

A Role Beyond Learning for NMDA Receptors in Reward-Based Decision-Making—a Pharmacological Study Using d-Cycloserine

Jacqueline Scholl^{*1}, Jan Günthner¹, Nils Kolling², Elisa Favaron¹, Matthew FS Rushworth^{2,3}, Catherine J Harmer¹ and Andrea Reinecke¹

¹Department of Psychiatry, University of Oxford, Warneford Hospital, Oxford, UK; ²Department of Experimental Psychology, University of Oxford, Oxford, UK; ³Centre for Functional Magnetic Resonance Imaging of the Brain (FMRIB), University of Oxford, Nuffield Department of Clinical Neurosciences, John Radcliffe Hospital, Oxford, UK

N-methyl-D-aspartate (NMDA) receptors are known to fulfill crucial functions in many forms of learning and plasticity. More recently, biophysical models, however, have suggested an additional role of NMDA receptors in evidence integration for decision-making, going beyond their role in learning. We designed a task to study the role of NMDA receptors in human reward-guided learning and decision-making. Human participants were assigned to receive either 250 mg of the partial NMDA agonist d-cycloserine ($n = 20$) or matching placebo capsules ($n = 27$). Reward-guided learning and decision-making were assessed using a task in which participants had to integrate learnt and explicitly shown value information to maximize their monetary wins and minimize their losses. To tease apart the effects of NMDA on learning and decision-making we used simple learning models. D-cycloserine shifted decision-making towards a more optimal integration of the learnt and the explicitly shown information, in the absence of a direct learning effect. In conclusion, our results reveal a distinct role for NMDA receptors in reward-guided decision-making. We discuss these findings in the context of NMDA's roles in neuronal super-additivity and as crucial for evidence integration for decisions.

Neuropsychopharmacology (2014) **39**, 2900–2909; doi:10.1038/npp.2014.144; published online 9 July 2014

INTRODUCTION

Animals and humans live in ever changing and complex environments. They need to continuously track and learn about the changing properties of their environment and use them to behave adaptively. The neural mechanisms engaged in learning about the environment depend on the type of information being tracked. Nevertheless, the plasticity underpinning several of these various forms of learning has been argued to depend on the same molecular mechanism, ie, N-methyl-D-aspartate (NMDA) subtypes of glutamate receptors. Accordingly a number of studies have investigated whether NMDA receptor manipulations affect learning and memory (Bannerman *et al*, 2012; Kuriyama *et al*, 2011; Bohn *et al*, 2003).

Most of the studies looking at the role of NMDA receptors in reward learning have focused on simple tasks, such as a single association between a stimulus and an outcome or its reversal. However, in more ecological, and thus complex, scenarios, learning and decision-making often involve

consideration of positive and negative aspects of potential outcomes as well as the integration of learnt information with information that is explicitly cued. Importantly, it has recently been shown that integration of information conveyed by different dimensions during decision-making is an active process recruiting particular neural mechanisms (Burke *et al*, 2013; Stein and Stanford, 2008). Thus, to understand the full role of NMDA receptors in reward learning and decision-making, it might not be sufficient to only study a single simple component in isolation.

Interestingly, there is some, albeit indirect, evidence that NMDA receptors are not only important for learning, but also for integration of information: NMDA receptor blockade has been reported to affect the integration of multisensory information in cat superior colliculus (Binns and Salt, 1996) or the integration of reward and delay in rats (Floresco *et al*, 2008). However, to our knowledge, it has not been investigated whether NMDA receptors play a role in human value-based decision-making and learning.

To examine the influence of changes in NMDA receptor activity on complex learning and decision-making behaviors in humans, we used the partial NMDA agonist d-cycloserine. D-cycloserine binds to the glycine site of the NMDA receptor. Glycine is a co-agonist of the NMDA receptor, meaning that NMDA receptors only open when both glycine and glutamate bind. D-cycloserine can thus

*Correspondence: Jacqueline Scholl, Department of Psychiatry, University of Oxford, Warneford Hospital, Oxford OX3 7JX, UK, Tel: +44 01865 223 606, Fax: +44 01865 251 076, E-mail: Jacqueline.scholl@psy.ox.ac.uk

Received 13 March 2014; revised 9 May 2014; accepted 30 May 2014; accepted article preview online 13 June 2014

increase the probability of glutamate release opening NMDA receptors, which in turn enhances NMDA receptor-mediated activation. We designed a multi-attribute decision-making task in which participants learnt changing accepted article preview 13 June 2014 of gains and losses of two options. They made choices between options by integrating those learnt probabilities with explicitly cued information about gain and loss magnitudes. To assess potential effects of d-cycloserine on learning and decision-making, we used reinforcement-learning models.

Surprisingly, we did not find any evidence for a change in the rate at which participants learned about reward or punishment outcomes. However, we found that d-cycloserine improved decision-making. D-cycloserine led to a more optimal integration of the learnt probability information with the explicitly cued magnitude information.

METHODS

Participants

The study was approved by the local ethics committee. In total, 52 healthy volunteers (age 18–30) took part in the study (inclusion details in Supplementary Methods). The groups were well-matched on sociodemographic and personality parameters (Table 1). Five participants were excluded (Supplementary Methods). There remained 20 participants in the d-cycloserine and 27 participants in the placebo group.

Procedure

In a double-blind, placebo-controlled design, participants were randomly allocated to a single dose of d-cycloserine (250 mg) or matching placebo capsule. They fasted 2 h before the testing visit. Approximately, 250 mg was chosen in agreement with recent studies (Klumpers *et al*, 2012; Onur *et al*, 2010). Participants were tested 3 h after drug administration. According to product information (King's Pharmaceutical), plasma peak levels are reached within 3–4 h; other studies (van Berckel *et al*, 1997, 1998; Patel *et al*, 2011) found that peak levels are reached within ~1 h. However, given d-cycloserine's half-life of 8–12 h (product information) or 15 h (Patel *et al*, 2011), plasma levels would have been close to peak levels during testing, given either time-to-peak information. To assess potential subjective changes following d-cycloserine, participants completed questionnaires (Table 1) before capsule intake and before testing.

Probabilistic Instrumental Learning Task

Participants performed a probabilistic learning task with monetary wins and losses (Figure 1a). Participants made repeated choices between two options with the aim to maximize their monetary pay-off.

On each trial, participants had a choice between two options. Each option had four independent attributes: a reward and a loss magnitude, a reward and a loss probability. The magnitude determined how many points could be won (and lost) on this trial, while the probabilities determined how probable winning and losing was respec-

tively. After participants selected one of the options, they were shown the outcomes for both options. However, only the option they had chosen contributed to the participants' earnings. In trials where the chosen option incurred both, wins and losses, the participants' earnings in that trial were the sum of both. Therefore, to maximize the overall gains, participants should be choosing the option on each trial with the highest reward utility (reward magnitude \times reward probability) and with the lowest loss utility (loss magnitude \times loss probability).

Reward and loss magnitudes were explicitly cued at the time of choice and were randomly drawn from a flat distribution between 1 and 100 points. In contrast, the probabilities were not explicitly shown and had to be learnt across trials by observing the outcomes. The outcomes for one option could either be a win and a loss, only a win, only a loss, or neither win nor loss. The independent reward and loss probabilities determined the probability of these outcomes. The probabilities varied over the course of the experiment between 20 and 80%, with only one of the four probabilities varying at any given time (Figure 1c).

Before the experiment, participants were instructed about the task (see Supplementary Methods). Participants first performed 30 training trials, followed by 381 test trials, which were included in the analysis. Each participant was tested using the same task schedule to allow for better group comparisons. At the start of the task participants were given 4£ (400 points) to ensure that they had sufficient funds to sustain losses incurred even at the beginning of the experiment.

Analysis

The behavioral analysis compared the effects of d-cycloserine *vs* placebo on (a) learning of reward and loss probabilities, and (b) integration of learnt (probabilities) and explicitly cued (magnitudes) information for guiding complex decisions. All analyses were performed in Matlab and SPSS.

Logistic regression analysis. To ensure that participants learnt the probabilities, we first assessed the impact of past outcomes (reward and loss) and the explicitly cued magnitudes (reward and loss) on choice, using a logistic regression analysis and normalized regressor estimates. We included regressors for the last five trial differences in reward and loss outcomes between the two options, as well as the differences in the explicitly cued magnitudes.

To investigate whether the groups differed in their learning speeds, the resulting regression weights for the past outcomes were entered into an ANOVA with group (d-cycloserine *vs* placebo) as a between-subject factor and time (1, 2, 3, 4, or 5 trials in the past) and valence (reward or loss) as within-subject factors.

Modeling

To look at the learning effects more specifically and assess participants' strategies for the integration of information to make decisions, we used reinforcement-learning models to fit each participant's trial-by-trial behavior.

Table 1 Sociodemographic and Questionnaire Measurements

Demographics and questionnaire measurements							
	Pla (n = 27)		DCS (n = 20)		P		
Age	22.2 ± 0.6		22.3 ± 0.7		0.93		
Gender, F:M	15:12		11:9		0.97		
BDI	2.0 ± 0.5		0.9 ± 0.4		0.14		
Education years	16.9 ± 0.3		15.9 ± 0.5		0.10		
Trait anxiety	30.9 ± 1.6		30.5 ± 1.3		0.86		
BMI	22.0 ± 0.4		22.0 ± 0.5		0.99		
Neuroticism	5.2 ± 0.9		5.6 ± 0.9		0.33		
ACS focusing	26.7 ± 0.6		26.2 ± 0.9		0.64		
ACS shifting	34.7 ± 0.8		34.5 ± 1.0		0.31		
BIS	16.0 ± 0.8		16.7 ± 0.8		0.59		
BAS	24.4 ± 5.8		23.6 ± 4.8		0.59		

VAS items	Pla, before	Pla, after	DCS, before	DCS, after	P, before	P, after	P, diff score
Anxious	7.6 ± 1.6	3.6 ± 0.6	6.7 ± 1.5	5.7 ± 1.8	0.70	0.25	0.24
Sleepy	28.4 ± 4.0	21.3 ± 4.3	24.1 ± 3.3	13.7 ± 2.5	0.43	0.13	0.54
Flushed	9.1 ± 2.0	3.2 ± 0.8	7.1 ± 1.9	3.7 ± 0.7	0.50	0.62	0.31
Tearful	3.3 ± 0.9	2.9 ± 0.8	3.1 ± 0.7	2.8 ± 0.5	0.84	0.87	0.95
Nauseous	3.3 ± 0.9	3.2 ± 0.9	2.9 ± 0.7	3.5 ± 0.6	0.72	0.83	0.41
Hopeless	3.3 ± 0.7	2.5 ± 0.5	4.1 ± 1.5	2.7 ± 0.5	0.61	0.85	0.65
Tremor	3.3 ± 0.9	3.2 ± 1.0	3.7 ± 1.0	3.1 ± 0.6	0.77	0.91	0.58
Sad	4.9 ± 0.9	2.8 ± 0.5	4.5 ± 1.3	2.8 ± 0.5	0.83	0.98	0.80
Dizzy	2.7 ± 0.7	3.5 ± 1.2	3.0 ± 0.7	5.4 ± 2.0	0.78	0.41	0.38
Depressed	2.9 ± 0.5	2.6 ± 0.5	3.1 ± 0.8	2.5 ± 0.5	0.85	0.90	0.65
Tachycardia	4.5 ± 1.4	3.5 ± 1.2	5.7 ± 1.5	3.5 ± 0.7	0.57	0.98	0.37
Alert	51.3 ± 5.0	48.4 ± 5.1	57.8 ± 5.0	52.3 ± 5.4	0.37	0.61	0.51

The following measurements were obtained before drug administration: age, gender, Beck's Depression Inventory (BDI, Beck et al, 1996), years of education at time of study, trait anxiety (Spielberger et al, 1983), body mass index (BMI, weight(kg)/height(cm)²), neuroticism (Eysenck and Eysenck, 1994), attention control scale (ACS, Derryberry and Reed, 2002), behavioral inhibition, behavioral activation (BIS/BAS Scale, Carver and White, 1994). The values reported are mean values with standard errors and the *P*-scores from between-subject *t*-tests (apart from the value for gender ratio, where a chi-squared test was used). Visual analog scales were given to participants before and after drug administration. For each of the listed items, they were asked to indicate how they were feeling by placing a tick mark on a 100 mm line, which was labeled 'not at all' on the left-hand side and 'extremely' on the right hand side. The values reported are the mean and standard error for the tick mark positions (in mm). *P*-values were calculated for the group differences in the first (at baseline, 'before') and second measurement (after drug administration, 'after') and on the difference scores (baseline–second measurement, 'diff score').

Each model consisted of three main components. First, the model had estimates about the probabilities underlying the outcomes of both options. These were updated on every trial using a reinforcement-learning algorithm. Second, the probability estimates were integrated with the explicitly cued magnitudes to calculate how valuable each of the two options was (ie, their utility). Third, these two utilities were compared to predict participants' choices. To determine the best parameter estimates for every participant, we used a standard log-likelihood maximization procedure.

When calculating how valuable each option is, participants might use different decision strategies for integrating learnt probabilities with explicit magnitudes. They could use a mathematically optimal strategy (utility as probability × magnitude). However, as this optimal strategy may be quite cognitively taxing, they could resort to a heuristic strategy (utility as a weighted sum of probability and magnitude). To test which decision strategy partici-

pants used, we fitted different models to the data. To test for differences in strategy, we then compared how well each of these models explained the groups' behavior. In addition, we fitted a third model, which directly estimated to what degree they used either decision strategy.

Optimal model. This model assumed that participants integrated the learnt probabilities *optimally* with the explicit magnitudes (magnitude × probability). The learning of probabilities was modeled using a standard reinforcement-learning rule. On each trial, the estimated probability of an attribute was updated based on the trial's outcome, as a function of the prediction error (PE):

$$\text{Prediction}_t = \text{Prediction}_{t-1} + \alpha \times \text{PE}_{t-1}$$

with

$$\text{PE}_{t-1} = \text{Outcome}_{t-1} - \text{Prediction}_{t-1}$$

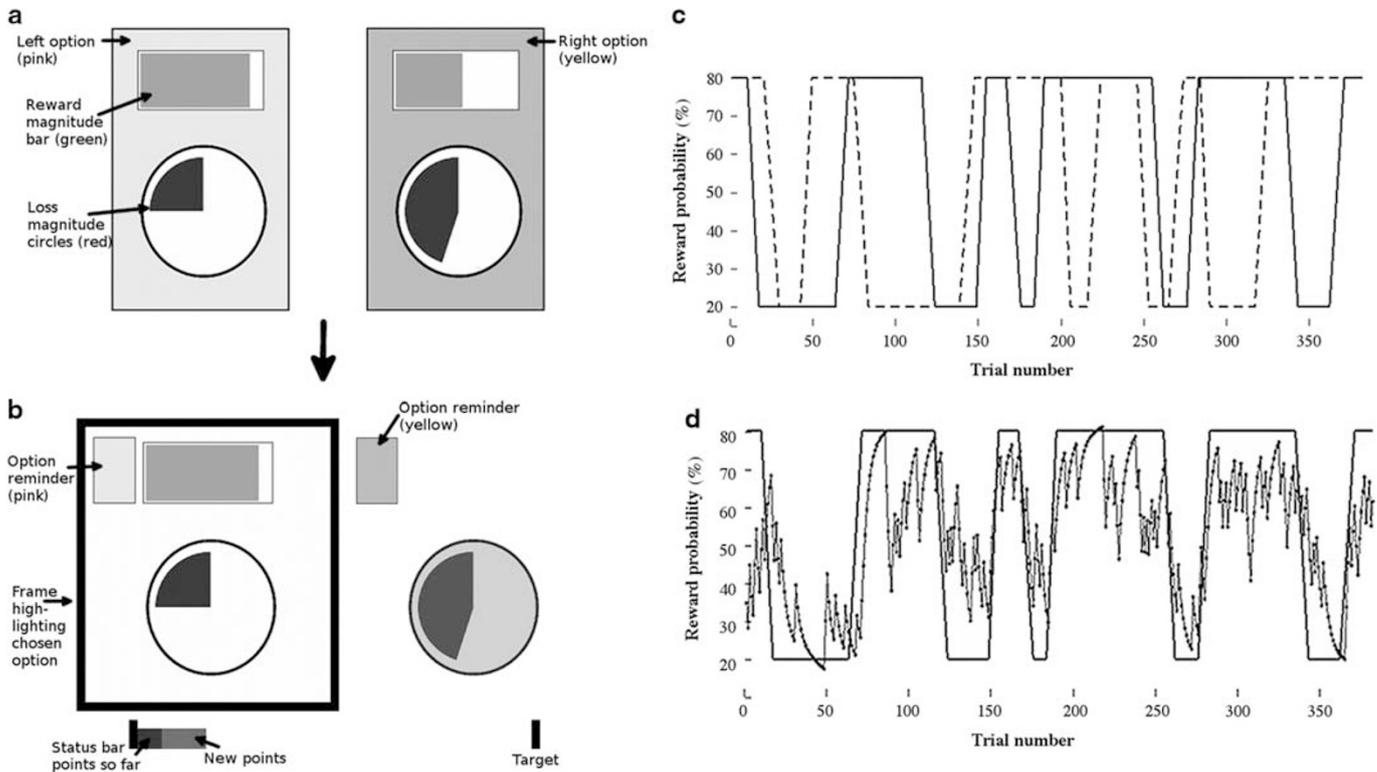


Figure 1 (a) At the beginning of a trial two options appeared on the left (pink square) and right side (yellow square) of the screen. Throughout the experiment, the pink square was always on the left side and the yellow square was always on the right side. Reward (bars at the top) and loss magnitudes (circles at the bottom) were presented overlaid on the option symbols. After 500 ms, a question mark appeared after which the participants chose an option. After participants made their selection (there was no time-out), the outcomes of the gambles were shown (b), first for the chosen option (duration: 2.5 s—left option in the example), then also for the unchosen option (duration: 2.5 s). If the gamble outcomes of the chosen option led to a reward, the reward bar was shown; otherwise the reward bar was not shown. Similarly, if the gamble outcome of the chosen option led to a loss, the loss circle was shown; otherwise, it was not shown. The sum of the reward and the loss incurred for the chosen option in a trial was added to a status bar at the bottom of the screen, allowing participants to keep track of their overall gains. Subsequently, the participants were shown the outcomes for the unchosen option in the same way, except that no points were added to the status bar. Importantly, presenting the outcomes of the chosen and the unchosen option ensured that participants had an equal chance to learn the probabilities of the chosen and the unchosen option. After an inter-trial interval of 1.5 s, the next trial started. (c) Example reward probabilities for the two options over the course of the experiment. The probabilities were either stable at 20% or 80%, or they drifted between 20 and 80%, taking between five and eight trials per drift. (d) Example reward probability for one of the options (solid line), together with the probability estimates from the Bayesian learner used (dotted line).

where α is the learning rate. Thus, the learning rate is a measure of how much participants updated their probability estimate when the outcome associated with an attribute differed from their expectation (eg, the probability that the left option yields a reward). Separate learning rates were used for learning about wins and losses.

These predictions were combined optimally with the shown magnitudes:

$$Utility_{\text{Reward}} = Probability_{\text{Reward}} \times Magnitude_{\text{Reward}}$$

The loss utility was computed in the same way and combined with the reward utility:

$$Utility_{\text{total}} = Utility_{\text{Reward}} - \lambda \times Utility_{\text{Loss}}$$

where λ determines how much participants weighted the prospect of rewards vs losses.

A standard soft-max decision rule was used to predict the probability of choosing an option.

$$P(\text{Option}_A) = \frac{e^{Utility_A \times \beta}}{e^{Utility_A \times \beta} + e^{Utility_B \times \beta}}$$

where β reflects a participant's ability to pick the option with higher utility.

To assess the effect of d-cycloserine on learning about wins and losses, we compared their respective learning rates between groups.

Heuristic model. This model differed from the optimal model only in the decision strategy for integrating learnt probabilities with explicit magnitudes. Instead of being a product of probability and magnitude, utility was computed as a weighted sum.

$$Utility_{\text{Reward}} = \mu \times Probability_{\text{Reward}} + Magnitude_{\text{Reward}}$$

where μ is the probability weighting factor, describing the relative importance of the learnt probability compared to the explicit. The loss utility was computed in the same way, sharing the same μ .

Again, we compared the groups' learning rates. Additionally, a change in learning could also manifest as a changed reliance on learnt compared to explicit information, we therefore also compared the probability weighting factors between the groups.

Hybrid model. We used the hybrid model to examine whether d-cycloserine affected how participants integrated information, shifting them towards a more optimal, less heuristic decision strategy. The hybrid model computed utility as a weighted sum of the utilities from the optimal and the heuristic model:

$$\text{Utility} = \omega \times \text{Utility}_{\text{Heuristic}} + (1 - \omega) \times \text{Utility}_{\text{Optimal}}$$

where ω is the heuristic weight factor, determining how much the overall utility is like the heuristic or the optimal utility. The higher ω (between 0 and 1), the more a participant relied on a heuristic decision rule.

Model Comparison

If the groups differ in their decision strategy, this should also be reflected in how well the models incorporating the different strategies can explain behavior. To assess this, we compared the model fits using the Akaike information criterion (AIC). We calculated for each participant the AIC differences between the optimal and the heuristic model and also between the heuristic and the hybrid model. This allowed us to compare whether the d-cycloserine differed from the placebo group in how well their behavior was explained by the heuristic relative to the optimal and by the heuristic relative to the hybrid model.

To confirm the results from the modeling analysis using a different method, we performed an additional regression analysis with regressors analogous to the components of the hybrid model. We included regressors for the explicit magnitude differences, for the learnt probability estimate differences, as well as for the difference in optimal utilities (magnitude \times probability). The probability estimates were obtained using a Bayesian learner, like the one described in Behrens *et al*, 2007; also see Supplementary Methods. These Bayesian probability estimates are the most accurate estimates a participant could have given the past outcomes (Figure 1d). If participants' decision strategy is more heuristic, the main effects should have a larger impact on behavior. Conversely, if the decision strategy is more optimal, the interaction term (magnitude \times probability) should have a higher impact.

Learning About the Unchosen Option

There is some evidence (Boorman *et al*, 2011) that different brain areas are used to learn about the chosen and the unchosen option, thus there is a possibility they could be affected differently by d-cycloserine. We found that d-cycloserine did not affect the usage of the unchosen option's outcomes for decision-making or learning (see Supplementary Methods and Results).

RESULTS

General Performance

In the task, participants had to constantly track the independent win and loss probabilities of two options and integrate these with explicitly cued reward and loss magnitudes. The groups did not differ in the overall earnings or in the mean points won/lost (Table 2a).

Table 2a Results of a General Behavioral Analysis

Overall behavior	Pla	DCS	P
Total money won	12.9 \pm 1.1	13.0 \pm 0.9	0.66
Mean points won	35.4 \pm 1.1	35.4 \pm 1.1	0.89
Mean points lost	21.8 \pm 1.4	21.6 \pm 0.3	0.67
Mean reward magnitude chosen	60.7 \pm 0.4	60.7 \pm 0.6	0.92
Mean loss magnitude chosen	48.0 \pm 0.5	48.0 \pm 0.5	0.96
Mean reward gamble outcome	0.6 \pm 0.0	0.6 \pm 0.0	0.9
Mean loss gamble outcome	0.4 \pm 0.0	0.4 \pm 0.0	0.86

Table 2b Parameter Estimates (Mean and Standard Error) for both Groups and the Results of Significance Tests for the Optimal Decision Model

Optimal decision model	Pla	DCS	P
LR reward	0.16 \pm 0.03	0.17 \pm 0.04	0.83
LR loss	0.27 \pm 0.05	0.22 \pm 0.04	0.52
Beta	0.06 \pm 0.006	0.07 \pm 0.007	0.40
Loss weight (λ)	0.90 \pm 0.07	0.90 \pm 0.07	0.94
fval	165 \pm 5	154 \pm 6	0.15

Table 2c The Heuristic Decision Model

Heuristic decision model	Pla	DCS	P
LR reward	0.44 \pm 0.05	0.35 \pm 0.04	0.20
LR loss	0.43 \pm 0.05	0.38 \pm 0.06	0.52
Beta	0.04 \pm 0.00	0.04 \pm 0.00	0.81
Loss weight (λ)	0.88 \pm 0.05	0.87 \pm 0.05	0.86
Prob weight (μ)	71.7 \pm 9.6	77.8 \pm 15.1	0.72
fval	146 \pm 5	142 \pm 6	0.62

Table 2d The Hybrid Decision Model

Hybrid decision model	Pla	DCS	P
LR reward	0.46 \pm 0.05	0.32 \pm 0.04	0.07
LR loss	0.42 \pm 0.05	0.36 \pm 0.05	0.47
Beta	0.05 \pm 0.00	0.06 \pm 0.00	0.08
Loss weight (λ)	0.89 \pm 0.05	0.90 \pm 0.04	0.93
Prob weight (μ)	69.70 \pm 10.8	119.8 \pm 63.8	0.38
Heuristic weight (ω)	0.84 \pm 0.03	0.59 \pm 0.23	4×10^{-5}
fval	146 \pm 5	139 \pm 6	0.41

Logistic Regression

As a measure of learning, we assessed the impact of past reward/loss outcomes on choice, using a logistic regression analysis. We also included the currently displayed reward/loss magnitudes. Participants were more likely to pick options with higher reward ($t(46) = 18.6$, $P = 5 \times 10^{-23}$)

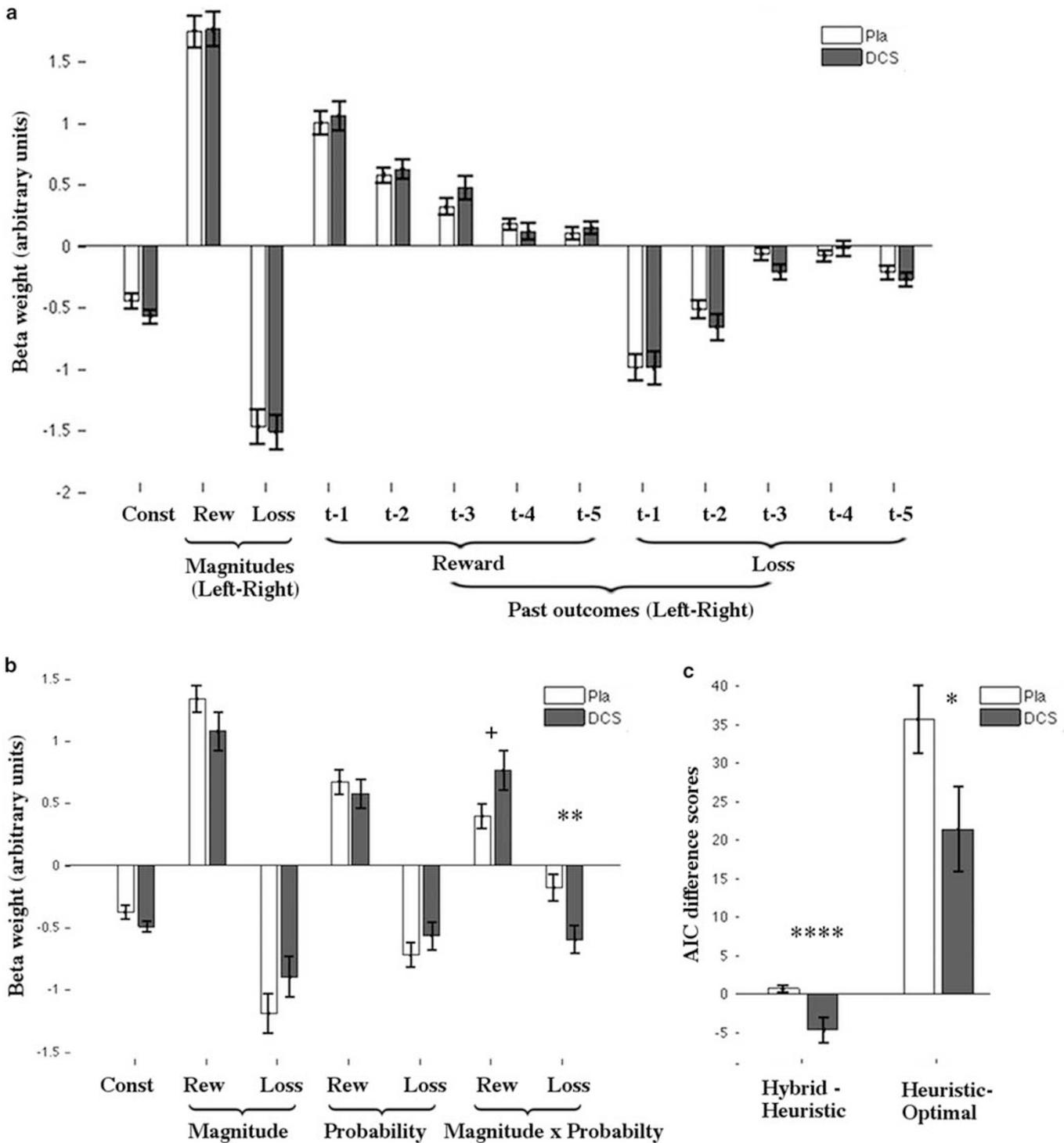


Figure 2 (a) Decision weights (beta) for placebo (white) and d-cycloserine (gray), showing the decision impact of current magnitude differences (left–right) and past gamble outcome differences (left–right), for one to five trials in the past, on choice. (b) Decision weights (beta) based on a regression using shown magnitude differences, probability prediction differences (estimated using a Bayesian model), and their interaction for both groups. (c) AIC difference scores comparing the relative fit of the hybrid to the heuristic and of the heuristic to the optimal model in both groups. Error bars indicate SE. $+P = 0.056$, $*P \leq 0.05$, $**P < 0.001$, $****P < 10^{-4}$.

and lower loss magnitudes ($t(46) = -14.9$, $P = 3 \times 10^{-19}$). They also chose options more frequently when they were associated with more past wins and fewer losses (Figure 2a), thus suggesting that they were able to learn from past outcomes.

To test whether learning differed between the groups, and maybe in dependence on reward and loss valence, we ran a 2 (group) \times 5 (time point) \times 2 (valence) ANOVA on the regression weights of the past outcomes (Figure 2a). That participants learnt the reward/loss probabilities over time

was evidenced by the fact that recent reward/loss outcomes influenced choices more than longer ago ones (main effect of time, $F(4,180) = 80.6$, $P < 10^{-6}$). This effect of time was stronger for losses than rewards (interaction effect, time \times valence, $F(4,180) = 6.2$, $P = 3.8 \times 10^{-4}$), suggesting that loss probabilities were learnt more quickly.

Importantly, the groups neither differed in overall learning speed (time \times group: $F(4,180) = 0.9$, $P = 0.42$), nor in their relative learning speeds for wins and losses (time \times valence \times group, $F(4,180) = 0.3$, $P = 0.89$).

Modeling

The aims of the modeling analyses were to assess whether the groups differed in their learning rate (all three models), as well as whether they differed in their decision strategies (model three). Each model used a standard prediction-error learning rule for learning about reward/loss probabilities. The models only differed in how learnt probabilities were integrated with explicit reward/loss magnitudes.

Optimal model. This model integrated magnitudes and probabilities optimally using a multiplicative rule. For assessing differences in learning, we compared the learning rate parameters between the groups. Surprisingly, the groups did not differ in their learning rates (reward: $t(45) = -0.2$, $P = 0.8$, loss: $t(45) = 0.7$, $P = 0.5$), nor did they differ in any other model parameter (Table 2b).

Heuristic model. Participants might use a more heuristic and less optimal decision strategy to reduce task complexity. This was accounted for in the heuristic model, in which the decision utility was calculated as a weighted sum of probability and magnitude. Here, we could detect a change in learning either as a changed learning rate or as a change in how much impact the learnt probabilities have on decisions relative to the explicit magnitudes. Again, we did not find any group differences for the two learning rates (reward: $t(45) = 1.3$, $P = 0.20$, loss: $t(45) = 0.7$, $P = 0.5$) or for the probability weighting factor ($t(45) = -0.36$, $P = 0.72$). The groups did not differ either in any other model parameter (Table 2c).

Hybrid model. This model combined the optimal and heuristic model's decision strategy using an additional weighting factor (ω , heuristic weight); ω indexed the degree to which participants used the heuristic relative to the optimal strategy. Again, the two groups did not differ in their learning rates (Table 2d). However, there was a strong difference in the heuristic weight between the groups, with the d-cycloserine group relying more on an optimal rather than heuristic decision strategy compared to the placebo group ($t(45) = 4.6$, $P = 3.6 \times 10^{-5}$).

Model comparison. To investigate this shift in decision strategy further we compared how well each model explained the behavior of each group, using the AIC (Figure 2c). We hypothesized that if participants in the d-cycloserine group used a more optimal decision strategy, models incorporating the optimal decision rule should

better explain their behavior compared to the placebo group.

First, we compared the optimal model relative to the heuristic model. In both groups, the heuristic model provided a better fit to the data than the optimal model. However, this effect was weaker in the d-cycloserine group than in the placebo group, as shown by a significant group difference ($t(45) = 2.1$, $P = 0.046$). This suggests that relative to the placebo, the d-cycloserine group used a more optimal strategy. We also compared the fits of the heuristic to the hybrid model. Similarly, we found that the difference in scores varied significantly between the groups ($t(45) = 3.6$, $P = 8.0 \times 10^{-4}$). For the d-cycloserine group, the hybrid model provided the best fit ($t(19) = 3.9$, $P = 0.01$). However, for the placebo group, both models explained their behavior equally well ($t(26) = 1.6$, $P = 0.11$). These findings are in line with those from the modeling analysis above and suggest that d-cycloserine shifted participants' behavior to rely more on an optimal decision strategy.

To investigate this further, we ran a logistic regression analysis, analogous to the hybrid model (Figure 2b). It included an interaction term for magnitude and probability (analogous to the optimal model), as well as each factor separately (analogous to the heuristic model). Similar to our hybrid model findings, the groups did not differ in the heuristic aspect: there was no difference between the main effects for magnitude and probability (magnitude: $F(1,45) = 2.2$, $P = 0.14$, probability: $F(1,45) = 0.98$, $P = 0.33$). In contrast, the groups differed on the optimal aspect of the model: a two-by-two ANOVA analysis on the interaction regressors of magnitude and probability for reward and loss revealed that those regressors had a stronger impact on the decisions in the d-cycloserine group ($F(1,45) = 10.8$, $P = 0.002$). *Post-hoc* tests revealed that this effect was significant for loss ($t(45) = -2.7$, $P = 0.01$) and was marginally significant for reward ($t(32.4) = -2.0$, $P = 0.056$).

DISCUSSION

Summary of the Results

We studied the effect of changes in NMDA activity on human reward-guided learning and decision-making using d-cycloserine, a partial NMDA agonist. Participants had to learn changing probabilities of positive and negative outcomes over time and integrate them with explicitly cued reward and loss magnitudes in order to make decisions. Although participants were able to track the changing probabilities, this learning was not modulated by d-cycloserine. If there had been an effect on learning, we would have expected either a change in learning rates or alternatively a changed reliance on the learnt compared to the explicit information, neither was the case.

However, d-cycloserine had a significant effect on decision-making. The drug shifted decision-making away from a heuristic approach towards a more optimal one. In the current context, more optimal behavior meant that participants multiplied the learnt probability and the displayed magnitude information rather than adding them (heuristic behavior). Note, however, that such multiplication might be more demanding especially since participants tracked several pieces of information at a time, making

heuristic strategies an appealing way of reducing task complexity.

Absence of a Learning Effect

To our knowledge, there are no studies in humans that have looked at the effects of d-cycloserine on reward-guided learning and decision-making. However, animal studies suggest that d-cycloserine improves fear extinction (Walker *et al*, 2002), drug use extinction (Flavell *et al*, 2011), and appetitive conditioning (Golden and Houpt, 2007). However, our task is most similar to instrumental reward reversal learning task paradigms and it is still unknown whether this neurally distinct form of learning is also improved by d-cycloserine, with Golden and Houpt, 2007 finding that at least simple reversal learning might not be.

Additionally, evidence for learning effects in human studies analogous to the animal studies above has been more inconsistent. Although there is evidence from clinical trials that d-cycloserine may accelerate the therapy of different anxiety disorders (reviewed in Hofmann *et al*, 2013), some experimental studies have not been able to replicate the basic fear extinction effects in humans (Guastella *et al*, 2007; Klumbers *et al*, 2012). Furthermore, clinical trials for alcohol or for cocaine use extinction have failed to find an enhancement by d-cycloserine (Kamboj *et al*, 2011, Price *et al*, 2013). This discrepancy between human and animal studies could be due to the strong dose dependency of d-cycloserine's effects (Walker *et al*, 2002): in general animal studies have used a higher dose (by about a factor of three) and injection rather than oral administration of the drug.

Changes in Evidence Integration for Decision-Making

We found that d-cycloserine led to a very selective shift in participants' decision strategy, making them integrate different sources of information in a more optimal manner, compared to the placebo group. More specifically, participants integrated the learnt probability and the explicitly cued magnitude more multiplicatively (ie, non-linearly) and thus less additively (ie, linearly).

Recent work has suggested functions for NMDA receptors beyond traditional long-term consolidation in learning. To our knowledge, nobody has yet tested whether changes in NMDA-mediated neural processing would affect evidence integration in a complex value-based decision-making task in humans. However, there is some evidence that NMDA receptors can modulate higher cognitive functions, like working memory (Wang *et al*, 2013), attention (Herrero *et al*, 2013; Meuwese *et al*, 2013; Self *et al*, 2012), or delay discounting (Floresco *et al*, 2008). However, it is important to note that it is unlikely that the decision-making effects in our task are a working-memory effect in disguise. Collins and Frank, 2012 have shown that in a standard human learning paradigm (similar to the one used here), learning rate captures a mixture of cortico-striatal learning systems as well as prefrontal working memory. Therefore, if the observed effect were driven by a working-memory mechanism, we would have expected to see a change in at least one of the different measures we used to probe for learning effects. However, we did not see any such effects.

There are at least two theoretical frameworks that are, however, in line with our results. First, NMDA receptors have been proposed as crucial for the integration of information over time during decision-making in a biophysical model (Wang, 2002). Although this model was originally based on animal recordings during perceptual decision-making tasks (reviewed in Wang, 2008), recent human brain imaging studies have shown that it is also relevant for human reward-based decision-making (Chau *et al*, 2014; Hunt *et al*, 2012; Jocham *et al*, 2012). Simulation studies have suggested that modulation of the NMDA receptor activity should enhance or impair decision-making, depending on the baseline activation of NMDA receptors (Standage and Paré, 2011). Following these findings, the decision-making improvement in our study could be explained as shifting the NMDA receptor activation into a more favorable range for decision-making.

Second, there is evidence that NMDA receptors are important for neural computations when more than one input is integrated non-linearly. For example, Binns and Salt, 1996 have shown that NMDA receptors are crucial for multisensory integration. More specifically, they found that the neuronal super-additivity between inputs of different modalities in the cat superior colliculus was strongly reduced by NMDA blockade. Furthermore, Self *et al*, 2012 have shown that NMDA receptors are crucial for the integration of top-down and bottom-up information during figure-ground separation. In the context of our task the optimal way to integrate the probabilities and the magnitudes presented, is non-linearly (ie, multiplicatively), rather than linearly (additively). Therefore d-cycloserine may have led to more optimal integration by biasing processing towards more non-linear neural interactions.

One possible explanation for the occurrence of an effect on decision-making in the absence of a learning effect is that while both are supported by NMDA receptors, there is a different optimal range of NMDA activation supporting both functions (Standage and Paré, 2011). In addition, the strength of the effect of d-cycloserine might vary across brain areas, due to different NMDA receptor subunit compositions or different local levels of glycine (see Yang and Svensson, 2008 for a review). Therefore, the behavioral dissociation could be explained as due to different d-cycloserine sensitivities in reward decision- compared to reward learning-related brain areas.

Future Directions

Future experiments will have to establish whether our results generalize to the integration of dimensions other than probability and magnitude, and to situations where all information to be integrated is explicitly shown rather than learnt.

It would also be interesting to manipulate the predicted choices for linear vs non-linear integration more widely to produce more distinct changes in choices and outcome between the groups, as well as testing whether other forms of non-linear integration or conditional reasoning (model-based) would also be affected.

Lastly, future decision-task designs could explicitly contrast which theoretical NMDA integration mechanism mentioned above drives the behavioral effect: while the

biophysical model explanation would predict decision enhancements in situations where attributes are integrated linearly or non-linearly, the super-additivity mechanism would predict enhancements only where attributes need to be integrated non-linearly.

CONCLUSION

We have found that enhancing the activation of NMDA receptors changed how humans integrated evidence in a multi-attribute decision-making and learning task in the absence of more direct effects on learning. This suggests a role for NMDA receptors beyond their well-studied role in long-term consolidation and plasticity. Although we cannot clearly dissociate between different candidate mechanisms for our observed effect, either biophysical attractor networks or basic super-additivity models could plausibly explain our data.

FUNDING AND DISCLOSURE

This research was funded by the Oxfordshire Health Services Research Committee (research grant awarded to AR, OHSRC 1021) and by the Wellcome Trust (JS and NK were awarded Four-Year PhD Studentships, 092759/Z/10/Z and 089280/Z/09/Z and MFSR holds a Senior Investigator Award, WT100973AIA). CJH has valueless shares in p1vital and serves on their advisory panel. She has received consultancy payments from p1vital, Servier, Eli Lilly, AstraZeneca, and Lundbeck and is a director of Oxford Psychologists. The remaining authors declare no conflict of interest.

ACKNOWLEDGEMENTS

We would like to thank Marco Wittmann for helpful comments on an earlier version of the manuscript and Bolton Chau, Rei Akaishi and David Bannerman for helpful discussions.

REFERENCES

- Bannerman DM, Bus T, Taylor A, Sanderson DJ, Schwarz I, Jensen V et al (2012). Dissecting spatial knowledge from spatial choice by hippocampal NMDA receptor deletion. *Nat Neurosci* 15: 1153–1159.
- Beck AT, Steer RA, Brown GK (1996). *Manual for the Beck Depression Inventory-II*. Psychological Corporation: San Antonio, TX.
- Behrens TEJ, Woolrich MW, Walton ME, Rushworth MFS (2007). Learning the value of information in an uncertain world. *Nat Neurosci* 10: 1214–1221.
- Binns KE, Salt TE (1996). Importance of NMDA receptors for multimodal integration in the deep layers of the cat superior colliculus. *J Neurophysiol* 75: 920–930.
- Bohn I, Gierler C, Hauber W et al (2003). NMDA receptors in the rat orbital prefrontal cortex are involved in guidance of instrumental behaviour under reversal conditions. *Cereb Cortex* 13: 968–976.
- Boorman ED, Behrens TE, Rushworth MF (2011). Counterfactual choice and learning in a neural network centered on human lateral frontopolar cortex. *PLoS Biol* 9: e1001093.
- Burke CJ, Brünger C, Kahnt T, Park SQ, Tobler PN (2013). Neural integration of risk and effort costs by the frontal pole: only upon request. *J Neurosci* 33: 1706–1713a.
- Carver CS, White TL (1994). Behavioral inhibition, behavioral activation, and affective responses to impending reward and punishment: the BIS/BAS Scales. *J Pers Soc Psychol* 67: 319–333.
- Chau BKH, Kolling N, Hunt LT, Walton ME, Rushworth MFS (2014). A neural mechanism underlying failure of optimal choice with multiple alternatives. *Nat Neurosci* 17: 463–470.
- Collins AGE, Frank MJ (2012). How much of reinforcement learning is working memory, not reinforcement learning? A behavioral, computational, and neurogenetic analysis. *Eur J Neurosci* 35: 1024–1035.
- Derryberry D, Reed MA (2002). Anxiety-related attentional biases and their regulation by attentional control. *J Abnorm Psychol* 111: 225–236.
- Eysenck HJ, Eysenck SBG (1994). *Manual for the Eysenck Personality Questionnaire (EPQ-R Adult)*. Educational Industrial Testing Service.
- Flavell CR, Barber DJ, Lee JLC (2011). Behavioural memory reconsolidation of food and fear memories. *Nat Commun* 2: 504.
- Floresco SB, Tse MTL, Ghods-Sharifi S (2008). Dopaminergic and glutamatergic regulation of effort- and delay-based decision making. *Neuropsychopharmacology* 33: 1966–1979.
- Golden GJ, Houpt TA (2007). NMDA receptor in conditioned flavor-taste preference learning: blockade by MK-801 and enhancement by D-cycloserine. *Pharmacol Biochem Behav* 86: 587–596.
- Guastella AJ, Lovibond PF, Dadds MR, Mitchell P, Richardson R (2007). A randomized controlled trial of the effect of D-cycloserine on extinction and fear conditioning in humans. *Behav Res Ther* 45: 663–672.
- Herrero JL, Gieselmann Ma, Sanayei M, Thiele A (2013). Attention-induced variance and noise correlation reduction in macaque V1 is mediated by NMDA receptors. *Neuron* 78: 729–739.
- Hofmann SG, Wu JQ, Boettcher H (2013). D-Cycloserine as an augmentation strategy for cognitive behavioral therapy of anxiety disorders. *Biol Mood Anxiety Disord* 3: 11.
- Hunt LT, Kolling N, Soltani A, Woolrich MW, Rushworth MFS, Behrens TEJ (2012). Mechanisms underlying cortical activity during value-guided choice. *Nat Neurosci* 15: 470–476, S1–3.
- Jocham G, Hunt LT, Near J, Behrens TEJ (2012). A mechanism for value-guided choice based on the excitation-inhibition balance in prefrontal cortex. *Nat Neurosci* 15: 960–961.
- Kamboj SK, Massey-Chase R, Rodney L, Das R, Almahdi B, Curran HV et al (2011). Changes in cue reactivity and attentional bias following experimental cue exposure and response prevention: a laboratory study of the effects of D-cycloserine in heavy drinkers. *Psychopharmacology (Berl)* 217: 25–37.
- Klumpers F, Denys D, Kenemans JL, Grillon C, van der Aart J, Baas JM (2012). Testing the effects of Δ^9 -THC and D-cycloserine on extinction of conditioned fear in humans. *J Psychopharmacol* 26: 471–478.
- Kuriyama K, Honma M, Soshi T, Fujii T, Kim Y (2011). Effect of D-cycloserine and valproic acid on the extinction of reinstated fear-conditioned responses and habituation of fear conditioning in healthy humans: a randomized controlled trial. *Psychopharmacology (Berl)* 218: 589–597.
- Meuwese JDI, van Loon AM, Scholte HS, Lirk PB, Vulink NC, Hollmann MW et al (2013). NMDA receptor antagonist ketamine impairs feature integration in visual perception. *PLoS One* 8: e79326.
- Onur OA, Schlaepfer TE, Kukulja J, Bauer A, Jeung H, Patin A et al (2010). The N-methyl-D-aspartate receptor co-agonist D-cycloserine facilitates declarative learning and hippocampal activity in humans. *Biol Psychiatry* 67: 1205–1211.
- Patel DS, Sharma N, Patel MC, Patel BN, Shrivastav PS, Sanyal M (2011). Development and validation of a selective and sensitive

- LC-MS/MS method for determination of cycloserine in human plasma: application to bioequivalence study. *J Chromatogr B Analyt Technol Biomed Life Sci* **879**: 2265–2273.
- Price KL, Baker NL, McRae-Clark AL, Saladin ME, Desantis SM, Santa Ana EJ et al (2013). A randomized, placebo-controlled laboratory study of the effects of D-cycloserine on craving in cocaine-dependent individuals. *Psychopharmacology (Berl)* **226**: 739–746.
- Self MW, Kooijmans RN, Supèr H, Lamme Va, Roelfsema PR (2012). Different glutamate receptors convey feedforward and recurrent processing in macaque V1. *Proc Natl Acad Sci USA* **109**: 11031–11036.
- Spielberger CD, Gorsuch RL, Lushene PR, Vagg PR, Jacobs AG (1983). *Manual for the State-Trait Anxiety Inventory (Form Y). Man State-Trait Anxiety Invent STAI* 4–6.
- Standage D, Paré M (2011). Persistent storage capability impairs decision making in a biophysical network model. *Neural Netw* **24**: 1062–1073.
- Stein BE, Stanford TR (2008). Multisensory integration: current issues from the perspective of the single neuron. *Nat Rev Neurosci* **9**: 255–266.
- van Berckel BN, Lipsch C, Gispen-de Wied C, Wynne HJ, Blankenstein MA, van Ree JM et al (1998). The partial NMDA agonist D-cycloserine stimulates LH secretion in healthy volunteers. *Psychopharmacology (Berl)* **138**: 190–197.
- van Berckel BN, Lipsch C, Timp S, Gispen-de Wied C, Wynne H, van Ree JM et al (1997). Behavioral and neuroendocrine effects of the partial NMDA agonist D-cycloserine in healthy subjects. *Neuropsychopharmacology* **16**: 317–324.
- Walker DL, Ressler KJ, Lu K-T, Davis M (2002). Facilitation of conditioned fear extinction by systemic administration or intra-amygdala infusions of D-cycloserine as assessed with fear-potentiated startle in rats. *J Neurosci* **22**: 2343–2351.
- Wang M, Yang Y, Wang C-J, Gamo NJ, Jin LE, Mazer JA et al (2013). NMDA receptors subserve persistent neuronal firing during working memory in dorsolateral prefrontal cortex. *Neuron* **77**: 736–749.
- Wang X-J (2002). Probabilistic decision making by slow reverberation in cortical circuits. *Neuron* **36**: 955–968.
- Wang X-J (2008). Decision making in recurrent neuronal circuits. *Neuron* **60**: 215–234.
- Yang CR, Svensson KA (2008). Allosteric modulation of NMDA receptor via elevation of brain glycine and D-serine: the therapeutic potentials for schizophrenia. *Pharmacol Ther* **120**: 317–332.

Supplementary Information accompanies the paper on the Neuropsychopharmacology website (<http://www.nature.com/npp>)