Journal of Affective Disorders 264 (2020) 400–406

Contents lists available at ScienceDirect

Journal of Affective Disorders

journal homepage: www.elsevier.com/locate/jad

Research paper

Aberrant probabilistic reinforcement learning in first-degree relatives of individuals with bipolar disorder

Julia O. Linke[a,*], Georgia Koppe[b], Vanessa Scholz[a], Philipp Kanske[c,d], Daniel Durstewitz[b], Michèle Wessa[a]

[a] Department of Neuropsychology and Clinical Psychology, Psychological Institute, Johannes-Gutenberg University of Mainz, Germany
[b] Department of Theoretical Neuroscience, Central Institute of Mental Health, Medical Faculty Mannheim/Heidelberg University, Mannheim, Germany
[c] Clinical Psychology and Behavioral Neuroscience, Faculty of Psychology, Technische Universität Dresden, Dresden, Germany
[d] Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany

ARTICLE INFO

Keywords:
Bipolar disorder
First-degree relatives
Behavioral activation system (BAS)
Reinforcement learning
Computational modeling

ABSTRACT

Background: Motivational dysregulation represents a core vulnerability factor for bipolar disorder. Whether this also comprises aberrant learning of stimulus-reinforcer contingencies is less clear.
Methods: To answer this question, we compared healthy first-degree relatives of individuals with bipolar disorder (n = 42) known to convey an increased risk of developing a bipolar spectrum disorder and healthy individuals (n = 97). Further, we investigated the effects of the behavioral activation system (BAS) on reinforcement learning across the entire sample. All participants were assessed with a probabilistic learning task that distinguishes learning from positive and negative feedback. Main outcome measures included choice frequencies and learning rate parameters generated by computational reinforcement learning algorithms.
Results: First-degree relatives choose more rewarding stimuli more consistently and showed marginally reduced learning rates from unexpected negative feedback. Further, first-degree relatives had lower BAS scores than controls, which were negatively associated with learning rates from unexpected negative feedback.
Limitations: However as probands also reported other mental disorders such as Attention-Deficit/Hyperactivity Disorder and substance abuse among their first-degree relatives, we cannot know, whether these findings are specific to the risk for bipolar disorder.
Conclusion: The behavior of first-degree relatives of individuals with bipolar disorder, who also display increased BAS sensitivity, is less influenced by unexpected negative feedback. This reduced learning from unexpected negative feedback biases subsequent choices towards stimuli with higher probabilities for a reward. In sum, our results confirm the role of aberrant reinforcement learning in the pathophysiology of bipolar disorder.

Bipolar disorder (BD), which is characterized by episodes of aberrant affect, motivation, and cognition (American Psychiatric Association, 2013), has a heritability of 60–80% (McGuffin et al., 2003). Thus, a family history of BD is the strongest predictor of developing BD (Paaren et al., 2014). However, to date, we cannot predict which relatives will manifest the disorder themselves. For early and precise diagnosis and the development of targeted preventions, it is crucial to enhance our understanding of how an increased familial risk is expressed at a mechanistic level.

Motivational aberrancies and more specifically, the aberrant processing of action outcomes might be one of the mechanisms through which familial risk is expressed (Wessa et al., 2013). More specifically, it has been suggested that lower learning rates for negative than positive outcomes might lead to overly optimistic expectations. These excessively optimistic expectations are thought to increase the frequency and magnitude of unexpected, adverse outcomes, leading to low mood. When the mood eventually returns to baseline levels, the pessimistic expectations that developed during depressed mood may now lead to increased positive surprises and improved mood (Eldar and Niv, 2015). Indeed, heightened responses to positive feedback (Linke et al., 2012; Singh et al., 2014), and reduced sensitivity to unexpected negative feedback (Linke et al., 2012) have been observed in unaffected first-degree relatives of individuals with BD.

Of note, models of BD also emphasize the role to the behavioral

* Corresponding author: Julia Linke, Department of Clinical and Neuropsychology, Institute for Psychology, Johannes Gutenberg-University of Mainz, Mainz, Germany
E-mail address: linkej@uni-mainz.de (J.O. Linke).

https://doi.org/10.1016/j.jad.2019.11.063
Received 17 June 2019; Received in revised form 22 September 2019; Accepted 10 November 2019
Available online 12 November 2019
0165-0327/

activation system (BAS), which presumably mediates individual differences in the sensitivity and reactivity to appetitive stimuli (Gray, 1987). Higher BAS-scores were shown to predict the onset of mania (Alloy et al., 2012a; Meyer et al., 2001). Further, high levels of BAS and the behavioral inhibition system (BIS), which is thought to be sensitive to signals of punishment and non-reward, predict the progression from bipolar spectrum disorders to BD I disorder (Alloy et al., 2012b). On a neurobiological level, the BAS has been associated with the mesolimbic and mesocortical dopamine system (Depue and Iacono, 1989), which codes prediction errors referring to expectation violations (e.g., feedback that is better or worse than expected (Montague et al., 1996)) relevant for reinforcement learning. While learning from unexpected positive feedback has been linked to striatal neurons expressing mostly dopamine-D1 receptors, learning from unexpected negative feedback relies more on dopamine-D2 receptors (Maia and Frank, 2011). Interestingly, high BAS scores have been associated with the disequilibrium between dopamine-D2 receptor density and the availability of the enzyme catechol-O-methyltransferase (COMT) (Reuter et al., 2006) both relevant for learning from negative outcomes (Maia and Frank, 2011).

To date, little is known how the familial risk for BD and individual differences in BAS and BIS interact in terms of aberrancies in reinforcement learning. Thus, the primary goal of the present study was to parse common versus specific abnormalities in reinforcement learning associated with (a) familial risk for BD and (b) individual differences in the BAS and BIS. We used a probabilistic learning task (Frank et al., 2004) and applied computational reinforcement learning algorithms which make use of the individual trial-by-trial choice behavior to estimate parameters reflecting learning from unexpected positive and negative feedback. Based on the literature, we predicted that both familial risk for BD and high BAS scores would be associated with impaired learning from negative prediction errors. We further hypothesized that relatives of individuals with bipolar disorder would show a stronger preference for stimuli with a higher probability to be followed by positive feedback.

1. Methods

1.1. Participants

The study sample consisted of 42 unaffected first-degree relatives of individuals with bipolar I disorder (REL, 17 siblings, 25 children) and 97 healthy volunteers (HV) without a family history of mental disorder. REL were recruited through individuals with bipolar I disorder that participated in previous and ongoing studies conducted by our group. Of the 82 REL interested in participating, 40 were excluded during an initial phone screening due to mental illness (major depressive disorder: $n = 18$, Attention-deficit/Hyperactivity Disorder (ADHD): $n = 8$, anxiety disorder: $n = 10$, Substance use disorders: $n = 3$, anorexia nervosa: $n = 1$). HVs were recruited from an existing subject pool. REL and HVs were comparable regarding sex, years of education, socioeconomic status, intelligence, working memory capacity and BIS-score. However, HVs were younger and showed higher BAS-scores than REL. For details, please see Table 1.

The study was approved by the Ethics Committee of the Medical Faculty Mannheim of the University of Heidelberg and complied with the Declaration of Helsinki. Participants gave written informed consent before study participation.

1.2. Clinical assessment

The absence of affective, psychotic, anxiety, eating, and substance use and addictive, and Attention-Deficit/Hyperactivity Disorder (ADHD) was confirmed in all participants using the Structured Clinical Interview for DSM-IV for Axis I (Wittchen et al., 1997) and ADHD diagnostics (Retz-Junginger et al., 2002). All index cases met criteria for bipolar I disorder. Twelve index cases met criteria for alcohol use and addictive disorder, and 2 cases fulfilled criteria of panic disorder. Moreover, REL reported additional cases of major depression ($n = 18$), ADHD ($n = 1$), anxiety ($n = 5$) and alcohol use and addictive disorder ($n = 6$) among their first-degree relatives. Seventeen REL reported ≥2 cases of BD within their family. HVs did not endorse any affective, psychotic, substance use and addictive disorder, or ADHD for their first-degree relatives.

Table 1
Demographic characteristics, cognitive abilities and questionnaire data

	Relatives ($n = 42$)	HV ($n = 97$)	Statistics	p-value
Demographics				
Sex, female/ male	24/18	57/40	$\chi^2_{(1)} = 0.03$	.86
Age, y, mean (SD)	33.6 (14.2)	27.5 (9.3)	$t_{(137)} = -2.52$	.02
Years of Education, mean (SD)	15.5 (2.7)	16.1 (1.7)	$t_{(137)} = 1.23$	.23
ISEI score, mean (SD)	56.5 (16.3)	58.4 (14.3)	$t_{(137)} = 0.69$	.50
Cognitive abilities				
Intelligence score, mean (SD)[b]	104 (11)	101 (10)	$t_{(137)} = -1.07$	.29
Working memory, mean (SD) [c]	58 (30)	57 (28)	$t_{(137)} = -0.29$	.77
Questionnaires				
BAS score, mean (SD)	2.9 (0.4)	3.1 (0.3)	$t_{(137)} = 3.41$	.001
BIS score, mean (SD)	2.6 (0.3)	2.6 (0.4)	$t_{(137)} = 0.59$	.55

Abbreviations: ISEI, international socio-economic index of occupational status ranging from 16 (low) to 90 (high) according to Ganzeboom et al., (1992); *n*, sample size; SD, Standard deviation; y, years.
[b] Raw scores were standardized by IQ-transformation.
[c] Raw scores were standardized by percentile rank transformation.

2. Questionnaires

2.1. BAS, BIS

We assessed the German version of the BIS/BAS scales (Strobel et al., 2001) that measure participant's urge to approach reward-related cues or to pursue goals (BAS) with 13 items and their sensitivity to signals of punishment or non-reward (BIS) with seven items. All items are assessed on a four-point Likert scale, ranging from 1 (strongly disagree) to 4 (strongly agree). Both scales have good psychometric properties (BAS: Cronbach's $\alpha = 0=.81$; BIS: Cronbach's $\alpha = 0.78$ (Strobel et al., 2001).

2.2. Neuropsychological assessment

We used the German version of the Multiple Choice Word Vocabulary Test (Lehrl, 2005) that has a 6-month retest-reliability of $r = 0.95$ and a criterion validity of $r = 0.81$ (Lehrl, 2002) to measure intelligence. Working memory performance was assessed with the digit span subtest from the Wechsler Adult Intelligence Scale (Wechsler, 1997).

2.3. Probabilistic Selection Task

The probabilistic selection task consists of an acquisition phase and a test phase that implicitly measures how well the reinforcement contingencies have been learned (Frank et al., 2004). During every trial of the acquisition phase, participants see one of three different stimulus pairs characterized by different feedback probabilities in random order (pair “AB” – positive feedback ratio: 80/20; pair “CD” – positive feedback ratio: 70/30, pair “EF” – positive feedback ratio: 60/40). Participants must choose one stimulus from the presented pair, which is followed by positive (smiling cartoon face) or negative feedback (frowning cartoon face). During 300 trials, participants learn to choose stimuli

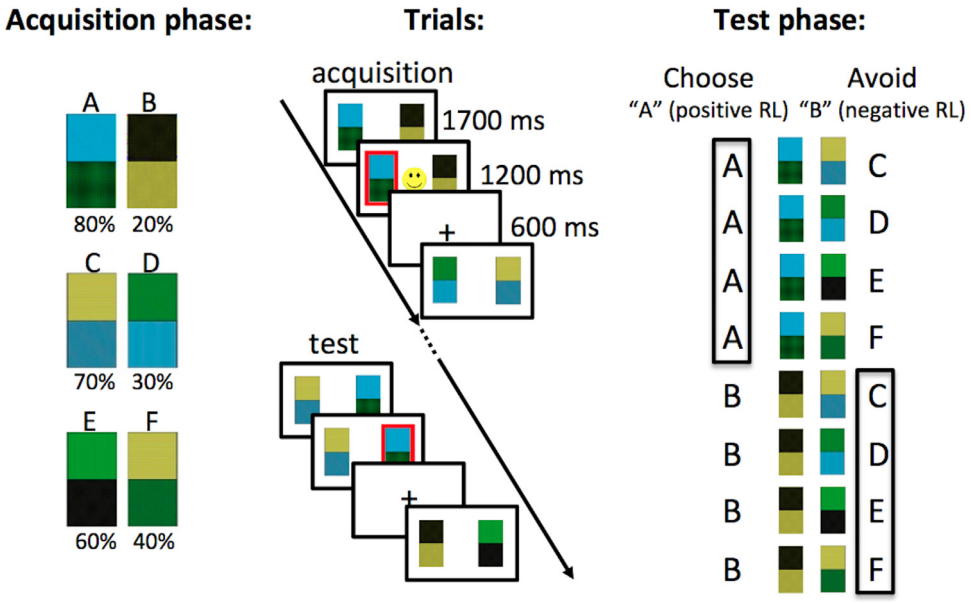

Fig. 1. Illustration of the task and stimuli.

that are probabilistically associated with positive feedback over the stimuli associated with negative feedback.

In the test phase, novel stimulus combinations are additionally presented to test the generalization of the previously learned stimulus-specific reinforcement contingencies (Fig. 1). Each pair is shown 12 times in random order without feedback. The frequency of A-choices in all 48 trials of the test phase containing "A" serves as a measure for the sensitivity to positive feedback/ reward since "A" had previously been followed by positive feedback with a probability of 80%. The frequency of B-avoidance in all 48 trials of the test phase containing "B" measures sensitivity to negative feedback since "B" had been previously followed by negative feedback with a probability of 80% (Fig. 1).

2.4. Computational modeling

Learning of reward contingencies was assessed through a reinforcement learning model (Chase et al., 2010; Frank et al., 2007; Solomon et al., 2011). The model uses reward prediction errors defined as the difference between the actual feedback and the expected feedback, to update reward expectations about a stimulus, also referred to as values. The magnitude of each value update depends on the model's learning rate parameters. Importantly, the present model distinguishes between learning from positive prediction errors meaning that the actual feedback is better than expected, and negative reward prediction errors meaning that the actual feedback is worse than expected, by introducing separate learning rates for each error type. Specifically, values are updated according to:

$$Q_i(t+1) = Q_i(t) + \alpha_+[r(t) - Q_i(t)]_+ + \alpha_-[r(t) - Q_i(t)]_-$$

where $Q_i(t)$ represents the value of stimulus i on trial t, $r(t)$ denotes the reward feedback (with $r = =1$ for positive, and $r = =0$ for negative feedback), and α_+, and α_- are the learning rates for outcomes that are better or worse than expected. Note that a strong focus on immediate rewards, indicated by high trial-by-trial fluctuations of expectancies and behavioral choices, would result in the estimation of a high learning rate. In contrast, low learning rates reflect more gradual but stable changes in values and have been related to higher performance during the test phase of this task (Frank et al., 2007).

Within the computational model, values (i.e., reward expectations) are translated into choices via a softmax function, which determines the probability for choosing stimulus i compared to the simultaneously presented stimulus j during any given trial as:

$$p(a_i(t)|Q(t)) = \frac{e^{\frac{Q_i(t)}{\beta_k}}}{e^{\frac{Q_i(t)}{\beta_k}} + e^{\frac{Q_j(t)}{\beta_k}}}$$

The parameter β_k governs the exploitation/exploration trade-off, where k indexes the acquisition or test phase respectively (Khamassi et al., 2015). Low β_k-values imply a strong tendency toward exploiting the current strategy (e.g. continue to choose stimulus "A"), while high β-values indicate exploration of both actions (e.g. choosing "A" and "B").

The reinforcement learning model applied in this study thus consists of four parameters. α_+ and α_- capture the immediate impact of positive and negative reward prediction errors on expectancies modeled by the algorithm. β_{acq} and β_{test} formalize the trade-off between exploiting a choice policy and exploring new strategies during either the acquisition or the test phase. These parameters are estimated for each subject separately based on its individual history of choices and feedbacks during acquisition and decisions during testing. To examine whether differences in learning were already present during the acquisition phase, we additionally estimated the model on the acquisition phase alone (Khamassi et al., 2015).

Estimation was performed by maximizing the log-likelihood $L(\theta) = \sum_t logp(a_i(t)|Q(t))$, where index t runs across all trials of either the entire data set or the acquisition phase only. Parameters were constrained with the range [0.01 1] and estimation was repeated for parameter initializations within this range using a step size of 0.05, to rule out local minima in the log-likelihood solution. The supplement provides information on additional performance measures previously reported for this task (S1).

2.5. Statistical analysis

All statistical analyses were performed using the Statistical Packages for Social Sciences version 23.0. Demographic data conforming to the assumptions of parametric analysis were analysed using Students t-test. For nominal demographic data Chi-square tests were computed.

We first conducted a multivariate analysis of variance (MANOVA) that included the learning parameters from the learning and test phase as well as the frequency of A-choices and B-avoidance as dependent variables and group as the independent variable for an overall

assessment of effects. This was followed by a multivariate analysis of covariance (MANCOVA) with age and task version as between-subjects covariates of no interest (Schutte et al., 2017).

To determine the magnitude of the effect in the different dependent variables, learning parameters from the computational approach were first contrasted between groups via Mann-Whitney U tests. Data from the test phase were initially analysed using Students *t*-test, followed by univariate analyses of covariance (ANCOVA) including age and task version as nuisance variables. For within-group differences concerning the choice of "A" or the avoidance of "B" repeated measures ANCOVAs with task version as a between-subjects covariate of no interest were computed. Across all ANOVAs, the Greenhouse-Geisser correction was used when applicable. Finally, correlations between learning parameters, and choice behavior during the task were analyzed with Spearman rank correlations within groups and compared using Fisher's z transformation. We report 95% confidence intervals and results were considered significant with $p < .05$.

3. Results

3.1. Validation of the computational model in the complete sample

In the whole sample, we observed that learning rates for unexpected positive (α_+) and negative (α_-) feedback were uncorrelated ($r_S = -.040$, $p = .637$). Further, learning rates for positive feedback were negatively correlated with "A" choices ($r_S = -.325$, $p < .001$) and learning rates for negative feedback were negatively correlated with avoidance of "B" ($r_S = -.23$, $p = .007$). This finding is consistent with previous reports that low learning rates, which indicate that behavior is only to a minor extent driven by immediate reward and punishment, facilitate the generalization of the learned stimulus-specific reinforcement contingencies probed during the test phase (Frank et al., 2007).

3.2. Group differences

The initial multivariate tests indicated a significant group effect in the data (MANOVA: $F_{(6,132)} = 2.47$; $p = .027$; $\eta^2 = 0.101$; MANCOVA: $F_{(6,132)} = 2.28$; $p = .040$; $\eta^2 = 0 = .095$). The initial set of post-hoc tests showed that REL and HV did not differ in any of the learning parameters α_+, α_-, β_{acq} and β_{test} estimated on both acquisition and test phase. A second model estimated on the acquisition data alone confirmed the absence of group differences (Table 2). However, when we added age and task version as nuisance variables, we observed marginally lower learning rates from unexpected negative feedback in REL ($F_{(1,135)} = 3.35$; $p = .069$; $\eta^2 = 0.024$).

During the test-phase, REL chose the most richly rewarded symbol "A" more consistently than HV. This result remained significant, when age and task version were added as covariates ($F_{(1,135)} = 6.71$; $p = .011$; $\eta^2 = 0.047$). There was no age effect, but task version appeared to have influenced the choice of 'A' ($F_{(1,135)} = 4.48$; $p = .036$; $\eta^2 = 0.032$; for details, see Supplement S2). There were no group differences in the avoidance of symbol "B", which was probabilistically associated with negative feedback, or the reaction times during the choice of "A" and the avoidance of "B" (Table 2).

Intra-group comparisons showed that HV ($t_{(96)} = 5.29$, $p < .001$), but not REL ($t_{(41)} = -0.25$, $p = .80$) avoided stimulus "B" significantly more often than they chose stimulus "A" (Fig. 2c). This effect remained when task version and age were added as nuisance variables (REL: $F_{(1,39)} = 0.03$, $p = .861$; HV: $F_{(1,94)} = 4.14$, $p = .045$).

In the whole sample, we also observed larger learning rates for unexpected positive feedback (α_+) than negative feedback (α_-) ($Z = 2.08$, $p = .038$; Table 3). This effect was observable in HV ($Z = 2.20$, $p = .028$), but absent in REL ($Z = 0.31$, $p = .76$). Given that small learning rates should yield greater transfer of learned reward-contingencies observable as higher accuracies in the test phase, HV should be more accurate in avoiding "B" than choosing "A". To test this, the difference of both learning parameters ($\alpha_+ - \alpha_-$) was correlated with the difference between "A"-choices and 'B'-avoidance, revealing a significant association ($r = -0.45$, $p < .001$, Fig. 2B).

Table 2
Parameters of the computational modeling approach and behavioral data from test phase for relatives, individuals with hypomanic temperament and healthy volunteers.

	Relatives ($n = 42$)	Controls ($n = 97$)	Statistic	*p*-value
α_+, median (SD)	0.10 (0.34)	0.08 (0.28)	$U = 2033$	.99
α_-, median (SD)	0.07 (0.34)	0.03 (0.22)	$U = 1739$	.16
β_{acqu}, median (SD)	0.26 (0.33)	0.32 (0.30)	$U = 1943$	.67
β_{test}, median (SD)	0.19 (0.37)	0.26 (0.29)	$U = 1792$	.26
Learning parameters estimated for the acquisition phase				
α_+, median (SD)	0.06 (0.28)	0.09 (0.27)	$U = 2003$	.87
α_-, median (SD)	0.12 (0.32)	0.05 (0.25)	$U = 1831$	.34
β, median (SD)	0.32 (0.30)	0.29 (0.29)	$U = 1991$	.84
Test phase				
Choice of "A", mean percent (SD)	71 (21)	58 (26)	$t_{(137)} = 2.90$	.004
Avoidance of "B", mean percent (SD)	71 (18)	77 (19)	$t_{(137)} = -1.56$	.12
"A"chosen in msec, mean (SD)	933 (375)	876 (288)	$t_{(137)} = 0.87$	.39
"B"avoided in msec, mean (SD)	944 (353)	834 (232)	$t_{(137)} = 1.84$	.07

Abbreviations: α_+, learning rate from unexpected positive feedback; α_-, learning rate from unexpected negative feedback; β_{acqu} exploitation/ exploration trade-off during the acquisition phase; β_{test} exploitation/ exploration trade-off during the test phase; msec, milliseconds, SD, standard deviation.

We also explored differences between individuals, who had more than one first-degree relative with BD, and those, who had only one case in their family, as the first might have a higher vulnerability for BD. However, there were no significant differences between these subgroups in the learning parameters (all *p*-values > .14) and the choice behavior during the test-phase (all *p*-values > .52).

3.3. BAS/ BIS scales

REL showed lower BAS-scores than HV (see Table 1). Across the entire sample, there was a significant negative association between BAS scores and learning rates from unexpected negative feedback (α_-) ($r_S = -0.21$, $p = .011$). This effect was driven by REL (REL: $r_S = -0.34$, $p = .029$; HV: $r_S = -0.07$, $p = .481$; $Z = -1.49$, $p = .14$; see Fig. 2). There were no additional significant associations between BAS or BIS scores and other parameters of the learning or test phase across the entire sample (all $r_S < 0.12$, $p > .14$).

3.4. Exploration of differential associations between learning parameters and performance during the test phase

While learning rates of positive feedback (α_+) were negatively related to the choice of "A" in both groups, learning rates of negative feedback (α_-) were positively associated with the choice of "A" in HV only (Table 3). This indicates that REL might rely more on positive feedback, which is consistent with the observation that REL rated the positive feedback as more positive than HV ($t_{(137)} = 2.22$, $p = .028$; for details, see Supplement S3).

Further, in HV only, avoidance of "B" was negatively associated with learning rates of negative feedback and positively related to learning rates of positive feedback (Table 3). This suggests a complex interaction of learning rate effects on avoidance of "B" in HV. However, in REL, avoidance of "B" was associated with the exploitation of the choice strategy (e.g., choosing "A") as opposed to exploring different options. Thus, REL might learn through positive prediction errors to

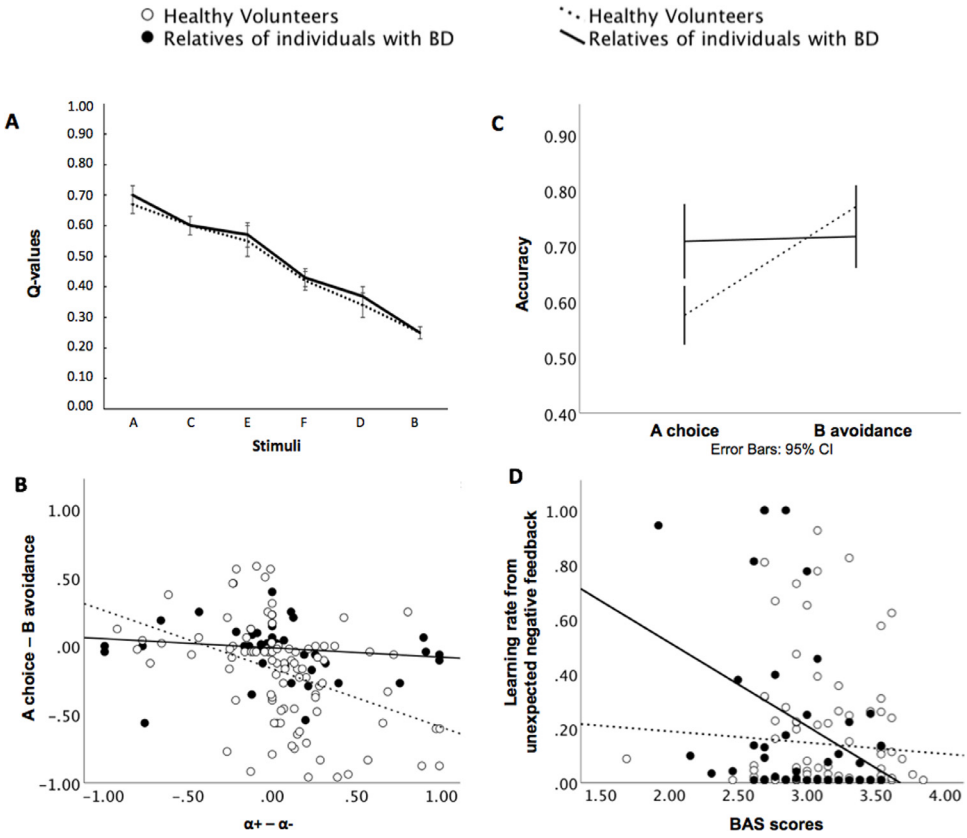

Fig. 2. Main results of computational reinforcement learning models and test phase. A Shows comparable learning of Q-values for each stimulus at the end of acquisition across groups. B Correlation between the difference scores of the learning parameters (α_+, α_-) and test parameters ("A" choices, "B" avoidance). C Mean percentage of accurately choosing the most often rewarded stimulus "A" and correctly avoiding the least often rewarded stimulus "B" separately for the two groups. Error bars represent standard errors. D Association between BAS scores and learning rate from unexpected negative feedback

Table 3
Correlations between parameters of the acquisition phase (learning from unexpected positive feedback [α_+], learning from unexpected negative feedback [α_-], exploitation/ exploration trade-off during acquisition phase [β_{acqu}], exploitation/ exploration trade-off during testing phase [β_{test}]) and test phase (choice of "A", avoidance of 'B') separately for both groups and comparison of correlations coefficients between groups.

	(r_S) Relatives	Controls	Relatives vs. Controls z	p
"A" chosen				
Learning parameters estimated for learning and test phase				
α_+	−0.39**	−0.30***	−0.54	.59
α_-	−0.07	0.30***	−1.99	.047
β_{acqu}	−0.54***	−0.40***	−0.95	.34
β_{test}	−0.43***	−0.29***	−0.85	.40
Learning parameters estimated for learning phase				
α_+	−0.16	−0.09	−00.37	.71
α_-	−0.17	0.15	−1.69	.09
β	−0.47***	−0.27***	−1.22	.22
"B" avoided				
Learning parameters estimated for learning and test phase				
α_+	−0.15	0.28***	−2.30	.02
α_-	−0.09	−0.23**	0.76	.45
β_{acqu}	−0.44***	0.06	−2.79	.005
β_{test}	−0.20	−0.20*	0	1
Learning parameters estimated for learning phase				
α_+	0.01	−0.00	0.05	.96
α_-	−0.12	0.09	−1.11	.27
β	−0.30†	−0.14	−0.89	.37

* $p < .10$.
** $p < .05$.
*** $p < .01$.

choose "A" and exploit this strategy thereby also avoiding "B" as both stimuli are paired during acquisition.

4. Discussion

The present study presents several new insights regarding the use of negative and positive feedback during reinforcement learning and subsequent approach and avoidance behavior associated with (a) the familial risk for BD and (b) elevated levels of BAS. Individuals with a familial risk for BD showed an increased approach of highly rewarding stimuli compared to HV and marginally reduced learning rates from unexpected negative feedback when controlling for age and task version. Higher BAS scores were also associated with reduced learning rates from unexpected negative feedback, and this association was more pronounced in REL. Exploratory analyses additionally showed differential associations between learning rate parameters and choice behavior during the test phase. While in HV the choice of "A" and the avoidance of "B" were related to both learning rates from positive and negative feedback, in REL, the choice of "A" was only related to learning from unexpected positive feedback in REL, and avoidance of "B" was unrelated to the learning rates.

4.1. Differences between first-degree relatives and healthy volunteers

Consistent with previous reports of increased reward sensitivity (Linke et al., 2012; Singh et al., 2014) in REL, we observed a more consistent choice of the most highly rewarding stimulus "A". Learning rates from unexpected positive feedback were comparable between groups and negatively associated with the choice of "A" in both groups. However, we observed marginally lower learning rates from unexpected negative feedback in REL. Moreover, while HV showed a positive association between learning rates from unexpected negative feedback and the choice of "A", no such relationship could be observed in REL. Of note, low learning rates from unexpected positive feedback

have been previously associated with a more consistent choice of "A" during the test phase (Frank et al., 2007). The striatum, which is thought to integrate the long-term probabilities of positive and negative outcomes continuously (Jog et al., 1999), is thought to be the neurobiological correlate of this slow habitual learning. In contrast, prefrontal cortical regions might allow learning on a shorter timescale by actively maintaining recent reinforcement experiences in a working memory-like state (Frank and Claus, 2006). These representations in the prefrontal cortex are thought to modify ongoing behavior by top-down influences on subcortical structures (Miller and Cohen, 2001), a process that would be reflected in high learning rates. So, in HV, behavior appears to be modified by immediate negative consequences, whereas REL consider unexpected negative feedback less for their choices.

Further, it is of interest that, in REL, the avoidance of the least rewarding stimulus "B" was not associated with learning rates but with the exploitation of the current strategy (i.e., continue to choose stimulus "A"). So, it is possible that REL only indirectly learn to avoid "B" as "A" and "B" are always presented together during the acquisition phase, which is consistent with the observation of comparable accuracies in choosing "A" and avoiding "B" during the test phase. Future studies are needed to replicate this finding.

4.2. BAS sensitivity and reinforcement learning

As a group, REL showed lower BAS scores than HV. However, in REL, we observed a significant negative association between BAS scores and the learning rate from negative feedback (α-). In other words, REL with high BAS scores showed particularly low learning rates from negative feedback. Lower learning rates for negative than positive feedback are thought to promote overly positive expectations, which will result in more frequent negative surprises leading to a more depressed mood state (Eldar and Niv, 2015). It has been further hypothesized that during this depressed mood state, more pessimistic expectations might develop that once the mood returns to baseline increase the likelihood for positive surprises leading to an elevated mood state (Eldar and Niv, 2015). Thus, REL with a hypersensitive BAS might be at an increased risk to develop an affective disorder themselves, whereas low BAS scores might indicate resilience. Future studies should test this hypothesis with a longitudinal design.

Of note, previous reports showed that high BAS scores are associated with a disequilibrium between the availability of the enzyme COMT and dopamine-D2 receptor density (Reuter et al., 2006) both relevant for learning from negative feedback (Maia and Frank, 2011). Notably, the enzyme COMT, which catalyzes the degradation of dopamine (Mannisto and Kaakkola, 1999) has been associated with bipolar disorder. Previous research focused on rs4680, a single nucleotide polymorphism of the COMT gene characterized by a substitution of guanine by adenine resulting in a valine (Val)-to-methionine (Met) transition, which leads to a significant reduction of catalytic activity in vitro (Moskovitz et al., 2015). Of note, the *Met* allele appears to confer elevated risk for mania (Goghari and Sponheim, 2008; Zhang et al., 2009) and higher severity of manic symptoms in BD (Benedetti et al., 2010, 2011; Lelli-Chiesa et al., 2011; Soeiro-de-Souza et al., 2012), but also independent of this genotype, COMT activity in the striatum might be related to the severity of manic symptoms (Bortolato et al., 2017). Thus, future studies might want to test, whether the negative association between BAS and learning from unexpected negative feedback is mediated by the COMT polymorphism.

4.3. Intragroup differences

Although learning rates for unexpected positive and negative feedback are uncorrelated, our findings implicate that avoidance and approach behavior (differentially) depend on both learning rates for unexpected positive and negative feedback, such that it may be necessary to consider their relation to one another. Indeed, we observed that HV exhibited larger learning rates from positive than from negative feedback, while REL did not. This difference predicted higher relative performance in the avoidance of "B" compared to the choice of "A". This imbalance between learning rates may account for the fact that HV show better avoidance as compared to approach performance while REL show similar performance in these measures.

Besides, we like to point out that the bias to avoid "B" more reliably than to choose "A", which we observed in HV is unusual as HV are often reported to develop no bias (Frank et al., 2004). However, HV might show biased choices in the test phase depending on their genotype (Frank et al., 2007) and the properties of the stimulus material (Schutte et al., 2017). While we do not have information regarding the genotype, we provide additional analyses regarding the stimulus material in Supplement S2 (available online) to guide future studies, which might investigate whether individual differences in the preference for certain stimuli differentially interact with reinforcement learning in individuals at risk for BD.

4.4. Limitations

Some aspects of this study limit the scope of conclusions. First, we did not include a group of patients with BD, which would have been essential to parse risk versus resilience markers. Further, we did not correct for multiple testing; and therefore, especially the results regarding the differential associations between parameters from the acquisition and test phase as well as between parameters of the computational modeling approach and the test phase must be cautiously interpreted, as strong a-priori hypotheses did not guide them. We decided to report these results because we believe that they aid in the interpretation of our main findings. Moreover, we are unable to determine whether the increased approach of highly rewarded stimuli is specific for BD or might also increase the risk for ADHD or substance use disorders. To ultimately judge the significance of our findings, additional studies of motivational processes and reward learning are warranted that compare BD patients and individuals with an increased risk for BD as well as other diagnostic groups with partially overlapping neurobiology such as ADHD.

5. Conclusion

Individuals with a family history of BD chose positive stimuli more reliably than HV and showed marginally lower learning rates from unexpected negative feedback and lower BAS scores. Notably, BAS scores and learning rates from unexpected negative feedback were negatively associated; and this effect was particularly strong among REL. Exploratory analyses on associations between learning rate parameters, subsequent choice behaviour and BAS scores indicate that future studies should focus on moderators of individual differences in reinforcement learning in individuals at risk for BD.

Funding

German Research Foundation (SFB636/C6, WE3638/3-1, DU 354/8-2 within the SPP-1665).

CRediT authorship contribution statement

Julia O. Linke: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Validation, Visualization, Writing - original draft. **Georgia Koppe:** Formal analysis, Methodology, Visualization, Writing - original draft. **Vanessa Scholz:** Methodology, Writing - original draft. **Philipp Kanske:** Conceptualization, Writing - original draft. **Daniel Durstewitz:** Methodology, Writing - original draft. **Michèle Wessa:** Conceptualization, Funding acquisition, Project administration,

Resources, Supervision, Validation, Writing - original draft.

Declaration of Competing Interest

None of the authors has any conflict of interest.

Acknowledgements

We wish to thank Eliza Eckhardt for her assistance in data acquisition.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.jad.2019.11.063.

References

Alloy, L.B., Bender, R.E., Whitehouse, W.G., Wagner, C.A., Liu, R.T., Grant, D.A., Jager-Hyman, S., Molz, A., Choi, J.Y., Harmon-Jones, E., Y., A.L., 2012a. High behavioral approach system (BAS) sensitivity, reward responsiveness, and goal-striving predict first onset of bipolar spectrum disorders: a prospective behavioral high-risk design. J. Abnorm. Psychol. 12, 339–351.

Alloy, L.B., Urosevic, S., Abramson, L.Y., Jager-Hyman, S., Nusslock, R., Whitehouse, W.G., Hogan, M., 2012b. Progression along the bipolar spectrum: a longitudinal study of predictors of conversion from bipolar spectrum conditions to bipolar I and II disorders. J. Abnorm. Psychol. 121, 16–27.

American Psychiatric Association, 2013. Diagnostic and Statistical Manual of Mental Disorders, fifth ed. American Psychiatric Publishing, Arlington, VA.

Benedetti, F., Dallaspezia, S., Colombo, C., Lorenzi, C., Pirovano, A., Smeraldi, E., 2010. Association between catechol-O-methyltransferase Val(108/158)Met polymorphism and psychotic features of bipolar disorder. J. Affect. Disord. 125, 341–344.

Benedetti, F., Dallaspezia, S., Locatelli, C., Radaelli, D., Poletti, S., Lorenzi, C., Pirovano, A., Colombo, C., Smeraldi, E., 2011. Recurrence of bipolar mania is associated with catechol-O-methyltransferase Val(108/158)Met polymorphism. J. Affect. Disord. 132, 293–296.

Bortolato, M., Walss-Bass, C., Thompson, P.M., Moskovitz, J., 2017. Manic symptom severity correlates with COMT activity in the striatum: a post-mortem study. World J. Biol. Psychiatry 18, 247–254.

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. Psychol. Med. 40, 433–440.

Depue, R.A., Iacono, W.G., 1989. Neurobehavioral aspects of affective disorders. Ann. Rev. Psychol. 40, 457–492.

Eldar, E., Niv, Y., 2015. Interaction between emotional state and learning underlies mood instability. Nat. Commun. 6, 6149.

Frank, M.J., Claus, E.D., 2006. Anatomy of a decision: striato-orbitofrontal interactions in reinforcement learning, decision making, and reversal. Psychol. Rev. 113, 300–326.

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. Proc. Natl. Acad. Sci. U. S. A. 104, 16311–16316.

Frank, M.J., Seeberger, L.C., O'Reilly, R.C., 2004. By carrot or by stick: cognitive reinforcement learning in parkinsonism. Science 306, 1940–1943.

Ganzeboom, H.B.G., De Graaf, P.M., Treiman, D.J., 1992. A standard international socio-economic index of occupational status. Soc Sci Res 21, 1–56.

Goghari, V.M., Sponheim, S.R., 2008. Differential association of the COMT Val158Met polymorphism with clinical phenotypes in schizophrenia and bipolar disorder. Schizophr. Res. 103, 186–191.

Gray, J.A., 1987. The Psychology of Fear and Stress. Cambridge University Press, Cambridge.

Jog, M.S., Kubota, Y., Connolly, C.I., Hillegaart, V., Graybiel, A.M., 1999. Building neural representations of habits. Science 286, 1745–1749.

Khamassi, M., Quilodran, R., Enel, P., Dominey, P.F., Procyk, E., 2015. Behavioral regulation and the modulation of information coding in the lateral prefrontal and cingulate cortex. Cereb. Cortex 25, 3197–3218.

Lehrl, S., 2002. Brickenkamp Handbuch Psychologischer Tests, third ed. Hogrefe, Göttingen.

Lehrl, S., 2005. Mehrfachwahl-Wortschatz-Intelligenztest MWT-B, fifth ed. Spitta Verlag, Balingen.

Lelli-Chiesa, G., Kempton, M.J., Jogia, J., Tatarelli, R., Girardi, P., Powell, J., Collier, D.A., Frangou, S., 2011. The impact of the Val158Met catechol-O-methyltransferase genotype on neural correlates of sad facial affect processing in patients with bipolar disorder and their relatives. Psychol. Med. 41, 779–788.

Linke, J., King, A.V., Rietschel, M., Strohmaier, J., Hennerici, M., Gass, A., Meyer-Lindenberg, A., Wessa, M., 2012. Increased medial orbitofrontal and amygdala activation: evidence for a systems-level endophenotype of bipolar I disorder. Am. J. Psychiatry 169, 316–325.

Maia, T.V., Frank, M.J., 2011. From reinforcement learning models to psychiatric and neurological disorders. Nat. Neurosci. 14, 154–162.

Mannisto, P.T., Kaakkola, S., 1999. Catechol-O-methyltransferase (COMT): biochemistry, molecular biology, pharmacology, and clinical efficacy of the new selective COMT inhibitors. Pharmacol. Rev. 51, 593–628.

McGuffin, P., Rijsdijk, F., Andrew, M., Sham, P., Katz, R., Cardno, A., 2003. The heritability of bipolar affective disorder and the genetic relationship to unipolar depression. Arch. Gen. Psychiatry 60, 497–502.

Meyer, B., Johnson, S.L., Winters, R., 2001. Responsiveness to threat and incentive in bipolar disorder: relations of the BIS/BAS scales with symptoms. J. Psychopathol. Behav. Assess 23, 133–143.

Miller, E.K., Cohen, J.D., 2001. An integrative theory of prefrontal cortex function. Annu. Rev. Neurosci. 24, 167–202.

Montague, P.R., Dayan, P., Sejnowski, T.J., 1996. A framework for mesencephalic dopamine systems based on predictive Hebbian learning. J. Neurosci. 16, 1936–1947.

Moskovitz, J., Walss-Bass, C., Cruz, D.A., Thompson, P.M., Hairston, J., Bortolato, M., 2015. The enzymatic activities of brain catechol-O-methyltransferase (COMT) and methionine sulphoxide reductase are correlated in a COMT Val/Met allele-dependent fashion. Neuropathol. Appl. Neurobiol. 41, 941–951.

Paaren, A., Bohman, H., von Knorring, L., Olsson, G., von Knorring, A.L., Jonsson, U., 2014. Early risk factors for adult bipolar disorder in adolescents with mood disorders: a 15-year follow-up of a community sample. BMC Psychiatry 14, 363.

Retz-Junginger, P., Retz, W., Blocher, D., Weijers, H.-G., Trott, G.-E., Wender, P.H., Rösler, M., 2002. Wender utah ratin scale (WURS-k). Nervenarzt 73, 830–838.

Reuter, M., Schmitz, A., Corr, P., Hennig, J., 2006. Molecular genetics support Gray's personality theory: the interaction of COMT and DRD2 polymorphisms predicts the behavioural approach system. Int. J. Neuropsychopharmacol. 9, 155–166.

Schutte, I., Slagter, H.A., Collins, A.G.E., Frank, M.J., Kenemans, J.L., 2017. Stimulus discriminability may bias value-based probabilistic learning. PLoS One 12, e0176205.

Singh, M.K., Chang, K.D., Kelley, R.G., Saggar, M., Reiss, A.L., Gotlib, I.H., 2014. Reward processing in healthy offspring of parents with bipolar disorder. JAMA 71, 1148–1156.

Soeiro-de-Souza, M.G., Machado-Vieira, R., Soares, B., D., Do Prado, C.M., Moreno, R.A., 2012. COMT polymorphisms as predictors of cognitive dysfunction during manic and mixed episodes in bipolar I disorder. Bipolar Disord. 14, 554–564.

Solomon, M., Smith, A.C., Frank, M.J., Ly, S., Carter, C.S., 2011. Probabilistic reinforcement learning in adults with autism spectrum disorders. Autism Res. 4, 109–120.

Strobel, A., Beauducel, A., Debener, S., Brocke, B., 2001. Eine deutschsprachige version des BIS/BAS-Fragebogens von Carver und White. Z. Differ. Diagn. Psychol. 22, 216–227.

Wechsler, D., 1997. Wechsler Adult Intelligence Scale, third ed. Harcourt Assessment, San Antonio, TX.

Wessa, M., Kanske, P., Linke, J., 2013. Bipolar disorder: a neural network perspective on a disorder of emotion and motivation. Restor. Neurol. Neuroscie.

Wittchen, H.U., Zaudig, M., Fydrich, T., 1997. Strukturiertes Klinisches Interview für DSM-IV [Structural Clinical Interview for DSM-IV Axis I Disorders]. Hogrefe, Göttingen.

Zhang, Z., Lindpaintner, K., Che, R., He, Z., Wang, P., Yang, P., Feng, G., He, L., Shi, Y., 2009. The Val/Met functional polymorphism in COMT confers susceptibility to bipolar disorder: evidence from an association study and a meta-analysis. J. Neural. Transm. 116, 1193–1200.