs Control Individuals

A Systematic Review and Meta-analysis

Alexandra C. Pike, DPhil; Oliver J. Robinson, PhD

IMPORTANCE Computational psychiatry studies have investigated how reinforcement learning may be different in individuals with mood and anxiety disorders compared with control individuals, but results are inconsistent.

OBJECTIVE To assess whether there are consistent differences in reinforcement-learning parameters between patients with depression or anxiety and control individuals.

DATA SOURCES Web of Knowledge, PubMed, Embase, and Google Scholar searches were performed between November 15, 2019, and December 6, 2019, and repeated on December 3, 2020, and February 23, 2021, with keywords (*reinforcement learning*) AND (*computational* OR *model*) AND (*depression* OR *anxiety* OR *mood*).

STUDY SELECTION Studies were included if they fit reinforcement-learning models to human choice data from a cognitive task with rewards or punishments, had a case-control design including participants with mood and/or anxiety disorders and healthy control individuals, and included sufficient information about all parameters in the models.

DATA EXTRACTION AND SYNTHESIS Articles were assessed for inclusion according to MOOSE guidelines. Participant-level parameters were extracted from included articles, and a conventional meta-analysis was performed using a random-effects model. Subsequently, these parameters were used to simulate choice performance for each participant on benchmarking tasks in a simulation meta-analysis. Models were fitted, parameters were extracted using bayesian model averaging, and differences between patients and control individuals were examined. Overall effect sizes across analytic strategies were inspected.

MAIN OUTCOMES AND MEASURES The primary outcomes were estimated reinforcement-learning parameters (learning rate, inverse temperature, reward learning rate, and punishment learning rate).

RESULTS A total of 27 articles were included (3085 participants, 1242 of whom had depression and/or anxiety). In the conventional meta-analysis, patients showed lower inverse temperature than control individuals (standardized mean difference [SMD], −0.215; 95% CI, −0.354 to −0.077), although no parameters were common across all studies, limiting the ability to infer differences. In the simulation meta-analysis, patients showed greater punishment learning rates (SMD, 0.107; 95% CI, 0.107 to 0.108) and slightly lower reward learning rates (SMD, −0.021; 95% CI, −0.022 to −0.020) relative to control individuals. The simulation meta-analysis showed no meaningful difference in inverse temperature between patients and control individuals (SMD, 0.003; 95% CI, 0.002 to 0.004).

CONCLUSIONS AND RELEVANCE The simulation meta-analytic approach introduced in this article for inferring meta-group differences from heterogeneous computational psychiatry studies indicated elevated punishment learning rates in patients compared with control individuals. This difference may promote and uphold negative affective bias symptoms and hence constitute a potential mechanistic treatment target for mood and anxiety disorders.

JAMA Psychiatry. doi:10.1001/jamapsychiatry.2022.0051
Published online March 2, 2022.

Supplemental content

Author Affiliations: Anxiety Lab, Neuroscience and Mental Health Group, Institute of Cognitive Neuroscience, University College London, London, United Kingdom (Pike, Robinson); Research Department of Clinical, Educational and Health Psychology, University College London, London, United Kingdom (Robinson).

Corresponding Author: Alexandra C. Pike, DPhil, Anxiety Lab, Neuroscience and Mental Health Group, Institute of Cognitive Neuroscience, University College London, 17-19 Queen Sq, Alexandra House, Bloomsbury, London WC1N 3AR, United Kingdom (alex.pike02@gmail.com).

Downloaded From: https://jamanetwork.com/ by a University College London User on 03/09/2022

Anxiety and depression are major individual and public health burdens.[1-3] However, current treatment options have relatively low recovery rates (ranging from 28% to 52%[4-6]), and there are limited novel treatment prospects on the horizon. Part of the difficulty in developing and improving treatments is that we have an incomplete understanding of the mechanisms underlying anxiety and depression. However, a growing number of proposed mechanisms of mood and anxiety symptoms have emerged from a field known as computational psychiatry.[7-11]

The basic premise of computational psychiatry is that variations in how the brain performs computations (eg, in learning, perception, and decision-making) may, over time, generate emergent symptoms that are observed in various psychiatric disorders. For example, one theory is that individuals with a higher learning rate for negative stimuli (ie, punishment learning rate) might learn more from each negative event they experience, producing the negative affective bias that is frequently associated with depressive and anxious disorders.[12,13] This theory situates the computational approach within clinical psychology concepts that date back to the 1960s[14] and provides a mechanistic and falsifiable hypothesis for how clinical phenomena like negative affective bias may emerge.

The most common class of computational models tested in this field to date are reinforcement-learning models. Reinforcement learning can be defined as learning to obtain rewards and avoid punishments,[15] and this type of computational model has some notable strengths. Namely, quantities computed by these models may be encoded in the phasic firing of dopamine neurons,[16] providing a bridge between brain and behavior.[15] Moreover, reinforcement-learning models can accurately mimic highly complex human behaviors.[17,18] Further, there is a large body of evidence[19,20] suggesting that those with depression and anxiety may show differences in processing rewards and/or punishments. Reinforcement-learning models may allow us to better understand this phenomenon.

We are now at the point where the body of case-control research investigating reinforcement-learning parameters in mood and anxiety disorders is sufficiently extensive that looking for overall patterns is possible. However, findings are varied. For example, different studies have argued that anxiety or depression may be associated with increased punishment learning rates[21-23] or reduced reward sensitivity.[24] While either of these differences would produce a negative bias toward the processing of punishments rather than rewards, the specifics can actually have considerable implications for how we treat such symptoms. For instance, reduced reward sensitivity in patients would require treatments that focused on how much the individual liked experiencing positive events, while treatments for elevated punishment learning rates would seek to encourage individuals to avoid immediately changing their behavior in response to negative outcomes.

The aim of this meta-analysis is therefore to assess consistencies across these reinforcement-learning studies and generate more highly powered estimates of the underlying group differences,[25] hypothesizing that there will be a difference in reinforcement-learning parameters across groups. We first present the results of a conventional meta-analysis. However, this analysis was unsatisfactory for the modeling approaches used in computational psychiatry, as studies use both different tasks and different models to obtain their results.[26] Is there a more principled way to combine the effect sizes from different tasks, models, and parameters?

To this end, a benefit of the modeling approach is that rather than simply taking a summary statistic over participants, individual-level trial-by-trial data are used to generate a proposed model of the underlying mechanisms. This generative model also provides precise predictions about how each individual's behavior might generalize outside of the specific reported context. It is therefore possible to invert a reported model and simulate data for participants, even on tasks that they did not perform in the original study. As such, we can simulate performance for participants across studies on standardized benchmarking tasks, removing task inconsistencies across studies. We can then compare parameters across consistent models in our newly standardized data, removing model inconsistencies across studies. This method can increase the generalizability of these parameters, as we obtain model parameter estimates in this common space and use them to meta-analytically estimate parameter differences across groups. Thus, the aim of this article is to leverage the unique advantages of computational modeling to create a novel simulation-based meta-analytic method, which can be used to test the hypothesis that there are case-control differences in reinforcement learning across mood and anxiety disorders.

Key Points

Question Are there differences in reinforcement learning between patients with mood and anxiety disorders and control individuals?

Findings In this systematic review and meta-analysis, a novel computational simulation method showed differences in reinforcement learning between patients and control individuals. In particular, patients showed elevated punishment learning rates compared with control individuals.

Meaning These findings show that patients may be more likely than control individuals to modify their behavior in response to punishments, suggesting a possible mechanistic treatment target for negative affective bias symptoms.

Methods

The procedure used in our meta-analysis is summarized below and in **Figure 1**, and explained in more detail in the eMethods in the Supplement.

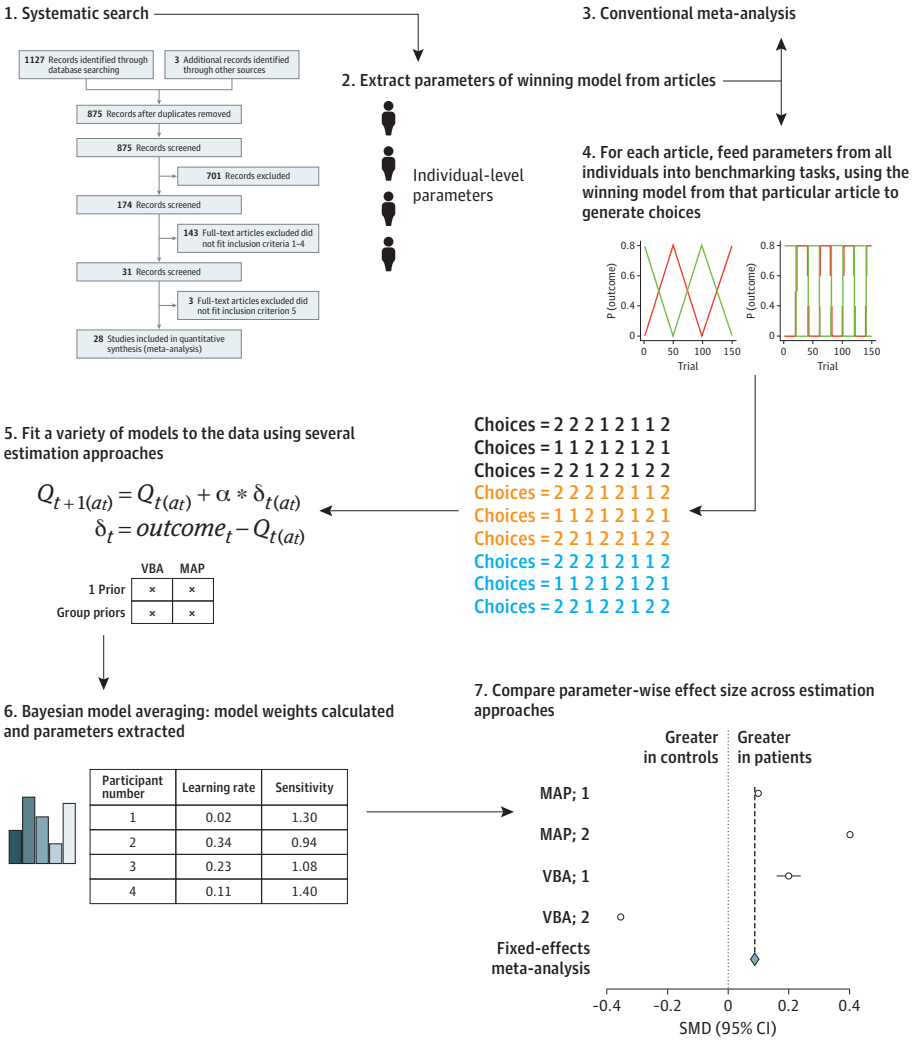
Article Selection and Extraction of Model Parameters

Articles were included if they met the following criteria:

1. Human participants.
2. Reported choice data from a cognitive task with monetary or point-based rewards or punishments.

Downloaded From: https://jamanetwork.com/ by a University College London User on 03/09/2022

Figure 1. Study Procedure



Note that the methods used in (5) were maximum a posteriori (MAP) with either 1 prior across all participants (MAP; 1) or separate priors for each group (MAP; 2); or variational bayesian analysis (VBA) with 1 prior (VBA; 1) or separate group priors (VBA; 2).

3. Fit reinforcement-learning models to choice data.
4. Used a case-control design comparing control individuals with individuals with mood and/or anxiety disorders.
5. Reported sufficient statistics of all parameters in the models or made individual-level parameters available.

A standard systematic search was performed according to the Meta-analysis of Observational Studies in Epidemiology (MOOSE) reporting guideline (eAppendix 1 in the Supplement) between November 15, 2019, and December 6, 2019 (A.C.P.), and independently verified (O.J.R.). The search was repeated on December 3, 2020, and February 23, 2021. Further details may be found in the eMethods in the Supplement.

Extraction of Winning Model Parameters From Articles

We used individual parameter estimates where available (eMethods in the Supplement) from the best-fitting model reported by the studies. Where these were not available in the article or a repository, we contacted the corresponding author to request them.

Conventional Meta-analysis

We used a modified version of the Newcastle-Ottawa scale to assess study quality, with details and results reported in the eMethods in the Supplement. We performed a series of random-effects meta-analyses on raw values of the most commonly reported reinforcement-learning parameters from the included studies. Heterogeneity and publication bias were assessed and are reported in the eResults in the Supplement.

Simulation Meta-analysis

In parallel with the conventional meta-analysis, we also performed a novel simulation meta-analysis. We describe this approach briefly here and in Figure 1 and in more detail in the eMethods in the Supplement.

In brief, we took the originally reported models from each article (eAppendix 2 in the Supplement) and used the model parameters reported for each participant to simulate choice behavior on 5 new benchmarking tasks. In other words, we used the generative models reported in the articles to anticipate the

Downloaded From: https://jamanetwork.com/ by a University College London User on 03/09/2022

choices participants might have made if they had done the same 5 tasks without adjusting behavioral strategy (ensuring that all choices were in the same task space). We then fit a selection of reinforcement-learning models (overlapping with the models across all the original articles) to this new choice data set, and then extracted parameters using bayesian model averaging according to the strength of fit of each model to the data. This ensured all results were in the same model space. This enabled us to test the primary hypothesis: whether any parameters differed between groups.

There are a number of different methods of parameter estimation commonly used in computational psychiatry. To determine whether our findings were robust to analytic method, we used 4 different analytic strategies to estimate parameters. The 4 analytic strategies were combinations of maximum a posteriori and variational bayesian estimation and assumed either that all patients and control individuals came from the same underlying population (a single empirical prior was used for each parameter) or that they came from 2 separate underlying populations (2 priors were used, 1 for each group). We performed fixed-effects meta-analyses on the most common parameters for illustrative purposes to allow us to visualize consistency of effect sizes across approaches.

Results

Systematic Search

After a systematic search, 27 articles were included.[21,23,24,27-50] The total number of participants included was 3085, 1242 of whom were patients with mood and anxiety disorders. A PRISMA diagram and a summary of the studies included, as well as an assessment of study quality, can be found in the eResults in the Supplement.

Conventional Meta-analysis

After individual-level parameters had been extracted, we performed a series of random-effects meta-analyses to examine whether any of the most commonly reported parameters showed differences between patients and control individuals. There was no parameter that all articles had in common, highlighting the importance of our simulation approach for inference across all included studies. There was no significant standardized mean difference in a single learning rate parameter (9 of 27 articles reported this parameter: standardized mean difference, 0.196 [95% CI, −0.044 to 0.437]; **Figure 2**A). However, inverse temperature (or temperature, which we converted to inverse temperature) was elevated in control individuals (represented in 19 of 27 articles: standardized mean difference, −0.215 [95% CI, −0.354 to −0.077]; Figure 2B). Some articles reported learning rates that were separated by valence: there was no significant standardized mean difference in these parameters (reward learning rate was represented in 14 of 27 articles: standardized mean difference, −0.152 [95% CI, −0.310 to 0.006]; **Figure 3**A; punishment learning rate was also represented in 14 of 27 articles: standardized mean difference, −0.037 [95% CI, −0.306 to 0.232]; Figure 3B). There was evidence of moderate to substantial heterogeneity based on the values of the between-study variance of true effect sizes (τ^2) and the approximate proportion of total variability (I^2) (eResults in the Supplement).

Simulation Meta-analysis

We used bayesian model averaging to obtain parameter estimates from each model in proportion to empirically determined model weights based on bayesian information criterion values. Subsequently, we performed 4 multivariate analyses of variance, 1 corresponding to each different parameter estimation method (dependent variables included all parameters for which there was at least 1 estimate), including group, study, and task as main effects. Each of these indicated that there was a main effect of group (**Table**), suggesting that there was a general difference in reinforcement learning between patients and control individuals regardless of estimation method.

There were also effects of study and task. We describe the effect of task further in the eResults in the Supplement. Briefly, recovery for separate reward and punishment learning rates was notably worse in benchmarking tasks in which rewards and punishments were nonindependent. In a supplementary analysis, we show that our findings held when only including the benchmarking tasks with good recovery. However, it is also possible that this issue is present in the raw parameter data that we used in this meta-analysis: not all tasks in the original article had orthogonal rewards and punishments.

We examined the effect of group for the parameters that were represented most frequently after bayesian model averaging. Statistics are shown in the eResults in the Supplement, and a summary of the effect sizes for each approach can be seen in **Figure 4**.

Here, we report the standardized mean differences for the 4 most highly represented parameters from our simulation meta-analysis using bayesian model averaging, combined using a fixed-effects meta-analysis. Across analysis methods (Figure 4), we saw a meaningful increase in punishment learning rates (standardized mean difference, 0.107 [95% CI, 0.107 to 0.108]) in patients vs control individuals and a slight decrease in reward learning rates (standardized mean difference, −0.021 [95% CI, −0.022 to −0.020]) with single learning rates also showing a slight increase (standardized mean difference, 0.041 [95% CI, 0.040 to 0.042]). Inverse temperature, which appeared to be different in a conventional meta-analysis, showed only a negligible difference (standardized mean difference, 0.003 [95% CI, 0.002 to 0.004]).

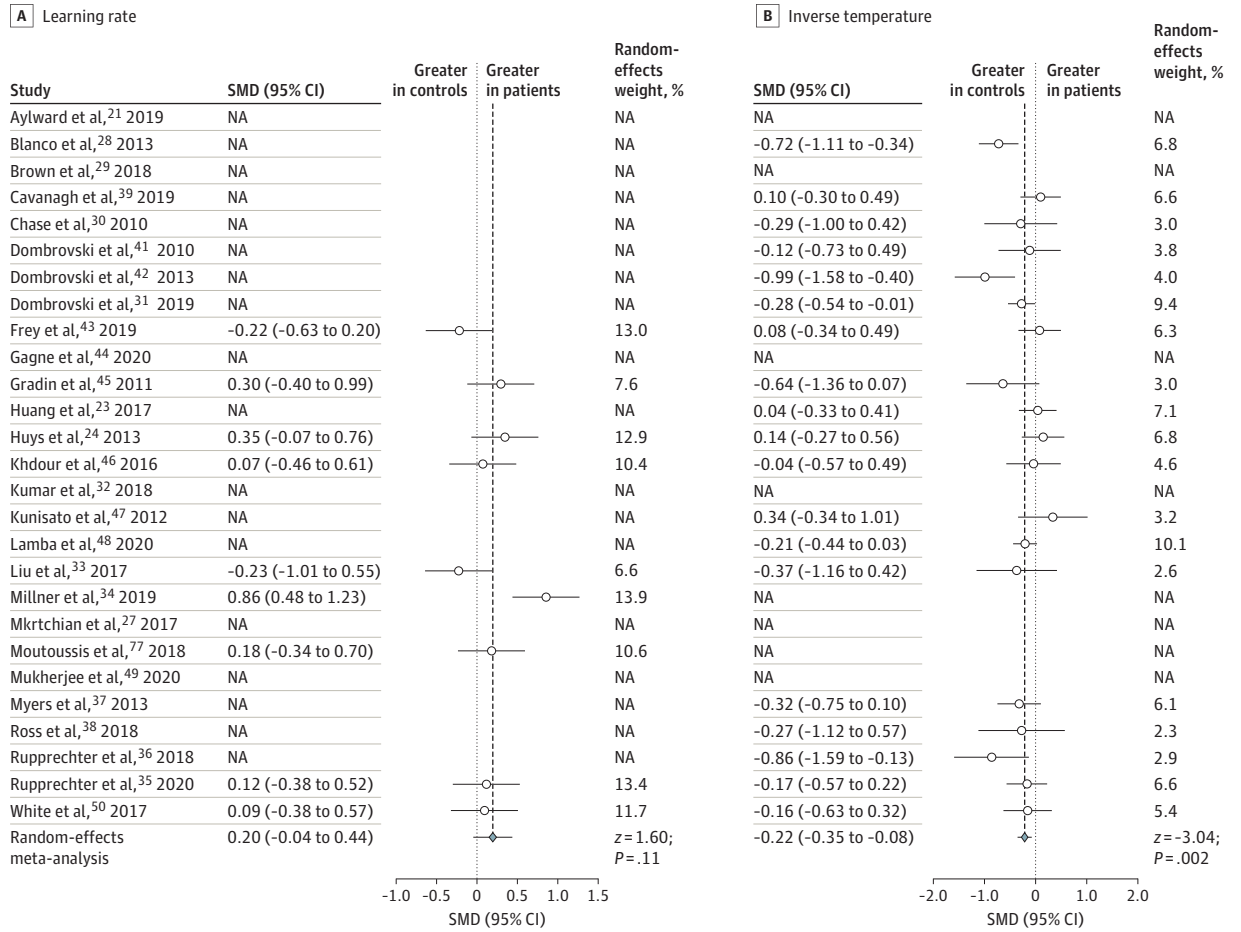
Subgroup analyses that investigate how these results varied by participant group, and meta-regressions controlling for study quality, year of publication, and parameter-level uncertainty, are in the eResults in the Supplement.

Discussion

Our conventional meta-analysis suggested the only difference in reinforcement-learning parameters between patients and control individuals was in inverse temperature, with

Downloaded From: https://jamanetwork.com/ by a University College London User on 03/09/2022

Figure 2. Forest Plots for the Conventional Meta-analysis Comparing Learning Rate and Inverse Temperature



The figure shows standardized mean differences (Hedges *g*) between patients and control individuals. Where the relevant parameter was not included in the original article, the standardized mean difference is marked as not applicable (NA). SMD indicates standardized mean difference.

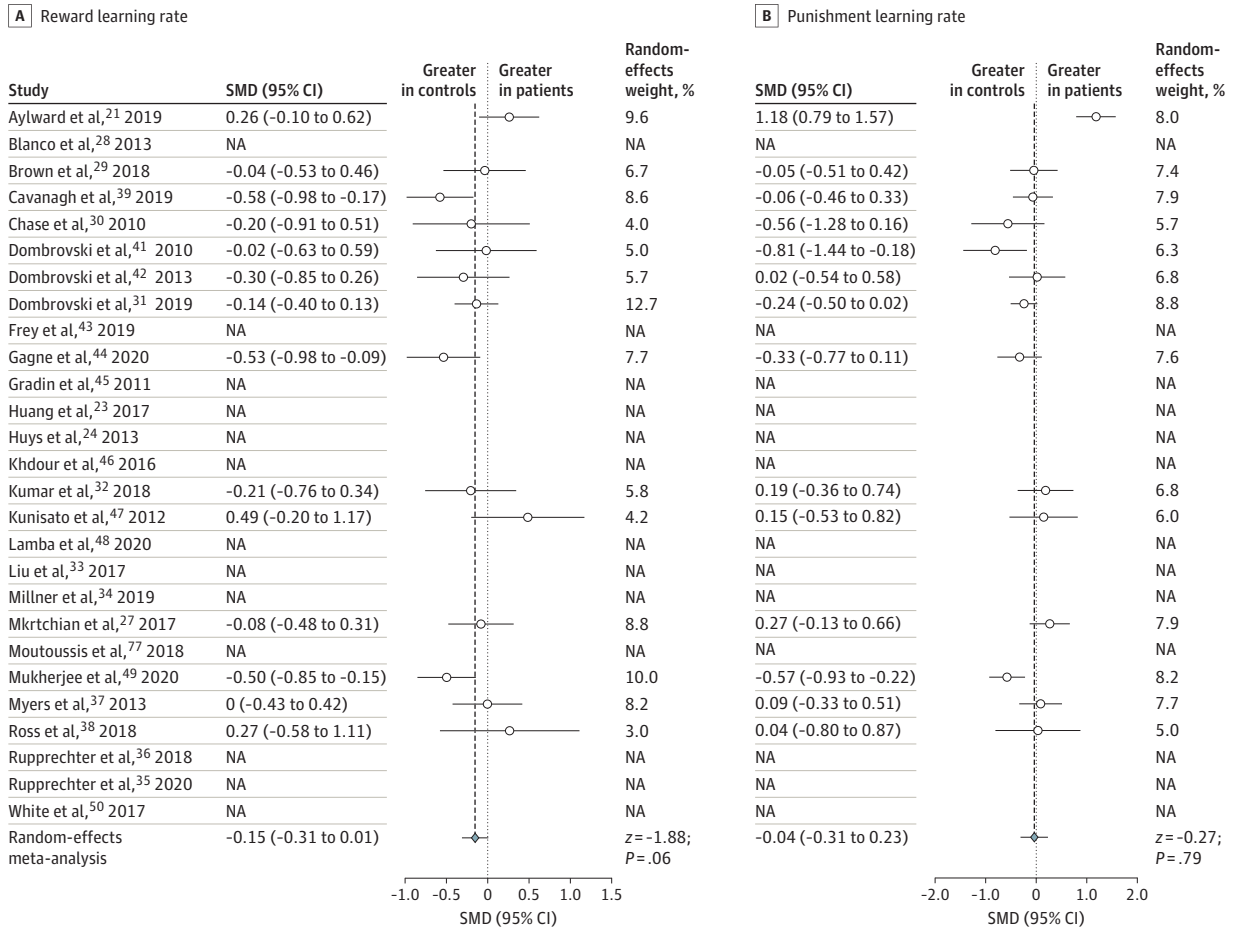
patients showing lower inverse temperature. However, the limitations of conventional methods when applied to computational modeling research were apparent: many articles did not have parameters in common. Using our novel meta-analytic method to estimate parameters for all articles across consistent task space and model space, we found meaningfully higher punishment learning rates and slightly lower reward learning rates in patients than in control individuals. This was seen alongside negligible group differences in inverse temperature.

The primary finding from our simulation meta-analysis was that those with mood and anxiety disorders showed a different balance between reward and punishment learning rates compared with control participants. Specifically, patients updated their learned values meaningfully more than control individuals after receiving a punishment and slightly less than control individuals after receiving a reward outcome. This association with learning rates was not apparent in our conventional meta-analysis; however, only half (14 of 27) of the original studies included the parameters (separate reward and punishment learning rates) that we required to test this using conventional methods. This highlights a key strength of our new simulation approach, as we were able to test for differences in these parameters across all studies.

The second key finding is that we did not observe any robust evidence of meaningful differences in inverse temperature or outcome sensitivity across patients and control individuals using our novel simulation method. It is worth bearing in mind that these parameters incorporate noisiness, participant exploration, and sensitivity to outcomes and thus are perhaps not pure estimations of either choice stochasticity or outcome sensitivity. This null finding is also made unclear by the poorer recovery of sensitivity effect sizes using our pipeline (eResults in the Supplement), and the fact that a larger association with inverse temperature was observed when controlling for study quality in a meta-regression. However, on the basis of our results, we tentatively suggest that how individuals learn and change their behavior to outcomes may be more important than other factors, such as how much individuals like or dislike outcomes. Further replication using tasks and models designed to robustly estimate these parameters will be necessary to confirm this interpretation, but it is interest-

Downloaded From: https://jamanetwork.com/ by a University College London User on 03/09/2022

Figure 3. Forest Plots for the Conventional Meta-analysis Comparing Reward Learning Rate and Punishment Learning Rate



The figure shows standardized mean differences (Hedges *g*) between patients and control individuals. Where the relevant parameter was not included in the original article, the standardized mean difference is marked as not applicable (NA). SMD indicates standardized mean difference.

Table. Results From the Multivariate Analysis of Variance[a]

Analysis method	Main association	Approximate *F*	*df*		*P* value
			Numerator	Denominator	
VBA					
1[b]	Group	971.30	16	15 273 788	<.001
	Task	60 255.67	64	61 095 164	<.001
	Study	14 795.04	416	244 380 848	<.001
2[c]	Group	10 590.50	16	15 273 788	<.001
	Task	47 039.51	64	61 095 164	<.001
	Study	11 683.04	416	244 380 848	<.001
MAP					
1[b]	Group	672.91	16	15 273 788	<.001
	Task	20 178.88	64	61 095 164	<.001
	Study	3636.98	416	244 380 848	<.001
2[c]	Group	1707.41	16	15 273 788	<.001
	Task	20 696.29	64	61 095 164	<.001
	Study	6560.50	416	244 380 848	<.001

Abbreviations: MAP, maximum a posteriori; VBA, variational bayesian analysis.

[a] Full univariate results for all parameters can be found in the eResults in the Supplement.

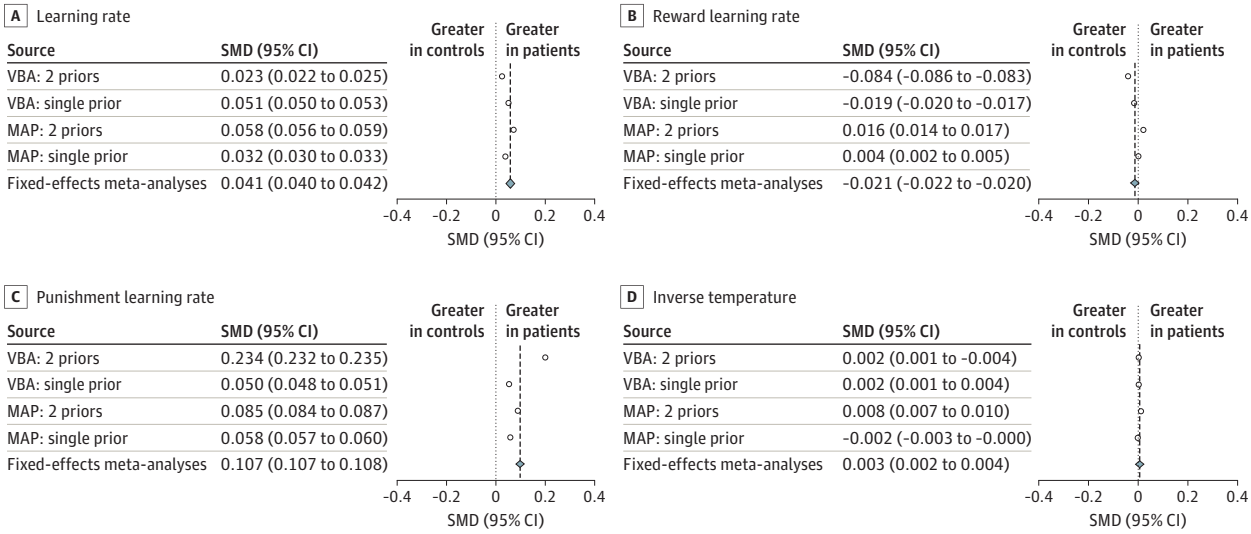
[b] Parameters were estimated using a hierarchical bayesian approach with only a single prior over groups.

[c] Parameters were estimated using a different prior for each group.

ing that our findings contrast with some accounts of depression and anxiety,[24,51] which propose that depression is associated with reduced reward sensitivity (eg, anhedonia as a diagnostic criterion) and that anxiety is associated with increased punishment sensitivity (eg, biased attention or memory for threats).

Downloaded From: https://jamanetwork.com/ by a University College London User on 03/09/2022

Figure 4. Forest Plots of the Cohen *d* Effect Sizes for the 4 Most Highly Represented Parameters From the Simulation Meta-analysis



Fixed-effects meta-analyses were performed over the different analytic approaches. Each effect size represents a different analytic approach. Note that 95% CIs are not visible here as they overlap with the effect sizes. MAP indicates maximum a posteriori analysis; VBA, variational bayesian analysis.

Our findings may help refine our understanding of negative affective bias,[12] in which patients focus on negative outcomes or occurrences. Specifically, our results allow us to tentatively distinguish between 2 potential causes of negative affective bias: greater subjective valuation of negative outcomes and different learning in response to negative outcomes.[13] This meta-analysis found that patients with mood and anxiety disorders learned more from each instance of a negative outcome and showed no differences in how much they disliked these outcomes.

Furthermore, the results we have presented may be relevant to how clinicians target cognitive interventions. Rather than encouraging individuals with depression or anxiety to downweigh the subjective experience of negative outcomes or to sit with and tolerate resulting distress (as a necessary prerequisite for subsequently altering behavioral responses to distress, ie, in dialectical behavior therapy),[52] interventions should focus directly on modifying how an individual changes their behavior in response to that negative outcome.[21] For example, a therapist could try to encourage the individual to pause and not immediately change their behavior after something bad happens to provide the space to contextualize the negative outcome. This focus on punishment learning rates rather than punishment sensitivity could also help us understand how some common interventions work. For instance, flooding in exposure therapy (eg, where an agoraphobic individual is taken to a busy city center) may be effective through preventing the individual from performing their habitual behavioral responses to a negative outcome, rather than by tuning down their aversive response (ie, punishment sensitivity) to the exposure (eg, the crowd of people).

One of the strengths of reinforcement-learning models is that key quantities (ie, reward prediction errors) predicted by these models are thought to be reflected by neural activity.[16,53-57] Notably, the learning rate may be an emergent property of neuromodulators, and in particular catecholamines, such as dopamine and noradrenaline.[58,59] This has implications for drug interventions for depression and anxiety: if the balance of catecholamines modulates learning rates, pharmacological agents that affect learning rates may be of benefit to patients. Much previous work has focused on neuromodulators in depression and anxiety, following articulation of the monoamine hypothesis in the 1960s.[60-62] Indeed, many of the first-line treatments for these disorders are selective serotonin reuptake inhibitors,[63,64] which are associated with both serotonin and dopamine.[65] However, many of these agents were discovered serendipitously, and the mechanisms by which they act on mood and anxiety symptoms are still unclear.[66] As a result, there are few intermediate end points that have been validated for use in drug discovery. The results of this meta-analysis may point to a genuine intermediate end point: learning rates. In particular, individualized measures of learning rate balance could be obtained using straightforward behavioral tasks, thus allowing dose personalization and early indications of drug efficacy for individuals. This end point is also translationally valuable, as learning rates can also be measured in animals, potentially allowing preclinical drug discovery work.[67] This meta-analysis therefore provides a possible first step toward connecting different levels of analysis in mental health research, from behavioral symptoms to the underlying neurobiology and pharmacology.

Limitations

This study has several limitations. A core assumption of the field of computational psychiatry is that parameters and models generalize across tasks, samples, and model parameterizations (highlighted by the use of the same terms, such as learning rate, in different studies and models).[26] We relied on this

Downloaded From: https://jamanetwork.com/ by a University College London User on 03/09/2022

assumption for our conventional meta-analysis, although for our simulation meta-analysis we only required that parameters generalized across tasks and samples. Specifically, we assumed that it was possible to use a model that captured behavior on one task to simulate behavior on another. It is unlikely, however, that this is straightforwardly possible,[26] which is reflected in the consistent effect of study we found on all parameter values. Particularly concerning is recent empirical evidence suggesting that parameters may not be stable within individuals either over time[68] or across different tasks.[69] In addition, parameters may not be stable even within a task; learning rates are known to adjust with the volatility of the environment,[13,70,71] which also changes between tasks, along with other variables (eg, responses, timing, outcomes, and contingency structure). Furthermore, parameters defined for different tasks with different underlying statistics may play different roles in new tasks. Understanding how the parameters derived from one task relate to another—perhaps by establishing task-specific parameter norms or by testing the same individuals across multiple tasks—will allow stronger inferences about parameters to be drawn in this kind of meta-analytic approach. If parameters from one task do not relate at all to each other, this will be a serious issue for the use of reinforcement-learning models in computational psychiatry. An implicit assumption is that these parameters are representative of underlying generative processes across tasks and are thus more relevant to real-world behavior than summary statistics, such as mean accuracy. Perhaps a more realistic assumption is that parameters do generalize, but imperfectly, between tasks. This can be observed in related families of models, such as drift diffusion models, the parameters of which do mostly generalize across tasks,[72] including in clinical populations.[73] It is nevertheless promising that we were able to observe consistent parameter effect sizes across different analytic methods despite this substantial source of noise.

The results from our conventional meta-analysis differ from our findings using our novel simulation method. Our conventional meta-analysis found differences in inverse temperature, but not with learning rate, namely reduced inverse temperature in patients. Reduced inverse temperature may reflect noisy choice data, which encompasses strategies not based on reinforcement learning, exploratory behavior, or simply nonspecific difficulties in attention and concentration[8,74] that are often observed clinically. Future work might therefore measure general executive function alongside reinforcement learning to disentangle the contribution of learning-specific associations from overall cognitive function.[75,76] However, the conventional meta-analysis was also limited by the lack of commonality across parameters, reducing the amount of usable data, and by lack of generalizability across task space and model space, which was the motivation behind the development of our novel method. Future work will be needed to fully assess and compare the conventional method with our novel one (although see the eResults in the Supplement for simulations indicating the effect sizes from our novel meta-analysis are generally underestimates).

On a related note, it is possible that our findings were driven, at least in part, by our selection of benchmarking tasks. As illustrated in the eResults in the Supplement, recovery depends on the structure of the task. Future work might adopt additional benchmarking tasks to further probe the robustness of meta-analytic differences to task specification.

Additionally, in this novel method, we did not carry forward all the information we had about parameter-level uncertainty to our final inference. However, the results of a meta-regression using parameter-level uncertainty (eResults in the Supplement) showed that the effect of group on punishment learning rate was robust to this source of noise.

Moreover, there was considerable heterogeneity in the included studies. This may be driven by the different participant groups, tasks, and models included in these studies, but another important source of variance might be the variety of different methods used in parameter estimation. Parameters estimated in original studies may have been subject to hierarchical fitting or not, may be regularized or not, may have been constrained or not, and the original authors may or may not have tested parameter and model recovery and stability. This should not prohibit meta-analytic inference, but is an additional source of noise that should temper confidence in meta-analytic estimates.

Conclusions

Overall, this study provides support for the hypothesis that reinforcement learning differs across patients with mood and anxiety disorders and control individuals. Specifically, we demonstrated elevated punishment learning rates and reduced reward learning rates in patients. We concluded that negative affective bias in mood and anxiety disorders may be driven by patients being too quick to update their behavior in response to negative outcomes. Moreover, by providing a formal computational account of this process, we were able to associate these symptoms with different levels of analysis (eg, neurobiological and pharmacological) and gain a mechanistic insight into how psychological therapy may work.

ARTICLE INFORMATION

Accepted for Publication: December 23, 2021.

Published Online: March 2, 2022. doi:10.1001/jamapsychiatry.2022.0051

Open Access: This is an open access article distributed under the terms of the CC-BY License. © 2022 Pike AC et al. *JAMA Psychiatry*.

Author Contributions: Drs Pike and Robinson had full access to all the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis.
Concept and design: All authors.
Acquisition, analysis, or interpretation of data: All authors.
Drafting of the manuscript: All authors.
Critical revision of the manuscript for important intellectual content: All authors.
Statistical analysis: All authors.
Obtained funding: Robinson.
Administrative, technical, or material support: All authors.
Supervision: Robinson.

Conflict of Interest Disclosures: Drs Pike and Robinson report salary support from a Medical Research Council senior nonclinical fellowship during the conduct of the study, awarded to Dr Robinson, and a Medical Research Council Proximity-to-Discovery award with Roche, who have provided in-kind contributions regarding work

Downloaded From: https://jamanetwork.com/ by a University College London User on 03/09/2022

on heart rate variability and anxiety and have sponsored travel outside of the submitted work. Dr Robinson reports that his senior nonclinical fellowship is partially in collaboration with Cambridge Cognition (who plan to provide in-kind contribution); he is also running an investigator-initiated trial with medication donated by Lundbeck (escitalopram and placebo; no financial contribution); has completed consultancy work on affective bias modification for Peak, online cognitive behavioral therapy for Ieso Digital Health, and on randomized clinical trials for anxiety for Roche; Dr Robinson also sits on the committee of the British Association of Psychopharmacology. No other disclosures were reported.

Funding/Support: This research was funded by a UK Research and Innovation Medical Research Council senior nonclinical fellowship to Dr Robinson (MR/R020817/1).

Role of the Funder/Sponsor: The funders had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

Disclaimer: The contents and conclusions do not necessarily represent the views of the US Department of Veterans Affairs or the US government.

Additional Contributions: We thank the Neuroscience and Mental Health Group, especially Vincent Valton, PhD, and Anahita Talwar, MSc, Neuroscience and Mental Health Group, Institute of Cognitive Neuroscience, University College London, London, United Kingdom, for their suggestions and troubleshooting advice during the creation of this article, and Quentin Huys, MD, PhD, Division of Psychiatry and Max Planck Centre for Computational Psychiatry and Ageing Research, University College London, London, United Kingdom, and Peter Dayan, PhD, Max Planck Institute for Biological Cybernetics, 72076 Tübingen, Germany and University of Tübingen, Tübingen, Germany, for their helpful comments and discussion. These individuals were not compensated for their contributions. We also thank the authors of the articles in this analysis who responded to requests for raw parameter values and without whom this article would be much less comprehensive. The raw parameter values are shared as open data with the agreement of the corresponding authors, for which we are also grateful. The acknowledgments from these original articles are also therefore relevant here.

Additional Information: Open data and code to replicate all analyses are available at https://doi.org/10.17605/OSF.IO/GTPFV.

REFERENCES

1. Kyu HH, Abate D, Abate KH, et al; GBD 2017 DALYs and HALE Collaborators. Global, regional, and national disability-adjusted life-years (DALYs) for 359 diseases and injuries and healthy life expectancy (HALE) for 195 countries and territories, 1990-2017: a systematic analysis for the Global Burden of Disease Study 2017. *Lancet*. 2018;392(10159):1859-1922. doi:10.1016/S0140-6736(18)32335-3

2. Baxter AJ, Vos T, Scott KM, Ferrari AJ, Whiteford HA. The global burden of anxiety disorders in 2010. *Psychol Med*. 2014;44(11):2363-2374. doi:10.1017/S0033291713003243

3. Lopez AD, Mathers CD, Ezzati M, Jamison DT, Murray CJ. Global and regional burden of disease and risk factors, 2001: systematic analysis of population health data. *Lancet*. 2006;367(9524):1747-1757. doi:10.1016/S0140-6736(06)68770-9

4. NHS Digital. Psychological therapies, annual report on the use of IAPT services 2018-19. Published July 11, 2019. Accessed May 14, 2020. https://digital.nhs.uk/data-and-information/publications/statistical/psychological-therapies-annual-reports-on-the-use-of-iapt-services/annual-report-2018-19

5. Trivedi MH, Rush AJ, Wisniewski SR, et al; STAR*D Study Team. Evaluation of outcomes with citalopram for depression using measurement-based care in STAR*D: implications for clinical practice. *Am J Psychiatry*. 2006;163(1):28-40. doi:10.1176/appi.ajp.163.1.28

6. Insel TR, Wang PS. The STAR*D trial: revealing the need for better treatments. *Psychiatr Serv*. 2009;60(11):1466-1467. doi:10.1176/ps.2009.60.11.1466

7. Adams RA, Huys QJM, Roiser JP. Computational psychiatry: towards a mathematically informed understanding of mental illness. *J Neurol Neurosurg Psychiatry*. 2016;87(1):53-63.

8. Eshel N, Roiser JP. Reward and punishment processing in depression. *Biol Psychiatry*. 2010;68(2):118-124. doi:10.1016/j.biopsych.2010.01.027

9. Friston KJ, Stephan KE, Montague R, Dolan RJ. Computational psychiatry: the brain as a phantastic organ. *Lancet Psychiatry*. 2014;1(2):148-158. doi:10.1016/S2215-0366(14)70275-5

10. Montague PR, Dolan RJ, Friston KJ, Dayan P. Computational psychiatry. *Trends Cogn Sci*. 2012;16(1):72-80. doi:10.1016/j.tics.2011.11.018

11. Huys QJM, Maia TV, Frank MJ. Computational psychiatry as a bridge from neuroscience to clinical applications. *Nat Neurosci*. 2016;19(3):404-413. doi:10.1038/nn.4238

12. Mathews A, MacLeod C. Cognitive vulnerability to emotional disorders. *Annu Rev Clin Psychol*. 2005;1:167-195. doi:10.1146/annurev.clinpsy.1.102803.143916

13. Pulcu E, Browning M. Affective bias as a rational response to the statistics of rewards and punishments. *Elife*. 2017;6(6):e27879. doi:10.7554/eLife.27879

14. Beck AT. Thinking and depression: I. idiosyncratic content and cognitive distortions. *Arch Gen Psychiatry*. 1963;9(4):324-333. doi:10.1001/archpsyc.1963.01720160014002

15. Maia TV, Frank MJ. From reinforcement learning models to psychiatric and neurological disorders. *Nat Neurosci*. 2011;14(2):154-162. doi:10.1038/nn.2723

16. Schultz W, Dayan P, Montague PR. A neural substrate of prediction and reward. *Science*. 1997;275(5306):1593-1599. doi:10.1126/science.275.5306.1593

17. Mnih V, Kavukcuoglu K, Silver D, et al. Human-level control through deep reinforcement learning. *Nature*. 2015;518(7540):529-533. doi:10.1038/nature14236

18. Silver D, Huang A, Maddison CJ, et al. Mastering the game of Go with deep neural networks and tree search. *Nature*. 2016;529(7587):484-489. doi:10.1038/nature16961

19. Halahakoon DC, Kieslich K, O'Driscoll C, Nair A, Lewis G, Roiser JP. Reward-processing behavior in depressed participants relative to healthy volunteers: a systematic review and meta-analysis. *JAMA Psychiatry*. 2020;77(12):1286-1295. doi:10.1001/jamapsychiatry.2020.2139

20. Katz BA, Matanky K, Aviram G, Yovel I. Reinforcement sensitivity, depression and anxiety: a meta-analysis and meta-analytic structural equation model. *Clin Psychol Rev*. 2020;77:101842. doi:10.1016/j.cpr.2020.101842

21. Aylward J, Valton V, Ahn WY, et al. Altered learning under uncertainty in unmedicated mood and anxiety disorders. *Nat Hum Behav*. 2019;3(10):1116-1123. doi:10.1038/s41562-019-0628-0

22. Beevers CG, Worthy DA, Gorlick MA, Nix B, Chotibut T, Todd Maddox W. Influence of depression symptoms on history-independent reward and punishment processing. *Psychiatry Res*. 2013;207(1-2):53-60. doi:10.1016/j.psychres.2012.09.054

23. Huang H, Thompson W, Paulus MP. Computational dysfunctions in anxiety: failure to differentiate signal from noise. *Biol Psychiatry*. 2017;82(6):440-446. doi:10.1016/j.biopsych.2017.07.007

24. Huys QJ, Pizzagalli DA, Bogdan R, Dayan P. Mapping anhedonia onto reinforcement learning: a behavioural meta-analysis. *Biol Mood Anxiety Disord*. 2013;3(1):12. doi:10.1186/2045-5380-3-12

25. Button KS, Ioannidis JPA, Mokrysz C, et al. Power failure: why small sample size undermines the reliability of neuroscience. *Nat Rev Neurosci*. 2013;14(5):365-376. doi:10.1038/nrn3475

26. Eckstein MK, Wilbrecht L, Collins AGE. What do reinforcement learning models measure? interpreting model parameters in cognition and neuroscience. *Curr Opin Behav Sci*. 2021;41:128-137. doi:10.1016/j.cobeha.2021.06.004

27. Mkrtchian A, Aylward J, Dayan P, Roiser JP, Robinson OJ. Modeling avoidance in mood and anxiety disorders using reinforcement learning. *Biol Psychiatry*. 2017;82(7):532-539. doi:10.1016/j.biopsych.2017.01.017

28. Blanco NJ, Otto AR, Maddox WT, Beevers CG, Love BC. The influence of depression symptoms on exploratory decision-making. *Cognition*. 2013;129(3):563-568. doi:10.1016/j.cognition.2013.08.018

29. Brown VM, Zhu L, Wang JM, Frueh BC, King-Casas B, Chiu PH. Associability-modulated loss learning is increased in posttraumatic stress disorder. *Elife*. 2018;7:e30150. doi:10.7554/eLife.30150

30. Chase HW, Frank MJ, Michael A, Bullmore ET, Sahakian BJ, Robbins TW. Approach and avoidance learning in patients with major depression and healthy controls: relation to anhedonia. *Psychol Med*. 2010;40(3):433-440. doi:10.1017/S0033291709990468

31. Dombrovski AY, Hallquist MN, Brown VM, Wilson J, Szanto K. Value-based choice, contingency learning, and suicidal behavior in mid- and late-life depression. *Biol Psychiatry*. 2019;85(6):506-516. doi:10.1016/j.biopsych.2018.10.006

32. Kumar P, Goer F, Murray L, et al. Impaired reward prediction error encoding and striatal-midbrain connectivity in depression. *Neuropsychopharmacology*. 2018;43(7):1581-1588. doi:10.1038/s41386-018-0032-x

33. Liu WH, Valton V, Wang LZ, Zhu YH, Roiser JP. Association between habenula dysfunction and motivational symptoms in unmedicated major depressive disorder. *Soc Cogn Affect Neurosci*. 2017;12(9):1520-1533. doi:10.1093/scan/nsx074

34. Millner AJ, den Ouden HEM, Gershman SJ, et al. Suicidal thoughts and behaviors are

Downloaded From: https://jamanetwork.com/ by a University College London User on 03/09/2022

associated with an increased decision-making bias for active responses to escape aversive states. *J Abnorm Psychol*. 2019;128(2):106-118. doi:10.1037/abn0000395

35. Rupprechter S, Romaniuk L, Series P, et al. Blunted medial prefrontal cortico-limbic reward-related effective connectivity and depression. *Brain*. 2020;143(6):1946-1956. doi:10.1093/brain/awaa106

36. Rupprechter S, Stankevicius A, Huys QJM, Steele JD, Seriès P. Major depression impairs the use of reward values for decision-making. *Sci Rep*. 2018;8(1):13798. doi:10.1038/s41598-018-31730-w

37. Myers CE, Moustafa AA, Sheynin J, et al. Learning to obtain reward, but not avoid punishment, is affected by presence of PTSD symptoms in male veterans: empirical data and computational model. *PLoS One*. 2013;8(8):e72508. doi:10.1371/journal.pone.0072508

38. Ross MC, Lenow JK, Kilts CD, Cisler JM. Altered neural encoding of prediction errors in assault-related posttraumatic stress disorder. *J Psychiatr Res*. 2018;103:83-90. doi:10.1016/j.jpsychires.2018.05.008

39. Cavanagh JF, Bismark AW, Frank MJ, Allen JJB. Multiple dissociations between comorbid depression and anxiety on reward and punishment processing: evidence from computationally informed EEG. *Comput Psychiatr*. 2019;3:1-17. doi:10.1162/CPSY_a_00024

40. Dombrovski AY, Szanto K, Clark L, et al. Corticostriatothalamic reward prediction error signals and executive control in late-life depression. *Psychol Med*. 2015;45(7):1413-1424. doi:10.1017/S0033291714002517

41. Dombrovski AY, Clark L, Siegle GJ, et al. Reward/punishment reversal learning in older suicide attempters. *Am J Psychiatry*. 2010;167(6):699-707. doi:10.1176/appi.ajp.2009.09030407

42. Dombrovski AY, Szanto K, Clark L, Reynolds CF, Siegle GJ. Reward signals, attempted suicide, and impulsivity in late-life depression. *JAMA Psychiatry*. 2013;70(10):1. doi:10.1001/jamapsychiatry.2013.75

43. Frey AL, Frank MJ, McCabe C. Social reinforcement learning as a predictor of real-life experiences in individuals with high and low depressive symptomatology. *Psychol Med*. 2021;51(3):408-415. doi:10.1017/S0033291719003222

44. Gagne C, Zika O, Dayan P, Bishop SJ. Impaired adaptation of learning to contingency volatility in internalizing psychopathology. *Elife*. 2020;9:e61387. doi:10.7554/eLife.61387

45. Gradin VB, Kumar P, Waiter G, et al. Expected value and prediction error abnormalities in depression and schizophrenia. *Brain*. 2011;134(Pt 6):1751-1764. doi:10.1093/brain/awr059

46. Khdour HY, Abushalbaq OM, Mughrabi IT, et al. Generalized anxiety disorder and social anxiety disorder, but not panic anxiety disorder, are associated with higher sensitivity to learning from negative feedback: behavioral and computational investigation. *Front Integr Neurosci*. 2016;10:20. doi:10.3389/fnint.2016.00020

47. Kunisato Y, Okamoto Y, Ueda K, et al. Effects of depression on reward-based decision making and variability of action in probabilistic learning. *J Behav Ther Exp Psychiatry*. 2012;43(4):1088-1094. doi:10.1016/j.jbtep.2012.05.007

48. Lamba A, Frank MJ, FeldmanHall O. Anxiety impedes adaptive social learning under uncertainty. *Psychol Sci*. 2020;31(5):592-603. doi:10.1177/0956797620910993

49. Mukherjee D, Filipowicz ALS, Vo K, Satterthwaite TD, Kable JW. Reward and punishment reversal-learning in major depressive disorder. *J Abnorm Psychol*. 2020;129(8):810-823. doi:10.1037/abn0000641

50. White SF, Geraci M, Lewis E, et al. Prediction error representation in individuals with generalized anxiety disorder during passive avoidance. *Am J Psychiatry*. 2017;174(2):110-117. doi:10.1176/appi.ajp.2016.15111410

51. Robinson OJ, Chase HW. Learning and choice in mood disorders: searching for the computational parameters of anhedonia. *Comput Psychiatr*. 2017;1(1):208-233. doi:10.1162/CPSY_a_00009

52. Linehan MM, Armstrong HE, Suarez A, Allmon D, Heard HL. Cognitive-behavioral treatment of chronically parasuicidal borderline patients. *Arch Gen Psychiatry*. 1991;48(12):1060-1064. doi:10.1001/archpsyc.1991.01810360024003

53. Niv Y. Reinforcement learning in the brain. *J Math Psychol*. 2009;53(3):139-154. doi:10.1016/j.jmp.2008.12.005

54. Watabe-Uchida M, Eshel N, Uchida N. Neural circuitry of reward prediction error. *Annu Rev Neurosci*. 2017;40(1):373-394. doi:10.1146/annurev-neuro-072116-031109

55. Lee D, Seo H, Jung MW. Neural basis of reinforcement learning and decision making. *Annu Rev Neurosci*. 2012;35(1):287-308. doi:10.1146/annurev-neuro-062111-150512

56. Dayan P, Daw ND. Decision theory, reinforcement learning, and the brain. *Cogn Affect Behav Neurosci*. 2008;8(4):429-453. doi:10.3758/CABN.8.4.429

57. Kable JW, Glimcher PW. The neurobiology of decision: consensus and controversy. *Neuron*. 2009;63(6):733-745. doi:10.1016/j.neuron.2009.09.003

58. Silvetti M, Vassena E, Abrahamse E, Verguts T. Dorsal anterior cingulate-brainstem ensemble as a reinforcement meta-learner. *PLoS Comput Biol*. 2018;14(8):e1006370. doi:10.1371/journal.pcbi.1006370

59. Jepma M, Murphy PR, Nassar MR, Rangel-Gomez M, Meeter M, Nieuwenhuis S. Catecholaminergic regulation of learning rate in a dynamic environment. *PLOS Comput Biol*. 2016;12(10):e1005171. doi:10.1371/journal.pcbi.1005171

60. Schildkraut JJ. The catecholamine hypothesis of affective disorders: a review of supporting evidence. *Am J Psychiatry*. 1965;122(5):509-522. doi:10.1176/ajp.122.5.509

61. Delgado PL. Depression: the case for a monoamine deficiency. *J Clin Psychiatry*. 2000;61(suppl 6):7-11.

62. Bunney WE Jr, Davis JM. Norepinephrine in depressive reactions. a review. *Arch Gen Psychiatry*. 1965;13(6):483-494. doi:10.1001/archpsyc.1965.01730060001001

63. National Institute for Health and Care Excellence. Depression in adults: recognition and management (clinical guideline CG90). Accessed October 12, 2021. https://www.nice.org.uk/guidance/cg90

64. National Institute for Health and Care Excellence. Generalised anxiety disorder and panic disorder in adults: management (clinical guideline CG113). Accessed October 12, 2021. https://www.nice.org.uk/guidance/cg113

65. Dremencov E, El Mansari M, Blier P. Effects of sustained serotonin reuptake inhibition on the firing of dopamine neurons in the rat ventral tegmental area. *J Psychiatry Neurosci*. 2009;34(3):223-229.

66. Hyman SE. Revolution stalled. *Sci Transl Med*. 2012;4(155):155cm11. doi:10.1126/scitranslmed.3003142

67. Pike AC, Lowther M, Robinson OJ. The importance of common currency tasks in translational psychiatry. *Curr Behav Neurosci Rep*. 2021;8(1):1-10. doi:10.1007/s40473-021-00225-w

68. Shahar N, Hauser TU, Moutoussis M, Moran R, Keramati M, Dolan RJ; NSPN consortium. 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*. 2019;15(2):e1006803. doi:10.1371/journal.pcbi.1006803

69. Enkavi AZ, Eisenberg IW, Bissett PG, et al. Large-scale analysis of test-retest reliabilities of self-regulation measures. *Proc Natl Acad Sci USA*. 2019;116(12):5472-5477. doi:10.1073/pnas.1818430116

70. Browning M, Behrens TE, Jocham G, O'Reilly JX, Bishop SJ. Anxious individuals have difficulty learning the causal statistics of aversive environments. *Nat Neurosci*. 2015;18(4):590-596. doi:10.1038/nn.3961

71. Behrens TEJ, Woolrich MW, Walton ME, Rushworth MFS. Learning the value of information in an uncertain world. *Nat Neurosci*. 2007;10(9):1214-1221. doi:10.1038/nn1954

72. Schmitz F, Wilhelm O. Modeling mental speed: decomposing response time distributions in elementary cognitive tasks and correlations with working memory capacity and fluid intelligence. *J Intell*. 2016;4(4):13. doi:10.3390/jintelligence4040013

73. Powell G, Jones CRG, Hedge C, et al. Face processing in autism spectrum disorder re-evaluated through diffusion models. *Neuropsychology*. 2019;33(4):445-461. doi:10.1037/neu0000524

74. Blanchette I, Richards A. The influence of affect on higher level cognition: a review of research on interpretation, judgement, decision making and reasoning. *Cogn Emot*. 2010;24(4):561-595. doi:10.1080/02699930903132496

75. Collins AGE, Frank MJ. How much of reinforcement learning is working memory, not reinforcement learning? a behavioral, computational, and neurogenetic analysis. *Eur J Neurosci*. 2012;35(7):1024-1035. doi:10.1111/j.1460-9568.2011.07980.x

76. Niv Y, Daniel R, Geana A, et al. Reinforcement learning in multidimensional environments relies on attention mechanisms. *J Neurosci*. 2015;35(21):8145-8157. doi:10.1523/JNEUROSCI.2978-14.2015

Downloaded From: https://jamanetwork.com/ by a University College London User on 03/09/2022