Contextual influence of reinforcement learning performance of depression: evidence for a negativity bias?

Henri Vandendriessche, Amel Demmou, Sophie Bavard, Julien Yadak, Cédric Lemogne, Thomas Mauras, Stefano Palminteri

Contextual influence of reinforcement learning performance in depression: evidence for a negativity bias?

Authors

Henri Vandendriessche*[1,2], Amel Demmou*[3], Sophie Bavard[1,2,4], Julien Yadak[3], Cédric Lemogne[5,6], Thomas Mauras[7], Stefano Palminteri[1,2]
*co-first author

1- Laboratoire de Neurosciences Cognitives Computationnelles, Institut National de la Santé et Recherche Médicale, Paris, France
2- Département d'Etudes Cognitives, Ecole Normale Supérieure, PSL Research University, Paris, France
3- Hôpital Cochin Port Royal, Paris, France
4- Department of Psychology, University of Hamburg, Von-Melle-Park 11, 20146 Hamburg, Germany
5- Université de Paris, INSERM U1266, Institute de Psychiatrie et Neurosciences de Paris, F-75014 Paris, France
6- Service de Psychiatrie de l'adulte, AP-HP, Hôpital Hôtel-Dieu, F-75004 Paris, France
7- Groupe Hospitalier Universitaire, GHU paris psychiatrie neurosciences, 1 rue Cabanis 75014 Paris.

Corresponding authors

Stefano Palminteri, Laboratoire de Neurosciences Cognitives et Computationelles 29 rue d'ULM 75005 Paris, stefano.palminteri@ens.fr
Henri Vandendriessche, Laboratoire de Neurosciences Cognitives et Computationelles 29 rue d'ULM 75005 Paris, henri.vandendriessche@ens.fr

Key Words

Depression, reward processing, reinforcement learning, context dependency, negativity bias

Abstract (250 words)
Main text (5517 words)
Figures (4)
Tables (2)

Abstract

Backgrounds:

Value-based decision-making impairment in depression is a complex phenomenon: while some studies did find evidence of blunted reward learning and reward-related signals in the brain, others indicate no effect. Here we test whether such reward sensitivity deficits are dependent on the overall value of the decision problem.

Methods

We used a two-armed bandit task with two different contexts: one ‘rich’, one ‘poor’ where both options were associated with an overall positive, negative expected value, respectively. We tested patients (N=30) undergoing a major depressive episode and age, gender and socio-economically matched controls (N=26). Learning performance followed by a transfer phase, without feedback, were analysed to distangle between a decision or a value-update process mechanism. Finally, we used computational model simulation and fitting to link behavioural patterns to learning biases.

Results

Control subjects showed similar learning performance in the ‘rich’ and the ‘poor’ contexts, while patients displayed reduced learning in the ‘poor’ context. Analysis of the transfer phase showed that the context-dependent impairment in patients generalized, suggesting that the effect of depression has to be traced to the outcome encoding. Computational model-based results showed that patients displayed higher learning rate for negative compared to positive outcomes (the opposite was true in controls).

Conclusions

Our results illustrate that reinforcement learning performances in depression depend on the value of the context. We show that depressive patients have a specific trouble in contexts with an overall negative state value, which in our task is consistent with a negativity bias at the learning rates level.

Main Text

Introduction:

Depression is a common debilitating disease that is a worldwide leading cause of morbidity and mortality. According to the latest estimates from World Health Organization, in 2015 more than 300 million people are now living with depression (World Health Organization, 2017)and anhedonia are core symptoms of major depressive disorder. Those two symptoms are key criteria to the diagnostic of Major Depressive Disorder (MDD) in the Diagnostic and Statistical Manual of Mental Disorders (DSM-5) (American Psychiatric Association, 2013). Anhedonia is broadly defined as a decreased ability to experience pleasure from positive stimuli. Specifically, it is described as a reduced motivation to engage in daily life activities (motivational anhedonia) and reduced enjoyment of usually enjoyable activities (consummator anhedonia).

Depression is a complex and heterogeneous disorder implying instinctual, emotional and cognitive dysfunctions. Although its underlying mechanisms remain unclear, it has been proposed - based on the importance of anhedonia and low mood in depression - that reduced reward processing, both in terms of incentive motivation and reinforcement learning, plays a key role in the clinical manifestation of depression (Admon & Pizzagalli, 2015; Chen et al., 2015; Eshel & Roiser, 2010; Q. J. Huys et al., 2013; Safra et al., 2019; Whitton et al., 2016). This hypothesis implies that subjects with depression should display reduced reward sensitivity both at the behavioral and neural levels in value-based learning. On the long term, a better understanding of these processes could help for the prevention and management of depression.

Following up on this assumption, numerous studies have tried to identify and characterize such reinforcement learning deficits, however the results have been mixed so far. Indeed, while some studies did find evidence of blunted reward learning and reward-related signals in the brain, others indicate limited or no effect (Brolsma et al., 2020; Chung et al., 2017; Hägele et al., 2015; Rothkirch et al., 2017; Rutledge et al., 2017; Shah et al., 1999). Outside the learning domain, other recent studies showed no disrupted valuation during decision-making under risk (Chung et al., 2017; Moutoussis et al., 2018). It is also worth noting that many of previous studies identifying value-related deficits in depression, only included one valence domain (i.e., only rewards or only punishments) and did not directly contrast between rewards and punishments nor separate the two valence domains in different experimental sessions (Admon & Pizzagalli, 2015; Elliott et al., 1996, 1997; Forbes & Dahl, 2012; Gradin et al., 2011; Kumar et al., 2008; Pizzagalli, 2014; Vrieze et al., 2013; Zhang et al., 2013). A recent study (Pike & Robinson, 2022), where reward and punishment sensitivity has been computationally quantified by assuming different learning rate parameters for positive or negative outcomes show that, compared to controls, patients; contrary to what is generally found in healthy subjects (Chambon et al., 2020; Palminteri et al., 2017) are generally better explained assuming blunted reward compared to punishment learning.

Here we speculate that the lack of concordant results may be in part explained by the fact that reinforcement learning impairment in depression is dependent on the overall value of the learning context. In fact, computational studies clearly illustrate that the behavioural consequences of blunted reward and punishment sensitivity depend on the underlying distribution of outcome. More specifically, Cazé and Van Der Meer (Cazé & van der Meer, 2013) showed that greater sensitivity to reward compared to punishment (positivity bias; as proxied by different learning rates; Pike & Robinson,

2022) advantages learning in contexts with poor overall reward expectation (i.e., ‘poor’ contexts) compared those with high overall reward expectation (‘rich’ contexts). Conversely, greater sensitivity to punishment compared to reward (negativity bias) should advantage learning in ‘rich’ context. As a consequence, if depressive patients present blunted reward compared to punishment sensitivity (i.e., a negativity bias) this should induce a difference in performance, specifically in ‘poor’ contexts, where displaying a positivity bias is optimal.

To test the hypothesis , we adapted a standard protocol composed by a learning and a post-learning transfer phase. The learning phase included two different contexts: one defined as ‘rich’ (in which the two options have an overall positive expected value) and the other as ‘poor’ (two options with an overall negative expected value). In contrast with the learning phase, there was no feedback in the transfer phase, in order to probe the subjective values of the options without modifying it (Bavard et al., 2018; Frank et al., 2004; Palminteri et al., 2015). In similar tasks, healthy subjects are generally reported to be able to learn equally from rewards and punishments (Palminteri et al., 2015; Pessiglione et al., 2006). However, based on the idea that depression blunts reward sensitivity and that a positivity bias is advantageous in the ‘poor’ contexts, we expected a learning asymmetry in MDD patients. More precisely, learning rate differences should induce lower performance in the ‘poor’ context in MDD patients.

In addition to choice data, we also analysed reaction times and outcome observation times as ancillary measures of attention and performance. Previous findings suggest that negative value contexts are associated with overall slower responses (Fontanesi, Gluth, et al., 2019; Fontanesi, Palminteri, et al., 2019). However, previous studies did not find any specific reaction time signatures in patients (Brolsma et al., 2021; Chase et al., 2010; Douglas et al., 2009; Knutson et al., 2008)

Methods

Participants and inclusion criteria

Fifty-six subjects were recruited in a clinical center (the Ginette Amado psychiatric crisis center) in Paris between May 2016 and July 2017. Inclusion criteria were a diagnosis of major unipolar depression diagnosed by a psychiatrist and an age between 18 and 65 years old (see **Table 1**). A clear, oral and written explanation was also delivered to all participants. All procedures contributing to this work comply with the ethical standards of the relevant national and institutional committees on human experimentation and with the Helsinki Declaration of 1975, as revised in 2008. In total, we tested N=30 patients undergoing a Major Depressive Episode (MDE) and N=26 age-, gender- and socioeconomically-matched controls. For patients, exclusion criteria were the presence of psychotic symptoms or a diagnosis of chronic psychosis, severe personality disorder, neurological or any somatic disease that might cause cognitive alterations, neuroleptic treatment, electro-convulsive therapy in the past 12 months and current substance use. Psychiatric co-morbidities were established by a clinician with a semi-structured interview based on the Mini International Neuropsychiatric Interview (MINI) (Sheehan et al., 1998). Our final sample, some patients (n=13) presented anxiety-related disorders. Among them, some (n=6) presented isolated anxiety-related disorders (social anxiety n=2; panic disorder n=2; agoraphobia n=1 ; claustrophobia n=1) and the rest of the group (n=7) presented several associated anxiety-related disorders (agoraphobia n= 4; panic disorder n=4; social anxiety n=3; generalized anxiety n=3; OCD n=1; PTSD n=1). Others (n=8) presented substance abuse disorder (cannabis n=3; alcohol n=4; cocaine n=2). All patients were undertaking medication (see Table 2 for details). Participants included in the healthy volunteer group had no past or present psychiatric diagnosis and were not taking any psychoactive treatment

Behavioral testing

Patients volunteering to take part in the experiment were welcomed in a calm office away from the center's activity where they were given information about the aim and the procedure of the study. The study was verbally described as an evaluation of cognitive functions through a computer « game ». The diagnostic of MDE and the presence of psychiatric co-morbidities were assessed with the MINI screener completed in a semi-structured interview with a psychiatrist by the MINI. The subjects were then asked to complete several questionnaires assessing their level of optimism (Life Orientation Test-Revised (LOT-R), an optimism analogue scale (created for this study to contrast usual and current level of optimism) and the severity of depression (Beck Depression Inventory – II) (Beck et al., 1996). The participants were told they were going to play a simple computer game, whose goal was to earn as many points as possible. Written instructions were provided and verbally reformulated if necessary. There was no monetary compensation as patients did the task alongside a psychiatric assessment. To match patients' conditions, controls did not receive any compensation either.

As in previous studies of reinforcement learning the behavioral protocol was divided into a learning phase and a transfer phase (Chase et al., 2010; Frank et al., 2004; Palminteri & Pessiglione, 2013) (**Figure 1A**). Options were materialized by abstract symbols (agathodaimon font). Symbols appeared in pairs of abstract symbols displayed on a black screen., During the learning phase, options were presented in fixed pairs, while during the transfer phase they were presented in all possible combinations (**Figure 1B**). Before the subjects were told that one of the two options was more advantageous than the other and encouraged to identify it to maximize their (fictive) reward. The

reward probability attached to each symbol was never explicitly given and the subjects had to learn it through trial and error. Each symbol was associated to a fixed reward probability. Reward probabilities were inspired by previous empirical and theoretical studies (Cazé & van der Meer, 2013; Chambon et al., 2020; Palminteri & Pessiglione, 2017) and distributed across symbols as follows: 10% / 40% (‘poor’ context), 60% / 90% (“rich context”). The reward probabilities were decided in order to have the same choice difficulty (as indexed by the difference in expected value between the two options) across choice contexts. The learning phase was further divided in two sessions of 100 trials each (each involving both the ‘rich’ and the ‘poor’ context repeated 50 times).

In the transfer phase the 8 different symbols were presented by pairs in all binary combinations four times (including pairing that had never been displayed together in the previous phase; 112 trials). The subjects had to choose which symbol deemed the more rewarding, however, in the transfer phase, no feedback was provided in order not to interfere with subjects’ final estimates of option values (Chase et al., 2010; Frank et al., 2004; Palminteri & Pessiglione, 2017).The subjects were told to use instinct when doubting. The aim of the transfer phase was to assess the participants’ learning process on a longer time scale than the learning phase, which is supposed to mainly rely on working memory (Collins & Frank, 2012) capacity to remember and extrapolate the symbols’ subjective values out of their initial context (generalization).

When the symbols appeared on the screen, the subject had to choose between the two symbols by pushing a right or a left key on a keyboard. In rewarded/punished trials a green/red smiley/sad face and “+1pts” / “-1pts” appeared on screen. In order to be sure that the subjects paid attention to the feedback, they had to push the up key after a win and the down key after a loss to move to the next trial (**Figure 1C**; top). Trials in the transfer phase different in that the feedback was not displayed (**Figure 1C**; bottom).

Dependent variables

The main behavioral variables of our study are the correct choice rates, as measured in the learning and the transfer phase. A choice is defined ‘correct’ (coded as ‘1’) if the participant picks the reward maximizing option, incorrect (coded as ‘0’) otherwise. In the learning phase, the correct choice is, therefore picking ‘A’ in the ‘rich’ context and ‘B’ in the ‘poor’ contexts (**Figure 1B**). For display purposes, the learning curves were smoothed (five trials sliding average) (**Figure 2A**). In the transfer phase, the correct choice was defined in a trial-by-trial basis and depended on the particular presented combination (note that in some trials, a correct choice could not be defined, as the comparison involved two symbols with the same value, originally presented in different sessions). (**Figure 1B**). For display purpose, concerning the transfer phase, we also considered the choice rate, defined as how many time an options have been chosen, divided by the numbers of time an option have been presented (calculated across all possible combinations except the similar option ones) (**Figure 2B**). As ancillary exploratory dependent variables we also looked at two different measures of response times. More precisely, we extracted the reaction times (i.e., the time spent between symbols’ onset and choice; **Figure 3A**) and the outcome observation time (i.e., the time spent between reward onset and key press to next trial; **Figure 3B**). For display purposes, also response time curves were smoothed (five trials sliding average).

Statistical analyses

The dependent variables were analyzed using Generalized Linear Mixed Models (GLMM) as implemented by the function glmer of the software R (R version 3.6.3 (2020-02-29) R Core Team (2022)) and the package lme4 (lme4 version: 1.1-27.1; Bates et al., 2015). The GLMMs of correct choice rates (both in the learning and the transfer test) used a binomial linking function, while those of response times (both reaction times and outcome observation time) used a gamma linking function (Yu et al., 2022). All GLMMs were similarly constructed and included 'subject' number as a random effect and 'group' (between-subject variable: controls versus patients), 'context' (within-subject variable) and interaction between the two as fixed-effects. For dependent variables extracted from the learning phase the 'context' within subject variable corresponded to whether the measure was taken from the 'rich' or the 'poor' context. In the GLMM of the correct choice rate in the transfer test the variable 'condition' took three levels that corresponded to whether or not the choice under consideration involved the best possible option in the 'rich' condition ('A present)'; whether or not the choice under consideration involved the worst possible option in the 'poor' condition ('D present') and all the other trials ('other') (see **Figure 1B**). Post hoc comparisons were assessed by comparing the marginal means of the contrast of interest to zero. All p-values are reported after Tukey's correction for multiple comparisons.

Model fitting and model simulations

To link the behavioral performance in our task to computational processes, we performed some simulations. More specifically, to assess the behavioral consequences of learning rate biases, we simulated a variant of a standard cognitive model of reinforcement learning literature. The model assumes that subjective option values (Q-values) are learnt from reward prediction errors (RPE) that quantify the difference between expected and obtained outcome (Sutton & Barto, 2018). In this model Q-values are calculated for each combination of states (*s;* in our task the four contexts; **Figure 1B**) and actions (*a*; in our task the symbols). Most of those models assume that subjective options values are updated following a Rescorla-Wagner rule (Rescorla & Wagner, 1972). However, to assess the behavioral consequences of a positivity and negativity bias, based on previous studies (Chambon et al., 2020; Frank et al., 2007; Niv et al., 2012), we modified the standard model by including different learning rates for positive and negative prediction errors (that in our design are respondent to positive and negative outcomes):

$$Q(s,a) \Leftarrow Q(s,a) + \begin{cases} \alpha_+ \times \left(r - Q(s,a)\right), if r > 0 \\ \alpha_- \times \left(r - Q(s,a)\right), if r < 0 \end{cases}$$

The model decision rule was implemented as a softmax function, that calculate the probability of choosing a given option as a function of the difference between the Q-values of the two options, as follows:

$$P_t(s,a) = \frac{1}{1 + e^{\left(\frac{Q_t(s,b) - Q_t(s,a)}{\beta}\right)}}$$

To assess the effect of the positivity and negativity bias on learning performance of our task we ran extensive model simulation where artificial agents played our learning task (i.e., a ‘rich’ and a ‘poor’ context, for 50 trials each). More specifically, we simulated two different sets of learning rates (10000 virtual agents each). One set represented agent with a positivity bias (i.e., $\alpha_+ > \alpha_-$) the other set agents with a negativity bias ($\alpha_+ < \alpha_-$). The value of the parameters (learning rates and temperatures) were randomly drown from uniform distributions; the temperature was drawn from $\in U(0,1)$ the learning rates (for example in the positivity bias case) were drawn from $\alpha_+ \in U(0,1)$ and $\alpha_- \in U(0, \alpha_+)$.

After running the simulations, we also fitted the on the empirical data. More specifically, we focused on fitting the transfer phase choices, because it allows to estimate learning rates involved in long term learning, which are not contaminated by working memory or choice perseveration biases (Collins & Frank, 2012; Frank et al., 2007; Katahira et al., 2017). The model free parameters (temperature and learning rates) were fitted at the individual level using the fmincon function (Optimization Toolbox R2021b**.** MATLAB. (2021). 9.11.0.1809720 (R2021b). 2021B, Natick, Massachusetts: The MathWorks, Inc.) via log model evidence maximization as previously described (Daw et al., 2011; Wilson & Collins, 2019).

Results:

Demographics.

Patients and controls were matched in age ($t(51)=-1.1$, $p=0.28$), gender ($t(53)=1.15$, $p=0.29$) and years of education ($t(54)=-1.59$, $p=0.12$). Concerning the optimism measures, patients with depression were found to be less optimistic in all scales (LOT-R: $t(47)=-7.42$, $p=1.76e-09$; usual optimism: $t(51)=-2.29$, $p=0.03$; current optimism: $t(50)=-10.34$, $p=4.19e-14$). Furthermore, the comparison between usual vs. current optimism in patients and controls, revealed that only patients were significantly less optimistic than usual at the moment of the test (patients: $t(29)=8.26$, $p=4.21e-09$; controls $t(25)=-1.53$, $p=0.14$), consistent with the fact that they were undergoing an MDE. All patients were taking at least one psychotropic medication at the moment of test. Their average BDI was: 29.37 and they had, on average, 1.8 previous MDE in the past.

Learning phase results

Global inspection of the learning curves (**Figure 2A**) suggests that, overall, participants were able to learn to respond correctly. Indeed, all the learning curves are above chance whatever the group or the context. A more detailed inspection reveals that controls' learning curves were unaffected by the choice context ('rich' vs. 'poor'), while patients' learning curves were different depending on the choice context (with a lower correct response rate in the 'poor' context).

Correct response rate (as proxied by the intercept of our GLMM) in the learning phase (**Figure 2A**) indicated that overall performance is significantly above chance ($\chi2(1,56)=16.17$, $p<0.001$) which reflects the fact that accuracy was, in average, well above chance level (0.5). There was no significant effect of context ($\chi2(1,56)=0.046$, $p=0.83$) and no main effect of group ($\chi2(1,56)=2.86$, $p=0.091$) meaning that there were no overall significant differences between the patients and controls and between the 'rich' and 'poor' contexts. However, there was a significant interaction between context and group ($\chi2(1,56)=5.88$, $p=0.015$). Concerning the interaction context and group, post hoc tests indicated that it was driven by an effect of context present in patients (slope=-0.72 , SE=0.24 , $p<0.0027$), but not in controls (slope=-0.063 , SE=0.29 , $p=0.83$).

These results therefore show a specific impact of the context on the two groups. Patients displayed higher accuracy in the 'rich' compared to the 'poor' contexts, while controls were affected this factor as expected from previous articles in the literature (Palminteri et al., 2015; Pessiglione et al., 2006).

Critically, learning phase results cannot establish whether the performance asymmetry observed in patients stems from the learning (i.e., how values are updated) or a decision effect (i.e., how options are selected) processes. To tease apart these interpretations we turned to the analysis of the transfer phase performance.

Transfer phase analysis

The visual inspection of the option-by-option choice rate in the transfer phase, showed that subjects were able to retrieve the values of the options and express meaning preferences among them (**Figure 2B**). In fact, in all groups, the options 'A' (overall highest value) were chosen much more frequently

compared to options ‘D’ (overall lowest value) in both groups. Intermediate value options (‘B’ and ‘C’) scored in between the extreme one (with a pattern reminiscent of relative value encoding; Klein et al., 2017; Palminteri & Lebreton, 2021).

Before assessing whether the learning asymmetry observed in patients in the learning phase replicated in the transfer phase, one has to keep in mind that there were not only two fixed choices contexts in the transfer phase, but rather options were presented in all possible combinations. Accordingly, the context factor used for the transfer test contained three levels, defined by the presence of particular options: 1) trials involving the ‘A’ options (and not ‘D’); 2) trials involving the ‘D’ options (and not ‘A’); 3) other trials. Also in the transfer test, average correct response rate (as proxied by the intercept of our GLMM) shows that overall performance was significantly above chance ($\chi2(1,56)=15.9$, $p<0.001$). We also found a significant effect of group ($\chi2(1,56)=6.83$, $p=0.009$), no effect of context ($\chi2(1,56)=2.23$, $p=0.327$) and a very strong and significant group by context interaction ($\chi2(1,56)=53.21$, $p<0.001$). Post-hoc tests reveal that controls were equally able to make the correct decision in contexts involving seeking ‘A’ or those involving avoiding ‘D’ (slope=-0.004 , SE=0.1 , p=0.999) whereas patients were strikingly better at seeking ‘A’ than avoiding ‘D’ (slope=1.06 , SE=0.1 , p<0.001).

These results are consistent with the learning phase results. The context-specific asymmetry in patients that we found in the learning phase was also present in the transfer phase where all the different options were extracted from their initial context and paired with other options. It allows us to conclude that the performance asymmetry can be traced back to the learning asymmetry, where negative outcomes (more frequent following the worst possible option ‘D’) seem to exert a smaller effect on patients’ learning performances than positive ones (more frequent following the best possible option ‘A’) (Frank et al., 2004).

Modelling results

Model simulations indicate that learning biases affect performance in a context-dependent manner (**Figure 3A**). More specifically in our task, a positivity bias ($\alpha_+ > \alpha_-$) is associated to similar accuracy in the ‘rich’ and ‘poor’ contexts, while a negativity bias ($\alpha_+ < \alpha_-$) is associated much higher accuracy in the ‘rich’ compared to the ‘poor’ context. The reason for this result can be traced down to the idea that it is rational to preferentially learn from rate outcomes (Cazé & van der Meer, 2013). Comparing model simulations to our data, we note that the ‘positivity bias’ pattern is closer to the result obtained from the healthy control participants, while the ‘negativity bias’ pattern is closer to the result obtained from the patients. Intriguingly, the ‘positivity bias’ behavioral pattern closely resembles that observed in healthy participants, while the ‘negativity bias’ pattern closely reminds the one observed in patients, thus indicating that patients are better explained by a computational negativity bias.

To formally substantiate this intuition, we submitted the learning rates fitted from transfer phase choices to a 2x2 ANOVA, with group (patients vs. controls) and valence (positive or negative learning rate), as between- and within-subject variables, respectively (**Figure 3B**). The results showed a main effect of patient group ($F(1,107)=5.26$, $p=0.024$; Eta2 (partial)=0.05, 95% CI [3.37e-03, 1.00]), no main effect of valence other ($F(1,107)=3.27e-03$, $p=0.954$; Eta2 (partial)=3.06e-05, 95% CI [0.00, 1.00]), and, crucially, a significant valence-by-group interaction ($F(1,107)=7.58$, $p=0.007$; Eta2 (partial)=0.07, 95% CI [0.01, 1.00]). Finally, we detected no significant different in the choice temperature ($t(48)=1.64$, $p=0.11$).

Response time analysis

As an exploratory analysis, to assess how learning performance reflected into response times (both at the decision and the learning phase), we looked at reaction and outcome observation times during the learning phase. Reaction time (defined as the difference between stimuli onset and button pressing to make a decision) showed a main effect of the context ($\chi2(1,56)=9.83$, $p=0.002$), with reaction tiles being higher in the 'poor' compared to the 'rich' condition, which is consistent with previous studies showing valence induced slowing in reinforcement learning (Fontanesi, Palminteri, et al., 2019; **Figure 4A**). Reaction times showed is no significant main effect of the group ($\chi2(1,56)=0.03$, $p=0.86$) nor interaction between context and group ($\chi2(1,56)=0.12$, $p=0.73$). Post hoc tests showed that the effect of context was significant in both controls (slope=0.047, SE=0.016 , $p<0.003$) and patients (slope=-0.043 SE=0.0067 , $p<0.001$).

Outcome observation time (defined as the difference between the outcome onset and button pressing to move to the next trial) also displayed no significant effect of the context ($\chi2(1,56)=10.39$, $p<0.123$) but no effect of the group ($\chi2(1,56)=2.17$, $p=0.14$) nor interaction ($\chi2(1,56)=0.39$, $p=0.53$) (**Figure 4B**).

Taken together reaction and outcome observation time analysis, suggest that learning performance asymmetry in patients could not be accounted for by reduced engagement and outcome processing during the learning task.

Discussion:

Summary

In the present study, we assessed reinforcement learning with a behavioral paradigm involving two different reward contexts - one 'rich' with a positive overall expected value and one 'poor' with a negative overall expected value - in patients undergoing a major depressive episode and age-, gender- and education-matched healthy volunteers.

We used reinforcement learning task featuring two different learning contexts: one with an overall positive expected value ('rich' context) and one with a overall negative expected value ('poor' context). Coherent with previous studies, healthy subjects learned equally well in both contexts (Palminteri & Pessiglione, 2017). On the other hand, patients with depression displayed reduced correct response rate in the 'poor' context. This context-dependent learning asymmetry found in the learning phase was confirmed in the analysis of the transfer phase, where subjects were asked to retrieve and generalize the values learned during the learning sessions.

In standard reinforcement learning tasks, a participant has to learn the value of the options select among them. A deficit in reinforcement learning can therefore arise from two possible causes. On one hand, it can be caused by a learning impairment, i.e., failing to accurately update the value of the stimulus. On the other hand, it can be the result of a decision impairment. In this scenario, a participant could still end up selecting the wrong stimuli even though the learning process in itself is intact. Our design, coupling a learning phase with feedback and a transfer phase, where we shuffled all options without any feedback, allows us to separate these two possible sources of error. Indeed, a decision-related problem would lead to a specific impairment during the learning phase but in the transfer phase, there should be none or only an unspecific impairment. On the other side, an valence-specific update-related deficit would originate in the learning phase (when feedback is provided) and would therefore propagate in the transfer phase and be associated only to the concerned specific options (Frank et al., 2007).

Our results are consistent with this second scenario, as we showed that patients were less able to identify the correct response of the 'poor' context both in the learning and the transfer phase. Hence, this suggests that the asymmetrical performance observed in patients, stems from the learning process *per se* and not from the decision process. Therefore, we suppose that this asymmetric learning pattern is the consequence of a more complex mechanism, embedded in the learning process and triggered by affectively negative situations or less frequent affectively positive situations ('poor' context).

Our results suggest that learning performances in depression are dependent on the valence of the context. More specifically, patients undergoing a major depressive episode seem to perform worst at learning in negative value context, compared to positive one. This was true despite the fact that the two contexts are matched in difficulty. Accordingly, control participants on the contrary show no difference in performance between the two contexts. *Prima facie*, this observation challenges some formulations of the negative bias hypothesis described in the literature. Some studies describe negative affective biases in several cognitive processes, such as emotion, memory and perception, as an increased and aberrant saliency of negative affective stimuli (for review see Gotlib & Joormann, 2010; Joormann & Quinn, 2014). From this view, one could extrapolate that, contrary to what we observed in our data, MDD patients should display, if anything, higher performance in the 'poor' contexts. This prediction

contrasts with a computational definition of negativity bias, as a difference between learning rates for positive and negative outcomes (or reward prediction errors). In fact, model simulations studies clearly show that learning positivity or negativity biases affect performance in a context-dependent manner that in our case is consistent with the idea of a negativity bias in depression (Bavard & Théro, 2018; Cazé & van der Meer, 2013). The results was confirmed by model simulations and analysis of learning rates that were fitted from transfer phase choices and, even if it is hard to find in the literature a systematic pattern, it is consistent with a recent computational meta analyses by Pike and co (Beck, 1967; Brolsma et al., 2020; Chase et al., 2010; Eshel & Roiser, 2010; Gradin et al., 2011; Henriques et al, 1994; Q. J. Huys et al., 2013; Knutson et al., 2008; Kumar et al., 2008; Murphy et al., 2003; Pike & Robinson, 2022; Pizzagalli et al., 2005; Steele et al., 2007; Ubl et al., 2015; Whitton et al., 2016). Crucially, consistent with our simulations, The overall good performance of patients and more specifically in the 'rich' context indicated that patients displayed not generic impairments. Overall good performance of patients in some control conditions is actually not uncommon and can be explained by the fact that patients in general are more focused and more involved than controls in this type of study (the so-called Hawthorne effect), because the result of this experiment is much more "meaningful" for them than it is for controls (Frank et al., 2004).

In addition to choice data, in our studies we collected two different response time measures. The first one, reaction time, was classically defined as the time between the stimuli onset the choice button press. Reaction times were not different between our groups of participants, indicating that in our experiment we were not able to provide support for the idea of a generalized sensorimotor slowing in patients (Byrne, 1976). On the other hand, reaction times were strongly affected by the experimental condition, being significantly slower in the 'poor' context in both groups. This finding is at apparent odds with the fact that objective difficulty (as quantified by the difference in value between the two options) was matched across contexts (note that this effect was also present in healthy controls, who displayed equal performance in both conditions). However, slower reaction times in the 'poor' context are consistent with recent findings (Fontanesi, Palminteri, et al., 2019). Indeed, previous studies coupling behavioral decision diffusion model analyses with reinforcement learning paradigms indicate that reaction times are tend to be slower in negative valence contexts, compared to positive valence ones. This effect is well captured by a combination of increased non-decision time (a possible manifestation of Pavlovian-to-instrumental transfer; (Guitart-Masip et al., 2012) and increased cautiousness (a possible manifestation of loss attention; Yechiam & Hochman, 2014). we also recorded the outcome observation times, that quantify the time separating the onset of the outcome from the button press necessary to move to the subsequent trial. Overall, outcome observation times were not significantly modulated by our factors, therefore indicating that the learning asymmetry observed in patients could not be explained by not processing outcome information.

Our study of course suffers from few important limitations. One limitation is the relatively small sample size, which is of course due to the fact that our study was monocentric and went for a relatively short time period. We note, however, that several meaningful insights concerning impairment of reinforcement learning in psychiatric diseases has been obtained until very recently from studies with sample size comparable to our (Chase et al., 2010; Frank et al., 2004; Henriques & Davidson, 2000; Q. J. M. Huys et al., 2016; Moutoussis et al., 2018; Murphy et al., 2003; Rothkirch et al., 2017; Rupprechter et al., 2018). Future, multi-centric, studies will be required to overcome this issue and probe the replicability and generalizability of our findings. Furthermore, by openly sharing our data, our study may contribute to (computational) meta-analysis (Pike & Robinson, 2022).Another limitation of our study is that patients were medicated at the time of the experiment. Even though studies have found effects on performance on medicated and unmedicated patients (Douglas et al., 2009; Steele et

al., 2007), it is always difficult to control for this effect, especially when certain patients take medications for other comorbidities. Additionally, the role of serotonin in reward and punishment learning is far from being understood (Palminteri & Pessiglione, 2013). In some tasks, it has been shown to improve performance in a valence-independent manner, making unlikely that the observed effect was a consequence of medication (Palminteri et al., 2012). So, under the theory that serotonin drives punishment avoidance learning, we would observe the opposite effect. Finally, as MDD is a polysemic condition, and even though we tried to monitor and control the inclusion of patients to avoid interference with other mental conditions, some patients had other symptoms, especially addictive disorders, that should be considered in future studies.

In the literature, is has been repeatedly shown that controls perform equally when they have to choose a reward or avoid a punishment. It is also frequent that patients with mental or neurological disorders other than MDD show an imbalance behavior when implicated in a task with a reward selection and a punishment avoidance (Frank et al., 2004). Studying several aspects of reward processing that correspond to different neurobiological circuits and exploring dysregulation across different psychiatric disorders could be a very efficient way to unfold abnormalities in reward-related decision making. It could be interesting to apply our task to other psychiatric disorders in order to identify neurobiological signatures and develop more targeted and promising treatments (Brolsma et al., 2020; Insel et al., 2010; Whitton et al., 2015).

Data availability

Data collected for this paper, a R script presenting the main figures of the paper as well as some Matlab simulation files are available here https://github.com/hrl-team/Data_depression

Acknowledgments

We thank Magdalena Soukupova for her bright insights on statistical analysis. H.V is supported by the Institut de Recherche en Santé Publique (IRESP, grant number : 20II171-00). SP is supported by the Institut de Recherche en Santé Publique (IRESP, grant number : 20II138-00), and the Agence National de la Recherche (CogFinAgent: ANR-21-CE23-0002-02; RELATIVE: ANR-21-CE37-0008-01; RANGE : ANR-21-CE28-0024-01). The Departement d'études cognitives is funded by the Agence National de la Recherche (FrontCog ANR-17-EURE-0017). The funding agencies did not influence the content of the manuscript.

Disclosures

Dr. Lemogne reports personal fees and non-financial support from Boehringer Ingelheim, Janssen-Cilag, Lundbeck, Otsuka Pharmaceutical, outside the submitted work. The other authors declare not competing conflict of interest concerning the related work.

References

Admon, R., & Pizzagalli, D. A. (2015). Dysfunctional reward processing in depression. *Current Opinion in Psychology*, *4*, 114–118. https://doi.org/10.1016/j.copsyc.2014.12.011

American Psychiatric Association. (2013). *Diagnostic and Statistical Manual of Mental Disorders (DSM-5®)*. American Psychiatric Pub.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, *67*, 1–48. https://doi.org/10.18637/jss.v067.i01

Bavard, S., Lebreton, M., Khamassi, M., Coricelli, G., & Palminteri, S. (2018). Reference-point centering and range-adaptation enhance human reinforcement learning at the cost of irrational preferences. *Nature Communications*, *9*(1), 4503. https://doi.org/10.1038/s41467-018-06781-2

Bavard, S., & Théro, H. (2018). *[Re] Adaptive Properties Of Differential Learning Rates For Positive And Negative Outcomes*. https://doi.org/10.5281/ZENODO.1289889

Beck, A. T. (1967). *Depression: Clinical, Experimental, and Theoretical Aspects*. Hoeber Medical Division, Harper & Row.

Beck, A. T., Steer, R. A., Ball, R., & Ranieri, W. F. (1996). Comparison of Beck Depression Inventories-IA and-II in Psychiatric Outpatients. *Journal of Personality Assessment*, *67*(3), 588–597. https://doi.org/10.1207/s15327752jpa6703_13

Brolsma, S. C. A., Vassena, E., Vrijsen, J. N., Sescousse, G., Collard, R. M., van Eijndhoven, P. F., Schene, A. H., & Cools, R. (2021). Negative Learning Bias in Depression Revisited: Enhanced Neural Response to Surprising Reward Across Psychiatric Disorders. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, *6*(3), 280–289. https://doi.org/10.1016/j.bpsc.2020.08.011

Brolsma, S. C. A., Vrijsen, J. N., Vassena, E., Kandroodi, M. R., Bergman, M. A., Eijndhoven, P. F. van, Collard, R. M., Ouden, H. E. M. den, Schene, A. H., & Cools, R. (2020). Challenging the negative learning bias hypothesis of depression: Reversal learning in a naturalistic psychiatric sample. *Psychological Medicine*, 1–11. https://doi.org/10.1017/S0033291720001956

Byrne, D. G. (1976). Choice Reaction Times in Depressive States. *British Journal of Social and Clinical Psychology*, *15*(2), 149–156. https://doi.org/10.1111/j.2044-8260.1976.tb00020.x

Cazé, R. D., & van der Meer, M. A. A. (2013). Adaptive properties of differential learning rates for positive and negative outcomes. *Biological Cybernetics*, *107*(6), 711–719. https://doi.org/10.1007/s00422-013-0571-5

Chambon, V., Théro, H., Vidal, M., Vandendriessche, H., Haggard, P., & Palminteri, S. (2020). Information about action outcomes differentially affects learning from self-determined versus imposed choices. *Nature Human Behaviour*, *4*(10), 1067–1079. https://doi.org/10.1038/s41562-020-0919-5

Chase, H. W., Frank, M. J., Michael, A., Bullmore, E. T., Sahakian, B. J., & Robbins, T. W. (2010). Approach and avoidance learning in patients with major depression and healthy controls: Relation to anhedonia. *Psychological Medicine*, *40*(3), 433–440. https://doi.org/10.1017/S0033291709990468

Chen, C., Takahashi, T., Nakagawa, S., Inoue, T., & Kusumi, I. (2015). Reinforcement learning in depression: A review of computational research. *Neuroscience & Biobehavioral Reviews*, *55*, 247–267. https://doi.org/10.1016/j.neubiorev.2015.05.005

Chung, D., Kadlec, K., Aimone, J. A., McCurry, K., King-Casas, B., & Chiu, P. H. (2017). Valuation in major depression is intact and stable in a non-learning environment. *Scientific Reports*, *7*, 44374. https://doi.org/10.1038/srep44374

Collins, A. G. E., & Frank, M. J. (2012). How much of reinforcement learning is working memory, not reinforcement learning? A behavioral, computational, and neurogenetic analysis. *European Journal of Neuroscience*, *35*(7), 1024–1035. https://doi.org/10.1111/j.1460-9568.2011.07980.x

Daw, N. D., Gershman, S. J., Seymour, B., Dayan, P., & Dolan, R. J. (2011). Model-Based Influences on Humans' Choices and Striatal Prediction Errors. *Neuron*, *69*(6), 1204–1215. https://doi.org/10.1016/j.neuron.2011.02.027

Douglas, K. M., Porter, R. J., Frampton, C. M., Gallagher, P., & Young, A. H. (2009). Abnormal response to failure in unmedicated major depression. *Journal of Affective Disorders*, *119*(1), 92–99. https://doi.org/10.1016/j.jad.2009.02.018

Elliott, R., Sahakian, B. J., Herrod, J. J., Robbins, T. W., & Paykel, E. S. (1997). Abnormal response to negative feedback in unipolar depression: Evidence for a diagnosis specific impairment. *Journal of Neurology, Neurosurgery & Psychiatry*, *63*(1), 74–82. https://doi.org/10.1136/jnnp.63.1.74

Elliott, R., Sahakian, B. J., McKay, A. P., Herrod, J. J., Robbins, T. W., & Paykel, E. S. (1996). Neuropsychological impairments in unipolar depression: The influence of perceived failure on subsequent performance. *Psychological Medicine*, *26*(5), 975–989. https://doi.org/10.1017/S0033291700035303

Eshel, N., & Roiser, J. P. (2010). Reward and Punishment Processing in Depression. *Biological Psychiatry*, *68*(2), 118–124. https://doi.org/10.1016/j.biopsych.2010.01.027

Fontanesi, L., Gluth, S., Spektor, M. S., & Rieskamp, J. (2019). A reinforcement learning diffusion decision model for value-based decisions. *Psychonomic Bulletin & Review*, *26*(4), 1099–1121. https://doi.org/10.3758/s13423-018-1554-2

Fontanesi, L., Palminteri, S., & Lebreton, M. (2019). Decomposing the effects of context valence and feedback information on speed and accuracy during reinforcement learning: A meta-analytical approach using diffusion decision modeling. *Cognitive, Affective, & Behavioral Neuroscience*, *19*(3), 490–502. https://doi.org/10.3758/s13415-019-00723-1

Forbes, E. E., & Dahl, R. E. (2012). Research Review: Altered reward function in adolescent depression: what, when and how? *Journal of Child Psychology and Psychiatry*, *53*(1), 3–15. https://doi.org/10.1111/j.1469-7610.2011.02477.x

Frank, M. J., Moustafa, A. A., Haughey, H. M., Curran, T., & Hutchison, K. E. (2007). Genetic triple dissociation reveals multiple roles for dopamine in reinforcement learning. *Proceedings of the National Academy of Sciences*, *104*(41), 16311–16316. https://doi.org/10.1073/pnas.0706111104

Frank, M. J., Seeberger, L. C., & O'Reilly, R. C. (2004). By Carrot or by Stick: Cognitive Reinforcement Learning in Parkinsonism. *Science*, *306*(5703), 1940–1943. https://doi.org/10.1126/science.1102941

Gotlib, I. H., & Joormann, J. (2010). Cognition and Depression: Current Status and Future Directions. *Annual Review of Clinical Psychology*, *6*(1), 285–312. https://doi.org/10.1146/annurev.clinpsy.121208.131305

Gradin, V. B., Kumar, P., Waiter, G., Ahearn, T., Stickle, C., Milders, M., Reid, I., Hall, J., & Steele, J. D. (2011). Expected value and prediction error abnormalities in depression and schizophrenia. *Brain: A Journal of Neurology*, *134*(Pt 6), 1751–1764. https://doi.org/10.1093/brain/awr059

Guitart-Masip, M., Huys, Q. J. M., Fuentemilla, L., Dayan, P., Duzel, E., & Dolan, R. J. (2012). Go and no-go learning in reward and punishment: Interactions between affect and effect. *NeuroImage*, *62*(1), 154–166. https://doi.org/10.1016/j.neuroimage.2012.04.024

Hägele, C., Schlagenhauf, F., Rapp, M., Sterzer, P., Beck, A., Bermpohl, F., Stoy, M., Ströhle, A., Wittchen, H.-U., Dolan, R. J., & Heinz, A. (2015). Dimensional psychiatry: Reward dysfunction and depressive mood across psychiatric disorders. *Psychopharmacology*, *232*(2), 331–341. https://doi.org/10.1007/s00213-014-3662-7

Henriques et al. (1994). *Reward fails to alter response bias in depression.* https://psycnet.apa.org/buy/1994-45308-001

Henriques, J. B., & Davidson, R. J. (2000). Decreased responsiveness to reward in depression. *Cognition and Emotion*, *14*(5), 711–724. https://doi.org/10.1080/026999300501176 84

Huys, Q. J. M., Gölzer, M., Friedel, E., Heinz, A., Cools, R., Dayan, P., & Dolan, R. J. (2016). The specificity of Pavlovian regulation is associated with recovery from depression. *Psychological Medicine*, *46*(5), 1027–1035. https://doi.org/10.1017/S0033291715002597

Huys, Q. J., Pizzagalli, D. A., Bogdan, R., & Dayan, P. (2013). Mapping anhedonia onto reinforcement learning: A behavioural meta-analysis. *Biology of Mood & Anxiety Disorders*, *3*(1), 12. https://doi.org/10.1186/2045-5380-3-12

Insel, T., Cuthbert, B., Garvey, M., Heinssen, R., Pine, D. S., Quinn, K., Sanislow, C., & Wang, P. (2010). Research Domain Criteria (RDoC): Toward a New Classification Framework for Research on Mental Disorders. *American Journal of Psychiatry*, *167*(7), 748–751. https://doi.org/10.1176/appi.ajp.2010.09091379

Joormann, J., & Quinn, M. E. (2014). Cognitive Processes and Emotion Regulation in Depression. *Depression and Anxiety*, *31*(4), 308–315. https://doi.org/10.1002/da.22264

Katahira, K., Yuki, S., & Okanoya, K. (2017). Model-based estimation of subjective values using choice tasks with probabilistic feedback. *Journal of Mathematical Psychology*, *79*, 29–43. https://doi.org/10.1016/j.jmp.2017.05.005

Klein, T. A., Ullsperger, M., & Jocham, G. (2017). Learning relative values in the striatum induces violations of normative decision making. *Nature Communications*, *8*(1), 16033. https://doi.org/10.1038/ncomms16033

Knutson, B., Bhanji, J. P., Cooney, R. E., Atlas, L. Y., & Gotlib, I. H. (2008). Neural Responses to Monetary Incentives in Major Depression. *Biological Psychiatry*, *63*(7), 686–692. https://doi.org/10.1016/j.biopsych.2007.07.023

Kumar, P., Waiter, G., Ahearn, T., Milders, M., Reid, I., & Steele, J. D. (2008). Abnormal temporal difference reward-learning signals in major depression. *Brain*, *131*(8), 2084–2093. https://doi.org/10.1093/brain/awn136

Moutoussis, M., Rutledge, R. B., Prabhu, G., Hrynkiewicz, L., Lam, J., Ousdal, O.-T., Guitart-Masip, M., Fonagy, P., & Dolan, R. J. (2018). Neural activity and fundamental learning, motivated by monetary loss and reward, are intact in mild to moderate major depressive disorder. *PLOS ONE*, *13*(8), e0201451. https://doi.org/10.1371/journal.pone.0201451

Murphy, F. C., Michael, A., Robbins, T. W., & Sahakian, B. J. (2003). Neuropsychological impairment in patients with major depressive disorder: The effects of feedback on task performance. *Psychological Medicine*, *33*(3), 455–467. https://doi.org/10.1017/S0033291702007018

Niv, Y., Edlund, J. A., Dayan, P., & O'Doherty, J. P. (2012). Neural Prediction Errors Reveal a Risk-Sensitive Reinforcement-Learning Process in the Human Brain. *Journal of Neuroscience*, *32*(2), 551–562. https://doi.org/10.1523/JNEUROSCI.5498-10.2012

Palminteri, S., Clair, A.-H., Mallet, L., & Pessiglione, M. (2012). Similar Improvement of Reward and Punishment Learning by Serotonin Reuptake Inhibitors in Obsessive-Compulsive Disorder. *Biological Psychiatry*, *72*(3), 244–250. https://doi.org/10.1016/j.biopsych.2011.12.028

Palminteri, S., Khamassi, M., Joffily, M., & Coricelli, G. (2015). Contextual modulation of value signals in reward and punishment learning. *Nature Communications*, *6*(1), 8096. https://doi.org/10.1038/ncomms9096

Palminteri, S., & Lebreton, M. (2021). Context-dependent outcome encoding in human reinforcement learning. *Current Opinion in Behavioral Sciences*, *41*, 144–151. https://doi.org/10.1016/j.cobeha.2021.06.006

Palminteri, S., Lefebvre, G., Kilford, E. J., & Blakemore, S.-J. (2017). Confirmation bias in human reinforcement learning: Evidence from counterfactual feedback processing. *PLOS Computational Biology*, *13*(8), e1005684. https://doi.org/10.1371/journal.pcbi.1005684

Palminteri, S., & Pessiglione, M. (2013). Chapter Five—Reinforcement Learning and Tourette Syndrome. In D. Martino & A. E. Cavanna (Eds.), *International Review of Neurobiology* (Vol. 112, pp. 131–153). Academic Press. https://doi.org/10.1016/B978-0-12-411546-0.00005-6

Palminteri, S., & Pessiglione, M. (2017). Chapter 23 - Opponent Brain Systems for Reward and Punishment Learning: Causal Evidence From Drug and Lesion Studies in Humans. In J.-C. Dreher & L. Tremblay (Eds.), *Decision Neuroscience* (pp. 291–303). Academic Press. https://doi.org/10.1016/B978-0-12-805308-9.00023-3

Pessiglione, M., Seymour, B., Flandin, G., Dolan, R., & Frith, C. (2006). Dopamine-dependent prediction errors underpin reward-seeking behaviour in humans. *Nature*. https://doi.org/10.1038/nature05051

Pike, A. C., & Robinson, O. J. (2022). Reinforcement Learning in Patients With Mood and Anxiety Disorders vs Control Individuals: A Systematic Review and Meta-analysis. *JAMA Psychiatry*. https://doi.org/10.1001/jamapsychiatry.2022.0051

Pizzagalli, D. A. (2014). *Depression, Stress, and Anhedonia: Toward a Synthesis and Integrated Model | Annual Review of Clinical Psychology*. https://www.annualreviews.org/doi/abs/10.1146/annurev-clinpsy-050212-185606

Pizzagalli, D. A., Jahn, A. L., & O'Shea, J. P. (2005). Toward an objective characterization of an anhedonic phenotype: A signal-detection approach. *Biological Psychiatry*, *57*(4), 319–327. https://doi.org/10.1016/j.biopsych.2004.11.026

Rescorla, R., & Wagner, A. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In *Classical Conditioning II: Current Research and Theory: Vol. Vol. 2*.

Rothkirch, M., Tonn, J., Köhler, S., & Sterzer, P. (2017). Neural mechanisms of reinforcement learning in unmedicated patients with major depressive disorder. *Brain*, *140*(4), 1147–1157. https://doi.org/10.1093/brain/awx025

Rupprechter, S., Stankevicius, A., Huys, Q. J. M., Steele, J. D., & Seriès, P. (2018). Major Depression Impairs the Use of Reward Values for Decision-Making. *Scientific Reports*, *8*(1), 13798. https://doi.org/10.1038/s41598-018-31730-w

Rutledge, R. B., Moutoussis, M., Smittenaar, P., Zeidman, P., Taylor, T., Hrynkiewicz, L., Lam, J., Skandali, N., Siegel, J. Z., Ousdal, O. T., Prabhu, G., Dayan, P., Fonagy, P., & Dolan, R. J. (2017). Association of Neural and Emotional Impacts of Reward Prediction Errors With Major Depression. *JAMA Psychiatry*, *74*(8), 790–797. https://doi.org/10.1001/jamapsychiatry.2017.1713

Safra, L., Chevallier, C., & Palminteri, S. (2019). Depressive symptoms are associated with blunted reward learning in social contexts. *PLOS Computational Biology*, *15*(7), e1007224. https://doi.org/10.1371/journal.pcbi.1007224

Shah, P. J., O'carroll, R. E., Rogers, A., Moffoot, A. P. R., & Ebmeier, K. P. (1999). Abnormal response to negative feedback in depression. *Psychological Medicine*, *29*(1), 63–72. https://doi.org/10.1017/S0033291798007880

Sheehan, D. V., Lecrubier, Y., Sheehan, K. H., Amorim, P., Janavs, J., Weiller, E., Hergueta, T., Baker, R., & Dunbar, G. C. (1998). The Mini-International Neuropsychiatric Interview (M.I.N.I.): The development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. *The Journal of Clinical Psychiatry*, *59 Suppl 20*, 22-33;quiz 34-57.

Steele, J. D., Kumar, P., & Ebmeier, K. P. (2007). Blunted response to feedback information in depressive illness. *Brain*, *130*(9), 2367–2374. https://doi.org/10.1093/brain/awm150

Sutton, R. S., & Barto, A. G. (2018). *Reinforcement learning: An introduction* (Second edition). The MIT Press.

Ubl, B., Kuehner, C., Kirsch, P., Ruttorf, M., Diener, C., & Flor, H. (2015). Altered neural reward and loss processing and prediction error signalling in depression. *Social Cognitive and Affective Neuroscience*, *10*(8), 1102–1112. https://doi.org/10.1093/scan/nsu158

Vrieze, E., Pizzagalli, D. A., Demyttenaere, K., Hompes, T., Sienaert, P., de Boer, P., Schmidt, M., & Claes, S. (2013). Reduced Reward Learning Predicts Outcome in Major Depressive Disorder. *Biological Psychiatry*, *73*(7), 639–645. https://doi.org/10.1016/j.biopsych.2012.10.014

Whitton, A. E., Kakani, P., Foti, D., Van't Veer, A., Haile, A., Crowley, D. J., & Pizzagalli, D. A. (2016). Blunted Neural Responses to Reward in Remitted Major Depression: A High-Density Event-Related Potential Study. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, *1*(1), 87–95. https://doi.org/10.1016/j.bpsc.2015.09.007

Whitton, A. E., Treadway, M. T., & Pizzagalli, D. A. (2015). Reward processing dysfunction in major depression, bipolar disorder and schizophrenia. *Current Opinion in Psychiatry*, *28*(1), 7–12. https://doi.org/10.1097/YCO.0000000000000122

Wilson, R. C., & Collins, A. G. (2019). Ten simple rules for the computational modeling of behavioral data. *ELife*, *8*, e49547. https://doi.org/10.7554/eLife.49547

World Health Organization. (2017). *Depression and other common mental disorders: Global health estimates* (WHO/MSD/MER/2017.2). World Health Organization. https://apps.who.int/iris/handle/10665/254610

Yechiam, E., & Hochman, G. (2014). Loss Attention in a Dual-Task Setting. *Psychological Science*, *25*(2), 494–502. https://doi.org/10.1177/0956797613510725

Yu, Z., Guindani, M., Grieco, S. F., Chen, L., Holmes, T. C., & Xu, X. (2022). Beyond t test and ANOVA: Applications of mixed-effects models for more rigorous statistical analysis in neuroscience research. *Neuron*, *110*(1), 21–35. https://doi.org/10.1016/j.neuron.2021.10.030

Zhang, W.-N., Chang, S.-H., Guo, L.-Y., Zhang, K.-L., & Wang, J. (2013). The neural correlates of reward-related processing in major depressive disorder: A meta-analysis of functional magnetic resonance imaging studies. *Journal of Affective Disorders*, *151*(2), 531–539. https://doi.org/10.1016/j.jad.2013.06.039

Group	Patients	Controls	Significance
Gender (%female)	30 (53.33)	26 (61.53)	P = 0.54
Age (mean±sem)	36.5 ± 2.80	40.35 ± 2.09	P= 0.28
Education	1.97 ± 0.24	2.42 ± 0.21	P = 0.12
Usual Optimism	5.98 ± 0.42	7.16 ± 0.30	P= 0.03
Current Optimism	2.38 ± 0.40	7.46 ± 0.29	P= 4.19e-14
LOTR	9.1 ± 0.79	16 ± 0.49	P= 1.76e-09
BDI	29.37 ± 0.22	-	-
Previous MDE	1.8 ± 0.38	-	-

Table 1:
Descriptive statistics for age, gender, education, usual optimism (LOT-R: Life Orientation Test – Revised), current optimism, depression scores (BDI: Beck Depression Inventory) and number of major depressive episodes (MDE). Education: years after graduation For each sample, the mean of each variable is presented with its standard error of the mean.

Medication	Number of Patients
SSRI	22
Benzodiazepine	21
Tricyclic antidepressant	2
Tetracyclic antidepressant	1
Phenothiazine	2
Corticosteroïds	1
Others	2

Table 2:
Patients' treatments. 'SSRI': selective serotonin reuptake inhibitor; 'others': anti-arrhythmic agent or vitamins.

Figures legends

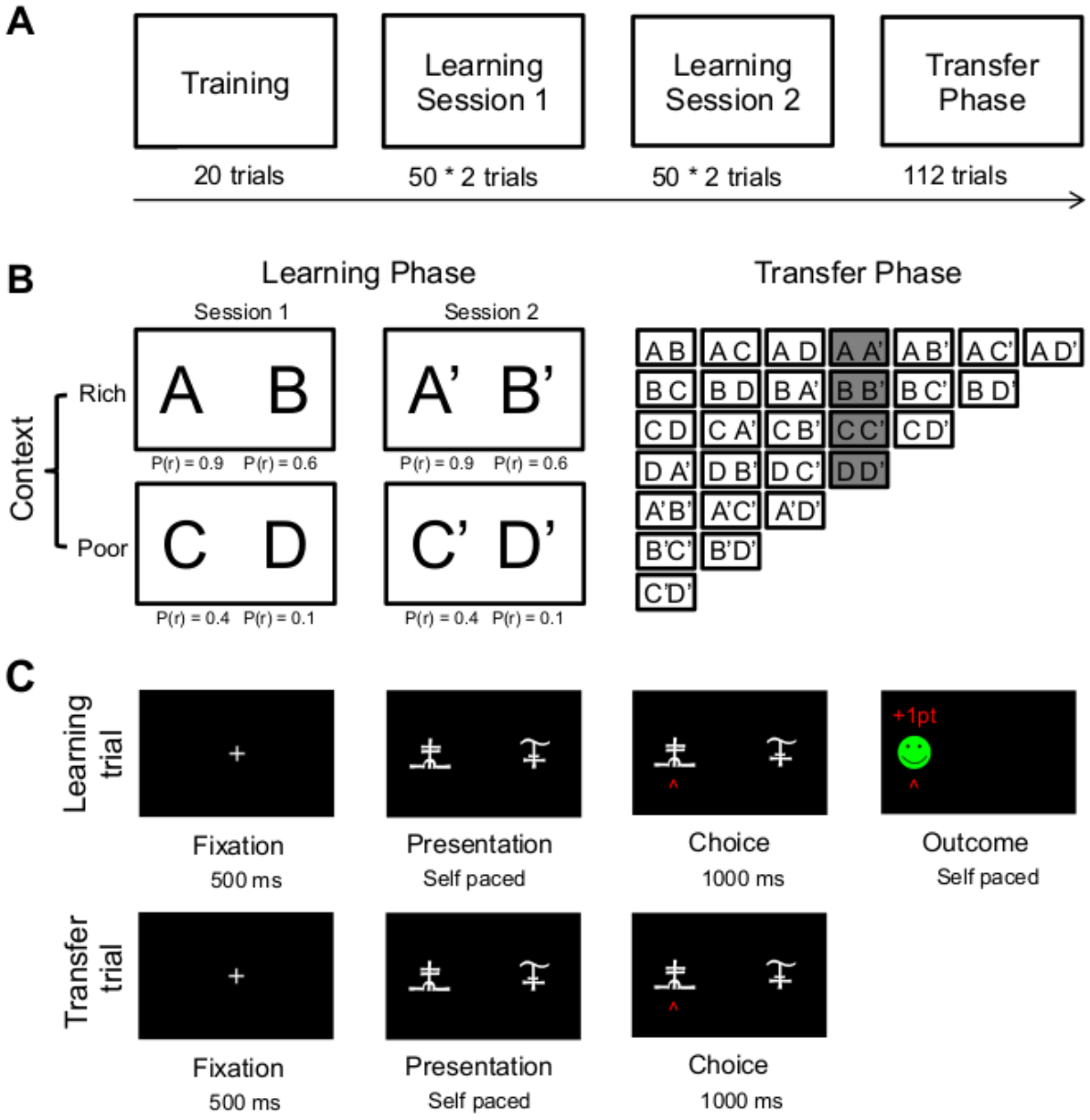

Figure 1: experimental methods. **(A)** Time course of the experiment: after written instruction the experiment started with a short training (16 trials) using different stimuli (letters). The training was followed by two learning sessions, each with 4 different stimuli, arranged in fixed pairs. Each pair was presented 50 times, learning to 200 trials in total. After the last session, participants were administered a transfer phase where all stimuli from the learning sessions were presented in all possible combinations. All pair-wise combinations (28) were presented 4 times, learning to 112 trials in total. **(B)** Option pairs. Each learning sessions featured two 2 fixed pairs of options (*contexts*), characterized by different outcomes values: a 'rich' one with an overall positive expected value (the optimal option with a 0.9 probability of reward) and a 'poor' context (the optimal option with a 0.4 probability of reward). The two contexts were presented in an interleaved manner during the learning phase. On the transfer phase all 8 symbols from the learning phase (2 symbols x 2 contexts x 2 learning sessions) were presented in every possible combination. Grey boxed indicate the comparisons between options with the same value (e.g., A vs A'), which were not included in the statistical analysis of the transfer test (because there is no accurate response). **(C)** Successive screen in the learning phrase (top) and the transfer phase (bottom). Durations are given in milliseconds.

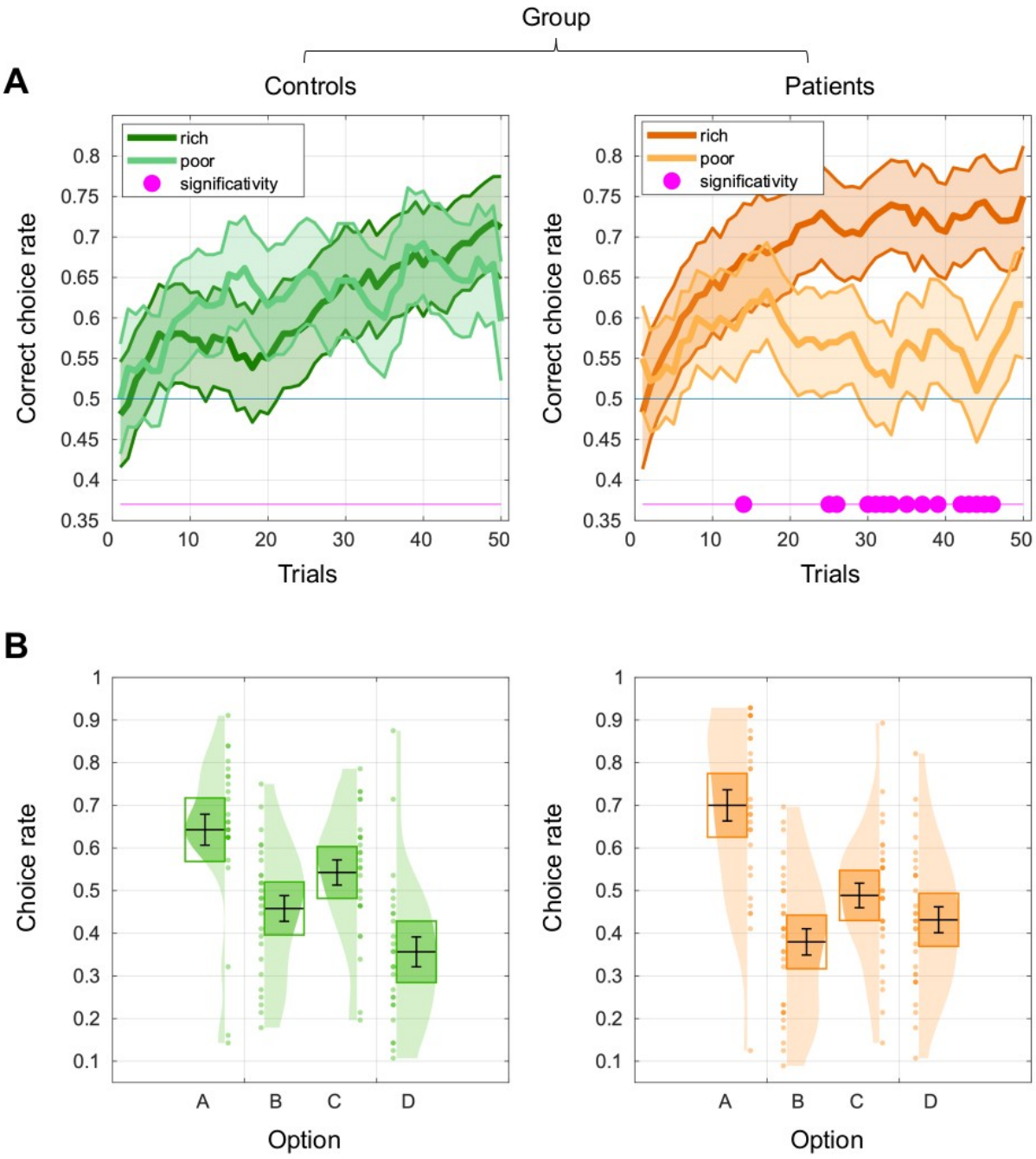

Figure 2: choice data. (A) 'Correct choice rate' is the probability of picking the most rewarding option. Thick lines represent smoothed running average (5 trials kernel) and shaded areas the standard error of the mean. The violet dots correspond to trials displaying a significant difference among conditions ($p<0.05$; calculated on the raw, unsmoothed, data points). **(B) '**Choice rate' is the probability of picking a given symbols in any given choice pair. The choice rates are averages across symbols belonging to the first and second session (in **Figure 1**, denoted A and A', respectively). Areas represent probability density functions. Boxes represent confidence interval (95%) and dots represent individual subjects.

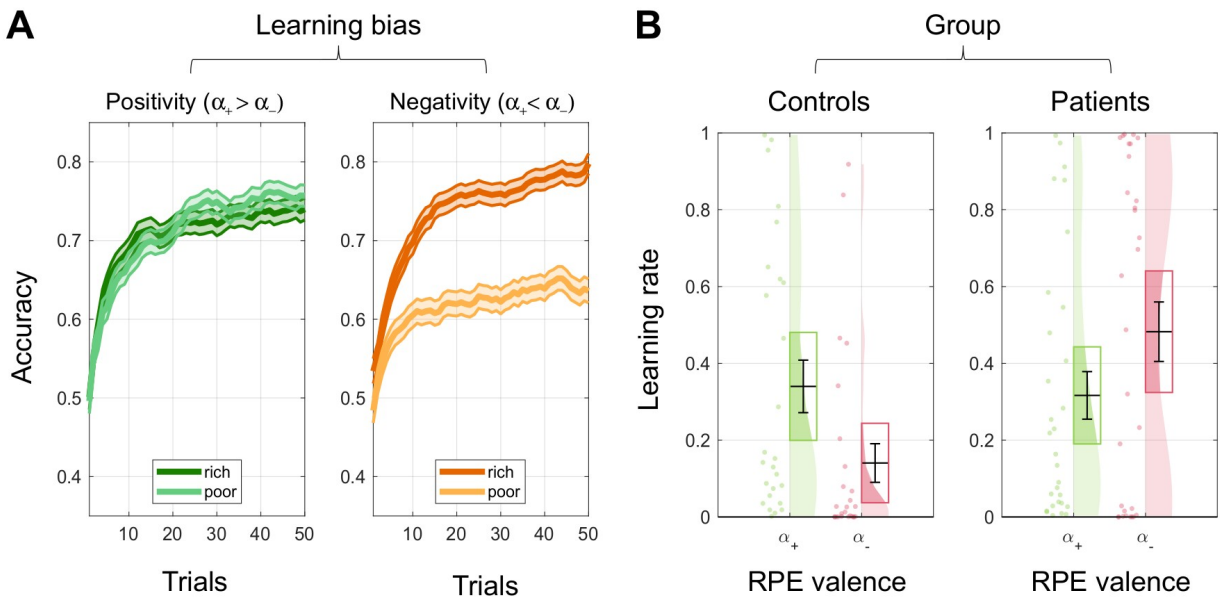

Figure 3: model-based results. (A) The panels depict the results of model simulations where agents are represented by a two learning rates model, featuring either a positivity or a negativity bias (N=1000 virtual subjects per group; see methods for more details about the simulations). The leftmost panel (green) show the simulations of agents displaying a positivity bias, while the rightmost panel (orange) displays the simulations of agents displaying a negativity bias. Thick lines represent smoothed running average (5 trials kernel) and shaded areas the standard error of the mean**. (B)** The panels represent learning rates for positive (green) and negative (red) prediction errors separately for healthy controls (leftmost panel) and patients (rightmost panel)

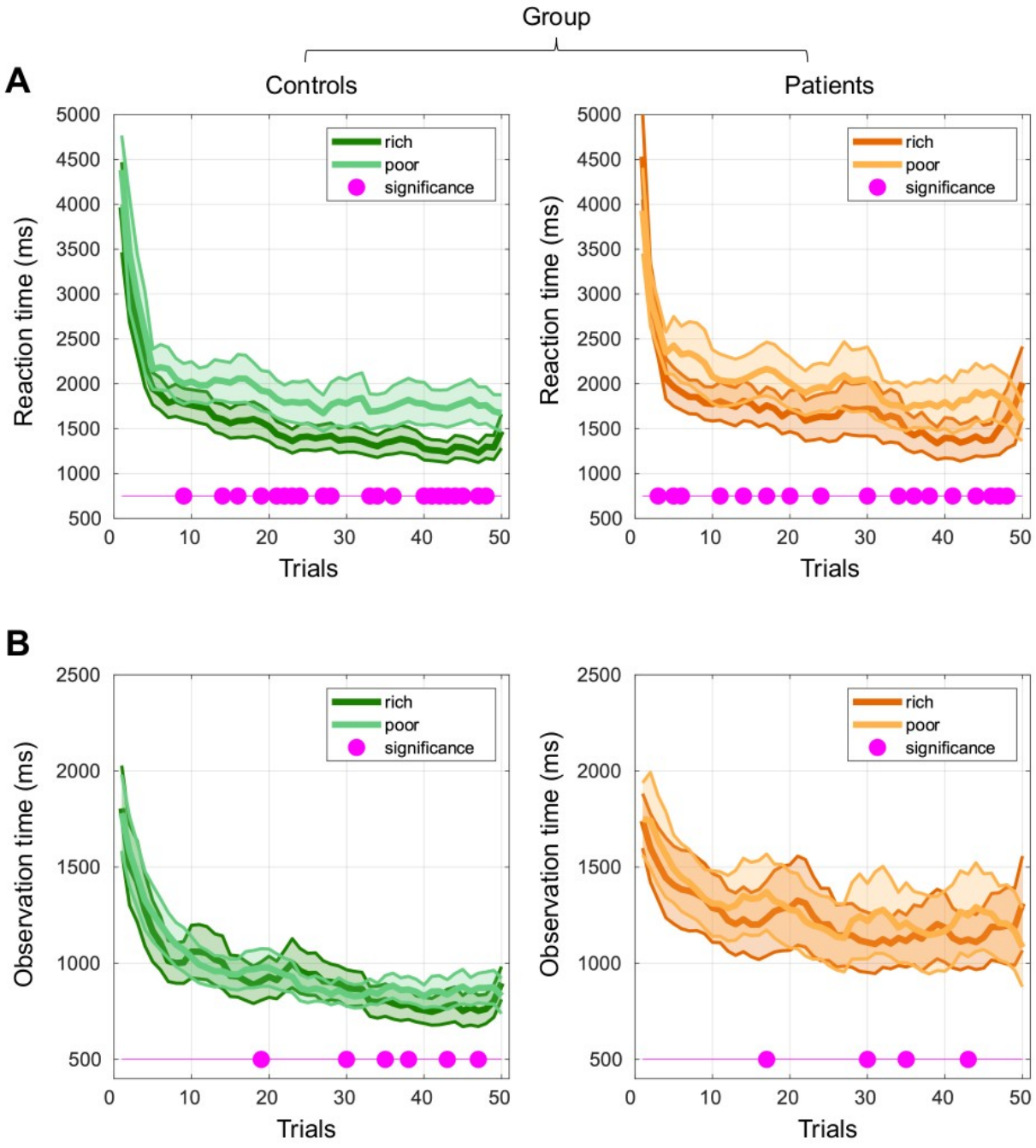

Figure 4: response times. (A) 'Reaction time' is the time separating the options onset from the moment the participant selects of one of the two options. Trials are grouped by condition and averaged across sessions. Durations are given in milliseconds. Thick lines represent smoothed running average (5 trials kernel) and shaded areas the standard error of the mean. The violet dots correspond to trials displaying a significant difference among conditions ($p<0.05$; calculated on the raw, unsmoothed, data points). **(B)** Outcome observation time is the time separating the outcome onset from the moment the participant confirms the outcome to move to the subsequent trial. Legend as in (**A**).