Impaired Punishment Learning in Conduct Disorder

Elster, Erik; Pauli, Ruth; Baumann, Sarah; De Brito, Stephane; Fairchild, Graeme; Freitag, Christine M.; Konrad, Kerstin; Roessner, Veit; Brazil, Inti A; Lockwood, Patricia; Kohls, Gregor

DOI:
10.1016/j.jaac.2023.05.032

Impaired Punishment Learning in Conduct Disorder

Abstract

Objective: Conduct disorder (CD) has been associated with deficits in the use of punishment to guide reinforcement learning (RL) and decision making. This may explain the poorly planned and often impulsive antisocial and aggressive behavior in affected youths. Here, we used a computational modeling approach to examine differences in RL abilities between CD youths and typically developing controls (TDCs). Specifically, we tested two competing hypotheses that RL deficits in CD reflect either reward dominance (also known as reward hypersensitivity) or punishment insensitivity (also known as punishment hyposensitivity).

Method: The study included 92 CD youths and 130 TDCs (ages 9-18, 48% girls) who completed a probabilistic RL task with reward, punishment, and neutral contingencies. Using computational modeling, we investigated the extent to which the two groups differed in their learning abilities to obtain reward and/or avoid punishment.

Results: RL model comparisons showed that a model with separate learning rates per contingency explained behavioral performance best. Importantly, CD youths showed lower learning rates than TDCs specifically for punishment, whereas learning rates for reward and neutral contingencies did not differ. Moreover, callous-unemotional (CU) traits did not correlate with learning rates in CD.

Conclusion: CD youths have a highly selective impairment in probabilistic punishment learning, regardless of their CU traits, while reward learning appears to be intact. In summary, our data suggest punishment insensitivity rather than reward dominance in CD. Clinically, the use of punishment-based intervention techniques to achieve effective discipline in patients with CD may be a less helpful strategy than reward-based techniques.

Introduction

Conduct disorder (CD) is a common psychiatric disorder in youths that is characterized by severe antisocial and aggressive behavior.[1] CD has long been hypothesized to be linked to deficits in reinforcement learning (RL), which may contribute to impaired social functioning, leading to CD behaviors and reduced quality of life.[2] In short, RL describes an individual's ability to learn the relationship between a particular stimulus, an action (i.e., a behavioral reaction), and a rewarding or punishing outcome conditional on the individual's action.[3] Accumulating evidence indicates that deficient RL in CD may partly be due to a problem in generating accurate estimations about the value of potential behavioral outcomes, such as punishment,[4] and that this may explain why youths with CD tend to make bad behavioral choices (e.g., decisions that lead to punishment rather than reward). In fact, deficits, particularly a failure to learn how to avoid choices that lead to punishment rather than reward, have consistently and repeatedly been shown in youths with CD.[4–6]

One theoretical explanation for this learning failure might be reward dominance (also referred to as reward hypersensitivity in recent literature).[7] For instance, O'Brien and Frick[8] used a probabilistic reward dominance task in which participants were asked to press a response key to see the reverse side of a stimulus, resulting in either a reward (gaining points) or a punishment (losing points), or they chose to quit the task and exchange the points earned for a prize. The ratio between reward and punishment changed with each 10 trials played, starting with a 90% chance of winning a reward and ending with 0% after 100 trials. Consequently, punishment eventually becomes dominant, and prolonged play is detrimental. Youths with conduct problems played more trials in this task than typically developing controls (TDC), suggestive of a strong tendency towards rewarding cues at the expense of punishing ones. However, because reward and punishment are presented intermixed within the task, it is not possible to clearly discern whether decisions in the CD group are due to atypical processing of reward, punishment, or both. Additionally, there is debate within the literature that aberrant reward processing, when present, may not be directly related to CD, but to other externalizing disorders that often accompany CD, such as ADHD.[9] For example, as with ADHD, behavioral studies have revealed that youths with CD problems prefer larger, immediate rewards while accepting the risk of loss.[10] Furthermore, neuroimaging studies in CD youths show rather inconsistent abnormalities in reward processing tasks with either hyper- or hypoactivation of reward-related brain circuits.[5] Taken together, the available literature provides a rather mixed picture regarding aberrant reward processing in CD.

Alternatively, but not necessarily mutually exclusive, youths with CD may show a primary deficit in punishment processing, such that they are less sensitive to cues of punishment and have difficulties learning from such cues.[5] According to Blair,[11] disrupted punishment processing in CD is limited to the use of punishment information in stimulus-reinforcement formation, which is learning to associate an aversive value with a particular stimulus. One prominent example of a stimulus-RL task is the passive avoidance learning (PAL) task, in which individuals learn through trial-and-error that a particular stimulus associated with a punishment (losing points) is 'bad' and should be avoided (by not pressing a response button), whereas the stimulus associated with a reward (gaining points) is 'good' and should be approached (pressing a response button). In a recent behavioral study using the PAL task in the largest sample of CD youths to date, more performance errors were found in the CD group in responding to punishment (i.e., difficulty in avoiding pressing the response button) but not in responding to reward contingencies, when compared to TDCs,[6,12] suggesting possible punishment-specific learning differences. However, computational models to precisely quantify learning ability were not applied in this particular study.

As 'learning' is a latent operation, it is best quantified directly and more precisely using computational RL models.[13] Such models are able to capture the trial-by-trial dynamics of the learning process as it unfolds over time – in contrast to 'traditional' indices, like accuracy, that provide only a rather coarse summary performance measure. Here, we used a prediction error based RL model[14] to investigate differences between learning performances in a well-powered sample (48% female participants) of youths with CD versus TDCs. This RL model is able to estimate learning rates α and exploration parameters β for each participant individually – two crucial computational indices underlying learning in choice situations (as in the probabilistic RL task used here). α reflects how quickly participants update their estimations of a particular outcome (i.e., reward, punishment, or neutral) by newer information from trial to trial, and β reflects the noisiness (or inconsistency) in picking the stimulus with the higher expected value while learning (higher β = more random choices of the best option).[15] Regarding α, a higher learning rate indicates a quicker and more efficient updating by more recent outcomes compared to older ones; it therefore combines temporal integration (i.e., reinforcer history) with reinforcer valuation.[16] Consistent with the relevant literature, we operationalized that a lower or higher learning rate per reinforcer reflects a lower or higher sensitivity to that particular reinforcer (e.g., reward, or punishment). We therefore chose α as our main computational learning index of interest.

Considering the 'reward dominance (or hypersensitivity)' hypothesis, one would predict that, compared to TDCs, youths with CD show a different pattern of learning particularly in the reward condition (i.e., higher learning rate α for reward), while considering 'punishment insensitivity (or hyposensitivity)' one would expect to find a learning deficit particularly in the punishment condition (e.g., lower learning rate α for punishment). Moreover, because some research implies a greater learning impairment particularly for punishment among CD youth with high callous-unemotional (CU) traits (i.e., reduced guilt and empathy, callousness, and uncaring attitudes),[17] we predicted a positive association between CU traits and aberrant punishment learning performance. Because the majority of relevant studies to date have investigated predominantly male- or female-only samples of youths with CD and/or ODD, but sex differences may or may not exist,[6] we also tested for sex-by-group interaction effects. However, we did not have a directional hypothesis for this analysis. Finally, we investigated whether there were group differences in the exploration parameter β with no directional hypothesis.

Method

Participants

248 participants, 9-18 years of age, were recruited through community outreach, mental health clinics and youth welfare institutions in Aachen (Germany) and Southampton (UK) as part of the FemNAT-CD study.[18] We excluded 26 individuals (11 CDs and 15 TDCs), because too many responses were missing in the experimental task. This left a final sample of 222 participants (Aachen n=112; Southampton n=110) including 92 youths with CD (37 girls) and 130 TDCs (69 girls). Exclusion criteria were autism spectrum disorder, psychosis or schizophrenia, mania or bipolar disorder, genetic syndromes, neurological disorders, and an IQ<70. The study protocol was approved by local ethics committees, and participants and their caregivers gave written informed consent. Participants were compensated for their participation, including the money they gained during the task.

The CD group had a current diagnosis of CD, and the TDCs had no current psychiatric diagnoses and no lifetime diagnoses of CD, ODD and ADHD. All diagnoses, including comorbidities, (or lack thereof) were based on DSM-IV-TR criteria[19] assessed with the Kiddie-Schedule for Affective Disorders and Schizophrenia for School-Age Children – Present and Lifetime Version (K-SADS-PL).[20] Full-scale IQs were estimated using the vocabulary and matrix reasoning subtests of the Wechsler Intelligence Scale for Children-Fourth Edition, the Wechsler Intelligence Scale for Adults-Fourth Edition,[21,22] or the

Wechsler Abbreviated Scale of Intelligence.[23] CU traits were assessed dimensionally with the total score of the subscales 'remorselessness', 'callousness' and 'unemotionality' of the self-reported Youth Psychopathic traits Inventory (YPI).[24] We also used the three CU traits subscales of the YPI to create a proxy for the "with limited prosocial emotions" (LPE) specifier in the DSM-5/ICD-11, following the procedure developed by Colins and Vermeiren.[25] A participant was considered to meet criteria for one of the CU traits when she/he reported that at least one item on the corresponding subscale applied "very well" to her/him [i.e., a score of 4 on a 4-point Likert scale, ranging from "Does not apply at all" (1) to "Applies very well" (4)]. Participants were considered to meet criteria for the LPE specifier if two or more CU traits were endorsed to threshold.

Groups did not differ in sex distribution, but age, IQ, CU traits as well as the presence of the LPE specifier differed between the groups, with the CD group being slightly older, having a lower IQ, higher CU traits and met more often the criteria for the LPE specifier than the TDCs (Table 1, and Table S3, available online, for sensitivity analyses).

[Table 1]

Task

We used a monetary probabilistic RL task[26] with reward, punishment, and neutral contingencies (Figure 1). Trials started with the presentation of a pair of cues (i.e., fractals) side-by-side. Each pair marked the onset of one of three conditions: reward (i.e., monetary gain), punishment (i.e., monetary loss), and neutral outcome (i.e., neither monetary gain nor loss). Participants were instructed to select one of the two cues by pressing the left or right key on a button box. The chosen cue increased in brightness and was followed by visual feedback 4s later, indicating whether participants received a reward (a picture of a 20 euro cent/20 pence coin, and the description: "You won 20 cent/20 pence"), a punishment (a picture of a 20 euro cent/20 pence coin overlaid with a red cross, and the description: "You lost 20 cent/20 pence”), a neutral outcome (a picture of a scrambled 20 euro cent/20 pence coin, and the description: “No change”), or nothing (a crosshair).

On each trial, participants could either select a high probability or a low probability cue. In reward trials, choosing the high probability cue either resulted in reward (+0.20 €/£) with a 70% probability, or in no feedback (i.e., no reward=crosshair) with a 30% probability. Conversely, choosing the low probability cue either resulted in reward with a 30% probability, or in no feedback with a 70% probability. In punishment trials, choosing the high probability cue either resulted in no feedback (i.e., no punishment=crosshair) with a 70% probability, or in punishment (-0.20 €/£) with a 30% probability. Conversely, choosing the low probability cue either resulted in no feedback with a 30% probability, or in punishment with a 70% probability. In neutral trials, participants either had a 70% or 30% chance of obtaining a neutral outcome (a scrambled coin), thereby receiving no feedback on the remaining trials.

The task was split into three runs with short breaks in-between. Each run had 45 trials (i.e., 15 trials per condition). The whole procedure lasted ~25 minutes. The order of trials was pseudo-randomized to ensure that the same condition was never presented twice or more consecutively and that all conditions were tested equally in total.

Prior to the task, participants were told that they would see three different pairs of unfamiliar cues, and on each trial, they had to choose one out of the two cues. Depending on their choices, they would win money, lose money, obtain a neutral outcome, or receive no feedback. The assignment of the three fractal pairs to the different conditions was counterbalanced across participants. It was explicitly stressed that they should try to win as much money as possible by always choosing the high probability cue. Each participant started the experiment with a fixed amount of money, and was told that any wins or losses would be added or subtracted, respectively, from this total.

[Figure 1]

Computational RL Modeling and Parameter Analyses

We used computational RL modeling to estimate the extent to which participants learned from different contingencies during the probabilistic RL task applying the Rescorla-Wagner learning rule.[13,14] The used here model is able to estimate learning rates α and exploration parameters β for each participant individually. These values are indices on how quickly participants update their estimations of a particular outcome (i.e., reward, punishment, or neutral), and the noisiness (or inconsistency) in picking the stimulus with the higher expected value, respectively. Hence, the learning rate α represents the speed at which an individual updates the expected outcome by new, more recent information (i.e., higher α=quicker update). The exploration parameter β represents an individual's random choices or invariance in choice behavior (i.e., higher β=more random choices). See also the introduction for more information on these computational indices.

Initially, we set up four possible candidate models, which we compared to determine which model fits the participants' choice behavior best. The models varied in terms of the number of learning rates and exploration parameters for the three different task conditions (i.e., shared or separate learning rates/exploration parameters). For model comparison, we calculated the Laplace approximation of the log model evidence (more positive values indicating a better model fit[27]) in a random-effects analysis using the spm_BMS routine (revision 7487). This calculates the exceedance probability, i.e., the posterior probability that each model is the most likely. An exceedance probability greater than 0.95 provides strong evidence for the best-fitting model. We also calculated the integrated Bayesian Information Criterion score (BIC_{int}) and R^2 for each model as additional measures of model fit. The BIC_{int} penalizes more complex models and indicates a better performance when BIC_{int} scores are lower. R^2 indicates which percentage of the variance can be explained by a model. The four candidate models were constructed as follows:

1. αβ: single learning rate α and single exploration parameter β for all conditions
2. 2α2β: combined reward and punishment α and β, neutral α and β
3. 3α3β: reward α and β, punishment α and β, neutral α and β
4. 3α1β: reward α, punishment α, neutral α and a single β for all conditions

We found that model 3 (i.e., 3α3β), which included separate learning rates and exploration parameters for each contingency, most accurately captured the learning behavior underlying the choices made by each participant (Figure 2). Of the four models, this model had the highest exceedance probability (> .99), the highest LME (-16681.96), and the lowest BIC_{int} value (33164). We further validated the winning model using parameter recovery and model identifiability procedures (see Supplement 1, Table S1 and Figure S1, available online).

[Figure 2]

The modeled parameters α and β from our winning model (3α3β) were then analyzed using two separate repeated-measures ANOVA (rmANOVA) models with group (CD vs. TDC), sex (male vs. female) and CU traits (LPE specifier present vs. absent) as between-subjects factor, and condition (reward vs. punishment vs. neutral) as within-subjects factor, followed by Holm-corrected post-hoc pairwise comparisons in case of significant main or interaction effects. As age and IQ did not significantly correlate with the dependent measures, these variables were not included as covariates in the main analyses. We also estimated correlations between CU traits and model parameters α and/or β in case there were between-group differences in any of these indices. The alpha level was set at 0.05. Effect sizes were calculated using partial eta squared (η^2_p), where 0.01, 0.06, and 0.14 represent small, medium and large effects, respectively, and Cohen's *d*, where 0.2, 0.5 and 0.8 represent small, medium and large effects, respectively. Analyses were conducted in R with RStudio (version 4.0.4) and the rstatix package. Bayes factors for non-significant results (i.e., BF_{01}) and Bayes factors for significant results (i.e., BF_{10}) were calculated in JASP (v 0.14) with the default prior. BF_{01} corresponds to how many times more likely the data are under the null hypothesis of no difference than under the alternative hypothesis that there is a difference. BF_{10} corresponds to how many times more likely the data are under the alternative hypothesis than under the null hypothesis. A $BF_{01}>3$ is considered substantial evidence in favor of the null hypothesis, while a $BF_{10}>3$ is considered substantial evidence in favor of the alternative hypothesis. A BF_{01} or BF_{10} between 1/3 and 3 indicates the data cannot clearly differentiate between the two hypotheses.[28]

Results

Differences in Learning Rates α

The rmANOVA for the learning rates α revealed a significant group by condition interaction effect [$F(2, 428)=6.15$, $p<.01$, $\eta^2_p=.03$]. The Holm-adjusted post-hoc comparisons revealed a significant lower punishment α in the CD group than the TDCs ($M_{Diff}=-0.09$, 95%-CI[-0.18, 0.01], $p=.04$, $BF_{10}=3.31$), but not group differences in reward α ($M_{Diff}=0.06$, $p=.24$, $BF_{01}=1.44$) and neutral α ($M_{Diff}=-0.02$, $p>.99$, $BF_{01}=5.16$). The sex by condition interaction effect ($p=.60$, $\eta^2_p<.01$) and the CU traits by condition interaction ($p=.25$, $\eta^2_p<.01$) were non-significant. All other interaction effects were non-significant as well ($ps>.26$, $\eta^2_ps<.01$). Additionally, we found a significant main effect of condition [$F(2, 428)=42.61$, $p<.001$, $\eta^2_p=.17$], but no significant main effects of group ($p=.48$, $\eta^2_p<.01$) or sex ($p=.58$, $\eta^2_p<.01$) or CU traits ($p=.72$, $\eta^2_p<.01$). The Holm-adjusted post-hoc comparisons for the condition effect showed higher punishment α compared to reward α and to neutral α ($M_{Diff(rew-pun)}=-0.17$, 95%-CI[-0.21, -0.13], $p<.001$, $BF_{10}>100$; $M_{Diff(pun-neut)}=0.17$, 95%-CI[0.13, 0.21], $p<.001$, $BF_{10}>100$), while reward α and neutral α did not differ ($M_{Diff(rew-neut)}<0.01$, 95%-CI[-0.04, 0.03], p=.91, BF_{01}=13.23). Finally, to confirm that a higher learning rate α was associated with better task performance (i.e., accuracy of choosing the high probability cue; see Supplement 1 and Figure S2, available online), we calculated a mean correlation between α and performance in the reward and punishment condition across both groups, which revealed a significant, moderate-sized positive association of r=.55 (p<.001).

[Figure 3]

Differences in Exploration Parameters β

The rmANOVA for the exploration parameters β revealed a significant main effect of condition [F(2, 428)=53.5, p<.001, η^2_p=.20], but no significant effects of group (p=.40, η^2_p<.01) or sex (p=.64, η^2_p<.01) or CU traits (p=.45, η^2_p<.01). The Holm-adjusted post-hoc comparisons for the condition effect showed a lower reward β compared to punishment β and neutral β ($M_{Diff(rew\text{-}pun)}$=-0.22, 95%-CI[-0.25, -0.19], p<.001, BF_{10}>100; $M_{Diff(rew\text{-}neut)}$=-0.21, 95%-$CI$[-0.26, -0.16], p<.001, BF_{10}>100), while punishment β and neutral β did not differ ($M_{Diff(pun\text{-}neut)}$=0.01, 95%-$CI$[-0.04, 0.05], p>.99 , BF_{01}=12.43). The group by condition interaction (p=.46, η^2_p<.01), the sex by condition interaction (p=.96, η^2_p<.01) and the CU traits by condition interaction (p=.98, η^2_p<.01) were non-significant. All other interaction effects were non-significant as well (ps>.16, η^2_ps<.01). To confirm, that a higher β would not explain group differences in α, we ran a correlation between both indices for punishment in either group, showing no significant correlation coefficients (CD: r=-0.03, p=0.8; TDC: r=0.03, p=0.74).

Correlations between CU Traits and Learning Rate α

The additional correlational analyses between CU traits as dimensional measure and learning rate α for punishment yielded no significant result in either group (rs<.06, ps>.5, BFs_{01}>3.61), suggesting that the group by condition interaction effect for the learning rate α for punishment were not influenced by CU traits (neither as categorical nor dimensional variable).

Discussion

The current study aimed to test two competing, but not necessarily mutually exclusive, hypotheses that aberrant RL in CD reflects either reward dominance/hypersensitivity or punishment insensitivity/hyposensitivity. To accomplish this, we used a probabilistic RL task with reward, punishment, and neutral contingencies and analyzed the acquired data using computational performance indices that capture learning in choice situations (e.g., learning rate α). Consistent with the punishment insensitivity hypothesis, we found significantly lower learning rates α for punishment in the CD group compared to the TDCs, but no between-group differences in learning rates for reward. Importantly, learning rates were not associated with IQ or age, suggesting that the punishment learning difficulties in CD cannot be explained simply by differences in general cognitive ability or age effects. Regarding the possible underlying mechanism(s) of punishment insensitivity in CD, research suggests that antisocial youths experience relatively little physiological arousal when they are actually punished and are therefore less able to form proper stimulus-punishment associations[29] – e.g., by not connecting disciplinary actions with one's own wrongdoing – which prevents them from modifying their behavior to avoid such scenarios in the future. Notably, we found no evidence for sex-specific effects, and CU traits also had no impact on learning rates, which is consistent with several other recent behavioral findings.[6] Taken together, punishment insensitivity appears to be observed in both sexes in CD and is not particularly related to a specific subgroup of youth with CD, namely those with high CU traits (also known as the "LPE specifier" in the DSM-5 and ICD-11). Although we have identified a primary insensitivity to punishment (but not to reward) in youths with CD, it remains difficult to disentangle whether this deficit is due to hyporeactivity to a cue (which triggers the expectation of potential punishment) or the actual receipt of punishment, or both. This needs to be investigated in follow-up studies.

This is the first larger-scale study to examine probabilistic RL between youth with a confirmed CD diagnosis and TDC using a computational modeling approach. Our approach extends the related work by White and colleagues who examined, e.g. prediction error signaling, but no individual learning rates in much smaller samples of youth with conduct problems or ODD who performed a probabilistic PAL task.[4,30,31] In the present study, we applied the Rescorla-Wagner learning rule to calculate how probabilistic RL processes occur in the context of monetary reward, punishment, and neutral contingencies. By comparing different hypothetical models of learning – i.e., learning may or may not be similar in all conditions – we were able to show that learning rates α (as well as exploration tendencies β) are best modeled with separate parameters for each condition and for each participant individually.

As expected, both groups learned from the reward and punishment contingencies, but not from neutral outcomes. Notably, compared to the reward condition, we found a higher learning rate α for punishment across the entire sample, suggesting a greater speed at which youths updated their estimates of punishment versus reward. This underscores findings from other research areas that learning from punishment and reward may involve qualitatively different latent neurocognitive processes.[32] There is a prevailing view that aversive outcomes have subjectively greater emotional value than pleasant ones,[33] eliciting relatively more on-task attention, mood changes and autonomic arousal, which may contribute to the fact that punishment learning in choice situations is computationally different – at least to some extent – from that of reward.[32] We can only speculate whether larger amounts of reward incentives would have triggered greater reward-driven learning behavior than documented here. However, our data of a higher learning rate α for punishment than for reward across the entire sample are in fact consistent with the so called 'learning rate asymmetry', meaning that learning rates are usually higher for punishment than reward contingencies.[34]

Our study had several strengths: We used a clinically well-characterized and adequately powered sample in terms of sex and group size to test two prominent hypotheses about RL differences in CD versus TDCs. In our analytical strategy, we used computational model-based indices (e.g., learning rate α), which are particularly sensitive to the effects of interest, because they are able to capture the temporal dynamics of RL processes that appear to be different for punishment and reward as demonstrated here. Finally, RL models are supported by neuroimaging findings linking, for example, prediction error signaling that drives RL to phasic activity, or suppression, of dopamine neurons in the midbrain and other reinforcement-sensitive brain regions such as striatum, amygdala, and prefrontal cortices,[3] all of which are thought to be implicated in CD.[11]

However, one limitation of our study is that our CD group had lower IQs (which is a typical finding in the CD literature),[35] were slightly older than the TDCs, and had additional co-occurring psychiatric disorders. But neither IQ nor age correlated significantly with the computational model parameters, and the presence of comorbidities did also not affect these parameters, making it unlikely that the between-group findings were influenced by these possible confounders.

In summary, our findings support the punishment insensitivity/hyposensitivity hypothesis of CD, but less so the hypothesis of reward dominance/hypersensitivity. Interestingly, punishment insensitivity/hyposensitivity appears to affect girls and boys with CD similarly and is largely unrelated to CU traits, which is consistent with an accumulating body of behavioral evidence.[5] These findings suggest that theoretical accounts of CD (e.g.,[36]) seem to apply equally to both sexes – at least with respect to RL.

Nevertheless, further studies with additional experimental manipulations (e.g., varying magnitudes of punishment and/or stimulus-outcome reversals) as well as other experimental paradigms (e.g., effort-based learning tasks) are needed to replicate the current findings and thus to substantiate our conclusion. Furthermore, since we only had behavioral data available in the current study, it would be interesting to investigate in follow-up studies the extent to which, for example, physiological markers (such as heart rate and/or electrodermal activity) are able to provide the necessary information about why youths with CD learn less efficiently from punishment than TDCs. And finally, because CD is a psychiatric disorder in which impairments in interpersonal, i.e., social, functioning are central,[2] the study of social reinforcement, such as social punishment,[37,38] rather than nonsocial monetary reinforcement as used in the current and most related studies, would clearly benefit this line of research and help to better understand the role of probabilistic RL deficits in the development and maintenance of CD. Clinically, our findings suggest that the use of punishment-based intervention techniques to modify behavior in order to achieve effective discipline in youths with CD may be a less helpful strategy than reward-based techniques.

References

1. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (DSM-5)*. American Psychiatric Association; 2013.

2. Pauli R, Lockwood PL. The computational psychiatry of antisocial behaviour and psychopathy. *Neurosci Biobehav Rev*. 2023;145:104995. doi:10.1016/j.neubiorev.2022.104995

3. Schultz W. Multiple reward signals in the brain. *Nat Rev Neurosci*. 2000;1(3):199-207. doi:10.1038/35044563

4. White SF, Pope K, Sinclair S, et al. Disrupted Expected Value and Prediction Error Signaling in Youths With Disruptive Behavior Disorders During a Passive Avoidance Task. *Am J Psychiatry*. 2013;170(3):315-323. doi:10.1176/appi.ajp.2012.12060840

5. Byrd AL, Loeber R, Pardini DA. Antisocial Behavior, Psychopathic Features and Abnormalities in Reward and Punishment Processing in Youth. *Clin Child Fam Psychol Rev*. 2014;17(2):125-156. doi:10.1007/s10567-013-0159-6

6. Kohls G, Baumann S, Gundlach M, et al. Investigating Sex Differences in Emotion Recognition, Learning, and Regulation Among Youths With Conduct Disorder. *J Am Acad Child Adolesc Psychiatry*. 2020;59(2):263-273. doi:10.1016/j.jaac.2019.04.003

7. Byrd AL, Hawes SW, Burke JD, Loeber R, Pardini DA. Boys with conduct problems and callous-unemotional traits: Neural response to reward and punishment and associations with treatment response. *Dev Cogn Neurosci*. 2018;30:51-59. doi:10.1016/j.dcn.2017.12.004

8. O'Brien BS, Frick PJ. Reward dominance: Associations with anxiety, conduct problems, and psychopathy in children. *J Abnorm Child Psychol*. 1996;24(2):223-240. doi:10.1007/BF01441486

9. Blair RJR. The neurobiology of psychopathic traits in youths. *Nat Rev Neurosci*. 2013;14(11):786-799. doi:10.1038/nrn3577

10. Fairchild G, van Goozen SHM, Stollery SJ, et al. Decision Making and Executive Function in Male Adolescents with Early-Onset or Adolescence-Onset Conduct Disorder and Control Subjects. *Biol Psychiatry*. 2009;66(2):162-168. doi:10.1016/j.biopsych.2009.02.024

11. Blair RJR. A cognitive neuroscience perspective on child and adolescent psychopathy. In: Salekin RT, Lynam DR, eds. *Handbook of Child and Adolescent Psychopathy*. The Guilford Press; 2010:156-178.

12. Kohls G, Fairchild G, Bernhard A, et al. Neuropsychological subgroups of emotion processing in youths with conduct disorder. *Front Psychiatry*. 2020;11:585052. doi:10.3389/fpsyt.2020.585052

13. Lockwood PL, Klein-Flügge MC. Computational modelling of social cognition and behaviour—a reinforcement learning primer. *Soc Cogn Affect Neurosci*. Published online March 30, 2020:nsaa040. doi:10.1093/scan/nsaa040

14. Rescorla RA, Wagner AR. A theory of Pavlovian conditioning: Variations on the effectiveness of reinforcement and non-reinforcement. In: Black AH, Prokasy WF, eds. *Classical Conditioning II: Current Research and Theory*. Appleton-Century-Crofts; 1972:64-99.

15. Jepma M, Schaaf JV, Visser I, Huizenga HM. Uncertainty-driven regulation of learning and exploration in adolescents: A computational account. Daunizeau J, ed. *PLOS Comput Biol*. 2020;16(9):e1008276. doi:10.1371/journal.pcbi.1008276

16. Zhang L, Lengersdorff L, Mikus N, Gläscher J, Lamm C. Using reinforcement learning models in social neuroscience: frameworks, pitfalls and suggestions of best practices. *Soc Cogn Affect Neurosci*. 2020;15(6):695-707. doi:10.1093/scan/nsaa089

17. Frick PJ, Viding E. Antisocial behavior from a developmental psychopathology perspective. *Dev Psychopathol*. 2009;21(4):1111-1131. doi:10.1017/S0954579409990071

18. Freitag CM, Konrad K, Stadler C, et al. Conduct disorder in adolescent females: current state of research and study design of the FemNAT-CD consortium. *Eur Child Adolesc Psychiatry*. Published online June 9, 2018:1-17. doi:10.1007/s00787-018-1172-6

19. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders: DSM-IV-TR.* American Psychiatric Association; 2000.

20. Kaufman J, Birmaher B, Brent D, et al. Schedule for Affective Disorders and Schizophrenia for School-Age Children-Present and Lifetime Version (K-SADS-PL): Initial Reliability and Validity Data. *J Am Acad Child Adolesc Psychiatry*. 1997;36(7):980-988. doi:10.1097/00004583-199707000-00021

21. Wechsler D. Wechsler Intelligence Scale for Children. San Antonio, TX: Psychological Corporation. Published online 2003.

22. Wechsler D. Wechsler Adult Intelligence Scale 4th-Edition (WAIS-IV). San Antonio, TX: Psychological Corporation. Published online 2008.

23. Wechsler D. Wechsler Abbreviated Scale of Intelligence (WASI). San Antonio, TX: Psychological Corporation. Published online 1999.

24. Andershed HA, Kerr M, Stattin H, Levander S. *Psychopathic Traits in Non-Referred Youths: A New Assessment Tool. Psychopaths: Current International Perspectives*. Elsevier, Den Haag; 2002.

25. Colins OF, Vermeiren RRJ. The Usefulness of DSM-IV and DSM-5 Conduct Disorder Subtyping in Detained Adolescents. *J Nerv Ment Dis*. 2013;201(9):736-743. doi:10.1097/NMD.0b013e3182a20e94

26. Baumann S, Hartz A, Scharke W, et al. Differentiating brain function of punishment versus reward processing in conduct disorder with and without attention deficit hyperactivity disorder. *World J Biol Psychiatry*. Published online November 15, 2021:1-12. doi:10.1080/15622975.2021.1995809

27. MacKay DJ, Mac Kay DJ, others. *Information Theory, Inference and Learning Algorithms*. Cambridge university press; 2003.

28. Marsman M, Wagenmakers EJ. Bayesian benefits with JASP. *Eur J Dev Psychol*. 2017;14(5):545-555. doi:10.1080/17405629.2016.1259614

29. Fairchild G, Van Goozen SH, Stollery SJ, Goodyer IM. Fear conditioning and affective modulation of the startle reflex in male adolescents with early-onset or adolescence-onset conduct disorder and healthy control subjects. *Biol Psychiatry*. 2008;63(3):279-285. doi:10.1016/j.biopsych.2007.06.019

30. White SF, Fowler KA, Sinclair S, et al. Disrupted Expected Value Signaling in Youth With Disruptive Behavior Disorders to Environmental Reinforcers. *J Am Acad Child Adolesc Psychiatry*. 2014;53(5):579-588.e9. doi:10.1016/j.jaac.2013.12.023

31. White SF, Tyler PM, Erway AK, et al. Dysfunctional representation of expected value is associated with reinforcement-based decision-making deficits in adolescents with conduct problems. *J Child Psychol Psychiatry*. 2016;57(8):938-946. doi:10.1111/jcpp.12557

32. Kubanek J, Snyder LH, Abrams RA. Reward and punishment act as distinct factors in guiding behavior. *Cognition*. 2015;139:154-167. doi:10.1016/j.cognition.2015.03.005

33. Tversky A, Kahneman D. Judgment under Uncertainty: Heuristics and Biases. *Science*. 1974;185(4157):1124-1131. doi:10.1126/science.185.4157.1124

34. Gershman SJ. Do learning rates adapt to the distribution of rewards? *Psychon Bull Rev*. 2015;22(5):1320-1327. doi:10.3758/s13423-014-0790-3

35. Fairchild G, Hawes DJ, Frick PJ, et al. Conduct disorder. *Nat Rev Dis Primer*. 2019;5(1):43. doi:10.1038/s41572-019-0095-y

36. Blair RJR, Leibenluft E, Pine DS. Conduct Disorder and Callous–Unemotional Traits in Youth. *N Engl J Med*. 2014;371(23):2207-2216. doi:10.1056/NEJMra1315612

37. Kohls G, Perino MT, Taylor JM, et al. The nucleus accumbens is involved in both the pursuit of social reward and the avoidance of social punishment. *Neuropsychologia*. 2013;51(11):2062-2069. doi:10.1016/j.neuropsychologia.2013.07.020

38. Hwang S, Meffert H, VanTieghem MR, et al. Dysfunctional Social Reinforcement Processing in Disruptive Behavior Disorders: An Functional Magnetic Resonance Imaging Study. *Clin Psychopharmacol Neurosci Off Sci J Korean Coll Neuropsychopharmacol*. 2018;16(4):449-460. doi:10.9758/cpn.2018.16.4.449

Tables

Table 1. Demographic and clinical characteristics of the sample.

	CD n=92	TDC n=130	Group comparison
Sex, n female/male	37/55	69/61	$\chi^2(1)=3.07$; $p=.08$
Age (years), mean (SD)	14.8 (2.3)	13.8 (2.8)	$t(215)=2.68$; $p<.01$
Estimated IQ, mean (SD)	93.1 (11.6)	102.4 (11.0)	$t(189)=-5.92$, $p<.001$
CU traits (YPI subscales), mean (SD)	33.1 (7.9)	28.5 (6.5)	$t(171)=4.63$, $p<.001$
LPE specifier, n (%)	37 (40.2)	28 (21.5)	$\chi^2(1)=9.2$, $p<.01$
Current comorbidities, n (%)			
ODD	66 (71.7)	N/A	
ADHD	47 (51.1)	N/A	
MDD	29 (31.5)	N/A	
PTSD	12 (13.0)	N/A	
SUD	7 (7.6)	N/A	
GAD	5 (5.4)	N/A	

Psychotropic medication, n (%)		
Methylphenidate	16 (17.4)	N/A
Antidepressants	5 (5.4)	N/A
Antipsychotics	5 (5.4)	N/A
Atomoxetine	3 (3.3)	N/A
Lisdexamfetamine	3 (3.3)	N/A

Note: ADHD = attention deficit and hyperactivity disorder; CD = conduct disorder; GAD = generalized anxiety disorder; IQ = intelligence quotient; LPE = with limited prosocial emotions; MDD = major depressive disorder; N/A = not applicable; ODD = oppositional defiant disorder; PTSD = post-traumatic stress disorder; SUD = substance use disorder; TDC = typically developing controls. All diagnoses are based on the Schedule for Affective Disorders and Schizophrenia for School-Age Children–Present and Lifetime version (K-SADS-PL). p-values are based on χ^2 or t-tests. Information on race and/or ethnicity was not collected in accordance with government policy in Germany.

Figure captions

Figure 1. Illustration of the probabilistic reinforcement learning task.

Figure 2. Model comparison. The 3α3β model (triangle) is the best model on LME, exceedance probability and integrated Bayesian Information Criterion.

Figure 3. Comparison of model parameters from the computational model. a Group and condition differences in learning rates α. **b** Condition, but no group, differences in exploration parameters β. *ns* = not significant; *$p \leq .05$; **$p \leq .01$; ***$p \leq .001$.