



## Brief article

## The influence of depression symptoms on exploratory decision-making



Nathaniel J. Blanco<sup>a</sup>, A. Ross Otto<sup>b</sup>, W. Todd Maddox<sup>c,\*</sup>, Christopher G. Beever<sup>c</sup>,  
Bradley C. Love<sup>d,\*</sup>

<sup>a</sup> Department of Psychology, University of Texas at Austin, United States

<sup>b</sup> Center for Neural Science, New York University, United States

<sup>c</sup> Department of Psychology, Institute for Mental Health Research, University of Texas at Austin, United States

<sup>d</sup> Department of Cognitive, Perceptual and Brain Sciences, University College London, UK

## ARTICLE INFO

## Article history:

Received 2 March 2013

Revised 13 August 2013

Accepted 14 August 2013

## Keywords:

Exploration vs. exploitation

Depression

Reinforcement learning

Dynamic decision making

Cognitive modeling

## ABSTRACT

People with symptoms of depression show impairments in decision-making. One explanation is that they have difficulty maintaining rich representations of the task environment. We test this hypothesis in the context of exploratory choice. We analyze depressive and non-depressive participants' exploration strategies by comparing their choices to two computational models: (1) an "Ideal Actor" model that reflectively updates beliefs and plans ahead, employing a rich representation of the environment and (2) a "Naïve Reinforcement Learning" (RL) model that updates beliefs reflexively utilizing a minimal task representation. Relative to non-depressive participants, we find that depressive participants' choices are better described by the simple RL model. Further, depressive participants were more exploratory than non-depressives in their decision-making. Depressive symptoms appear to influence basic mechanisms supporting choice behavior by reducing use of rich task representations and hindering performance during exploratory decision-making.

Crown Copyright © 2013 Published by Elsevier B.V. All rights reserved.

## 1. Introduction

Depression is a common condition linked to suicide attempts, interpersonal problems, unemployment, and substance abuse (Kessler et al., 2003). The World Health Organization estimates that 121 million people suffer from depression and many more have elevated depressive symptoms. Clarifying the relationship between depressive symptoms and cognition may be useful in understanding both depression and basic cognitive processes.

Depressive symptoms are associated with lower performance in working memory (Rogers et al., 2004), problem-solving (Elderkin-Thompson, Mintz, Haroon,

Lavretsky, & Kumar, 2006), planning (Elliott et al., 1997), and decision-making tasks (Clark, Chamberlain, & Sahakian, 2009; Gradin et al., 2011; Maddox, Gorlick, Worthy, & Beever, 2012; Murphy et al., 2001; Pizzagalli, Iosifescu, Hallett, Ratner, & Fava, 2008). Computational models have proven useful in understanding the basis for these impairments, particularly in decision-making (Eshel & Roiser, 2010; Montague, Dolan, Friston, & Dayan, 2012). Along these lines, Paulus and Yu (2012) suggest that depressive symptoms alter action-value computations, causing abnormal decision-making in people with higher depressive symptoms.

Here we examine how these computations in exploratory choice are affected by depressive symptoms. In decision-making, optimal choice often requires building a representation of the task that supports effective planning. Because depressive individuals exhibit deficits in planning and working memory, we expect that they will have difficulty maintaining a rich representation of the

\* Corresponding authors. Address: University College London, Department of Cognitive, Perceptual and Brain Sciences, 26 Bedford Way, London WC1H 0AP, UK. Tel.: +44 020 7679 2000 (B.C. Love).

E-mail addresses: [Maddox@psy.utexas.edu](mailto:Maddox@psy.utexas.edu) (W.T. Maddox), [b.love@ucl.ac.uk](mailto:b.love@ucl.ac.uk) (B.C. Love).

task environment (Otto, Gershman, Markman, & Daw, 2013). If so, they should tend to rely on simple strategies and impoverished task representations, resulting in suboptimal decision-making (Montague et al., 2012).

Supporting this notion, Huys et al. (2012) found that in sequential decision-making, depressive symptom severity correlated with a tendency to “prune” (i.e. avoid mentally searching) paths that included a large loss even when it was advantageous to consider such options. Huys et al. suggested that pruning is an inflexible strategy not adaptive to tasks demands that is *reflexively* applied in response to punishment. The greater the depressive symptoms, the stronger the tendency was to prune, as opposed to *reflectively* consider alternative strategies and plan.

Montague et al. (2012) further suggest depressive decision-makers should explore less. Indeed, people with depressive symptoms exhibit less switching between options in a choice task where reward contingencies change over time (Cella, Dymond, & Cooper, 2010). Moreover, the increased pruning (which reduces the number of solutions considered) by depressives in Huys et al. is akin to reduced exploration, though these participants sample more uniformly (i.e., are more exploratory) within this reduced choice set.

Here, we provide a finer-grained examination of exploration strategies in depressives’ decision-making, directly testing the hypothesis that individuals with depressive symptoms are less likely to use rich task representations to reflectively update their beliefs and plan choices. Our computational approach affords understanding the basis of deficits in those suffering from depressive symptoms and the nature of exploratory choice more generally.

### 1.1. The Leapfrog task

We examine the effects of depressive symptoms on exploratory strategies by using a paradigm termed the “Leapfrog” task (Knox, Otto, Stone, & Love, 2012), a variant of the “bandit” task (Sutton & Barto, 1998). In this task (Fig. 1), one of two options gives a higher reward than the other. On any trial the currently inferior option can increase in value, becoming the better option. This change happens with a fixed probability called the “volatility” of the environment. Because the relative superiority of the options shifts over time, on each trial the participant must choose between *exploiting* the option with the highest observed reward and *exploring* to see if the other option has surpassed it. Because each choice is effectively reduced to the decision to explore or exploit, the Leapfrog task is well-suited to investigating exploratory behavior.

### 1.2. Reflexive vs. reflective strategies

Recent work has sought to characterize the types of behavior and/or representations that give rise to exploratory choice (Badre, Doll, Long, & Frank, 2012; Otto, Markman, Gureckis, & Love, 2010). One basic theoretical distinction is whether exploration is guided by beliefs that evolve in a principled manner to reflect uncertainty in the environment (i.e., *reflective* updating) or by beliefs that only change as a result of direct feedback (i.e., *reflexive*

updating; Knox et al., 2012). The reflective and reflexive conceptualization echoes the distinction between “model-based” and “model-free” learning in Reinforcement Learning (RL; Daw, Niv, & Dayan, 2005).

A *reflexive* learner has no representation of its environment besides expected values for each option, which are updated only after receiving rewards. State uncertainty is not utilized to guide decisions. Exploratory choices are undirected, resulting from a purely stochastic decision process. In *reflective* choice, conversely, the learner has a richer representation of their environment. This representation can include (among other things) beliefs about the state of the environment, state transitions, and the probabilities of events. In the Leapfrog task, a reflective learner could direct its choices according to its belief as to whether the observed-to-be-superior option is still superior. With each successive exploitive choice, the probability that the relative superiority of the options has flipped increases, making the state of the environment less certain. In this way, exploratory behavior can be directed by uncertainty; as uncertainty increases, exploration becomes more valuable. By using this knowledge about the environment to plan exploration, reflective strategies should outperform reflexive strategies on the Leapfrog task.

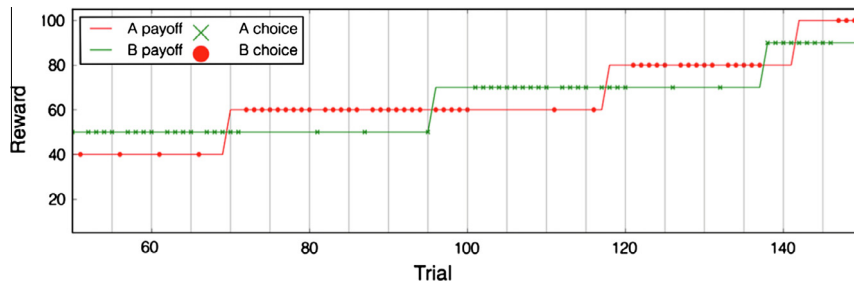
Through quantitatively comparing how well a reflexive vs. reflective account characterizes an individual’s choices, we assess whether the relative usage of reflexive and reflective strategies differs between depressive and non-depressive individuals. As previously discussed, we predict that depressive individuals will be less likely to use reflective strategies.

### 1.3. Models evaluated

We fit computational models that embody reflective and reflexive strategies to participants’ data to evaluate their strategies. The “Ideal Actor” reflectively updates beliefs and plans ahead, taking into account the information gained by each choice and making choices that maximize long-term payoffs. Action-values are a product of both expected rewards and the potential to reduce uncertainty about the state of the environment. In contrast, the Naïve RL model instantiates the reflexive account of choice, in which the values of actions are based only on the rewards experienced so far. Its beliefs are updated reflexively in response to observed changes in rewards.

Turning to the model details, both models incorporate a Softmax choice rule (Sutton & Barto, 1998), which chooses options as a function of the computed action-values. Critically, the action-values used in the Softmax choice rule differ between the two models, leading to qualitative differences in exploratory behavior. The Naïve RL model explores with equal probability on every trial. For the Ideal Actor model, the probability of exploring increases after each successive exploitive choice (see Fig. 4A).

For the Naïve RL model the value of each action is equal to the last observed reward for that action. Algorithmically, it is equivalent to the Softmax model used in Worthy, Maddox, and Markman (2007) with a learning rate of 1. The Ideal Actor computes action-values in two steps. First, it optimally updates its (Bayesian) beliefs about the state of



**Fig. 1.** The Leapfrog task: example choices over 100 trials. On any trial the lower option might, with a probability of 0.075, increase its reward by 20 points, surpassing the other option. The relative superiority of the two options alternates as their reward values “Leapfrog” over one another. The lines represent the true reward values, the dots a participant’s choices.

the environment based on observations and its estimate of the environment volatility—a free parameter denoted  $P(\text{flip})$ . It then optimally converts those beliefs into action-values using established methods in RL (Kaelbling, Littman, & Moore, 1996). For full formal descriptions of these models, see Knox et al. (2012).

## 2. Method

### 2.1. Participants

One-hundred-thirty-three University of Texas undergraduates participated for course credit and a small cash bonus tied to performance. Participants completed the Center for Epidemiological Studies Depression Scale (CES-D; Radloff, 1977). Following convention (Welsman, Sholomskas, Pottenger, Prusoff, & Locke, 1977), we classified participants as depressive who scored 16 or greater, which reflects mild or greater symptoms of depression (Radloff, 1977). 38 participants were classified as depressive, 95 as non-depressive.

### 2.2. Procedure

Participants performed 300 trials of the Leapfrog task. On each trial, they chose between two options. Initially, one option gave a reward of 10 points and the other 20 points. On any trial the currently lower option could, with fixed probability (i.e. volatility) of 0.075, increase by 20 points, becoming the higher option. In this way, the two options alternate as the best option over time (Fig. 1). On each trial, participants were given 1.5 s to choose between the options. Then the points received were displayed for 1 s. If a choice was not made within 1.5 s, the trial was repeated.

Before the choice task, participants passively viewed 500 training trials. To focus participants’ attention on the environment volatility rather than the true payoffs, during training the payoffs read either CHANGED or SAME indicating whether the reward increased or not. Before each block of 100 training trials, participants estimated the number of jumps they expected in that block.

## 3. Results

We classified each choice as exploratory or exploitive based on the rewards experienced up to that point.

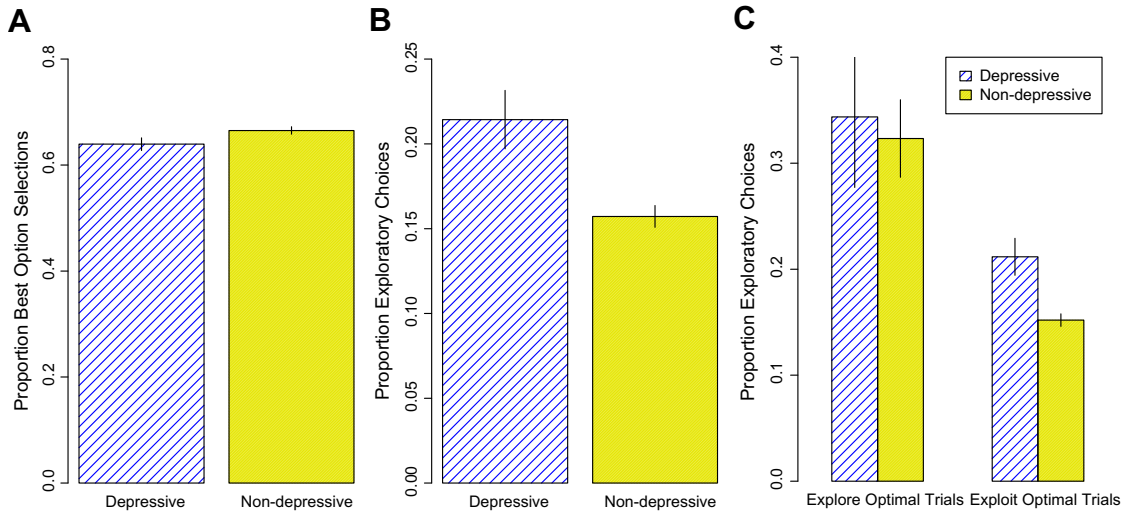
Choosing the option with the highest observed reward was an exploitive choice, and choosing the other option was an exploratory choice (cf. Daw, O’Doherty, Dayan, Seymour, & Dolan, 2006). Depressive participants explored reliably more often than non-depressives (Fig. 2B),  $t(131) = 3.84$ ,  $p < 0.001$ ,  $d = 0.74$ . We measured performance on the task as the proportion of trials that the higher payoff option was chosen, finding that depressives performed marginally worse (Fig. 2A),  $t(131) = 1.92$ ,  $p = 0.057$ ,  $d = 0.37$ .

### 3.1. Model-based analyses

We fit the Ideal Actor and the Naïve RL model to participants’ trial-by-trial choice data by conducting an exhaustive grid search to find the set of parameters that maximized the likelihood of each model for each participant (Table 1). Because the two models have different numbers of free parameters, we determined which model best fit each participant using the Bayesian Information Criterion (BIC; Schwarz, 1978). Fig. 3 shows that a greater proportion of depressive participants (0.47) were better fit by the Naïve RL model than non-depressives (0.21),  $[X^2(1, N = 133) = 9.21, p = 0.002]$ , suggesting a link between depressive symptoms and a lower use of reflective strategies.

Because approximately half of the depressive participants were best fit by each model, we sought to determine if these model fits could uncover subgroups that differed further in depressive symptoms and choice behavior. Indeed, depressives best fit by the Naïve RL model had a higher mean CES-D ( $M = 27.6$ ,  $SD = 7.73$ ) score than those best fit by the Ideal Actor ( $M = 22.7$ ,  $SD = 5.55$ ) [ $t(36) = 2.24, p = 0.03, d = 0.73$ ]. That is, even within depressive participants there is a relationship between decreased expression of reflective strategies and higher depressive symptoms. Unsurprisingly, depressives best fit by the Ideal Actor performed reliably better,  $t(36) = 2.32, p = 0.03, d = 0.75$ . Exploration rates did not differ significantly as a function of best-fitting model for depressives,  $t < 1$ .

Similarly, we compared non-depressive participants that were best fit by the different models. Again, participants best fit by the Ideal Actor performed better,  $t(93) = 5.80, p < 0.001, d = 1.46$ . Curiously, non-depressives best fit by the Ideal Actor model explored more often than those best fit the Naïve RL model,  $t(93) = 5.81, p < 0.001, d = 1.46$ . CES-D scores were not significantly different for



**Fig. 2.** Performance and exploration rates. (A) The proportion of trials that the higher payoff option was chosen. (B) The proportion of trials that an exploratory choice was made. Depressives explored more often than non-depressives. (C) The proportion of exploratory choices for explore-optimal and exploit-optimal trials. Trials were classified based on the prescription of the Ideal Actor given the history of rewards. On exploit-optimal trials, depressives explored more often than non-depressives. Error bars reflect standard errors.

**Table 1**  
BIC and best-fitting parameter values.

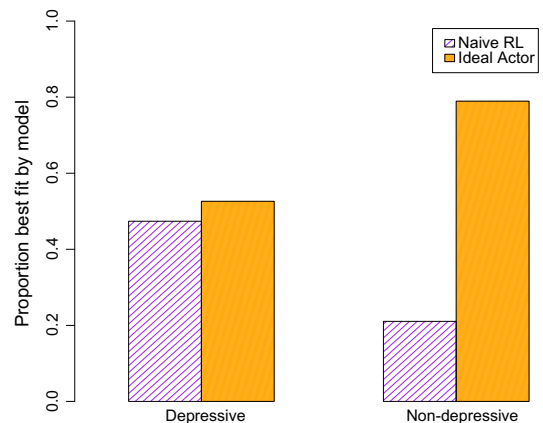
	<i>P</i> (flip) (SD)	Softmax parameter (SD)	Mean BIC (SD)
<i>Naïve RL</i>			
Depressive	NA	0.118 (0.044)	313 (62)
Non-depressive	NA	0.144 (0.030)	279 (57)
<i>Ideal Actor</i>			
Depressive	0.032 (0.056)	0.297 (0.184)	310 (57)
Non-depressive	0.023 (0.028)	0.431 (0.183)	264 (38)

non-depressives as a function of best-fitting model,  $t(93) = 1.38, p = 0.17, d = 0.35$ .

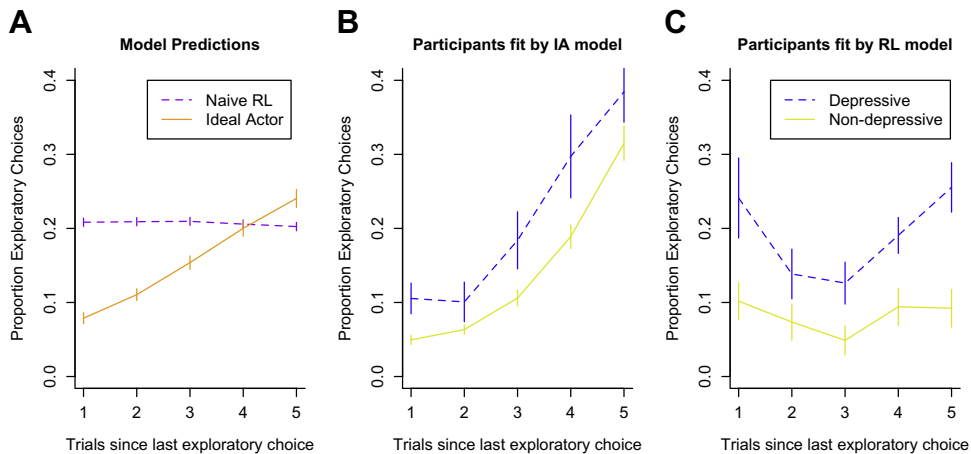
Further, we investigated how participants' choices deviated from the prescriptions of the Ideal Actor. Given the participant's history of choices and rewards, each trial was classified as "explore-optimal" or "exploit-optimal" based on which action the Ideal Actor attributed higher value. Participants' choices were compared to these prescriptions. A mixed-effects logistic regression on exploration rates revealed main effects of depression classification ( $z = 3.37, p < 0.001$ ) and optimal-selection type ( $z = 11.44, p < 0.001$ ), and a significant interaction ( $z = 3.43, p < 0.001$ ). Depressive participants primarily deviated from the optimal strategy by exploring when they should have been exploiting (Fig. 2C), exploring more frequently on exploit-optimal trials compared to non-depressives,  $t(131) = 4.14, p < 0.001, d = 0.80$ . Exploration rates were not significantly different for explore-optimal trials between the groups,  $t < 1$ .

A critical aspect of behavior for which the models make divergent predictions is the sequential structure of exploratory choices. With each successive exploitive choice, the probability that an unseen flip has occurred increases,

resulting in greater uncertainty in the state of the environment. Taking this uncertainty into account, the Ideal Actor's probability of exploring increases after each consecutive exploitive choice. The Naïve RL model does not track uncertainty, exploring with equal probability on every trial. Consistent with model predictions, participants best fit by the Ideal Actor show an increasing pattern in exploration rates as the number of consecutive exploitive trials increases, whereas exploration rates for those best fit by the Naïve RL model are relatively flat (Fig. 4). A logistic regression was performed on these exploration rates to calculate a slope for each subject. Slopes for depressives were reliably lower than for non-depressives,  $t(131) = 1.99, p < 0.05, d = 0.38$ . Supporting our modeling results, this indicates decreased use of reflective strategies in depressives. Consistent with our prior observations (Fig. 2A), depressives exhibit a greater base rate of exploration, regardless of best-fitting model.



**Fig. 3.** The proportion of participants best fit by each model. Depressives were more often best fit by the Naïve RL model than non-depressives.



**Fig. 4.** Exploration rates as a function of number of successive exploitive choices. (A) Predictions from the two models. (B) Behavioral results for participants best-fit by the Ideal Actor. (C) Results for participants best fit by the Naïve RL model. Error bars reflect standard errors.

#### 4. Discussion

We found that exploratory behavior differs between depressive and non-depressive participants. Relatively, depressives were better fit by the reflexive Naïve RL model, which utilizes a minimal representation of the environment and explores in an undirected way. Further, depressive individuals best described by the Naïve RL model reported the highest levels of depression symptoms. We suggest that depressive individuals may have difficulty maintaining the complex representation of the task structure necessary to perform optimally, and so rely more on parsimonious choice strategies.

We also found that depressive participants explored more, which contrasts with some previous findings (Cella et al., 2010; Huys et al., 2012). One simple explanation for this surface discrepancy is that our task only contained gains. Depressives tend to display enhanced processing of punishment and reduced processing of rewards (Maddox et al., 2012; Roiser, Elliott, & Sahakian, 2011). Decreased sensitivity to rewards may lead depressives to undervalue the benefit of obtaining the higher reward, thereby increasing the relative value of exploring. Our analysis comparing participants' choices to the Ideal Actor's prescriptions supports this explanation. Compared to non-depressives, depressive participants explored more often when the Ideal Actor indicated they should be exploiting. Complementarily, increased sensitivity to punishment could discourage depressives from exploring when large losses can occur (e.g., Cella et al., 2010; Huys et al., 2012). Other work, though, has found that depressed individuals show blunted responses to both positive and negative stimuli (Bylsma, Morris, & Rottenberg, 2008). Thus, future work should address how loss and gain framing impacts exploration in depressive decision-makers.

Our analysis is unusual in that we focus on the nature of exploration rather than its overall level. Reflective and reflexive strategies can produce equivalent exploration rates while leading to different performance and patterns of exploration. These two strategies differ critically in their

prescription of *when* to explore, not *how often* to explore. The greater exploration rates exhibited by depressive individuals and that overall they are characterized by a more reflexive account of choice are distinct phenomena. Indeed, exploration rates do not differ significantly between depressives best fit by different models, and among non-depressives, those fit by the Ideal Actor explored more often. This discrepancy highlights the importance of considering the possible mechanisms underlying exploratory behavior, rather than simply the overall level of exploration.

A limitation of the current study is that we did measure anxiety levels and so cannot determine whether anxiety may have contributed to the results. Another limitation is that we did not assess whether participants suffered from clinical levels of depression or took antidepressant medication. Instead, we focused on effects of elevated depressive symptoms and found meaningful differences in exploratory behavior between individuals with higher and lower levels of depressive symptoms. We believe this highlights the sensitivity of our choice paradigm to detect differences in exploration strategies and, more broadly, the importance of understanding choice in depression. Future work should seek to extend these findings to Major Depressive Disorder and should explore ways that decision-making can be modified and enhanced in depression. As demonstrated by the current study, insights and methods from cognitive and mathematical psychology may prove useful in this important endeavor.

#### Acknowledgements

This research was supported in part by NIMH grant MH091523-01A1 to BCL and NIDA grant DA032457 to WTM and CGB.

#### References

- Badre, D., Doll, B. B., Long, N. M., & Frank, M. J. (2012). Rostrolateral prefrontal cortex and individual differences in uncertainty-driven exploration. *Neuron*, 73, 595–607.



- Bylsma, L. M., Morris, B. H., & Rottenberg, J. (2008). A meta-analysis of emotional reactivity in major depressive disorder. *Clinical Psychology Review*, 28(4), 676–691.
- Cella, M., Dymond, S., & Cooper, A. (2010). Impaired flexible decision-making in major depressive disorder. *Journal of Affective Disorders*, 124(1), 207–210.
- Clark, L., Chamberlain, S. R., & Sahakian, B. J. (2009). Neurocognitive mechanisms in depression: Implications for treatment. *Annual Review of Neuroscience*, 32, 57–74.
- Daw, N. D., Niv, Y., & Dayan, P. (2005). Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. *Nature Neuroscience*, 8(12), 1704–1711.
- Daw, N. D., O'Doherty, J. P., Dayan, P., Seymour, B., & Dolan, R. J. (2006). Cortical substrates for exploratory decisions in humans. *Nature*, 441, 876–879.
- Elderkin-Thompson, V., Mintz, J., Haroon, E., Lavretsky, H., & Kumar, A. (2006). Executive dysfunction and memory in older patients with major and minor depression. *Archives of Clinical Neuropsychology*, 21(7), 669–676.
- Elliott, R., Baker, S. C., Rogers, R. D., O'leary, D. A., Paykel, E. S., Frith, C. D., et al. (1997). Prefrontal dysfunction in depressed patients performing a complex planning task: A study using positron emission tomography. *Psychological Medicine*, 27(04), 931–942.
- Eshel, N., & Roiser, J. P. (2010). Reward and punishment processing in depression. *Biological Psychiatry*, 68(2), 118–124.
- Gradin, V. B., Kumar, P., Waiter, G., Ahearn, T., Stickle, C., Milders, M., et al. (2011). Expected value and prediction error abnormalities in depression and schizophrenia. *Brain*, 134(6), 1751–1764.
- Huys, Q. J., Eshel, N., O'Nions, E., Sheridan, L., Dayan, P., & Roiser, J. P. (2012). Bonsai trees in your head: How the Pavlovian system sculpts goal-directed choices by pruning decision trees. *PLoS Computational Biology*, 8(3), e1002410.
- Kaelbling, L. P., Littman, M. L., & Moore, A. W. (1996). Reinforcement learning: A survey. *Journal of Artificial Intelligence Research*, 4, 237–285.
- Kessler, R. C., Berglund, P., Demler, O., Jin, R., Koretz, D., Merikangas, K. R., et al. (2003). The epidemiology of major depressive disorder: Results from the National Comorbidity Survey Replication (NCS-R). *JAMA*, 289(23), 3095–3105.
- Knox, W. B., Otto, A. R., Stone, P., & Love, B. C. (2012). The nature of belief-directed exploratory choice in human decision-making. *Frontiers in Psychology*, 2.
- Maddox, W. T., Gorlick, M. A., Worthy, D. A., & Beevers, C. G. (2012). Depressive symptoms enhance loss-minimization, but attenuate gain-maximization in history-dependent decision-making. *Cognition*, 125, 118–124.
- Montague, P. R., Dolan, R. J., Friston, K. J., & Dayan, P. (2012). Computational psychiatry. *Trends in Cognitive Sciences*, 16(1), 72–80.
- Murphy, F. C., Rubinsztein, J. S., Michael, A., Rogers, R. D., Robbins, T. W., Paykel, E. S., et al. (2001). Decision-making cognition in mania and depression. *Psychological Medicine*, 31(4), 679–693.
- Otto, A. R., Gershman, S. J., Markman, A. B., & Daw, N. D. (2013). The curse of planning: Dissecting multiple reinforcement learning systems by taxing the central executive. *Psychological Science*, 24(5), 751–761.
- Otto, A. R., Markman, A. B., Gureckis, T. M., & Love, B. C. (2010). Regulatory fit and systematic exploration in a dynamic decision-making environment. *Journal of Experimental Psychology Learning and Memory Cognition*, 36, 797–804.
- Paulus, M. P., & Yu, A. J. (2012). Emotion and decision-making: Affect-driven belief systems in anxiety and depression. *Trends in Cognitive Sciences*, 16(9), 476–483.
- Pizzagalli, D. A., Iosifescu, D., Hallett, L. A., Ratner, K. G., & Fava, M. (2008). Reduced hedonic capacity in major depressive disorder: Evidence from a probabilistic reward task. *Journal of Psychiatry Research*, 43(1), 76–87.
- Radloff, L. S. (1977). The CES-D scale: A self-report depression scale for research in the general population. *Applied Psychological Measurement*, 1(3), 385–401.
- Rogers, M. A., Kasai, K., Koji, M., Fukuda, R., Iwanami, A., Nakagome, K., et al. (2004). Executive and prefrontal dysfunction in unipolar depression: A review of neuropsychological and imaging evidence. *Neuroscience Research*, 50(1), 1–11.
- Roiser, J. P., Elliott, R., & Sahakian, B. J. (2011). Cognitive mechanisms of treatment in depression. *Neuropsychopharmacology*, 37(1), 117–136.
- Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, 6, 461–464.
- Sutton, R., & Barto, A. G. (1998). *Reinforcement learning*. Cambridge, MA: MIT Press.
- Wellsman, M. M., Sholomskas, D., Pottenger, M., Prusoff, B. A., & Locke, B. Z. (1977). Assessing depressive symptoms in five psychiatric populations: A validation study. *American Journal of Epidemiology*, 106, 203–214.
- Worthy, D. A., Maddox, W. T., & Markman, A. B. (2007). Regulatory fit effects in a choice task. *Psychonomic Bulletin & Review*, 14, 1125–1132.