Journal of the American Academy of Child and Adolescent Psychiatry

Impaired Punishment Learning in Conduct Disorder

--Manuscript Draft--

Manuscript Number:	JAACAP-D-23-00081R1
Full Title:	Impaired Punishment Learning in Conduct Disorder
Short Title:	Punishment Learning in Conduct Disorder
Article Type:	Research Article
Keywords:	conduct disorder; Decision Making; Punishment; Reinforcement, Psychology; Computational Modeling
Corresponding Author:	Erik Martin Elster, M.Sc. Dresden, Sachsen GERMANY
Corresponding Author Secondary Information:	
Corresponding Author's Institution:	
Corresponding Author's Secondary Institution:	
First Author:	Erik Martin Elster, M.Sc.
First Author Secondary Information:	
Order of Authors:	Erik Martin Elster, M.Sc.
	Ruth Pauli, Ph.D.
	Sarah Baumann, M.Sc.
	Stephane A. De Brito, Ph.D.
	Graeme Fairchild, Ph.D.
	Christine M. Freitag, M.D.
	Kerstin Konrad, Ph.D.
	Veit Roessner, M.D.
	Inti A. Brazil, Ph.D.
	Patricia L. Lockwood, Ph.D.
	Gregor Kohls, Ph.D.
Order of Authors Secondary Information:	
Manuscript Region of Origin:	GERMANY
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.

Klinik für Kinder- und Jugendpsychiatrie und -psychotherapie
Direktor: Prof. Dr. med. Veit Roessner
Bereich Forschung

Universitätsklinikum Carl Gustav Carus
DIE DRESDNER.

Universitätsklinikum Carl Gustav Carus · 01304 Dresden

Telefon: 0351 458-7168
Telefax: 0351 458-5754
KJPForschung@ukdd.de
Dresden, den 13. April 2023

Universitätsklinikum Carl Gustav Carus Dresden an der Technischen Universität Dresden
Anstalt des öffentlichen Rechts des Freistaates Sachsen

Hausanschrift:
Fetscherstraße 74
01307 Dresden
Telefon 0351 458-0
www.uniklinikum-dresden.de

RE: JAACAP-D-23-00081 "Impaired Punishment Learning in Conduct Disorder"

Dear Dr. Althoff, dear Miss Gambino,

Please find enclosed our revised manuscript entitled **"Impaired Punishment Learning in Conduct Disorder"** to be considered for publication in *The Journal of the American Academy of Child and Adolescent Psychiatry* (JAACAP). This submission was approved as part of JAACAP's rapid pre-submission review process. At the end of this cover letter, you will find once again the reviewers' comments on our original submission to the *American Journal of Psychiatry*, as well as our responses to each point. As requested, we have thoroughly incorporated our proposed changes into the revised manuscript submitted here. All changes have been completed and are marked in red in the revised manuscript.

In addition, as requested, we have included information on psychotropic medication use among the patients with conduct disorder in Table 1. We also added a statement to Table 1 that we are unable to report on race/ethnicity because this information was not collected in accordance to government policy in Germany.

Thank you for your time, and we hope that you consider the revised manuscript for publication in *JAACAP*. We look forward to your response.

Yours sincerely,

[redacted]
Erik M. Elster, M.Sc.
Corresponding author

[redacted]
Gregor Kohls, Ph.D.
Senior author

Response Letter

Please note that all proposed changes listed below in our replies to the reviewers' recommendations have been fully incorporated into the revised manuscript. They are highlighted in red in the revised manuscript.

Reviewer: 1

There was much to like about this paper, including the clearly laid out hypotheses, use of a large sample with CD (including males and females) and interesting analytic approach. I hope my comments (in chronological order) can be of assistance to the authors.

Reply: We thank the reviewer for this overall positive evaluation of our work. Please find below our replies to the questions and issues raised by the reviewer.

1) P.6, line 32 – deficits as demonstrated through which kinds of tasks?

1) Reply: We apologize for not being more precise here, and will add more information on the task(s) used, which was primarily the passive avoidance learning (PAL) task. The task is described in more detail later in the introduction.

2) P. 7, line 23-30: clarify that authors are suggesting that CU traits would be associated with aberrant learning across both punishment and reward trials? Not clear, but also not necessarily supported by existing data, where we might hypothesize reward learning to be intact but punishment learning to be particularly worse among youth high on CU traits?

2) Reply: We agree with the reviewer and will adjust the hypothesis accordingly: "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) (14), we predicted a positive correlation between CU traits and aberrant punishment learning performance."

3) What was the prevalence of other diagnoses within the CD group (p. 8, line 19) – e.g., ADHD, or mood disruption? I now see this in Table 1 – to what extent did the presence of comorbidities influence the pattern of findings? (especially since the authors claim in paragraph 1, p. 6, for example, that ADHD might be associated with differential reward learning: that could be up to 50% of the CD group?).

3) Reply: We thank the reviewer for pointing out this important aspect to consider. Our initial sensitivity analyses using rmANCOVAs (see Supplementary Table 3) did not include co-occurring psychopathological conditions because psychiatric comorbidities only applied to the CD group but not to the TDCs; therefore, there were no rmANCOVAs comparing the influence of comorbidities on both groups using 'group' as a factor. We are however pleased to add the results of additional regression analyses (i.e. mixed models), showing that there were no significant influences of comorbidities (incl. ADHD) on the learning rates alpha for punishment (or the two other contingencies) in the youths with CD. We will add the following information to the Supplement of the revised manuscript:

"To test the influence of comorbidities on the learning parameter α, we ran three linear mixed models for the CD group and compared the models with a likelihood ratio test. The models were specified as follows:

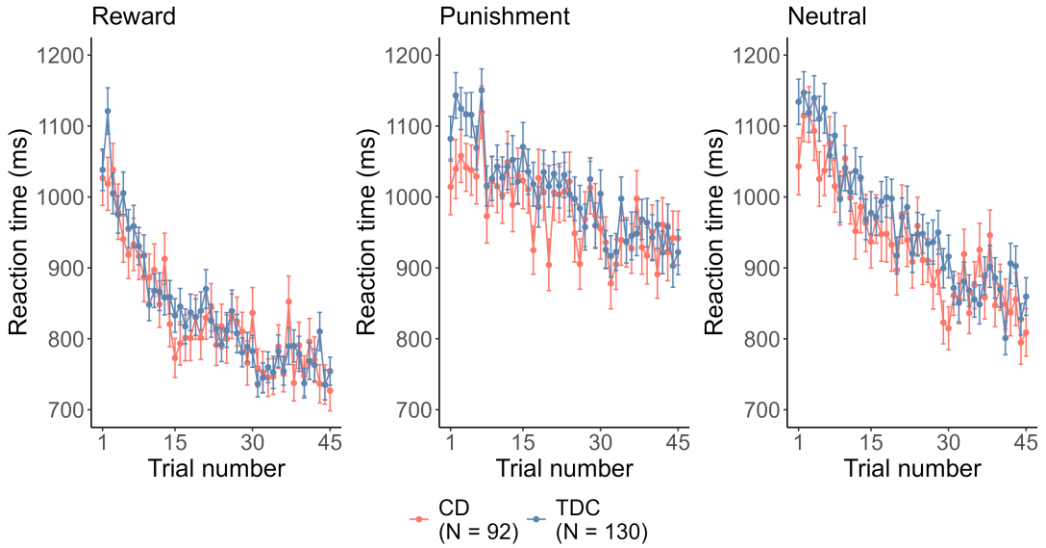
Model 0: α ~ 1 + (1|participant)
Model 1: α ~ condition + (1|participant)
Model 2: α ~ condition * comorbidity + (1|participant)

Model 0 is the intercept-only model. Model 1 contains the within-subject factor of task condition (i.e. reward, punishment, and neutral). Model 2 additionally includes the different main psychiatric comorbidities found in the CD group. The likelihood ratio test revealed that Model 1 explains the data best (χ^2 (2) = 21.77, p < .001), suggesting that the presence of comorbidities did not significantly affect the learning rates in the youths with CD."

4) 25 minutes for this task seems like a long time – does data quality become an issue later in the task, do people get bored, are there breaks (i.e., aside from the condition specific hypotheses, do people get less engaged and does beta increase overall across time for people?). What does "pseudo-random" order mean?

4) Reply: We apologize for not describing the task more clearly, which was partly due to the limited number of words allowed. Actually, the task was split into three runs with short breaks in-between; we will add this information to the revised manuscript. We also apologize for not being more precise about what we mean by "pseudo-random" order. Therefore, we will include the following description to the methods section of the revised manuscript: "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."

Notably, the length of the experiment was tolerated very well by both CD patients and typical controls. However, we did not collect explicit data on task engagement. Note that our computational approach was to model and estimate beta and also alpha values across the full range of collected behavioral data. Therefore, there is no possibility to provide parameter values, such as beta, as they may have changed over time. We nevertheless believe that boredom (or 'noisy' choice making) was not an issue here (Bench & Lench 2013, *Behav. Sci.*): Since participants knew exactly by instruction that they would receive the amount of money earned at the end of the experiment, it is reasonable to assume that this was highly motivating for them to perform well from start to finish. Also, if one were to use, for instance, reaction time as an indicator of task engagement, such as slowing down over time reflects boredom in performing the task at hand, the available data show the exact opposite: in all three conditions, participants responded faster over time (see below).



5) What was the rationale for the amount of money (20 cents/20 pence) used in the task – for example, why not larger sums (e.g., 1 Eur/1 Pound). Perhaps it was enough since it elicited sensitivity to punishment across the whole sample, but curious to know where that amount came from and whether larger amounts would have been more enticing/provoking when it came to reward-driven responding?

5) Reply: We thank the reviewer for this interesting question. In fact, in the original paper by Kim et al. (2006), from which our task design was derived, the authors used $1 as incentives. We decided against using such a large amount of money as incentives for ethical and practical reasons. According to the instructions, participants received the amount of money earned at the end of the experiment in addition to the compensation for participating in this study. Participants received an average amount of ~11€, which we believe is appropriate for participants in the age group tested here. Larger incentives would have resulted in an estimated four-fold amount of ~40-50€, which we wanted to avoid. Thus, we can only speculate whether larger incentives would have triggered greater reward-driven response 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 (Gershman 2015, *Psychon Bull Rev.*). This idea is discussed in the revised manuscript.

6) Figures 2 and 3 don't appear in the manuscript so I couldn't evaluate them or the results they present

6) Reply: We apologize for this inconvenience. For unknown reasons, Figures 2 and 3 were only available online on the journal's website for reviewers, but were not included with the downloadable manuscript. We will ensure that all figures appear in the revised manuscript.

7) It seemed like a lot of basic descriptive data for the whole sample was missing – whether from the main manuscript or from the supplement. I found myself wanting to understand overall, basic, responses both across the sample as a whole during the task (i.e., how do people perform, overall, across the task and during the different conditions), as well as within groups (perhaps that's what's in Figures 2 and 3?).

7) Reply: We kindly refer the interested reader (incl. the reviewer) to Supplementary Table 2 for descriptive behavioral data, including accuracy and reaction times, for each group and condition.

8) Could the authors expand more on the finding (supplement) that accuracy and reaction time did not vary as a function of group; I understand this is central to the premise of the paper, but it would be helpful to make it clearer what the difference is between accuracy and reaction time (which don't differ between groups) vs. learning rate parameter (which differs between groups for punishment). I guess I am asking, if accuracy (and beta) do not differ between groups, what explains the fact that CD youth are not learning to adjust the expected outcome based on new information.

8) Reply: We thank the reviewer for this intriguing question. First, however, we would like to emphasize that a lower learning rate alpha for punishment in the CD group cannot simply be interpreted as meaning that no learning occurred. Instead, it reflects that the learning process was slower and less efficient when it came to how youths

with CD updated their expectation of a given outcome (i.e. the value of punishment) with newer information from trial to trial; we will make this interpretation more explicit in the revised manuscript. The reviewer is however right in pointing out that accuracy and reaction times for punishment did not differ between the CD group and TDCs, which we expected though. Both accuracy and reaction times are only rather coarse summary measures of performance (i.e. they provide values averaged over trials per condition). Since reinforcement learning (RL) is a latent neurocognitive operation, it is best quantified directly and precisely by computational indices based on RL models (such as the learning rate alpha or the exploration parameter beta). These computational indices are more sensitive to the effects of interest as they are able to capture the trial-by-trial dynamics of the learning process as it unfolds over time. This information can be found in the revised manuscript.

However, because we only have the behavioral data available in the current study, we can just speculate on the underlying mechanism(s) (i.e. contributing factors) that might explain the lower learning rate alpha for punishment in the CD group compared to TDCs. In the manuscript, we currently write that "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." Therefore, 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. This suggestion will be included in the revised manuscript.

9) Related, p. 13, line 39 onwards: it would be helpful to understand what is happening in the CD group if not learning (i.e., based on lower learning rate, alpha) during the punishment condition. It would appear that higher beta does not explain group differences?

9) Reply: We thank the reviewer for raising this interesting question. However, we would like to emphasize again that a lower learning rate alpha for punishment in the CD group cannot simply be interpreted as meaning that no learning occurred. Instead, it reflects that the learning process was slower and less efficient when it came to how youths with CD updated their expectation of a given outcome (i.e. the value of punishment) with newer information from trial to trial; as mentioned above, we will make this interpretation more explicit in the revised manuscript. Beta represents an individual's randomness in choice behavior while learning. In fact, we found no correlation between alpha and beta for punishment in either group (CD: r = -0.03, p = 0.8; TDC: r = 0.03, p = 0.74), indicating that a higher beta would not explain group differences in alpha. We will add this information to the revised manuscript.

10) I didn't understand the analysis for CU traits; I was expecting to see CU traits entered in the model with a formal interaction effect tested. Can the authors justify this approach more and/or break down the findings more clearly?

10) Reply: We apologize for not describing the analysis for CU traits and its findings more clearly. We will include the following to the methods section of the revised manuscript:

"The modeled parameters α and β from our winning model (3α3β) were then analyzed using two separate repeated-measure ANOVA models with group (CD vs. TDC), sex (male vs. female) and CU traits (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 will then include the following to the results section of the revised manuscript:

"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, $BF_{01} = 13.23$).
The rmANOVA for the exploration parameters β revealed a significant main effect of condition [$F(2, 428) = 53.5$, $p < .001$, $\eta^2_p = .20$], but no significant effects of group ($p = .40$, $\eta^2_p < .01$) or sex ($p = .64$, $\eta^2_p < .01$) or CU traits ($p = .45$, $\eta^2_p < .01$). The Holm-adjusted post-hoc comparisons for the condition effect showed a lower reward β compared to punishment β and neutral β ($M_{Diff(rew-pun)} = -0.22$, 95%-$CI$[-0.25, -0.19], $p < .001$, $BF_{10} > 100$; $M_{Diff(rew-neut)} = -0.21$, 95%-$CI$[-0.26, -0.16], $p < .001$, $BF_{10} > 100$), while punishment β and neutral β did not differ ($M_{Diff(pun-neut)} = 0.01$, 95%-$CI$[-0.04, 0.05], p > .99 , $BF_{01} = 12.43$). The group by condition interaction ($p = .46$, $\eta^2_p < .01$), the sex by condition interaction ($p = .96$, $\eta^2_p < .01$) and the CU traits by condition interaction ($p = .98$, $\eta^2_p < .01$) were non-significant. All other interaction effects were non-significant as well ($ps > .16$, $\eta^2_ps < .01$)."

11) Can the authors explain more how this study, sample, and approach differs to their prior work in the same sample (p. 6, "using the PAL task in the largest sample of CD youths to date, we found more errors in responding to punishment"). What does the current approach do – both quantitatively (which I understand) and substantively (which I am less clear about), that adds to prior knowledge.

11) Reply: We would like to highlight that the current study differs from our prior work using the PAL task in two crucial aspects: First, in the current subsample of the FemNAT-CD cohort we used a reinforcement learning task with probabilistic feedback (e.g. incorrect responses lead to punishment only with a certain likelihood), while the PAL task traditionally uses deterministic feedback (i.e. incorrect responses always lead to punishment). Our approach of using probabilistic feedback here appears to be advantageous for studying reinforcement learning (and testing the hypotheses at hand), because this type of feedback provides a better model for the uncertainty of learning situations in real life, especially for youths who have a learning advantage in the case of probabilistic uncertainty, for instance, compared to adults or young children (e.g. Blankenstein et al. 2016, *Dev Neuropsychol.*). Uncertainty refers to ambiguity in how likely a punishing or rewarding outcome is to occur.

Second (see also our reply to point 8 above), previously we analyzed the reinforcement learning behavior in the PAL task using 'traditional' variables, such as the overall number of punishment- and reward-learning errors, but these provide only summary measures of performance. In contrast, computational indices based on reinforcement learning models (e.g. learning rate alpha) are able to capture the trial-by-trial dynamics of the learning process as it unfolds over time. By doing so, our analytic approach goes beyond the existing ones used in CD research, as we are able to estimate learning rates for punishment contingencies (versus reward or neutral ones) for each study participant individually – rather than using pre-set generic learning rates as done previously (e.g. the related work by White and colleagues). We plan to include the information in the revised manuscript on how our approach complements prior knowledge.

Reviewer: 2

The authors present a case-control behavioral study of probabilistic reward and punishment learning in 92 youths with conduct disorder (CD) vs. 130 controls, aged 9-18. They used a 70/30 probabilistic learning task without reversals, with blocked monetary reward, monetary punishment, and neutral conditions. They tested for group differences in meta-learning parameters by inverting a RW-type model using an iterative maximum a posteriori (MAP) algorithm. The dominant model was selected using Bayesian model comparison. The authors report that CD youths had a lower learning rate for punishments but not for rewards with no group differences in the [inverse] temperature or stochasticity parameter. Sampling, as well as modeling and statistical methods are appropriate and the paper is clearly written. I particularly appreciated the model and parameter identifiability analyses validating the modeling approach. The main finding is in line with earlier studies showing impaired punishment learning in CD.

Reply: We thank the reviewer for this overall positive evaluation of our work. Please find below our replies to the questions and issues raised by the reviewer.

I have only a few specific comments:

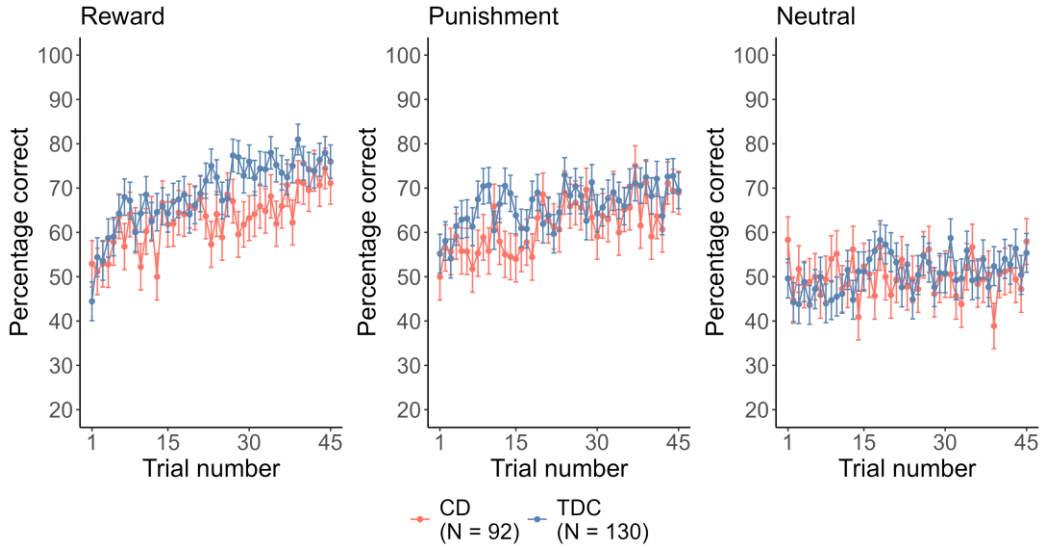
1) I did not entirely understand the task design. Were the conditions blocked or interleaved?

1) Reply: We apologize for not describing the task more clearly in the main text. We kindly refer the interested reader (incl. the reviewer) to the Supplementary where we describe the task in more detail. However, we will add the following explanation to the methods section of the revised manuscript: "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."

2) While modeling is useful, it would also be helpful to see learning curves by group for a model-free corroboration.

2) Reply: We appreciate the reviewer's request to provide more data, such as learning curves, and are pleased to add a figure showing learning curves to the Supplement of the revised manuscript. In our original submission, we did not present learning curves, but we did test whether participants' choices were random or whether learning actually occurred. We quantified 'learning' as choosing the high probability cue above chance level (>50%) per condition. As expected, participants across both groups were able to learn in both the reward and punishment conditions (all ts>14.2, all ps<.001, all Cohen's ds>2.14), but not in the neutral condition (t=0.82, p>.41).

This information can be found in the Supplement. We will also add learning curves (see below) to the revised Supplement as requested by the reviewer.



3) The inference about punishment insensitivity is broadly consistent with the data. However, the experiment and modeling are not entirely conclusive, because a lower learning rate reflects slower temporal integration of reinforcement (punishment in this case) and not necessarily a lower sensitivity to the reinforcer. For example, a model with lower punishment learning rate would also display slower extinction of the value of a punishing option. To fully dissociate punishment sensitivity (valuation) from its temporal integration (learning rate), one needs to experimentally manipulate both the magnitude and the temporal occurrence of punishment (e.g. a paradigm with varying size of punishments and reversals).

3) Reply: We thank the reviewer for bringing this interesting thought to our attention, and to some extent we agree with the reviewer. However, to be even more precise in terms of its practical meaning, a lower learning rate reflects slower updating of reinforcer value from recent outcomes; it therefore combines temporal integration (i.e. reinforcer history) with reinforcer valuation (Zhang et al. 2020, *Soc Cogn Affect Neurosci.*); we will include this definition to the revised manuscript. In line with Zhang et al. (2020, p. 699), we have therefore operationalized that a lower learning rate per reinforcer reflects a lower sensitivity to that particular reinforcer (e.g. punishment). However, we definitely agree that additional experimental manipulation(s), including varying magnitudes of punishments and stimulus-outcome reversals, could further substantiate our findings of punishment insensitivity in youths with CD. This however would add further length to an already quite long task (~25 minutes); this should be considered. Though, we will add the reviewer's suggestion to the discussion of the revised manuscript.

4) The authors present sensitivity analyses controlling for age, IQ and callous-unemotional traits. Lacking, however, is evidence that individual differences in learning rates are specifically attributable to CD as opposed to other forms of psychopathology, which were prevalent in the CD group.

4) Reply: We thank the reviewer for pointing out this important aspect to consider. Our initial sensitivity analyses using rmANCOVAs (see Supplementary Table 3) did not include co-occurring psychopathological conditions because psychiatric comorbidities only applied to the CD group but not to the TDCs; therefore, there were no

rmANCOVAs comparing the influence of comorbidities on both groups using 'group' as a factor. We are however pleased to add the results of additional regression analyses (i.e. mixed models), showing that there were no significant influences of comorbidities (incl. ADHD) on the learning rates alpha for punishment (or the two other contingencies) in the youths with CD as our primary finding. We will add the following information to the Supplement of the revised manuscript:

"To test the influence of comorbidities on the learning parameter α, we ran three linear mixed models for the CD group and compared the models with a likelihood ratio test. The models were specified as follows:

Model 0: $\alpha \sim 1 + (1|participant)$
Model 1: $\alpha \sim condition + (1|participant)$
Model 2: $\alpha \sim condition * comorbidity + (1|participant)$

Model 0 is the intercept-only model. Model 1 contains the within-subject factor of task condition (i.e. reward, punishment, and neutral). Model 2 additionally includes the different main psychiatric comorbidities found in the CD group. The likelihood ratio test revealed that Model 1 explains the data best (χ^2 (2) = 21.77, $p < .001$), suggesting that the presence of comorbidities did not significantly affect the learning rates in the youths with CD."

Reviewer: 3

This is a study using a commendably large sample of participants with CD and includes both males and females. Computational models that estimate learning rates and temperature parameters of learning are fitted separately for punishments and rewards. The findings replicate prior studies, including prior computational modelling studies, that indicate punishment learning difficulties in participants with CD. They also add to the more mixed picture of findings regarding reward processing in CD. A sample with both males and females enabled a comparison of sexes in punishment and reward processing, which was a welcome addition to the field. Although I think that the study makes an important added contribution to the literature of reinforcement learning and CD, I was not entirely convinced by the way in which the study was framed or the ultimate novelty of the findings.

Reply: We thank the reviewer for this honest summary of our work. Not surprisingly, we disagree with the assertion that our study ultimately provides no novel insights on the topic of punishment learning difficulties in youths with CD. Here are some of the main reasons that we actually see as a real strength of the current study compared to previously published work: To date, the vast majority of studies on reinforcement learning (RL) in youths with CD have been unsuited or underpowered for adequately testing for sex-by-group interaction effects as they primarily focused on predominantly male or female samples. Prior work has been further limited by relying on rather smaller-scale samples with varying selection criteria, very often including mixed groups of youths with CD (or subclinical levels of CD-related problems) and oppositional defiant disorder (ODD) or even attention deficit hyperactivity disorder (ADHD). Here, we tested a substantial sample of boys and girls with a confirmed CD diagnosis (vs. TDCs) that even included some girls with the rare form of childhood-onset CD. To enable clear data interpretation, we did not include a mixed clinical group of participants with CD or ODD (or even ADHD) as it is still premature (or even unlikely) to suggest that the same neurocognitive mechanisms underlie the etiology of these disorders. Moreover, the entire sample was comprehensively clinically assessed and reliably diagnosed using standardized, semi-structured interviews based on DSM-IV/DSM-5 criteria that enabled us to account statistically for common psychiatric comorbidities as potential confounding factors.

We were therefore able to demonstrate that the punishment learning problems are specifically attributable to CD as opposed to other forms of co-occurring psychopathology (incl. ODD, and ADHD). Finally, and crucially, is the contrast with the closely related work of White and colleagues (2013, 2014, and 2016), who also used a prediction error (PE)-based RL model type in youths with (subclinical) conduct problems or ODD performing a probabilistic passive avoidance learning (PAL) task. However, White et al. applied one single pre-set generic learning rate (alpha was set at 0.1) for all participants and contingency conditions, ignoring two crucial empirical facts, namely that learning rates can differ across individuals and the so-called "learning rate asymmetry", meaning that learning rates are typically higher for punishment than reward contingencies; this is exactly what we found here (the condition effect was at $p < .001$: punishment > reward = neutral). In fact, our analysis pipeline allowed us to set up and compare different hypothetical PE-based RL models (i.e. learning rates and exploration parameters may or may not be similar for different contingencies) to find a 'winning' model that fits the participants' learning behavior most optimally. To accomplish it, we applied cutting-edge analytic methods, including hierarchical expectation maximization and Bayesian model comparison as well as parameter recovery and model identifiability procedures. As expected, we found a 'winning' model that included three separate learning rates and exploration parameters, one for each contingency. Consequently, our analytic approach certainly goes beyond the existing ones used in this line of research (incl. White et al.), as we were able to estimate learning rates for punishment contingencies (versus reward or neutral ones) for each study participant individually – rather than using pre-set generic learning rates as done previously. Overall, we believe that our findings presented in this (revised) manuscript do actually replicate, but they are also novel and considerably and reliably extent, prior findings about punishment learning difficulties in youths with CD.

Please find below our additional replies to the questions and issues raised by the reviewer.

1) I think that it is somewhat of a 'red herring' to set the study up to, in part, test the 'reward dominance' hypothesis. The original study on which this hypothesis was based has already been criticised based on the fact that it is not possible to discern whether the performance on the task of that study was driven by reward or punishment processing. For example, Blair and colleagues have subsequently done a substantial amount of work to more precisely isolate aspects of reinforcement processing that are compromised in individuals with CD. Furthermore, a number of studies using different reward learning tasks have painted a fairly mixed picture regarding reward processing in CD. As such, I am not convinced that the 'reward dominance' hypothesis has quite the prominence in the field that the authors of the paper present.

1) Reply: While we agree with the reviewer's remark that the reward dominance studies by O'Brien et al. (1994, 1996) have been rightly criticized for the reasons stated above, we do not agree that the study of reward dominance can be considered a red herring or not legitimate. It may be that the term 'reward dominance' is no longer quite as prominent in the field, the premise behind the reward dominance hypothesis of a CD-related reward-seeking learning style in the presence of punishment cues is still quite alive in this line of research, but perhaps better known now under the term 'reward hypersensitivity' (e.g. Byrd et al. 2018, *Dev Cogn Neurosci.*). Therefore, in the revised manuscript, we will adjust the scope of our approach, that is, the study of reward dominance – or reward hypersensitivity – versus punishment hyposensitivity, to better highlight the use of the different terms in the current relevant literature. By doing so, we will broaden the cited studies by

including and discussing more (recent) literature on reward processing in CD. In this context, the reviewer is correct in pointing out that there is a rather mixed picture regarding aberrant reward processing in CD (i.e. there is evidence for or against the presence or absence of reward processing abnormalities). This could be due to several reasons (e.g. task-related; Richards et al. 2013, *Neurosci Biobehav Rev.*), one of which we mentioned in our very first response to this reviewer, namely the investigation of predominantly male, clinically mixed (i.e. CD and/or ODD and/or ADHD), and sometimes even subclinical (i.e. CD-problem) samples. Scientifically, this cannot mean that we should not test the idea of reward dominance or hypersensitivity, especially given our well-characterized sample of girls and boys with and without a confirmed CD diagnosis, which is certainly ideal for answering this particular research question and testing the hypotheses at hand.

2) The current study adds to the mixed picture of reward learning in CD, but because it only examines reward learning under particular task conditions, it is perhaps not appropriate to conclude that reward learning appears intact in CD. It might be more appropriate to conclude that under conditions where substantially different reward outcomes are associated with visually distinct stimuli and where these associations do not change, individuals with CD demonstrate comparable reward learning to TDCs.

2) Reply: It may be that the unimpaired reward learning ability in CD found in the current study is related to the probabilistic RL task used, although this is only speculation at this point, and follow-up studies using other task probes, e.g. a reversal learning task, are warranted. We will include this aspect in the discussion of the revised manuscript.

3) I am also not convinced that the task used in this study is optimised to test the 'reward dominance' hypothesis, if this is what the authors really want to test. Correctly discerning that one particular, clearly visually distinct object is less rewarding than another object is very different from having reward contingencies gradually change or choosing an object with a relatively low reward probability that e.g. requires effort to obtain. Either of these set ups would appear better suited to testing 'reward dominance' (as opposed to the ability to learn about rewards in the first place), although the former only if not in the context of punishment.

3) Reply: We do not entirely agree with the reviewer that our task may be unsuitable for testing reward dominance (or hypersensitivity) as we operationalized it in the current study: better reward learning at the expense of punishment learning (i.e. one would expect a higher learning rate for reward to be accompanied by a lower learning rate for punishment). As the reviewer rightly points out (see above), it was impossible to discern whether performance on the original 'reward dominance' task was driven by reward or punishment processing or both, but our task design does allow us to separate reward and punishment processing. However, we agree that additional experimental manipulation(s), including varying magnitudes of reward and punishment as well as stimulus-outcome reversals (or others, such as effort-based tasks), could further substantiate our findings of punishment insensitivity but not reward dominance in youths with CD. We will add this aspect to the discussion of the revised manuscript.

4) I was also surprised that the authors predicted a correlation between CU traits and greater learning impairment, while citing a 2009 review to support this claim. A number of studies, including several by Blair and colleagues, have been published after 2009 and have not found CU traits to be predictive of learning impairments or

atypical neural processing during learning, but have instead found these to characterise those with disruptive behaviour disorders overall (e.g. White et al., 2013; Hwang et al., 2018). Again, this is setting up the study to test a prediction that does not seem entirely warranted given the current evidence base.

4) Reply: We would like to emphasize that, like the literature on reward processing, the findings on greater RL impairments in CD youths with high CU traits are similarly mixed, both favoring and disfavoring (reviewed in Byrd et al. 2014, Child Fam Psychol Rev., or Frick et al. 2014, J Child Psychol Psychiatry., page 534: *"[…] children and adolescents with serious conduct problems and CU traits also show distinct cognitive characteristics compared to other youths with conduct problems. Specifically, children and adolescents with CU traits are more likely to show an insensitivity to punishment cues using tasks in which a reward-dominant response set is primed […]."*). However, the reviewer seems to cite studies from Blair's group exclusively, ignoring other relevant work in favor of our prediction. We believe that our prediction is warranted given all available (albeit mixed) evidence on the subject and should therefore be tested. Please note that we have adjusted this particular prediction in response to point 2 by reviewer #1 (see above). It now reads: "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) (14), we predicted a positive correlation between CU traits and aberrant punishment learning performance."

5) The authors suggest at the end of the study that the study of reinforcement learning should be extended to the social domain. No doubt more work in this area is needed, but there is already a study from Blair's group that has compared social and non-social reward processing in children with CD/ODD (Hwang et al., 2018). It seems like an omission not to cite this.

5) Reply: We certainly agree with the reviewer and will add this reference to the revised discussion. It should be noted, however, that the Hwang sample included only 14 youths with CD.

6) Although the large sample of individuals with CD in this study is welcome, the findings themselves might have been anticipated given several replications (across prior, smaller studies) of difficulties in punishment learning in CD (including the authors' own study on the same sample that did not include computational modeling). The reward learning findings of this study add to the currently mixed picture of reward learning and CD and underscore the importance of exploring reward processing using different task probes, but are currently discussed relatively simplistically.

6) Reply: We agree with the reviewer that our finding of unimpaired reward learning abilities in youths with CD adds to the mixed evidence currently available on this topic. Therefore, as mentioned above, in the revised manuscript we will expand on the cited studies by including and more thoroughly discussing additional (recent) literature on reward processing in CD, including the suggestion to examine reward processing in CD with different task probes than the one we used here.

7) Although the computational modeling approach is a helpful addition and provides a more nuanced picture of reinforcement processing in CD, I am not sure that in this case this approach has added a huge amount to what we already know.

7) Reply: Please see our first response to this reviewer, in which we explain in detail to which our approach complements previous knowledge. Thus, we consider this work

to be extremely important and anticipate that the novelty and reliability of our results as well as their clinical implications will be of considerably interest to researchers in the relevant field as well as to a broader readership.

Reviewer: 4

The authors investigated 92 youths with CD and 130 controls using computerized modelling to test reward and punishment contingencies within a probabilistic learning task. The results suggest a greater sensitivity to reward than punishment in youths with CD, independent of CU traits. The authors are to be commended on their well-designed and well-written paper. Also, they succeeded in collecting a fine number of youths with CD and control subjects, and also investigated sex differences and the influence of CU traits. Although the research is relevant and timely, my impression was that it better fits to a more specialized journal.

Reply: We thank the reviewer for this overall positive evaluation of our work. Although we still believe that our work, the novelty of our findings, and their clinical implications would be of considerably interest to the broad readership of *The American Journal of Psychiatry*, we now plan to submit a revised manuscript to the *Journal of the American Academy of Child & Adolescent Psychiatry* as a more specialized journal, as suggested by the reviewer.

Reviewer: 5

This study evaluates differential associations of reward dominance and punishment sensitivity among youth presenting with conduct disorder. The manuscript is well-written, of sound methodology, and addresses an interesting topic with important clinical implications. My only suggestion is that the study authors may wish to consider including a sentence or two that discusses implications that the study design (i.e., the monetary probabilistic RL task), which does not account for potential temporal effects on incentive processing (e.g., anticipation vs receipt), may have on the interpretation of the study findings. I thank the editor and study authors for allowing me to take part in this review.

Reply: We thank the reviewer for this overall positive evaluation of our work and the suggestion to discuss implications with regard to possible temporal effects on reinforcer processing (i.e. monetary reward anticipation vs. receipt) that may have influenced our findings. For example, while we have identified a primary insensitivity to punishment (but not to reward) in youths with CD, it remains difficult to disentangle whether such a deficit is due to hyporeactivity to a cue (which triggers the expectation of potential punishment) or the actual receipt of punishment or both. We will add this limitation to the revised discussion.

Reviewer: 1

There was much to like about this paper, including the clearly laid out hypotheses, use of a large sample with CD (including males and females) and interesting analytic approach. I hope my comments (in chronological order) can be of assistance to the authors.

Reply: We thank the reviewer for this overall positive evaluation of our work. Please find below our replies to the questions and issues raised by the reviewer.

1) P.6, line 32 – deficits as demonstrated through which kinds of tasks?

1) Reply: We apologize for not being more precise here, and will add more information on the task(s) used, which was primarily the passive avoidance learning (PAL) task. The task is described in more detail later in the introduction.

2) P. 7, line 23-30: clarify that authors are suggesting that CU traits would be associated with aberrant learning across both punishment and reward trials? Not clear, but also not necessarily supported by existing data, where we might hypothesize reward learning to be intact but punishment learning to be particularly worse among youth high on CU traits?

2) Reply: We agree with the reviewer and will adjust the hypothesis accordingly: "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) (14), we predicted a positive correlation between CU traits and aberrant punishment learning performance."

3) What was the prevalence of other diagnoses within the CD group (p. 8, line 19) – e.g., ADHD, or mood disruption? I now see this in Table 1 – to what extent did the presence of comorbidities influence the pattern of findings? (especially since the authors claim in paragraph 1, p. 6, for example, that ADHD might be associated with differential reward learning: that could be up to 50% of the CD group?).

3) Reply: We thank the reviewer for pointing out this important aspect to consider. Our initial sensitivity analyses using rmANCOVAs (see Supplementary Table 3) did not include co-occurring psychopathological conditions because psychiatric comorbidities only applied to the CD group but not to the TDCs; therefore, there were no rmANCOVAs comparing the influence of comorbidities on both groups using 'group' as a factor. We are however pleased to add the results of additional regression analyses (i.e. mixed models), showing that there were no significant influences of comorbidities (incl. ADHD) on the learning rates alpha for punishment (or the two other contingencies) in the youths with CD. We will add the following information to the Supplement of the revised manuscript:

"To test the influence of comorbidities on the learning parameter α, we ran three linear mixed models for the CD group and compared the models with a likelihood ratio test. The models were specified as follows:

Model 0: $\alpha \sim 1 + (1|\text{participant})$

Model 1: $\alpha \sim \text{condition} + (1|\text{participant})$

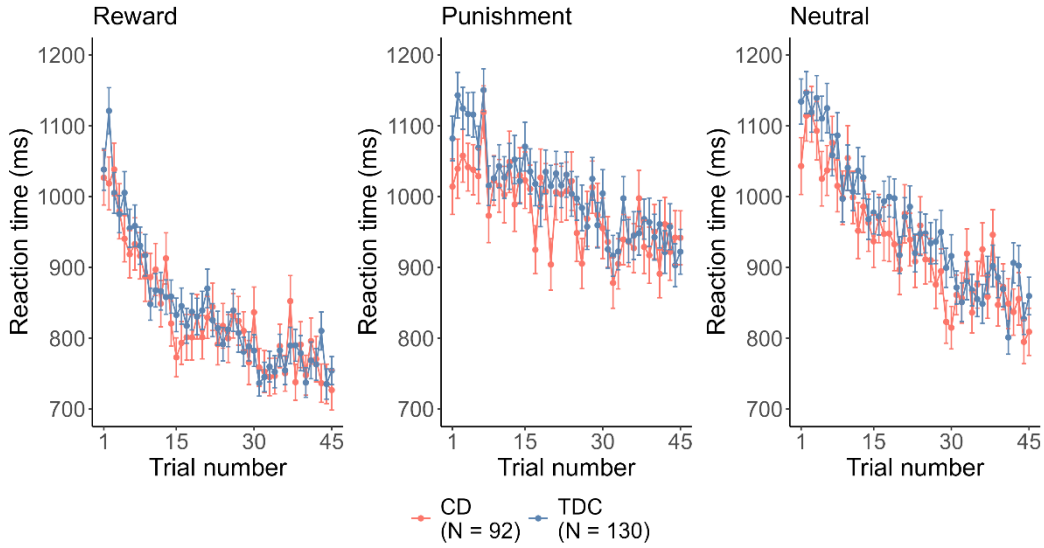
Model 2: $\alpha \sim \text{condition} * \text{comorbidity} + (1|\text{participant})$

Model 0 is the intercept-only model. Model 1 contains the within-subject factor of task condition (i.e. reward, punishment, and neutral). Model 2 additionally includes the different main psychiatric comorbidities found in the CD group. The likelihood ratio test revealed that Model 1 explains the data best (χ^2 (2) = 21.77, $p < .001$), suggesting that the presence of comorbidities did not significantly affect the learning rates in the youths with CD."

4) 25 minutes for this task seems like a long time – does data quality become an issue later in the task, do people get bored, are there breaks (i.e., aside from the condition specific hypotheses, do people get less engaged and does beta increase overall across time for people?). What does "pseudo-random" order mean?

4) Reply: We apologize for not describing the task more clearly, which was partly due to the limited number of words allowed. Actually, the task was split into three runs with short breaks in-between; we will add this information to the revised manuscript. We also apologize for not being more precise about what we mean by "pseudo-random" order. Therefore, we will include the following description to the methods section of the revised manuscript: "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."

Notably, the length of the experiment was tolerated very well by both CD patients and typical controls. However, we did not collect explicit data on task engagement. Note that our computational approach was to model and estimate beta and also alpha values across the full range of collected behavioral data. Therefore, there is no possibility to provide parameter values, such as beta, as they may have changed over time. We nevertheless believe that boredom (or 'noisy' choice making) was not an issue here (Bench & Lench 2013, *Behav. Sci.*): Since participants knew exactly by instruction that they would receive the amount of money earned at the end of the experiment, it is reasonable to assume that this was highly motivating for them to perform well from start to finish. Also, if one were to use, for instance, reaction time as an indicator of task engagement, such as slowing down over time reflects boredom in performing the task at hand, the available data show the exact opposite: in all three conditions, participants responded faster over time (see below).



5) What was the rationale for the amount of money (20 cents/20 pence) used in the task – for example, why not larger sums (e.g., 1 Eur/1 Pound). Perhaps it was enough since it elicited sensitivity to punishment across the whole sample, but curious to know where that amount came from and whether larger amounts would have been more enticing/provoking when it came to reward-driven responding?

5) Reply: We thank the reviewer for this interesting question. In fact, in the original paper by Kim et al. (2006), from which our task design was derived, the authors used $1 as incentives. We decided against using such a large amount of money as incentives for ethical and practical reasons. According to the instructions, participants received the amount of money earned at the end of the experiment in addition to the compensation for participating in this study. Participants received an average amount

of ~11€, which we believe is appropriate for participants in the age group tested here. Larger incentives would have resulted in an estimated four-fold amount of ~40-50€, which we wanted to avoid. Thus, we can only speculate whether larger incentives would have triggered greater reward-driven response 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 (Gershman 2015, *Psychon Bull Rev*.). This idea is discussed in the revised manuscript.

6) Figures 2 and 3 don't appear in the manuscript so I couldn't evaluate them or the results they present

6) Reply: We apologize for this inconvenience. For unknown reasons, Figures 2 and 3 were only available online on the journal's website for reviewers, but were not included with the downloadable manuscript. We will ensure that all figures appear in the revised manuscript.

7) It seemed like a lot of basic descriptive data for the whole sample was missing – whether from the main manuscript or from the supplement. I found myself wanting to understand overall, basic, responses both across the sample as a whole during the task (i.e., how do people perform, overall, across the task and during the different conditions), as well as within groups (perhaps that's what's in Figures 2 and 3?).

7) Reply: We kindly refer the interested reader (incl. the reviewer) to Supplementary Table 2 for descriptive behavioral data, including accuracy and reaction times, for each group and condition.

8) Could the authors expand more on the finding (supplement) that accuracy and reaction time did not vary as a function of group; I understand this is central to the premise of the paper, but it would be helpful to make it clearer what the difference is between accuracy and reaction time (which don't differ between groups) vs. learning rate parameter (which differs between groups for punishment). I guess I am asking, if accuracy (and beta) do not differ between groups, what explains the fact that CD youth are not learning to adjust the expected outcome based on new information.

8) Reply: We thank the reviewer for this intriguing question. First, however, we would like to emphasize that a lower learning rate alpha for punishment in the CD group cannot simply be interpreted as meaning that no learning occurred. Instead, it reflects that the learning process was slower and less efficient when it came to how youths with CD updated their expectation of a given outcome (i.e. the value of punishment) with newer information from trial to trial; we will make this interpretation more explicit in the revised manuscript. The reviewer is however right in pointing out that accuracy and reaction times for punishment did not differ between the CD group and TDCs, which we expected though. Both accuracy and reaction times are only rather coarse summary measures of performance (i.e. they provide values averaged over trials per condition). Since reinforcement learning (RL) is a latent neurocognitive operation, it is best quantified directly and precisely by computational indices based on RL models (such as the learning rate alpha or the exploration parameter beta). These computational indices are more sensitive to the effects of interest as they are able to capture the trial-by-trial dynamics of the learning process as it unfolds over time. This information can be found in the revised manuscript.

However, because we only have the behavioral data available in the current study, we can just speculate on the underlying mechanism(s) (i.e. contributing factors) that might explain the lower learning rate alpha for punishment in the CD group compared to TDCs. In the manuscript, we currently write that "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." Therefore, 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. This suggestion will be included in the revised manuscript.

9) Related, p. 13, line 39 onwards: it would be helpful to understand what is happening in the CD group if not learning (i.e., based on lower learning rate, alpha) during the punishment condition. It would appear that higher beta does not explain group differences?

9) Reply: We thank the reviewer for raising this interesting question. However, we would like to emphasize again that a lower learning rate alpha for punishment in the CD group cannot simply be interpreted as meaning that no learning occurred. Instead, it reflects that the learning process was slower and less efficient when it came to how youths with CD updated their expectation of a given outcome (i.e. the value of punishment) with newer information from trial to trial; as mentioned above, we will make this interpretation more explicit in the revised manuscript. Beta represents an individual's randomness in choice behavior while learning. In fact, we found no correlation between alpha and beta for punishment in either group (CD: r = -0.03, p = 0.8; TDC: r = 0.03, p = 0.74), indicating that a higher beta would not explain group differences in alpha. We will add this information to the revised manuscript.

10) I didn't understand the analysis for CU traits; I was expecting to see CU traits entered in the model with a formal interaction effect tested. Can the authors justify this approach more and/or break down the findings more clearly?

10) Reply: We apologize for not describing the analysis for CU traits and its findings more clearly. We will include the following to the methods section of the revised manuscript:

"The modeled parameters α and β from our winning model (3α3β) were then analyzed using two separate repeated-measure ANOVA models with group (CD vs. TDC), sex (male vs. female) and CU traits (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 will then include the following to the results section of the revised manuscript:

"The rmANOVA for the learning rates α revealed a significant group by condition interaction effect [F(2, 428) = 6.15, p < .01, η^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, η^2_p < .01) and the CU traits by condition interaction (p = .25, η^2_p < .01) were non-significant. All other interaction effects were non-significant as well (ps > .26, $\eta^2_p s$ < .01). Additionally, we found a significant main effect of condition [F(2, 428) = 42.61, p < .001, η^2_p = .17], but no significant main effects of group (p = .48, η^2_p < .01) or sex (p = .58, η^2_p < .01) or CU traits (p = .72, η^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).

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 ($M_{Diff(rew-pun)}$ = -0.22, 95%-*CI*[-0.25, -0.19], p < .001, BF_{10} > 100; $M_{Diff(rew-neut)}$ = -0.21, 95%-*CI*[-0.26, -0.16], p < .001, BF_{10} > 100), while punishment β and neutral β did not differ ($M_{Diff(pun-neut)}$ = 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)."

11) Can the authors explain more how this study, sample, and approach differs to their prior work in the same sample (p. 6, "using the PAL task in the largest sample of CD youths to date, we found more errors in responding to punishment"). What does the current approach do – both quantitatively (which I understand) and substantively (which I am less clear about), that adds to prior knowledge.

11) Reply: We would like to highlight that the current study differs from our prior work using the PAL task in two crucial aspects: First, in the current subsample of the FemNAT-CD cohort we used a reinforcement learning task with probabilistic feedback (e.g. incorrect responses lead to punishment only with a certain likelihood), while the PAL task traditionally uses deterministic feedback (i.e. incorrect responses always lead to punishment). Our approach of using probabilistic feedback here appears to be advantageous for studying reinforcement learning (and testing the hypotheses at hand), because this type of feedback provides a better model for the uncertainty of learning situations in real life, especially for youths who have a learning advantage in the case of probabilistic uncertainty, for instance, compared to adults or young children (e.g. Blankenstein et al. 2016, *Dev Neuropsychol.*). Uncertainty refers to ambiguity in how likely a punishing or rewarding outcome is to occur.

Second (see also our reply to point 8 above), previously we analyzed the reinforcement learning behavior in the PAL task using 'traditional' variables, such as the overall number of punishment- and reward-learning errors, but these provide only summary measures of performance. In contrast, computational indices based on reinforcement learning models (e.g. learning rate alpha) are able to capture the trial-by-trial dynamics of the learning process as it unfolds over time. By doing so, our analytic approach goes beyond the existing ones used in CD research, as we are able to estimate learning rates for punishment contingencies (versus reward or neutral ones) for each study participant individually – rather than using pre-set generic learning rates as done previously (e.g. the related work by White and colleagues). We plan to include the information in the revised manuscript on how our approach complements prior knowledge.

Reviewer: 2

The authors present a case-control behavioral study of probabilistic reward and punishment learning in 92 youths with conduct disorder (CD) vs. 130 controls, aged 9-18. They used a 70/30 probabilistic learning task without reversals, with blocked monetary reward, monetary punishment, and neutral conditions. They tested for group differences in meta-learning parameters by inverting a RW-type model using an iterative maximum a posteriori (MAP) algorithm. The dominant model was selected using Bayesian model comparison. The authors report that CD youths had a lower learning rate for punishments but not for rewards with no group differences in the [inverse] temperature or stochasticity parameter. Sampling, as well as modeling and statistical methods are appropriate and the paper is clearly written. I particularly appreciated the model and parameter identifiability analyses validating the modeling approach. The main finding is in line with earlier studies showing impaired punishment learning in CD.

Reply: We thank the reviewer for this overall positive evaluation of our work. Please find below our replies to the questions and issues raised by the reviewer.

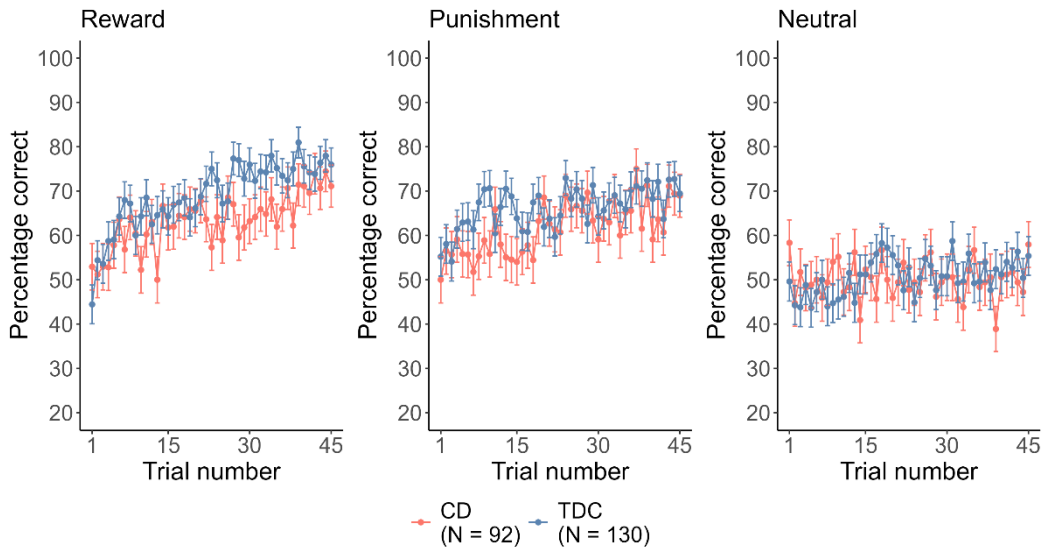
I have only a few specific comments:

1) I did not entirely understand the task design. Were the conditions blocked or interleaved?

1) Reply: We apologize for not describing the task more clearly in the main text. We kindly refer the interested reader (incl. the reviewer) to the Supplementary where we describe the task in more detail. However, we will add the following explanation to the methods section of the revised manuscript: "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."

2) While modeling is useful, it would also be helpful to see learning curves by group for a model-free corroboration.

2) Reply: We appreciate the reviewer's request to provide more data, such as learning curves, and are pleased to add a figure showing learning curves to the Supplement of the revised manuscript. In our original submission, we did not present learning curves, but we did test whether participants' choices were random or whether learning actually occurred. We quantified 'learning' as choosing the high probability cue above chance level (>50%) per condition. As expected, participants across both groups were able to learn in both the reward and punishment conditions (all *t*s>14.2, all *p*s<.001, all Cohen's *d*s>2.14), but not in the neutral condition (t=0.82, p>.41). This information can be found in the Supplement. We will also add learning curves (see below) to the revised Supplement as requested by the reviewer.



3) The inference about punishment insensitivity is broadly consistent with the data. However, the experiment and modeling are not entirely conclusive, because a lower learning rate reflects slower temporal integration of reinforcement (punishment in this case) and not necessarily a lower sensitivity to the reinforcer. For example, a model with lower punishment learning rate would also display slower extinction of the value of a punishing option. To fully dissociate punishment sensitivity (valuation) from its temporal integration (learning rate), one needs to experimentally manipulate both the magnitude and the temporal occurrence of punishment (e.g. a paradigm with varying size of punishments and reversals).

3) Reply: We thank the reviewer for bringing this interesting thought to our attention, and to some extent we agree with the reviewer. However, to be even more precise in terms of its practical meaning, a lower learning rate reflects slower updating of reinforcer value from recent outcomes; it therefore combines temporal integration (i.e. reinforcer history) with reinforcer valuation (Zhang et al. 2020, *Soc Cogn Affect Neurosci.*); we will include this definition to the revised manuscript. In line with Zhang et al. (2020, p. 699), we have therefore operationalized that a lower learning rate per reinforcer reflects a lower sensitivity to that particular reinforcer (e.g. punishment). However, we definitely agree that

additional experimental manipulation(s), including varying magnitudes of punishments and stimulus-outcome reversals, could further substantiate our findings of punishment insensitivity in youths with CD. This however would add further length to an already quite long task (~25 minutes); this should be considered. Though, we will add the reviewer's suggestion to the discussion of the revised manuscript.

4) The authors present sensitivity analyses controlling for age, IQ and callous-unemotional traits. Lacking, however, is evidence that individual differences in learning rates are specifically attributable to CD as opposed to other forms of psychopathology, which were prevalent in the CD group.

4) Reply: We thank the reviewer for pointing out this important aspect to consider. Our initial sensitivity analyses using rmANCOVAs (see Supplementary Table 3) did not include co-occurring psychopathological conditions because psychiatric comorbidities only applied to the CD group but not to the TDCs; therefore, there were no rmANCOVAs comparing the influence of comorbidities on both groups using 'group' as a factor. We are however pleased to add the results of additional regression analyses (i.e. mixed models), showing that there were no significant influences of comorbidities (incl. ADHD) on the learning rates alpha for punishment (or the two other contingencies) in the youths with CD as our primary finding. We will add the following information to the Supplement of the revised manuscript:

"To test the influence of comorbidities on the learning parameter α, we ran three linear mixed models for the CD group and compared the models with a likelihood ratio test. The models were specified as follows:

Model 0: $\alpha \sim 1 + (1|participant)$

Model 1: $\alpha \sim condition + (1|participant)$

Model 2: $\alpha \sim condition * comorbidity + (1|participant)$

Model 0 is the intercept-only model. Model 1 contains the within-subject factor of task condition (i.e. reward, punishment, and neutral). Model 2 additionally includes the different main psychiatric comorbidities found in the CD group. The likelihood ratio test revealed that Model 1 explains the data best (χ^2 (2) = 21.77, $p < .001$), suggesting that the presence of comorbidities did not significantly affect the learning rates in the youths with CD."

Reviewer: 3

This is a study using a commendably large sample of participants with CD and includes both males and females. Computational models that estimate learning rates and temperature parameters of learning are fitted separately for punishments and rewards. The findings replicate prior studies, including prior computational modelling studies, that indicate punishment learning difficulties in participants with CD. They also add to the more mixed picture of findings regarding reward processing in CD. A sample with both males and females enabled a comparison of sexes in punishment and reward processing, which was a welcome addition to the field. Although I think that the study makes an important added contribution to the literature of reinforcement learning and CD, I was not entirely convinced by the way in which the study was framed or the ultimate novelty of the findings.

Reply: We thank the reviewer for this honest summary of our work. Not surprisingly, we disagree with the assertion that our study ultimately provides no novel insights on the topic of punishment learning difficulties in youths with CD. Here are some of the main reasons that we actually see as a real strength of the current study compared to previously published work: To date, the vast majority of studies on reinforcement learning (RL) in youths with CD have been unsuited or underpowered for adequately testing for sex-by-group interaction effects as they primarily focused on predominantly male or female samples. Prior work has been further limited by relying on rather smaller-scale samples with varying selection criteria, very often including mixed groups of youths with CD (or subclinical levels of CD-

related problems) and oppositional defiant disorder (ODD) or even attention deficit hyperactivity disorder (ADHD). Here, we tested a substantial sample of boys and girls with a confirmed CD diagnosis (vs. TDCs) that even included some girls with the rare form of childhood-onset CD. To enable clear data interpretation, we did not include a mixed clinical group of participants with CD or ODD (or even ADHD) as it is still premature (or even unlikely) to suggest that the same neurocognitive mechanisms underlie the etiology of these disorders. Moreover, the entire sample was comprehensively clinically assessed and reliably diagnosed using standardized, semi-structured interviews based on DSM-IV/DSM-5 criteria that enabled us to account statistically for common psychiatric comorbidities as potential confounding factors. We were therefore able to demonstrate that the punishment learning problems are specifically attributable to CD as opposed to other forms of co-occurring psychopathology (incl. ODD, and ADHD). Finally, and crucially, is the contrast with the closely related work of White and colleagues (2013, 2014, and 2016), who also used a prediction error (PE)-based RL model type in youths with (subclinical) conduct problems or ODD performing a probabilistic passive avoidance learning (PAL) task. However, White et al. applied one single pre-set generic learning rate (alpha was set at 0.1) for all participants and contingency conditions, ignoring two crucial empirical facts, namely that learning rates can differ across individuals and the so-called "learning rate asymmetry", meaning that learning rates are typically higher for punishment than reward contingencies; this is exactly what we found here (the condition effect was at $p < .001$: punishment > reward = neutral). In fact, our analysis pipeline allowed us to set up and compare different hypothetical PE-based RL models (i.e. learning rates and exploration parameters may or may not be similar for different contingencies) to find a 'winning' model that fits the participants' learning behavior most optimally. To accomplish it, we applied cutting-edge analytic methods, including hierarchical expectation maximization and Bayesian model comparison as well as parameter recovery and model identifiability procedures. As expected, we found a 'winning' model that included three separate learning rates and exploration parameters, one for each contingency. Consequently, our analytic approach certainly goes beyond the existing ones used in this line of research (incl. White et al.), as we were able to estimate learning rates for punishment contingencies (versus reward or neutral ones) for each study participant individually – rather than using pre-set generic learning rates as done previously. Overall, we believe that our findings presented in this (revised) manuscript do actually replicate, but they are also novel and considerably and reliably extent, prior findings about punishment learning difficulties in youths with CD.

Please find below our additional replies to the questions and issues raised by the reviewer.

1) I think that it is somewhat of a 'red herring' to set the study up to, in part, test the 'reward dominance' hypothesis. The original study on which this hypothesis was based has already been criticised based on the fact that it is not possible to discern whether the performance on the task of that study was driven by reward or punishment processing. For example, Blair and colleagues have subsequently done a substantial amount of work to more precisely isolate aspects of reinforcement processing that are compromised in individuals with CD. Furthermore, a number of studies using different reward learning tasks have painted a fairly mixed picture regarding reward processing in CD. As such, I am not convinced that the 'reward dominance' hypothesis has quite the prominence in the field that the authors of the paper present.

1) Reply: While we agree with the reviewer's remark that the reward dominance studies by O'Brien et al. (1994, 1996) have been rightly criticized for the reasons stated above, we do not agree that the study of reward dominance can be considered a red herring or not legitimate. It may be that the term 'reward dominance' is no longer quite as prominent in the field, the premise behind the reward dominance hypothesis of a CD-related reward-seeking learning style in the presence of punishment cues is still quite alive in this line of research, but perhaps better known now under the term 'reward hypersensitivity' (e.g. Byrd et al. 2018, *Dev Cogn Neurosci.*). Therefore, in the revised manuscript, we will adjust the scope of our approach, that is, the study of reward dominance – or reward hypersensitivity – versus punishment hyposensitivity, to better highlight the use of the different terms in the current relevant literature. By doing so, we will broaden the cited studies by including and

discussing more (recent) literature on reward processing in CD. In this context, the reviewer is correct in pointing out that there is a rather mixed picture regarding aberrant reward processing in CD (i.e. there is evidence for or against the presence or absence of reward processing abnormalities). This could be due to several reasons (e.g. task-related; Richards et al. 2013, *Neurosci Biobehav Rev.*), one of which we mentioned in our very first response to this reviewer, namely the investigation of predominantly male, clinically mixed (i.e. CD and/or ODD and/or ADHD), and sometimes even subclinical (i.e. CD-problem) samples. Scientifically, this cannot mean that we should not test the idea of reward dominance or hypersensitivity, especially given our well-characterized sample of girls and boys with and without a confirmed CD diagnosis, which is certainly ideal for answering this particular research question and testing the hypotheses at hand.

2) The current study adds to the mixed picture of reward learning in CD, but because it only examines reward learning under particular task conditions, it is perhaps not appropriate to conclude that reward learning appears intact in CD. It might be more appropriate to conclude that under conditions where substantially different reward outcomes are associated with visually distinct stimuli and where these associations do not change, individuals with CD demonstrate comparable reward learning to TDCs.

2) Reply: It may be that the unimpaired reward learning ability in CD found in the current study is related to the probabilistic RL task used, although this is only speculation at this point, and follow-up studies using other task probes, e.g. a reversal learning task, are warranted. We will include this aspect in the discussion of the revised manuscript.

3) I am also not convinced that the task used in this study is optimised to test the 'reward dominance' hypothesis, if this is what the authors really want to test. Correctly discerning that one particular, clearly visually distinct object is less rewarding than another object is very different from having reward contingencies gradually change or choosing an object with a relatively low reward probability that e.g. requires effort to obtain. Either of these set ups would appear better suited to testing 'reward dominance' (as opposed to the ability to learn about rewards in the first place), although the former only if not in the context of punishment.

3) Reply: We do not entirely agree with the reviewer that our task may be unsuitable for testing reward dominance (or hypersensitivity) as we operationalized it in the current study: better reward learning at the expense of punishment learning (i.e. one would expect a higher learning rate for reward to be accompanied by a lower learning rate for punishment). As the reviewer rightly points out (see above), it was impossible to discern whether performance on the original 'reward dominance' task was driven by reward or punishment processing or both, but our task design does allow us to separate reward and punishment processing. However, we agree that additional experimental manipulation(s), including varying magnitudes of reward and punishment as well as stimulus-outcome reversals (or others, such as effort-based tasks), could further substantiate our findings of punishment insensitivity but not reward dominance in youths with CD. We will add this aspect to the discussion of the revised manuscript.

4) I was also surprised that the authors predicted a correlation between CU traits and greater learning impairment, while citing a 2009 review to support this claim. A number of studies, including several by Blair and colleagues, have been published after 2009 and have not found CU traits to be predictive of learning impairments or atypical neural processing during learning, but have instead found these to characterise those with disruptive behaviour disorders overall (e.g. White et al., 2013; Hwang et al., 2018). Again, this is setting up the study to test a prediction that does not seem entirely warranted given the current evidence base.

4) Reply: We would like to emphasize that, like the literature on reward processing, the findings on greater RL impairments in CD youths with high CU traits are similarly mixed, both favoring and

disfavoring (reviewed in Byrd et al. 2014, Child Fam Psychol Rev., or Frick et al. 2014, J Child Psychol Psychiatry., page 534: *"[…] children and adolescents with serious conduct problems and CU traits also show distinct cognitive characteristics compared to other youths with conduct problems. Specifically, children and adolescents with CU traits are more likely to show an insensitivity to punishment cues using tasks in which a reward-dominant response set is primed […]."*). However, the reviewer seems to cite studies from Blair's group exclusively, ignoring other relevant work in favor of our prediction. We believe that our prediction is warranted given all available (albeit mixed) evidence on the subject and should therefore be tested. Please note that we have adjusted this particular prediction in response to point 2 by reviewer #1 (see above). It now reads: "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) (14), we predicted a positive correlation between CU traits and aberrant punishment learning performance."

5) The authors suggest at the end of the study that the study of reinforcement learning should be extended to the social domain. No doubt more work in this area is needed, but there is already a study from Blair's group that has compared social and non-social reward processing in children with CD/ODD (Hwang et al., 2018). It seems like an omission not to cite this.

5) Reply: We certainly agree with the reviewer and will add this reference to the revised discussion. It should be noted, however, that the Hwang sample included only 14 youths with CD.

6) Although the large sample of individuals with CD in this study is welcome, the findings themselves might have been anticipated given several replications (across prior, smaller studies) of difficulties in punishment learning in CD (including the authors' own study on the same sample that did not include computational modeling). The reward learning findings of this study add to the currently mixed picture of reward learning and CD and underscore the importance of exploring reward processing using different task probes, but are currently discussed relatively simplistically.

6) Reply: We agree with the reviewer that our finding of unimpaired reward learning abilities in youths with CD adds to the mixed evidence currently available on this topic. Therefore, as mentioned above, in the revised manuscript we will expand on the cited studies by including and more thoroughly discussing additional (recent) literature on reward processing in CD, including the suggestion to examine reward processing in CD with different task probes than the one we used here.

7) Although the computational modeling approach is a helpful addition and provides a more nuanced picture of reinforcement processing in CD, I am not sure that in this case this approach has added a huge amount to what we already know.

7) Reply: Please see our first response to this reviewer, in which we explain in detail to which our approach complements previous knowledge. Thus, we consider this work to be extremely important and anticipate that the novelty and reliability of our results as well as their clinical implications will be of considerably interest to researchers in the relevant field as well as to a broader readership.

Reviewer: 4

The authors investigated 92 youths with CD and 130 controls using computerized modelling to test reward and punishment contingencies within a probabilistic learning task. The results suggest a greater sensitivity to reward than punishment in youths with CD, independent of CU traits. The authors are to be commended on their well-designed and well-written paper. Also, they succeeded in collecting a fine number of youths with CD and control subjects, and also investigated sex differences and the influence

of CU traits. Although the research is relevant and timely, my impression was that it better fits to a more specialized journal.

Reply: We thank the reviewer for this overall positive evaluation of our work. Although we still believe that our work, the novelty of our findings, and their clinical implications would be of considerably interest to the broad readership of *The American Journal of Psychiatry*, we now plan to submit a revised manuscript to the *Journal of the American Academy of Child & Adolescent Psychiatry* as a more specialized journal, as suggested by the reviewer.

Reviewer: 5

This study evaluates differential associations of reward dominance and punishment sensitivity among youth presenting with conduct disorder. The manuscript is well-written, of sound methodology, and addresses an interesting topic with important clinical implications. My only suggestion is that the study authors may wish to consider including a sentence or two that discusses implications that the study design (i.e., the monetary probabilistic RL task), which does not account for potential temporal effects on incentive processing (e.g., anticipation vs receipt), may have on the interpretation of the study findings. I thank the editor and study authors for allowing me to take part in this review.

Reply: We thank the reviewer for this overall positive evaluation of our work and the suggestion to discuss implications with regard to possible temporal effects on reinforcer processing (i.e. monetary reward anticipation vs. receipt) that may have influenced our findings. For example, while we have identified a primary insensitivity to punishment (but not to reward) in youths with CD, it remains difficult to disentangle whether such a deficit is due to hyporeactivity to a cue (which triggers the expectation of potential punishment) or the actual receipt of punishment or both. We will add this limitation to the revised discussion.

Impaired Punishment Learning in Conduct Disorder

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

Authors' Affiliations:

Mr. Elster, Prof. Roessner and Dr. Kohls are with the Department of Child and Adolescent Psychiatry, Faculty of Medicine, TU Dresden, Dresden, Germany.

Drs. Pauli, De Brito and Lockwood are with the Centre for Human Brain Health, School of Psychology, University of Birmingham, Birmingham, UK. Dr. Lockwood is also with the Department of Experimental Psychology, University of Oxford, Oxford, UK and with the Institute for Mental Health, School of Psychology, University of Birmingham, Birmingham, UK.

Mrs. Baumann and Prof. Konrad are with the Child Neuropsychology Section, Department of Child and Adolescent Psychiatry, Psychosomatics and Psychotherapy, University Hospital RWTH Aachen, Aachen, Germany. Prof. Konrad is also with the JARA-Brain Institute II, Molecular Neuroscience and Neuroimaging, RWTH Aachen and Research Centre Juelich, Juelich, Germany.

Dr. Fairchild is with the Department of Psychology, University of Bath, Bath, UK.

Prof. Freitag is with the Department of Child and Adolescent Psychiatry, Psychosomatics and Psychotherapy, University Hospital Frankfurt, Goethe University, Frankfurt am Main, Germany.

Prof. Brazil is with the Radboud University, Donders Institute for Brain, Cognition and Behaviour, Nijmegen, The Netherlands.

Corresponding Author: Erik M. Elster, M.Sc., Department of Child and Adolescent Psychiatry, Faculty of Medicine, TU Dresden, Fetscherstr. 74, 01307 Dresden, Germany.

Email: erik.elster@ukdd.de

Running Head: Punishment Insensitivity in Conduct Disorder

Disclosure of Interest: C.M.F receives royalties for books on attention-deficit/hyperactivity disorder and autism spectrum disorder. She has served as consultant to Desitin and Roche. The remaining authors have declared that they have no competing or potential conflicts of interest.

Funding: This study was funded by the European Commission's Seventh Framework Programme (FP7/2007-2013) under Grant Agreement no.602407 (FemNAT-CD, coordinator: Christine M. Freitag).

Keywords: Conduct Disorder; Decision Making; Punishment; Reinforcement, Psychology; Computational Modeling

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 & 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% females) 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 supplement 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 & β, neutral α & β
3. 3α3β: reward α & β, punishment α & β, neutral α & 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).

[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, η^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, η^2_p<.01) and the CU traits by condition interaction (p=.25, η^2_p<.01) were non-significant. All other interaction effects were non-significant as well (ps>.26, η^2_ps<.01). Additionally, we found a significant main effect of condition [F(2, 428)=42.61, p<.001, η^2_p=.17], but no significant main effects of group (p=.48, η^2_p<.01) or sex (p=.58, η^2_p<.01) or CU traits (p=.72, η^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), 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 ($M_{Diff(rew-pun)}$=-0.22, 95%-CI[-0.25, -0.19], p<.001, BF_{10}>100; $M_{Diff(rew-neut)}$=-0.21, 95%-CI[-0.26, -0.16], p<.001, BF_{10}>100), while punishment β and neutral β did not differ ($M_{Diff(pun-neut)}$=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, $\eta^2_p s$<.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,[37] the study of social reinforcement, such as social punishment,[38,39] 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. Pauli R, Lockwood P. *The Computational Psychiatry of Antisocial Behaviour and Psychopathy*. PsyArXiv; 2022. doi:10.31234/osf.io/mqbvu

38. 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

39. 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(f/m)	37/55	69/61	$\chi^2(1)=3.07$, $p=.08$
Age(years)	14.8 (2.3)	13.8 (2.8)	$t(215)=2.68$, $p<.01$
Estimated IQ	93.1 (11.6)	102.4 (11.0)	$t(189)=-5.92$, $p<.001$
CU traits (YPI subscales)	33.1 (7.9)	28.5 (6.5)	$t(171)=4.63$, $p<.001$
LPE specifier	37 (40,2%)	28 (21,5%)	$\chi^2(1)=9.2$, $p<.01$
Comorbidities n (%)			
ODD	66 (72)	N/A	
ADHD	47 (51)	N/A	
MDD	29 (32)	N/A	
PTSD	12 (13)	N/A	
SUD	7 (8)	N/A	

GAD	5 (5)	N/A

Note: CD=conduct disorder; TDC=typically developing controls; f/m=female/male; IQ=intelligence quotient; LPE=with limited prosocial emotions; N/A=not applicable; ODD=oppositional defiant disorder; ADHD=attention deficit and hyperactivity disorder; MDD=major depressive disorder; PTSD=post-traumatic stress disorder; SUD=substance use disorder; GAD=generalized anxiety disorder. 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.

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, *=$p \leq .05$, **=$p \leq .01$, ***=$p \leq .001$

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 & 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% females) 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 supplement 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 & β, neutral α & β
3. 3α3β: reward α & β, punishment α & β, neutral α & 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).

[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, η^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, η^2_p<.01) and the CU traits by condition interaction (p=.25, η^2_p<.01) were non-significant. All other interaction effects were non-significant as well (ps>.26, η^2_ps<.01). Additionally, we found a significant main effect of condition [F(2, 428)=42.61, p<.001, η^2_p=.17], but no significant main effects of group (p=.48, η^2_p<.01) or sex (p=.58, η^2_p<.01) or CU traits (p=.72, η^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), 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 ($M_{Diff(rew-pun)}$=-0.22, 95%-CI[-0.25, -0.19], p<.001, BF_{10}>100; $M_{Diff(rew-neut)}$=-0.21, 95%-CI[-0.26, -0.16], p<.001, BF_{10}>100), while punishment β and neutral β did not differ ($M_{Diff(pun-neut)}$=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, $\eta^2_p s$<.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,[37] the study of social reinforcement, such as social punishment,[38,39] 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. Pauli R, Lockwood P. *The Computational Psychiatry of Antisocial Behaviour and Psychopathy*. PsyArXiv; 2022. doi:10.31234/osf.io/mqbvu

38. 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

39. 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: CD=conduct disorder; TDC=typically developing controls; f/m=female/male; IQ=intelligence quotient; LPE=with limited prosocial emotions; N/A=not applicable; ODD=oppositional defiant disorder; ADHD=attention deficit and hyperactivity disorder; MDD=major depressive disorder; PTSD=post-traumatic stress disorder; SUD=substance use disorder; GAD=generalized anxiety disorder. 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, *=$p \leq .05$, **=$p \leq .01$, ***=$p \leq .001$

Figure 1

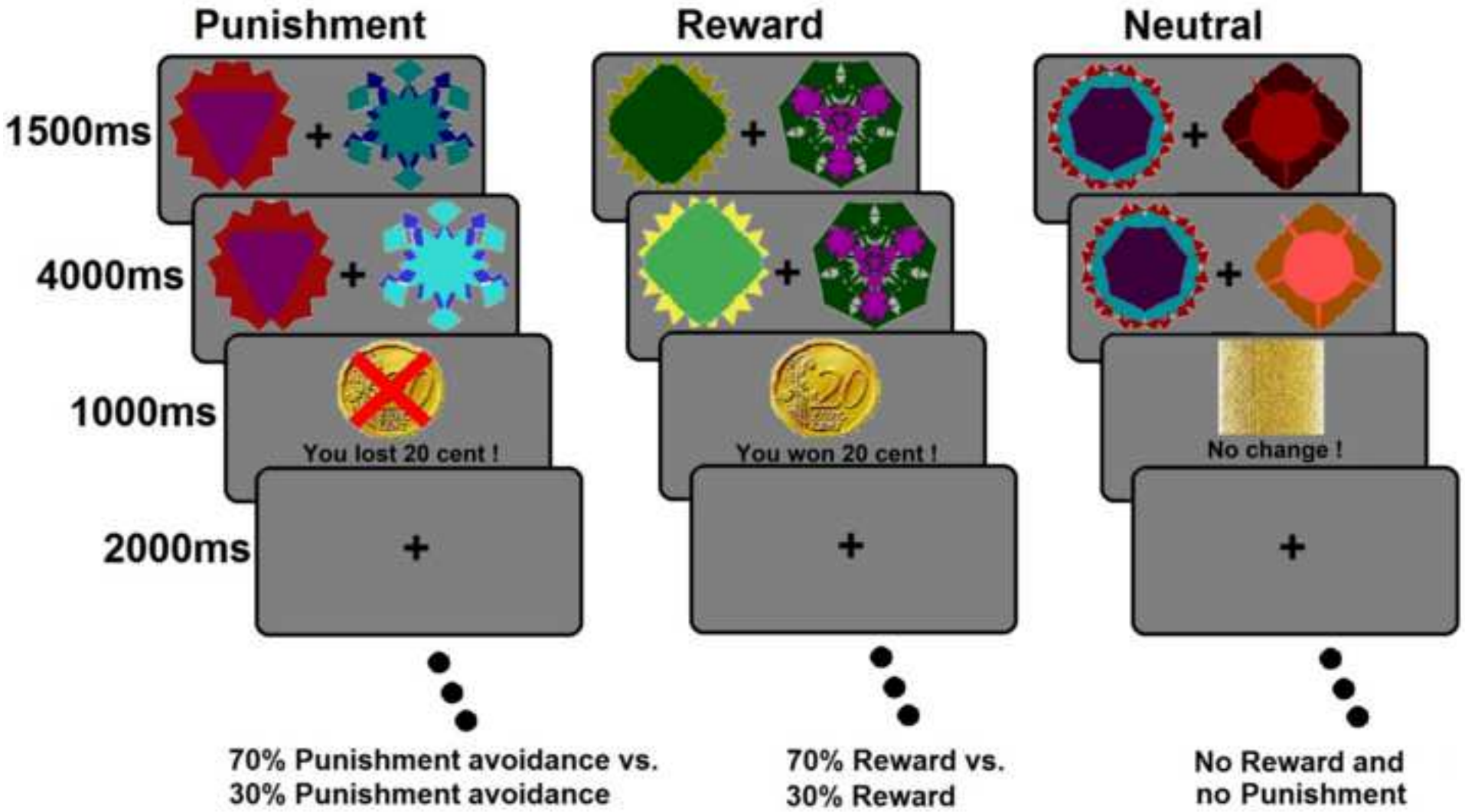


Figure 2

Click here to access/download;Figure;figure2.tiff

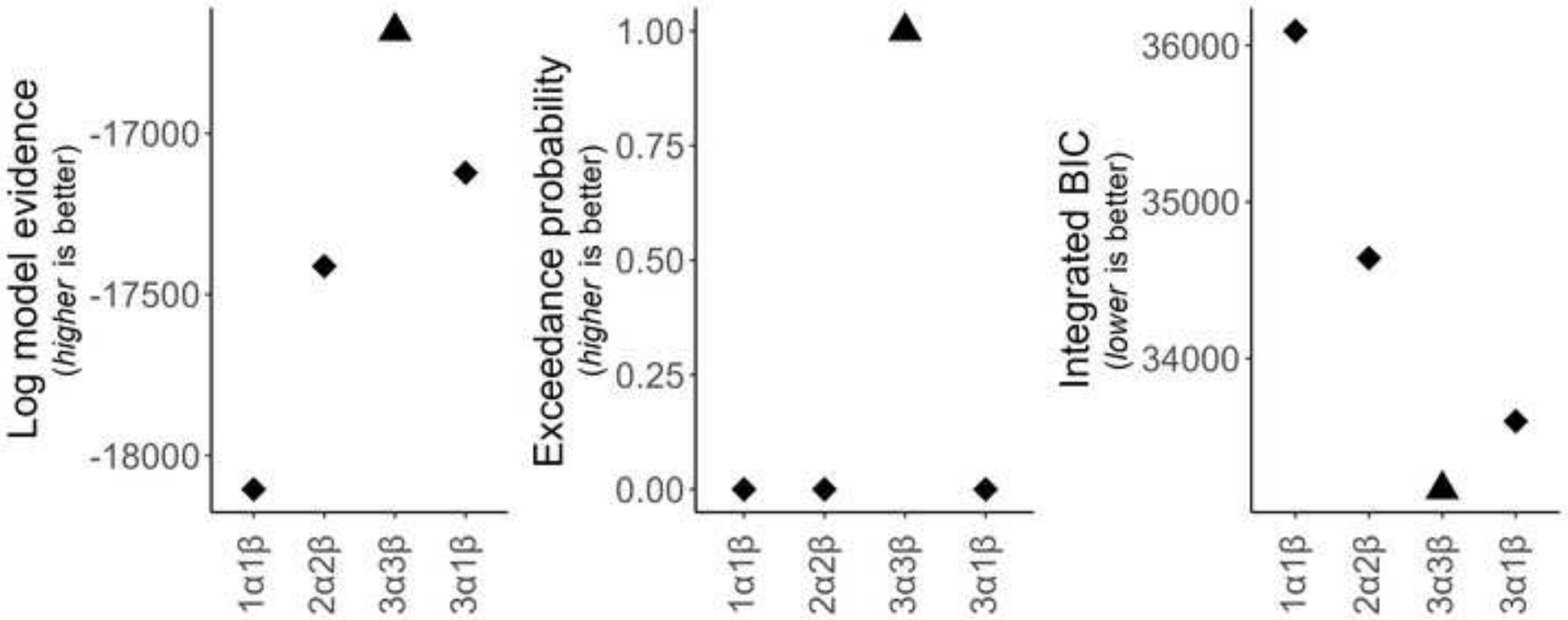
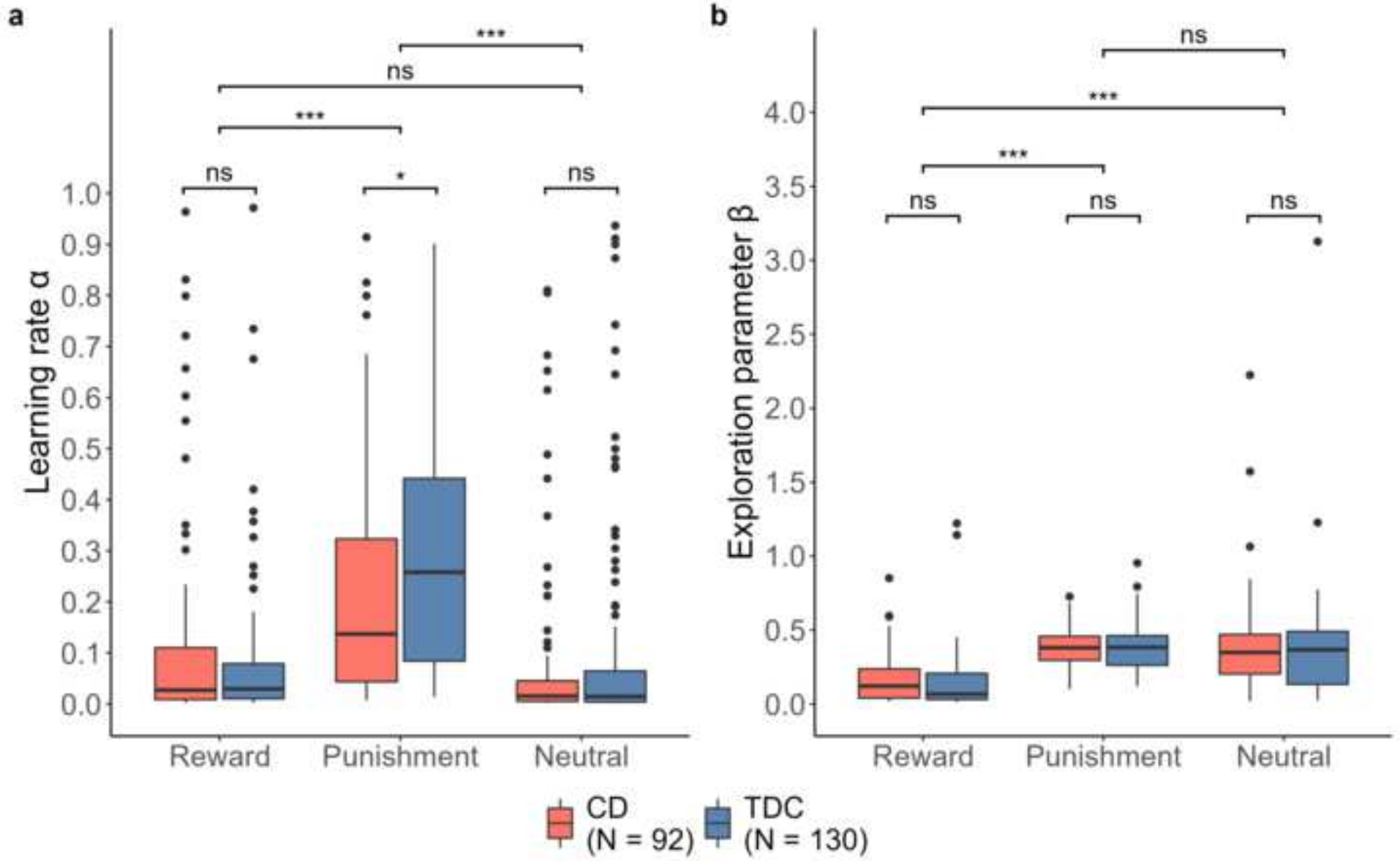


Figure 3

Click here to access/download;Figure;figure3.tiff



Supplementary Information for

"Impaired Punishment Learning in Conduct Disorder"

Method

Participants

The K-SADS-PL interview was administered by trained staff separately to participants and their caregivers, and clinical summary ratings were achieved to determine group allocation and to identify possible comorbid psychiatric diagnoses. Inter-rater reliability (IRR; N=75) of current CD was high (Cohen's kappa=.91, 95% agreement). Full-scale IQs were estimated using the vocabulary and matrix reasoning subtests of the Wechsler Intelligence Scale for Children-Fourth Edition, or the Wechsler Intelligence Scale for Adults-Fourth Edition.[1,2] The UK site used the Wechsler Abbreviated Scale of Intelligence.[3] The internal consistency and test-retest reliability for the two subtests were reported to be excellent (α>.90).[4] The CU trait subscale showed good internal consistency (Cronbach's α=.81).

Task

We used a monetary probabilistic RL task, originally described by Kim et al.[5] and modified by Baumann et al.,[6] to measure RL capabilities based on reward, punishment, and neutral, non-reinforcing contingencies (Figure 1). Behavioral data were collected as part of a parallel fMRI study. Note that we did not analyze the corresponding neuroimaging data here in order to have as large a sample as possible for this first study of its kind to date, as it is expected that ~25% or more of the participants will not provide analyzable (f)MRI data. We plan to address this important next scientific step in a follow-up project. Behavioral data collection and stimulus presentation were controlled by the MATLAB R2014a software.

Trials started with the presentation of a pair of cue stimuli (i.e., fractals) side-by-side. Each pair marked the onset of one of three different trial types (i.e., conditions): reward (i.e.,

monetary gain), punishment (i.e., monetary loss), and neutral outcome (i.e., neither monetary gain nor loss). Throughout the task, occurrence of trial types was pseudo-randomized (i.e., no trial type occurred twice or more in a row but all conditions were tested equally in sum), and the assignment of the three fractal pairs to the different conditions was fully counterbalanced across all participants. Participants were instructed to select one of the two simultaneously presented fractals by pressing the left or right key on a button box, placed in their right hand, and keys corresponded to the location of the two cues presented on the screen (i.e., left or right of a fixation cross). 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 blank screen with a crosshair in the center).

Statistical Analyses for Demographic Data, Clinical Variables, Reaction Time, and Accuracy

We compared the two groups on demographic and clinical variables using χ^2- and independent t-tests. At the task level, reaction times for correct choices (RTs in ms) and accuracy (in %, i.e. choosing the cue with the high probability of 70%) were analyzed using repeated-measure ANOVA models for each dependent variable with group (CD vs. TDC) and sex (male vs. female) as between-subjects factor, and condition (reward vs. punishment vs. neutral) as within-subjects factor, followed by Bonferroni-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 analyses. 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.[7]

Computational RL Modeling, Model Fitting and Comparison Procedure

Four different probabilistic reinforcement learning (RL) models were constructed. For each model, the preferable outcome was coded as *1* and the unpreferable outcome as *0*. First, we constructed a basic reinforcement learning model, in which learning was captured by a single learning rate parameter α and a single exploration parameter β, which captures random choices (or other invariances in choice behavior). In this model, the expected value *V* of a response on trial *t* is updated with a reward prediction error *PE* scaled by the learning rate *α*, where the prediction error is the experienced discrepancy between the outcome *r (1 or 0)* and the expected value.

If choosing stimulus A:	*If choosing stimulus B:*
$V_a(t{+}1) = V_a(t) {+} \propto \cdot PE_a(t)$	$V_b(t{+}1) = V_b(t) {+} \propto \cdot PE_b(t)$
$V_b(t{+}1) = V_b(t)$	$V_a(t{+}1) = V_a(t)$

with:

$$PE_i(t) = r(t) - V_i(t)$$

and i = a or b depending on the choice

Eq.1: basic model

The expected values are then converted to response probabilities using the Softmax equation, where the exploration parameter β adds noise:

$$probability\ of\ observed\ response\ P_i(t) = \frac{exp\frac{V_i(t)}{\beta}}{exp\frac{V_a(t)}{\beta} + exp\frac{V_b(t)}{\beta}}$$

Eq.2: softmax function

Model fitting and comparison were conducted in MATLAB 2020b.[8] We used an iterative maximum a posteriori (MAP) approach for all model fitting, in line with previous work using reinforcement learning models.[9–12] First, we initialised Gaussian distributions as uninformative priors with a mean of 0.1 (plus noise) and variance of 100. Next, during the expectation step, we estimated the model parameters for each participant using maximum likelihood estimation (MLE), calculating the log-likelihood of the participants' set of responses given the model being fitted. We then computed the maximum posterior probability estimate, given the participants' responses and the prior probability from the Gaussian distribution, and recomputed the Gaussian distribution over parameters during the maximisation step. These alternating expectation and maximisation steps were repeated iteratively until convergence of the posterior likelihood, or for a maximum of 800 iterations. Bounded free parameters were transformed from the Gaussian space into native model space using link functions (e.g., a sigmoid function for learning rates) to ensure accurate parameter estimation near the bounds.

For model comparison, we calculated the Laplace approximation of the log model evidence (more positive values indicating a better model fit[13]; in a random-effects analysis using the spm_BMS routine from SPM 12 (revision 7487, see https://www.fil.ion.ucl.ac.uk/spm/software/spm12/, and [14]). 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.

Supplementary Table 1. *Model comparison results*

		α			β			lme	BIC_{int}	XP	R^2
		R	P	N	R	P	N				
Model 1	1α1β		α			β		-18105	36091	0.00	16%
Model 2	2α2β	α_{RP}		α_N	β_{RP}		β_N	-17412	34641	0.00	21%
Model 3	**3α3β**	$\boldsymbol{\alpha_R}$	$\boldsymbol{\alpha_P}$	$\boldsymbol{\alpha_N}$	$\boldsymbol{\beta_R}$	$\boldsymbol{\beta_P}$	$\boldsymbol{\beta_N}$	**-16681**	**33164**	**> 0.99**	**26%**
Model 4	3α1β	α_R	α_P	α_N		β		-17122	33599	0.00	24%

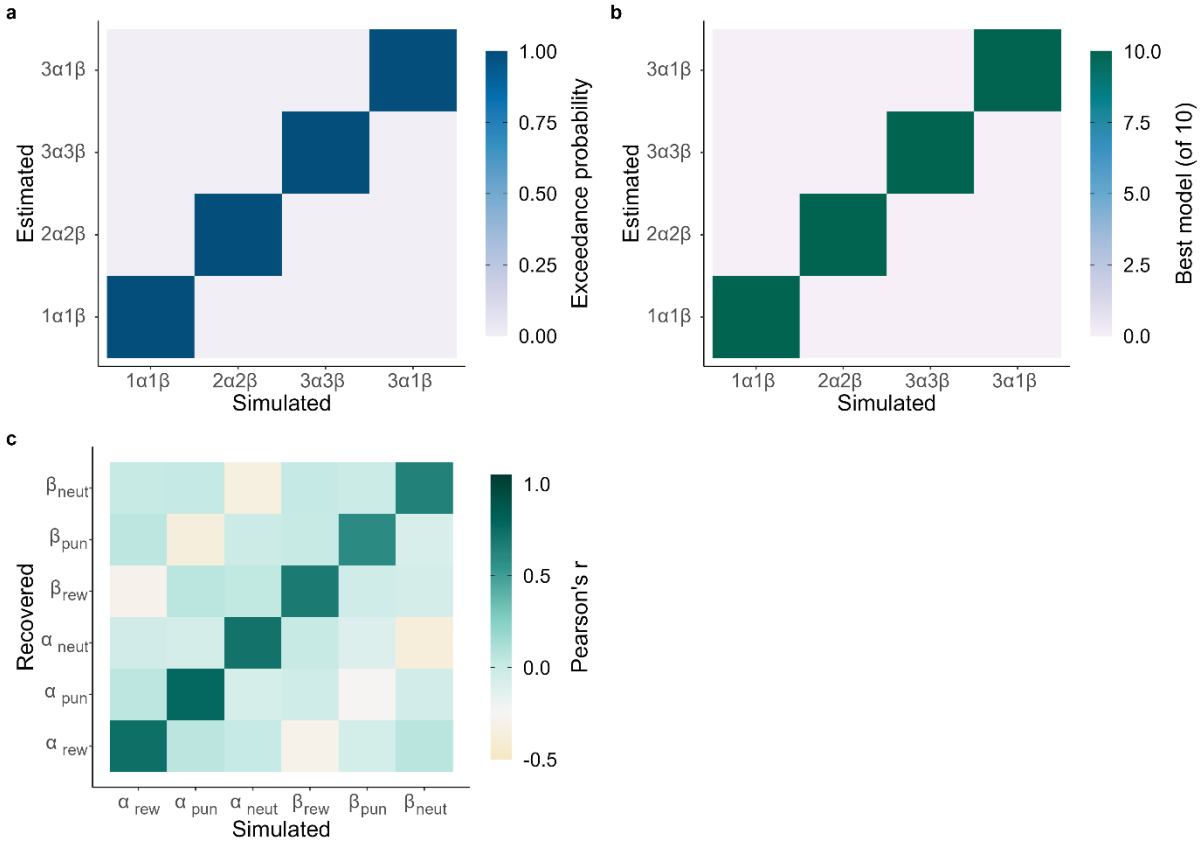
Note: R: Reward, P: Punishment, N: Neutral, lme: log model evidence, BIC_{int}: integrated Bayesian Information Criterion, XP: exceedance probability. The winning model is in ***bold****.*

Model Identifiability and Parameter Recovery

We used a model identifiability procedure to ensure that the reinforcement learning models were dissociable from each other, and a parameter recovery procedure to ensure that the parameters from the winning model were dissociable.[9] For the model identifiability procedure, we simulated participant's responses for each model, using a range of parameter values within the observed range from the real data. Our four models were then fitted to the simulated responses, using the MAP procedure, repeating each fitting 10 times. We then created confusion matrices for mean exceedance probability and for the number of times each model won, to check that for each model and its simulated data, the winning model was the

one that had been used to generate the data. This procedure confirms that the winning model (3α3β) is reliably associated with a different pattern of responses from the competing models.

For the parameter recovery procedure, we simulated participant response data only for the winning model, using a range of parameter values between the minimum and maximum possible values for that parameter. Data were simulated for 250 synthetic participants. The winning model was then fitted again to its simulated data using the MAP procedure, and correlations between the parameters used to simulate the data and the recovered parameters (estimated from the simulated data) were checked for correspondence. All parameters were recoverable, as indicated by positive correlations between true and fitted parameter values ranging from *r*=.59 to *r*=.78.



SFigure 1. a Model identifiability average exceedance probability confusion matrix. **b** Model identifiability best model selection confusion matrix. **c** Parameter recovery performed on data simulated by the wining model 3α3β.

Additional Results

Task Performance: Reaction Time, Accuracy, and Learning

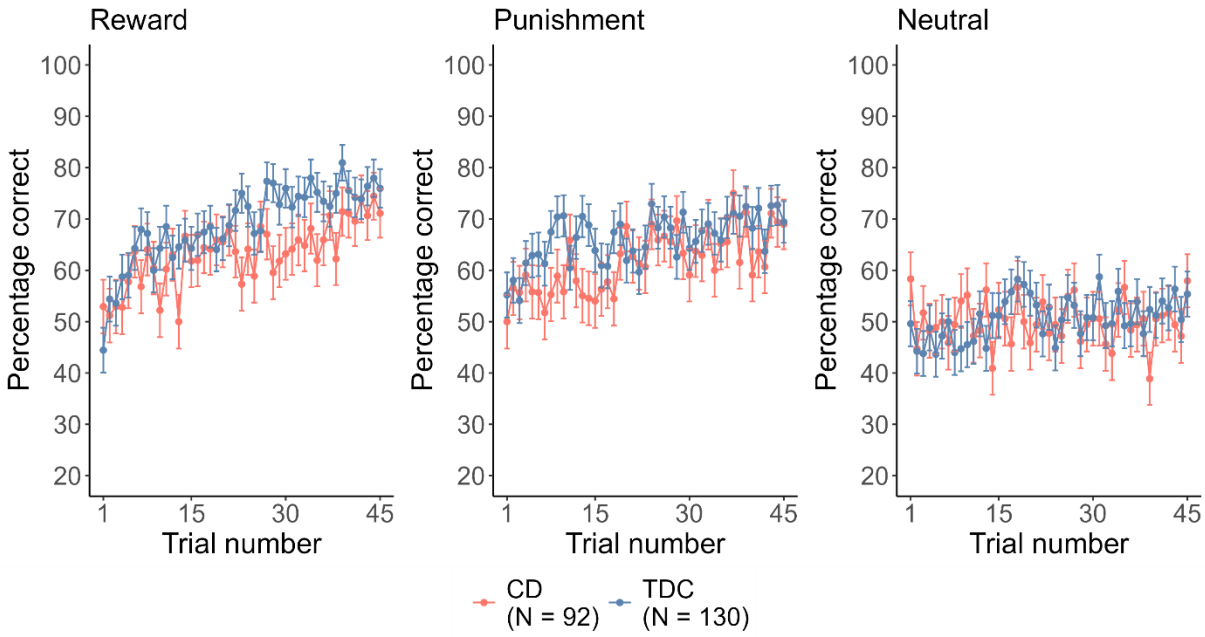
The rmANOVA for RTs (in msec) revealed a significant main effect of condition [F(2, 436)=137.5, p<.001, η^2_p=.39, BF_{10}>100], but no significant effects of group (p=.14, η^2_p=.01, BF_{01}=3.13) or sex (p=.051, η^2_p=.02, BF_{01}=0.77). The post-hoc comparisons showed the fastest RTs for reward ($M_{Diff(rew-pun)}$ =-161.0, *95%-CI*[-181.0, -141.0], p<.001, BF_{10}>100), followed by the neutral condition ($M_{Diff(rew-neut)}$=-114.0, *95%-CI*[-135.0, -93.9], p<.001, BF_{10}>100), and the slowest RTs for punishment ($M_{Diff(pun-neut)}$=46.4, *95%-CI*[30.4, 62.4], p<.001, BF_{10}>100).

The rmANOVA for accuracy (i.e., choosing the high probability cue in %) revealed a significant main effect of condition [F(2, 436)=23.3, p<.001, η^2_p=.10, BF_{10}>100], but again no effect for group (p=.10, η^2_p=.01, BF_{01}=2.66) or sex (p=.86, η^2_p<.01, BF_{01}=10.21). The post-hoc comparisons showed that accuracy was significantly higher for both the reward and punishment conditions relative to the neutral condition ($M_{Diff(rew-neut)}$=16%, *95%-CI*[11%, 21%], p<.001, BF_{10}>100; $M_{Diff(pun-neut)}$=13.8%, *95%-CI*[9.5%, 18.1%], p<.001, BF_{10}>100), while the reward and punishment conditions did not differ ($M_{Diff(rew-pun)}$=2.2%, *95%-CI*[-2.4%, 6.8%], p>.99, BF_{01}=8.6).

Note that all interaction effects were non-significant in both rmANOVAs (ps>.26, η^2_ps<.01, BFs_{01}>1.63), suggesting that the different reinforcement conditions similarly affected reaction time and accuracy in both groups, and regardless of sex, when using these 'traditional' indices of summary performance measures (see ST2).

We then examined whether participants were able to learn in the different task conditions. We quantified 'learning' as choosing the high probability cue above chance level (>50%) per condition. As expected, participants across both groups were able to learn in both the reward

and punishment conditions (all *t*s>14.2, all *p*s<.001, all Cohen's *d*s>2.14), but not in the neutral condition (*t*=0.82, *p*>.41).



SFigure 2. Learning curves for all conditions (i.e., reward, punishment and neutral contingency).

To test the influence of comorbidities on the learning parameter α, we ran three linear mixed models for the CD group and compared the models with a likelihood ratio test. The models were specified as follows:

Model 0: $\alpha \sim 1 + (1|\text{participant})$

Model 1: $\alpha \sim \text{condition} + (1|\text{participant})$

Model 2: $\alpha \sim \text{condition} * \text{comorbidity} + (1|\text{participant})$

Model 0 is the intercept-only model. Model 1 contains the within-subject factor of task condition (i.e. reward, punishment, and neutral). Model 2 additionally includes the different main psychiatric comorbidities found in the CD group. The likelihood ratio test revealed that Model 1 explains the data best ($\chi^2(2) = 21.77$, $p < .001$), suggesting that the presence of comorbidities did not significantly affect the learning rates in the youths with CD.

Supplementary Table 2. *Task performance of each group in the three different conditions of the probabilistic RL task.*

	CD	TDC
Reaction time (in msec.)		
Reward	832 (144)	841 (142)
Punishment	981 (169)	1010 (175)
Neutral	935 (162)	964 (162)
Accuracy (in %)		
Reward	63.1 (31.6)	68.8 (31.9)
Punishment	61.7 (16.0)	66.1 (15.5)
Neutral	49.9 (25.6)	50.8 (28.8)

Note: CD = conduct disorder; TDC = typically developing controls. Accuracy = correctly choosing the high probability cue

Supplementary Table 3. *Results of the additional sensitivity anlyses, i.e., rmANCOVA, controlling for potential dimensional confounders.*

	Age	IQ	CU traits
Learning parameter α		p (η^2_p)	
Condition[a]	< .001 (0.18)	< .001 (0.183)	< .001 (0.18)
Condition-by-group[a]	.002 (0.029)	.01 (0.021)	.002 (0.028)
Exploration parameter β		p (η^2_p)	
Condition[a]	< .001 (0.234)	< .001 (0.237)	< .001 (0.237)
Condition-by-group[b]	.817 (< 0.001)	.606 (0.002)	.695 (0.001)

Note: IQ = (estimated) Intelligence Quotient; CU traits = callous-unemotional traits (YPI).
[a]This effect was significant in the primary rmANOVA.
[b]This effect was non-significant in the primary rmANOVA.

Supplementary Table 4. *Additional R packages*

R package	Description
ggprism[15]	Adds significance indicators to ggplot
gridExtra[16]	Options to arrange graphics
pacman[17]	R package management tool
psych[18]	Provides more statistical functions
readxl[19]	For reading .xlsx files
tidyverse[20]	Contains useful functions and more packages, e.g., ggplot and dplyr

Supplemental References

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

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

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

4. McCrimmon AW, Smith AD. Review of the Wechsler Abbreviated Scale of Intelligence, Second Edition (WASI-II). *Journal of Psychoeducational Assessment*. 2013;31(3):337-341.

5. Kim H, Shimojo S, O'Doherty JP. Is Avoiding an Aversive Outcome Rewarding? Neural Substrates of Avoidance Learning in the Human Brain. *PLOS Biology*. 2006;4(8):e233. doi:10.1371/journal.pbio.0040233

6. 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. *The World Journal of Biological Psychiatry*. Published online November 15, 2021:1-12. doi:10.1080/15622975.2021.1995809

7. Marsman M, Wagenmakers EJ. Bayesian benefits with JASP. *European Journal of Developmental Psychology*. 2017;14(5):545-555. doi:10.1080/17405629.2016.1259614

8. MATLAB. *Version 9.9.0 (R2020b)*. Natick, Massachusetts: The MathWorks Inc.; 2020.

9. Cutler J, Wittmann MK, Abdurahman A, et al. Ageing is associated with disrupted reinforcement learning whilst learning to help others is preserved. *Nat Commun*. 2021;12(1):4440. doi:10.1038/s41467-021-24576-w

10. Huys QJM, Cools R, Gölzer M, et al. Disentangling the Roles of Approach, Activation and Valence in Instrumental and Pavlovian Responding. Rangel A, ed. *PLoS Comput Biol*. 2011;7(4):e1002028. doi:10.1371/journal.pcbi.1002028

11. Lockwood PL, Klein-Flügge MC, Abdurahman A, Crockett MJ. Model-free decision making is prioritized when learning to avoid harming others. *Proc Natl Acad Sci USA*. 2020;117(44):27719-27730. doi:10.1073/pnas.2010890117

12. Wittmann MK, Fouragnan E, Folloni D, et al. Global reward state affects learning and activity in raphe nucleus and anterior insula in monkeys. *Nat Commun*. 2020;11(1):3771. doi:10.1038/s41467-020-17343-w

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

14. Stephan KE, Penny WD, Daunizeau J, Moran RJ, Friston KJ. Bayesian model selection for group studies. *NeuroImage*. 2009;46(4):1004-1017. doi:10.1016/j.neuroimage.2009.03.025

15. Dawson C. *Ggprism: A "ggplot2" Extension Inspired by "GraphPad Prism."*; 2021. https://CRAN.R-project.org/package=ggprism

16. Auguie B. *GridExtra: Miscellaneous Functions for "Grid" Graphics*.; 2017. https://CRAN.R-project.org/package=gridExtra

17. Rinker TW, Kurkiewicz D. *Pacman: Package Management for R.*; 2018. http://github.com/trinker/pacman

18. Revelle W. *Psych: Procedures for Psychological, Psychometric, and Personality Research*. Northwestern University; 2021. https://CRAN.R-project.org/package=psych

19. Wickham H, Bryan J. *Readxl: Read Excel Files*.; 2019. https://CRAN.R-project.org/package=readxl

20. Wickham H, Averick M, Bryan J, et al. Welcome to the tidyverse. *Journal of Open Source Software*. 2019;4(43):1686. doi:10.21105/joss.01686

Taylor & Francis
Taylor & Francis Group

The World Journal of Biological Psychiatry

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/iwbp20

Differentiating brain function of punishment versus reward processing in conduct disorder with and without attention deficit hyperactivity disorder

Sarah Baumann, Arne Hartz, Wolfgang Scharke, Stephane A. De Brito, Graeme Fairchild, Beate Herpertz-Dahlmann, Kerstin Konrad & Gregor Kohls

To cite this article: Sarah Baumann, Arne Hartz, Wolfgang Scharke, Stephane A. De Brito, Graeme Fairchild, Beate Herpertz-Dahlmann, Kerstin Konrad & Gregor Kohls (2021): Differentiating brain function of punishment versus reward processing in conduct disorder with and without attention deficit hyperactivity disorder, The World Journal of Biological Psychiatry, DOI: 10.1080/15622975.2021.1995809

To link to this article: https://doi.org/10.1080/15622975.2021.1995809

THE WORLD JOURNAL OF BIOLOGICAL PSYCHIATRY
https://doi.org/10.1080/15622975.2021.1995809

ORIGINAL INVESTIGATION

Differentiating brain function of punishment versus reward processing in conduct disorder with and without attention deficit hyperactivity disorder

Sarah Baumann[a], Arne Hartz[b], Wolfgang Scharke[c], Stephane A. De Brito[d], Graeme Fairchild[e], Beate Herpertz-Dahlmann[a], Kerstin Konrad[b] and Gregor Kohls[b,f]

[a]Department of Child and Adolescent Psychiatry, Psychosomatics and Psychotherapy, University Hospital RWTH Aachen, Aachen, Germany; [b]Child Neuropsychology Section, Department of Child and Adolescent Psychiatry, Psychosomatics and Psychotherapy, University Hospital RWTH Aachen, Aachen, Germany; [c]Cognitive and Experimental Psychology, Institute of Psychology, RWTH Aachen University, Aachen, Germany; [d]Centre for Human Brain Health, School of Psychology, University of Birmingham, Birmingham, UK; [e]Department of Psychology, University of Bath, Bath, UK; [f]Department of Child and Adolescent Psychiatry, Faculty of Medicine, TU Dresden, Dresden, Germany

ABSTRACT

Objectives: Conduct disorder (CD) and attention-deficit/hyperactivity disorder (ADHD) are reported to co-occur in about 30–50% of affected individuals. Research suggests that poor reinforcement-based decision-making may contribute to impaired social functioning in both youths with CD and ADHD. Considering its frequent co-occurrence this raises the question whether decision-making deficits in both disorders have a disorder-specific and/or shared neurobiological basis.

Methods: 138 participants with CD, ADHD, or CD + ADHD, and typically developing controls (TDCs) aged 9–18 years (48% girls) were included in the study. Participants completed a reinforcement-based decision-making task in the fMRI scanner, investigating decision-making capabilities under different reinforcement contingencies (i.e. punishment vs. reward). Whole-brain and ROI analyses were used to test for potential group differences.

Results: For punishment versus reward contingencies, relative to TDCs, youths with CD + ADHD displayed lower brain activity in dorsal striatum (incl. caudate), middle temporal gyrus (MTG), inferior frontal gyrus (IFG) and lateral occipital cortex, and they showed lower activity in dorsal striatum (incl. putamen), orbitofrontal cortex (OFC) and IFG relative to participants with ADHD. All other group comparisons were found to be non-significant.

Conclusions: Participants with comorbid CD + ADHD are neurobiologically the most severely impaired group regarding reinforcement-based decision-making, particularly in response to punishment.

ARTICLE HISTORY
Received 8 March 2021
Revised 30 September 2021
Accepted 17 October 2021

KEYWORDS
Conduct disorder; attention-deficit/hyperactivity disorder; reinforcement learning; punishment; fMRI

Introduction

Conduct disorder (CD) and attention-deficit/hyperactivity disorder (ADHD) are two of the most prevalent externalising disorders in childhood and adolescence (Polanczyk et al. 2015). These disorders are reported to co-occur in about 30–50% of affected individuals (Banaschewski et al. 2005; Rubia et al. 2010). By acknowledging their frequent co-occurrence, the International Classification of Diseases (ICD-10) by the World Health Organisation (World Health Organization 1992) even gave the condition it's own diagnostic category known as hyperkinetic conduct disorder (ICD-10: F90.1). Compared to pure CD and pure ADHD, youths with hyperkinetic conduct disorder (CD + ADHD), are considered the more severe cases as they typically have an earlier age-of-onset of a more serious and persisting set of symptoms that require broader, i.e. cross-disorder, multimodal treatment approaches (Banaschewski et al. 2005; Connor and Doerfler 2008). However, it remains debateable whether CD + ADHD truly constitutes a distinct syndrome or whether it is simply a hybrid of CD and ADHD (Schachar and Tannock 1995). Research suggests that both CD and ADHD share certain behavioural (e.g. emotion dysregulation), cognitive (e.g. executive dysfunction), and neurobiological (e.g. ventral striatal dysfunctions) characteristics, but they also present with disorder-specific (brain) abnormalities (Rubia 2011). Disentangling

CONTACT Sarah Baumann sbaumann@ukaachen.de Department of Child and Adolescent Psychiatry, Psychosomatics and Psychotherapy, University Hospital RWTH Aachen, Neuenhofer Weg 21-22, Aachen D-52074, Germany

Supplemental data for this article can be accessed at https://doi.org/10.1080/15622975.2021.1995809.

disorder-specific from overlapping neural dysfunctions will help to better understand the aetiology of the two conditions as well as their comorbid presentation, and may thus inform theories of developmental psychopathology and treatment practices.

According to Sonuga-Barke and colleagues (2016) poor reinforcement-based decision-making contributes to impaired social functioning and reduced quality of life in youths with both CD and ADHD. However, it has been proposed that the mechanisms underlying decision-making deficits may differ between the two disorders: While CD is linked to reckless risk-taking and failure to learn from negative consequences (i.e. punishment), ADHD is associated with deficient (i.e. insufficiently reflective, and inconsistent) and impulsive actions (i.e. favouring immediate over delayed outcomes such as rewards) (Sonuga-Barke et al. 2016). Still, little is known about the neurobiological underpinnings of poor reinforcement-based decision-making in CD versus ADHD, and crucially, it is unclear to what extent decision-making deficits in both conditions have a disorder-specific and/or shared neurobiological basis (Banaschewski et al. 2005).

In an attempt to pinpoint particularly the distinct brain substrates of CD relative to ADHD, Rubia (2011) reviewed the relevant structural and functional magnetic resonance imaging (fMRI) studies and concluded that CD is associated with disorder-specific deficits in circuits known to regulate affective and motivational control processes (i.e. 'hot' executive functions), including regions such as orbitofrontal (OFC) and ventromedial prefrontal cortices (vmPFC), anterior cingulate cortex (ACC), striatum, and amygdala. The disorder-specific dysfunctions in ADHD, by contrast, appear in fronto-striato-parieto-cerebellar circuits that regulate motor, attentional, and cognitive control processes (i.e. 'cool' executive functions), most prominently the lateral inferior frontal cortex (for a more recent review, see also Puiu et al. 2018). Although both 'hot' and 'cool' control circuits are involved in the decision-making process (Ernst and Paulus 2005), the vast majority of fMRI studies reviewed by Rubia (2011) did not utilise experimental tasks that truly tap into reinforcement-based decision-making (Scholl and Klein-Flügge 2018). Thus, it still remains unclear to what extent the disorder-specific neural dysfunctions of CD versus ADHD, as highlighted by Rubia (2011), are linked to the differential decision-making deficits seen in both disorders.

More recently, fMRI studies have examined the neural substrates of reward and punishment processing as two pivotal computational mechanisms that may underlie the reinforcement-based decision-making deficits in CD and ADHD (Plichta and Scheres 2014; Blair et al. 2018). Dysfunctions in these two processes are thought to increase the risk of impulsiveness, frustration-induced reactive aggression and antisocial behaviour more generally (Alegria et al. 2016; Blair et al. 2018; Puiu et al. 2018). In comparison to typically developing controls (TDCs), youths with CD show reduced striatal and vmPFC responses to rewarding stimuli (e.g. monetary gains), whereas these two brain regions are overactive in CD in response to punishing stimuli (e.g. monetary loss) (Blair et al. 2018). A similar, but less consistent, pattern of neural dysfunction has been reported for ADHD (Plichta and Scheres 2014; Rubia 2018). Although these findings point to functional abnormalities concerning both reward and punishment processing that are partially shared by CD and ADHD, most studies have – either for practical or scientific reasons – grouped youths with different externalising disorders together, particularly CD and ODD, but also CD/ODD and ADHD, or have investigated externalising symptoms as a dimensional variable in high-risk samples (Fairchild et al. 2019). Thus, one has to be cautious in interpreting the available fMRI data in terms of any disorder-specific and/or shared pathophysiology of CD versus ADHD.

Notably, a recent fMRI meta-analysis on a variety of reinforcement-based decision-making paradigms, revealed that youths with disruptive behaviour and conduct problems versus TDCs have decreased activation in ventral and dorsal medial prefrontal cortex (including ACC), accompanied by increased dorsal striatal activation in caudate nucleus, even after ADHD comorbidity was statistically controlled (Alegria et al. 2016). Although these are the most thorough findings to date regarding a potential CD-specific neural dysfunction of reinforcement-based decision-making, this meta-analysis' conclusions are somewhat limited because (1) it did not separate youths with CD from youths with ODD, and (2) it meta-analysed reward and punishment processing in a combined fashion, rather than separately.

Thus, to address the above-mentioned research gaps, we directly compared reward and punishment processing in a group of youths with comorbid CD + ADHD and those with the individual disorder (i.e. CD only, and ADHD only) relative to TDCs, while they performed a reinforcement-based decision-making task in the MRI scanner (Kim et al. 2006). This design allowed us testing whether similar or different patterns of neural dysfunction characterise the two pure disorders and potentially identifying a profile that is unique to the comorbid group (i.e. distinctive vs. additive pathophysiology).

In line with the fMRI meta-analyses by Alegria et al. (2016) and Plichta and Scheres (2014), we predicted

that, compared to TDCs, (1) youths with CD would show atypical reinforcement signalling in ACC, and dorsal striatum (primarily caudate nucleus), (2) youths with ADHD would show atypical activation in the ventral striatum, and (3) the comorbid CD + ADHD group would show the most severe dysfunctions in prefrontal and striatal circuits. We additionally explored brain-behaviour associations between: (1) CD symptom severity and prefrontal as well as dorsal striatal brain activity, and (2) ADHD symptom severity and ventral striatal brain activity.

Method

Participants

180 participants, aged 9–18 years, were recruited through community outreach, mental health clinics and welfare institutions to participate in this cross-sectional fMRI study. Subsequently, 42 individuals were excluded because of excessive head movements, i.e. more than 3 mm of translational motion during the fMRI scan: CD = 2 (13.3%), ADHD = 10 (24.7%), CD + ADHD = 10 (18.2%), and TDC = 20 (34.5%) ($\chi^2(df=3) = 3.75$, $p = .30$). Thus, the final study sample comprised of 138 participants (CD: $n = 13$, ADHD: $n = 19$, CD + ADHD: $n = 45$, and TDC: $n = 61$) (Table 1). Overall exclusion criteria were autism spectrum disorder, psychosis or schizophrenia, mania or bipolar disorder, genetic syndromes, neurological disorders, an IQ < 70, and any MRI contraindications. The study protocol was approved by the local ethics committee, and participants and their caregivers gave written informed consent. Participants were compensated for their participation (50€ in addition to the money they gained during the task).

Table 1. Sample demographics and clinical characteristics.

	CD	ADHD	CD + ADHD	TDC	Group	Post-hoc comparisons
	N = 13	*N* = 19	*N* = 45	*N* = 61	(CD− vs. CD + vs. ADHD vs. TDC) F/X$^{2\#}$	*t*-tests$^{\#}$
Sex (f/m)	5/8	10/9	23/22	29/32	0.85	
Age (years)	14.8 (2.7)	12.8 (2.7)	13.8 (2.2)	14.2 (2.7)	2.0	
Estimated IQ	94.5 (12.1)	100.5 (10.6)	97.1 (12.0)	103.3 (11.6)	3.54*	TDC > CD + ADHD = CD; ADHD = TDC & CD & CD + ADHD
CD total symptoms (max. 15)	3.9 (2.0)	0.1 (0.3)	5.4 (2.3)	0.1 (0.2)	136.3***	CD + ADHD > CD > ADHD = TDC
CD subtype *n* (%)					3.64	
Childhood-onset	6 (46.2)	N/A	25 (55.6)	N/A		
Adolescent-onset	6 (46.2)	N/A	20 (44.4)	N/A		
Unspecified	1 (7.7)	N/A	0	N/A		
ADHD total symptoms (max. 18)	6.0 (4.2)	14.3 (3.5)	15.2 (3.2)	0.1 (0.4)	342.5***	CD + ADHD & ADHD > CD > TDC
ADHD subtype *n* (%)					4.31	
Inattentive	N/A	7 (36.8)	8 (17.8)	N/A		
Hyperactive	N/A	0	2 (4.4)	N/A		
Combined	N/A	12 (63.2)	32 (71.1)	N/A		
Unspecified	N/A	0	3 (6.7)	N/A		
Psychotropic medication *n* (%)	3.0 (23.1)	15 (78.9)	18.0 (40.0)	N/A	54.25***	
Stimulants	1 (7.7)	15 (78.9)	17 (37.8)	N/A		
Antidepressants	2 (15.4)	0	1 (2.2)	N/A		
Neuroleptics	0	0	2 (4.4)	N/A		
Comorbid Diagnoses *n* (%)						
CD	13 (100)	0	45 (100)	N/A	119.64***	
ODD	11 (84.6)	0	45 (100)	N/A	131.97***	CD + ADHD > CD; CD & CD + ADHD > TDC = ADHD
ADHD	0	19 (100)	45 (100)	N/A	139.00***	CD + ADHD = ADHD > CD = TDC
MDD	5 (38.5)	2 (10.5)	15 (33.3)	N/A	27.41***	CD = CD+ = ADHD, CD & CD + ADHD > TDC, ADHD = TDC
PTSD	1 (7.7)	0	9 (20.0)	N/A	17.34***	C CD + ADHD > TDC, CD = ADHD, ADHD = TDC; CD + ADHD > ADHD
Anxiety disorders	3 (23.1)	0	12 (26.7)	N/A	23.62***	CD = CD + ADHD > TDC = ADHD
SUD	1 (7.7)	0	9 (20.0)	N/A	17.34***	CD > TDC, CD = ADHD, CD+ > ADHD = TDC, CD+ = CD
YPI (CU total score)	29.3 (8.9)	24.8 (6.4)	32.4 (8.4)	28.4 (7.0)	4.97**	CD + ADHD > ADHD = TDC; ADHD & CD + ADHD = CD

ADHD: attention deficit hyperactivity disorder; CD: conduct disorder; TDC: typically developing controls; f/m: female/male; IQ: intelligence quotient; MDD: major depressive disorder; N/A: not applicable; ODD: oppositional defiant disorder; PTSD: post-traumatic stress disorder; SUD: substance use disorder (including substance abuse and dependence); YPI: youth psychopathic traits inventory.

Diagnoses and CD/ADHD symptoms and subtypes 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 *F*-tests (or χ^2-tests,) and follow-up pairwise comparisons with Bonferroni correction. *$p \leq .05$; **$p \leq .01$; ***$p \leq .001$

The four different groups were specified as follows: (1) CD: current diagnosis of CD but no current or past diagnosis of ADHD, (2) ADHD: current diagnosis of ADHD but no current or past diagnosis of CD or ODD (3) CD + ADHD: current diagnosis of CD and ADHD, and (4) TDC: no current psychiatric diagnoses and no lifetime diagnoses of CD, ODD and ADHD. All diagnoses were based on DSM-IV-TR criteria (American Psychiatric Association 2000). Participants who were taking psychotropic medication (Table 1) were tested while on medication.

All participants were clinically evaluated with the Kiddie-Schedule for Affective Disorders and Schizophrenia for School-Age Children – Present and Lifetime Version (K-SADS-PL) (Kaufman et al. 1997). The K-SADS-PL interview was administered by trained staff separately to participants and their caregivers, and clinical summary ratings were achieved to determine group allocation and to identify possible comorbid psychiatric diagnoses. Disorder severity of CD and ADHD was defined as the number of total symptoms endorsed in the K-SADS-PL interviews. Using the K-SADS-PL, we also determined CD-onset type (i.e. CO-CD: presence of at least one characteristic CD behaviour prior to age 10; AO-CD: absence of any CD behaviours prior to age 10) and ADHD subtypes. Full-scale IQs were estimated using the vocabulary and matrix reasoning subtests of the Wechsler Intelligence Scale for Children-Fourth Edition (Wechsler 2011), or the Wechsler Intelligence Scale for Adults-Fourth Edition (Wechsler 2012). The level of callous-unemotional traits was assessed by using the total score of the subscales 'remorselessness', 'callousness' and 'unemotionality' of the self-report version of the Youth Psychopathic traits Inventory (YPI) (Essau et al. 2006).

FMRI task

We used a monetary reinforcement-based decision-making task, originally described by Kim et al. (2006), to measure decision-making capabilities based on reward and punishment vs. neutral, non-reinforcing contingencies (Figure 1). In the scanner, the task was presented on a rear projection LCD screen and viewed by the participants through a mirror attached to the head coil. Behavioural data collection and stimulus presentation were controlled by the MATLAB R2014a software (The MathWorks Inc. 2014).

Trials started with the presentation of a pair of cue stimuli (i.e. fractals) side-by-side. Each pair marked the onset of one of three different trial types (i.e. conditions): reward (REW; i.e. monetary gain), punishment (PUN; i.e. monetary loss), and neutral outcome (NEUT; i.e. neither monetary gain nor loss). Throughout the task, occurrence of trial types was fully randomised, and the assignment of the three fractal pairs to the different conditions was fully counterbalanced across all participants. Participants were instructed to select one of the two simultaneously presented fractals by pressing the left or right key on a button box, placed in their right hand and keys corresponded to the location of the two cues presented on the screen (i.e. left or right of a fixation cross). The chosen cue increased in brightness and was followed by visual feedback 4 s later, indicating whether participants received a reward (a picture of a 20 Eurocent coin, and the description: 'You won 20 cent'), a

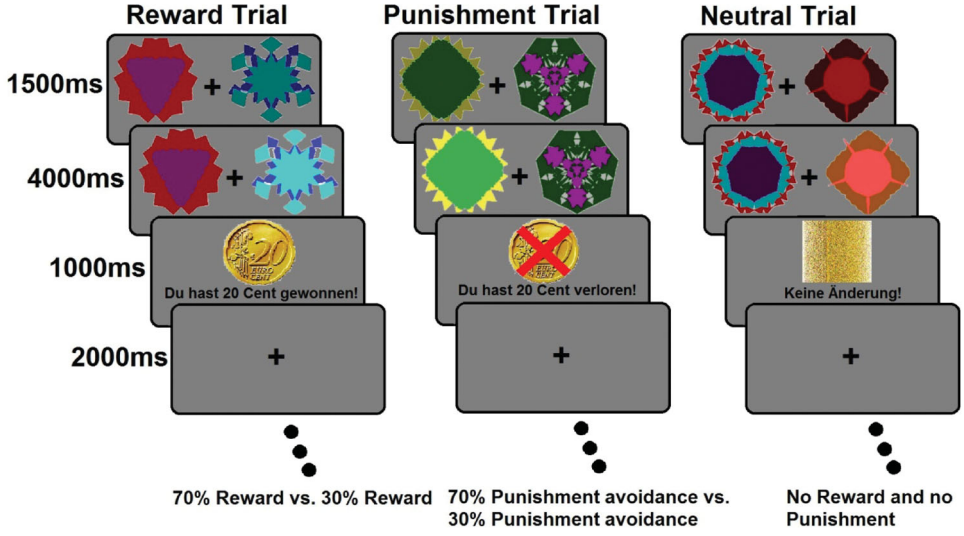


Figure 1. Illustration of the monetary instrumental task.

punishment (a picture of a 20 Eurocent coin overlaid with a red cross, and the description: 'You lost 20 cent'), a neutral outcome (a picture of a scrambled 20 Eurocent coin, and the description: 'No change'), or nothing (a blank screen with a crosshair in the centre).

On each trial, participants could either select a high probability or a low probability cue. In REW trials, choosing the high probability cue either resulted in reward (+0.20 €) with a probability of 70%, or in no feedback (i.e. no reward = crosshair) with a probability of 30%. Conversely, choosing the low probability cue either resulted in reward (+0.20 €) with a probability of 30%, or in no feedback (i.e. no reward = crosshair) with a probability of 70%. In PUN trials, choosing the high probability cue either resulted in no feedback (i.e. no punishment = crosshair) with a probability of 70%, or in punishment (−0.20 €) with a probability of 30%. Conversely, choosing the low probability cue either resulted in no feedback (i.e. no punishment = crosshair) with a probability of 30%, or in punishment (−0.20 €) with a probability of 70%. The NEUT trials served as baseline for the two other conditions, controlling for motor responses and simple visual effects. In this case, participants either had a 70% or 30% chance of obtaining a neutral outcome (a scrambled 20 Eurocent), thereby receiving no feedback on the remaining trials. Participants completed a total of 3 consecutive runs, each lasting approximately 6 minutes and containing a total of 135 trials with 45 trials per condition: 15x REW, 15x PUN, and 15x NEUT. The whole task procedure lasted approximately 25 minutes.

Prior to the scan, participants were told that they would see three different pairs of unfamiliar stimuli (i.e. fractals) as cues during the experiment, and on each trial they had to choose one out of the two simultaneously presented cues. Depending on their choices, they would win money, lose money, obtain a neutral outcome, or receive no feedback. It was, explicitly stressed that they should try to win as much money as possible. Each participant started the experiment with a fixed amount of 10€, and was told that any wins or losses would be added or subtracted, respectively, from this total. As per instructions, participants were paid according to their performance at the end of the experiment, receiving on average an amount of 11.35 ± 1.52€.

Behavioural data analyses

We compared the four groups on demographic and clinical variables using ANOVA and Chi-Square tests (SPSS v25.0; IBM Corp., Armonk, NY). At the task level, accuracy (in %, i.e. choosing the cue with a probability of 70%) and reaction times for correct choices (RTs in ms) on the decision-making task were analysed using a repeated-measures MANOVA model with group (CD vs. ADHD vs. CD + ADHD vs. TDC) as between-subjects factor, and condition (REW vs. PUN vs. NEUT) as within-subjects factor, followed by post-hoc pairwise comparisons in case of significant main or interaction effects using the Games–Howell procedure to control for multiple comparisons, as this procedure is recommended in case that sample sizes are very different and if one is uncertain whether the population variances are equivalent (Field 2009). As age and IQ did not correlate with the dependent measures, these variables were not included as covariates in the analyses. We decided to use ANOVA models rather than a 2×2 full-factorial design in all analyses, because the latter implicitly assumes that the combined behavioural as well as brain activation pattern are the sum of the single-disorder factors. Such an analysis would bias the results, while the present study aimed to test whether comorbidity of CD + ADHD is a unique disorder or simply the addition of the two individual clinical conditions. Thus, the ANOVA is the appropriate analysis approach here because it is blind to any direction of possible group differences. The alpha level was set at 0.05. Effect sizes were calculated using partial eta squared (η_p^2), where 0.01, 0.06, and 0.14 represent small, medium and large effects, respectively (Cohen 1988).

Image acquisition

T2* weighted BOLD images were obtained with echo-planar imaging using a Siemens Prisma fit 3.0 T scanner (Erlangen, Germany) and a 20-channel head coil. Whole-brain volumes of 41, 3-mm thick transversal slices (TR/TE = 2500/30 ms; flip angle = 83°; FOV = $192 \times 192\,mm^2$; matrix size = 64×64, and voxel size = $3.0 \times 3.0 \times 3.0\,mm^3$) were collected throughout three functional runs. A total of 465 functional volumes (plus 5 'dummy' scans per run allowing for T1 magnetic saturation) were acquired for each participant. Prior to the functional runs, 192 high-resolution T1-weighted structural images of the entire brain were acquired using a MPRAGE sequence (TR/TE = 1900/3.4 ms; flip angle = 9°; FOV = $192 \times 192\,mm^2$; matrix size = 256×256, and voxel size = $1 \times 1 \times 1\,mm^3$).

Image analysis

Data were preprocessed and analysed using Statistical Parametric Mapping (SPM12) software (http://www.fil.ion.ucl.ac.uk/spm), implemented in MATLAB. Prior to

image analysis, the first 5 volumes of each functional run were discarded because of the non-equilibrium state of magnetisation. Images were realigned to the first volume in the time series, anatomical scans were co-registered to the mean image and spatially normalised into a standard anatomical reference space, spatial smoothing was applied using a Gaussian kernel with a full-width at half-maximum (FWHM) of 6 mm. Regression analysis was carried out on the pre-processed functional time series of each participant using a general linear model (GLM) for an event-related design as implemented in SPM12. Motion parameters were entered as regressors, and simple contrasts were created for the following three conditions: (1) PUN, (2) REW, and (3) NEUT by modelling the whole trial duration, i.e. including cue onset, choice selection, and outcome presentation (Galvan et al. 2006). Given our particular interest in the most basic form of reward and punishment mechanisms, we did not distinguish between different components of the reinforcement process in the current study (e.g. anticipation vs. outcome; see (Knutson and Wimmer 2007). This approach would provide more of a common ground in terms of comparability with prior studies that each examined different aspects of reinforcement processing in youths with CD and/or ADHD (Plichta and Scheres 2014; Alegria et al. 2016).

At the second level, contrasts were entered into a full-flexible ANOVA with group (CD vs. ADHD vs. CD + ADHD vs. TDC) as between-subjects factor and condition (PUN vs. REW) as within-subjects factor. High-level contrast images were created for the comparisons (1) PUN > REW, and (2) REW > PUN to investigate whether the two contingencies differentially affected striatal and prefrontal brain regions (Delgado et al. 2000). Our main motivation to follow such approach (i.e. not modelling punishment vs. neutral, or reward vs. neutral) was that neutral outcomes that are intermixed with reward and punishment trials in the context of risk-taking tasks (incl. probabilistic reinforcement tasks) are actually not processed as neutral (i.e. neutral events can become affectively charged depending on the context in which they are presented) (see Grossberg and Gutowski 1987). For the whole-brain analyses across groups, Z-statistic maps were thresholded using clusters with $Z \geq 3.1$ (i.e. $p \leq .001$) at the voxel level and an FWE-corrected cluster-significance threshold at $p \leq .05$ to strictly control for type I errors (Eklund et al. 2016). Our a priori regions of interest (ROIs) comprised the caudate nucleus and the ACC (Alegria et al. 2016), the ventral striatum (Plichta and Scheres 2014), and the vmPFC/OFC (Blair 2016). Anatomical masks were created in standard MNI space using the FSL Harvard-Oxford cortical and subcortical structural atlas (Desikan et al. 2006). All ROI analyses were thresholded at $p < .05$ (voxel level), FWE-corrected for the specific ROI. Parameter estimates were extracted for all regions, and beta plots were generated for each group and both high-level contrasts. For the group comparisons, we will only refer to the results of the PUN > REW contrast, as the REW > PUN contrast only indicates the inverse of the group comparison results. ANOVAs with group as between-subjects factor were conducted on the beta values of the ROIs. Parameter estimates of the ROIs were correlated with ADHD and CD symptom severity (i.e. symptom counts from the K-SADS-PL interviews). We used the total counts of ADHD and CD symptoms, as well as symptoms of hyperactivity/impulsivity (ADHD) and aggression (CD), as specified in the DSM-5 (American Psychiatric Association 2013).

Results

Demographic characteristics

Mean age and sex distribution did not differ significantly between groups. However, the two groups of youths with CD had lower IQs than TDCs. The CD + ADHD group had the highest level of CD symptoms (K-SADS-PL) and CU traits (YPI), followed by the CD group, relative to both youths with ADHD and TDCs. Onset of CD symptomatology (childhood vs. adolescence) did not differ significantly between both CD groups. Level of ADHD symptoms (K-SADS-PL) was highest for both youths with CD + ADHD and ADHD, followed by the CD group, with the lowest symptom level for TDCs. Distribution of ADHD subtypes did not differ significantly between the two ADHD groups. Lastly, psychotropic medication use was highest for the ADHD group, followed by youths with CD and CD + ADHD, relative to TDCs.

Task performance

The repeated-measures MANOVA revealed a significant main effect of condition [$F(4, 536) = 44.9$, $p < .001$, $\eta^2_p = .25$], which was related to both accuracy ($p < .001$, $\eta^2_p = .10$) and RT ($p < .001$, $\eta^2_p = .43$) (Table 2). The post-hoc comparisons for accuracy revealed that the correct response rate (in %) was significantly lower for the NEUT condition compared to both the REW and PUN conditions (all ps $< .001$, all η^2_ps $> .10$; REW vs. PUN: $p = .07$, $\eta^2_p = .02$), which is in line with the findings by Kim et al. (2006). Regarding RTs, the post-

Table 2. Task performance between groups on the three different conditions of the monetary instrumental task.

	CD	ADHD	CD + ADHD	TDC
Accuracy (in %)				
Reward	72.6 (32.8)	70.6 (28.7)	69.3 (32.9)	69.2 (30.8)
Punishment	64.5 (17.3)	62.6 (11.9)	65.4 (12.3)	65.8 (13.3)
Neutral	58.1 (31.2)	53.6 (30.2)	48.2 (28.8)	48.9 (27.1)
Reaction time (in msec.)				
Reward	833.4 (135.8)	906.0 (197.0)	863.2 (150.2)	817.5 (119.8)
Punishment	1067.4 (178.5)	1084.0 (150.8)	1072.0 (142.9)	1011.3 (174.2)
Neutral	987.4 (143.1)	1015.5 (115.6)	1010.5 (150.6)	966.5 (166.4)

ADHD: attention deficit hyperactivity disorder; CD: conduct disorder; TDC: typically developing controls.

hoc comparisons showed the fastest RTs for REW, followed by the NEUT condition, and the slowest RTs for PUN (all $ps < .001$, all $\eta^2_p s > .15$); this, again, fits the data reported by Kim et al. (2006). The group by condition effect [$F(12, 536) = 0.34$, *ns*, $\eta^2_p = .01$] and the group effect [$F(6, 268) = 1.2$, *ns*, $\eta^2_p = .02$] were non-significant, suggesting that the different reinforcement conditions similarly affected task performance in all groups.

Whole-Brain between-group comparisons

Using whole-brain cluster thresholding that strictly controls against type I errors, the high-level PUN > REW contrast revealed significant differences in brain responses in the CD + ADHD group compared to both the TDC group and ADHD group (Table 3): For punishment versus reward contingencies, the youths with CD + ADHD displayed lower brain activity in the dorsal striatum (incl. caudate), medial temporal gyrus (MTG), inferior frontal gyrus (IFG) and lateral occipital cortex relative to TDCs (Figures 2 and 3), and they showed lower activity in the dorsal striatum (incl. putamen), orbitofrontal cortex (OFC; extending into insula) and IFG relative to the ADHD group (Figures 4 and 5). All other group comparisons (incl. CD vs. TDC, and CD vs. CD + ADHD) were found to be non-significant at the whole-brain cluster-corrected level.

Between-Group comparisons using a priori ROIs

For the high-level PUN > REW contrast, the extracted ß-values of our a priori anatomical ROIs (i.e. caudate nucleus, ventral striatum, ACC, and vmPFC/OFC) were entered into four separate ANOVA models with group as between-subjects factor. However, none of these analyses revealed significant group effects.

Table 3. Whole brain activation table for the group comparisons on the PUN > REW contrast.

Brain region	L/R	Cluster size (mm^3)	Z	MNI coordinates x	y	z
Punishment > Reward						
TDCs > CD + ADHD						
Caudate	R	195	4.64	10	12	8
			4.46	14	2	12
			4.00	10	4	4
Middle temporal gyrus	L	168	4.23	−42	14	36
Inferior frontal gyrus	L	108	4.31	−52	18	22
Lateral occipital cortex	L	152	4.18	−44	−60	54
ADHD > CD + ADHD						
Putamen	R	115	4.91	30	−10	−6
			3.92	24	0	0
			3.35	28	6	−6
Orbitofrontal cortex	R	198	4.56	40	22	−8

ADHD: attention deficit hyperactivity disorder; CD: conduct disorder; L/R: left/right, MNI: Montreal Neurological Institute. TDC: typically developing controls. Results were significant at $p < .05$ (FWE-corrected at cluster level, $p < .001$ voxel level, k = 10 voxels).

Correlations between ROI activity and clinical symptomatology

Our correlational analyses did not reveal any significant associations between brain activity and CD or ADHD symptom severity (as assessed with the K-SADS-PL interview) after correcting for multiple comparisons.

Discussion

To our knowledge, this is the first fMRI study investigating reinforcement-based decision-making in youths with CD and/or ADHD versus TDCs to reveal whether similar or different patterns of neural dysfunction characterise the two pure disorders and to potentially identify a profile that is unique to the comorbid group. Clinically, we found that patients with a comorbid condition of CD + ADHD were more severely impaired, including greater CD and ADHD symptoms and CU traits, than patients with either of the pure disorders. At the behavioural level, there were significant differences in task performance across groups depending on reinforcement type (accuracy: REW = PUN > NEU; reaction times: REW < NEU < PUN), which is likely attributable to differences in the cognitive processes required to execute the different trial conditions. For example, concerning punishment trials, individuals first have to inhibit the incorrect response, followed by selecting the correct one in order to avoid potential punishment. One can assume that this adds an intermediate processing step to proper choice selection, resulting in longer reaction times for such trials. At the whole-brain level, we were able to show that youths with CD + ADHD, in comparison to TDCs and youths with ADHD, demonstrated diminished

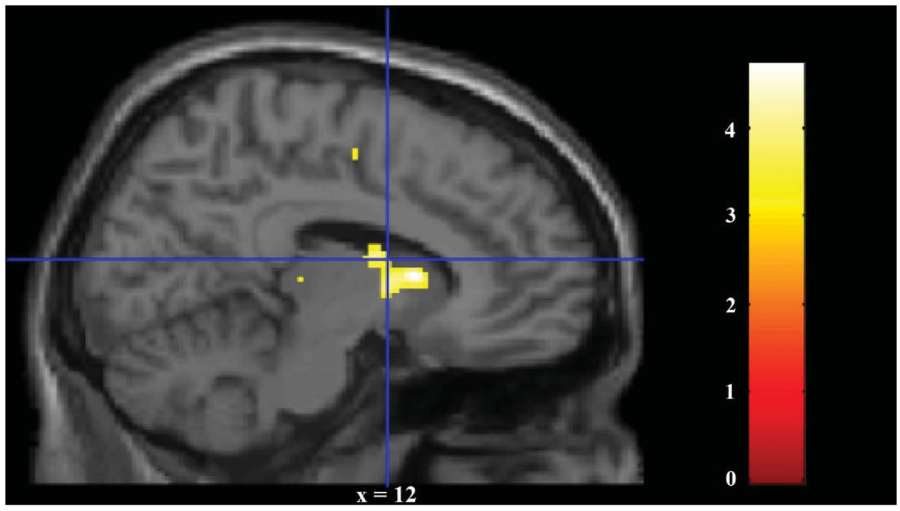


Figure 2. The caudate was more strongly activated in response to punishment versus reward in TDCs than CD + ADHD. Whole-brain results were significant at $p \leq .001$ at the voxel level, and for the cluster-level, a FWE-corrected cluster-significance threshold at $p \leq .05$ was set. For illustrative purposes, the uncorrected level is presented here, but results are reported for the cluster-level correction in the main text.

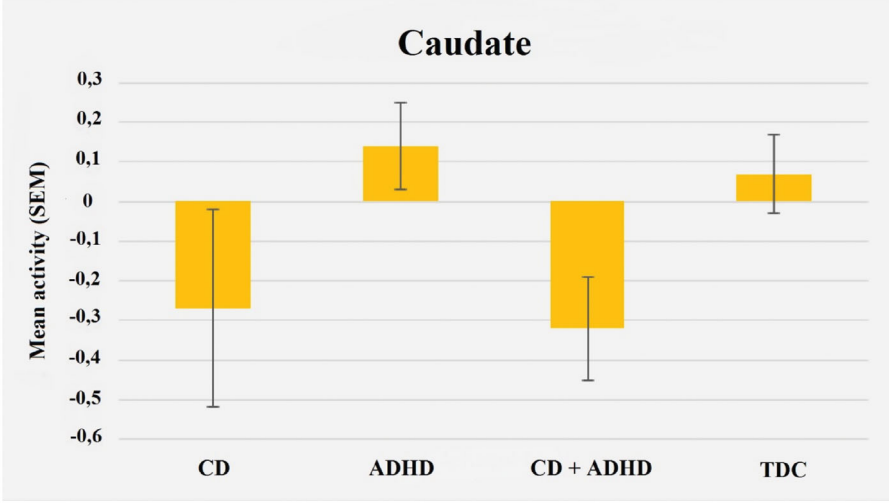


Figure 3. Beta plots (i.e. parameter estimates) generated for the differences in brain activity in the caudate in response to punishment versus reward separately for each group. Betas for the caudate were extracted based on the results of the whole-brain between-group comparisons for the significant difference between TDCs > CD + ADHD (see Table 3).

reinforcement signalling in dorsal striatal (i.e. caudate nucleus and putamen) and prefrontal brain regions (i.e. OFC, IFG), specifically in response to punishment in the form of monetary loss. There were, however, no significant activation differences between reinforcement conditions between TDCs and youths with ADHD, indicating that those with ADHD reacted equally well as TDCs during decision-making under different reinforcement contingencies (Rubia 2011).

Our group-specific predictions were only partially confirmed. In contrast to our hypothesis, we did not find atypical ventral striatal activity in patients with ADHD alone. This might be due to the fact that we did not analyse the different phases of reinforcement processing separately (e.g. anticipation, choice selection, and outcome). Notably, the previously reported diminished ventral striatal activity in patients with ADHD vs. TDCs in response to appetitive stimuli, were largely based on imaging data obtained with regard to the anticipation phase (see Plichta and Scheres 2014).

As predicted, patients with the comorbid condition of CD + ADHD were clinically and neurobiologically the most severely impaired group. Youths with a CD + ADHD diagnosis, in comparison to TDCs and youths with ADHD only, displayed lower brain responses in dorsal striatum (incl. caudate), OFC (extending into the anterior insula), IFG and MTG in

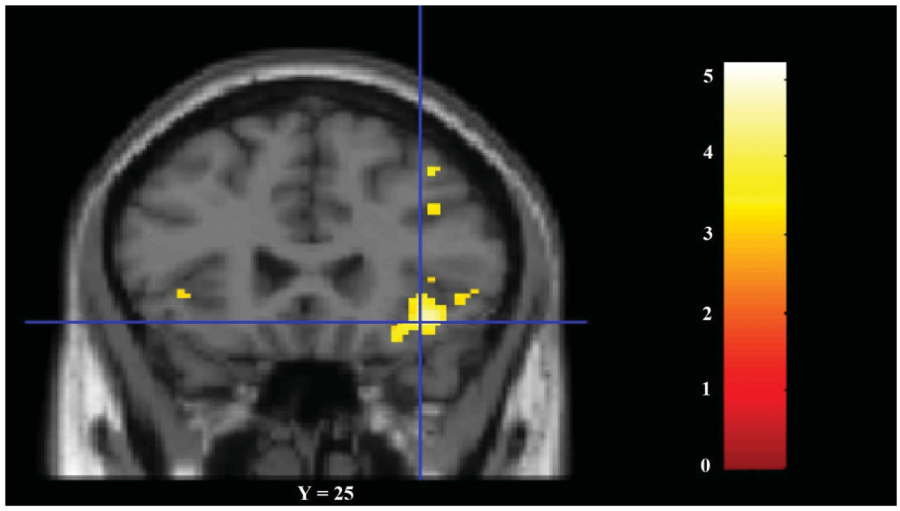


Figure 4. The OFC (extending into the insula) was more strongly activated in response to punishment versus reward in ADHD than CD + ADHD. Whole-brain results were significant at $p \leq .001$ at the voxel level, and for the cluster-level, a FWE-corrected cluster-significance threshold at $p \leq .05$ was set. For illustrative purposes, the uncorrected level is presented here, but results are reported for the cluster-level correction in the main text.

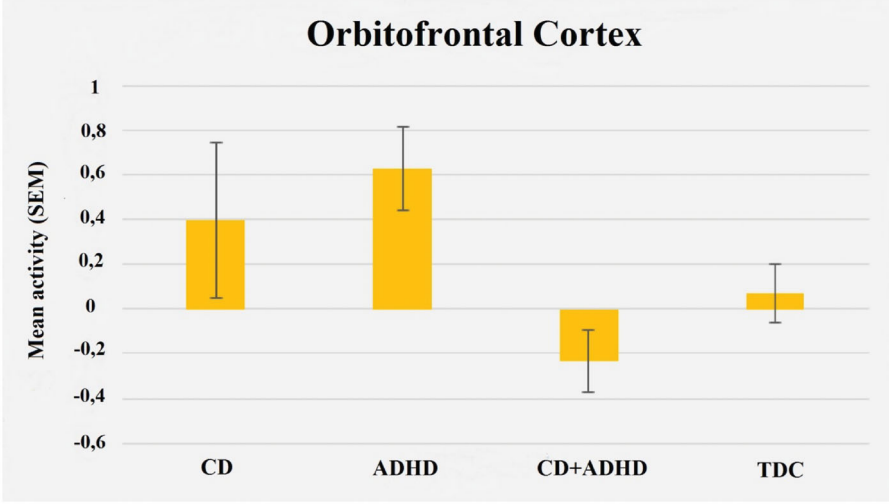


Figure 5. Beta plots (i.e. parameter estimates) generated for the differences in brain activity in the OFC in response to punishment versus reward seperately for each group. Betas for the OFC were extracted based on the results of the whole-brain between-group comparisons for the significant difference between ADHD > CD + ADHD (see Table 3).

response to punishment. Interestingly, we did not find any significant differences in brain responses during reinforcement processing between the groups of patients with pure CD and CD + ADHD. This might indicate that these two clinical groups share disorder-specific deficits in brain circuits related to the management of affective decision-making that is primarily associated with the CD phenotype. This is in line with the notion that particularly motivational and affective decision-making processes are impaired in youths with CD, primarily reflected in atypical brain responses in striatal and prefrontal brain regions as highlighted by Rubia 2011.

However, contrary to the meta-analysis of Alegria et al. (2016), we did not observe differential brain responses in the ACC between groups. Note, though, that Alegria and colleagues analysed reward and punishment processing in a combined fashion, rather than separately as done here. This makes it difficult to compare across study designs and might explain the different results regarding the ACC.

Our study had several strengths: We were able to test our hypotheses by using well-defined groups with participants who were extensively clinically assessed and reliably diagnosed. Our overall sample size of 138 participants is relatively large for an fMRI study

conducted with children and adolescents. Additionally, we were able to include a large number of girls with a CD (+ ADHD) diagnosis which is rare in fMRI studies of disruptive behaviour disordered samples (Fairchild et al. 2013, 2014; Alegria et al. 2016). Although the sample sizes of our four groups varied substantially, the sex ratio was similar across groups. Moreover, our CD and CD + ADHD groups consisted of participants who fulfilled diagnostic criteria for CD. Usually, most of the previous fMRI studies included mixed samples of CD or ODD cases or those with (subclinical) conduct problems (see Alegria et al. 2016).

However, our study had also several limitations: The four groups varied substantially in sample size which likely meant that some of our statistical analyses were underpowered (incl. ROI analyses), particularly with regard to the CD and ADHD groups versus the two other groups. Similarly, the lack of significant brain-behaviour correlations (i.e. CD and ADHD symptoms and brain activity in the pre-specified ROIs) in the present study fits well with recent experimental evidence that relatively small sample sizes might be insufficient for obtaining reproducible brain-behaviour correlations, regardless of analytic approach (Grady et al. 2021). It should be noted that recruiting a group of cases with CD without comorbid ADHD is a rather difficult task given the high co-occurrence rate of both disorders (see also: Rubia et al. 2009: noncomorbid CD: $n = 14$, noncomorbid ADHD: $n = 18$ which is comparable to our study). This is also reflected in the fact that many prior studies on reinforcement processing in youths with externalising problems often included mixed samples of youths with CD or ODD who had comorbid ADHD (e.g. Finger et al. 2008; White et al. 2013, 2014). Also, our two CD groups had significantly lower IQs than TDCs, which however is a typical finding in the CD literature (Murray and Farrington 2010), making our CD sample representative for this disorder. Noteworthy, the four groups did not differ in any task performance measures, and IQ did not correlate with these measures. This indicates that the reinforcement manipulations were similarly effective across groups and were not influenced by IQ, and, thus, it is unlikely that group differences in brain activation were confounded by IQ differences. Finally, for praticial reasons we neither excluded participants who took psychotropic medications nor asked them to withdraw them (e.g. stimulants) prior to being scanned. This, however, could have affected our findings. Thus future studies with medication-naïve youths are warranted.

In conclusion, the results of the current study provide new evidence for a disorder-related neural profile underlying impaired punishment processing in CD, but not ADHD, which supports the notion that deficient reinforcement-based decision-making is more closely related to CD than ADHD (Banaschewski et al. 2005). Moreover, patients with a comorbid condition showed the most severe dysfunctions in dorsal striatal and prefrontal circuits indicating 'additive' psychopathology that aggravates decision-making deficits that have been observed in both individual disorders (Finger et al. 2011). In clinical practice, CD without co-occurring ADHD has been shown to be extremely rare (Turgay 2004; Rubia et al. 2009), raising the question whether comorbid patients might need different treatment approaches than youths with CD only. Research has shown that psychostimulants (e.g. methylphenidate and amphetamines) reduce impulsive aggression (Pringsheim et al. 2015), a symptom which is commonly observed in CD and in ADHD. Deficient reinforcement-based decision-making, including impaired punishment processing at the neural level, as being observed in our sample of youths with CD + ADHD may contribute to frustration-based impulsive aggression (Blair 2016). It would be interesting to investigate whether the administration of psychostimulants to patients with CD with and without ADHD could reduce or even eliminate neural decision-making deficits, as previously shown with ADHD patients (Rubia et al. 2009). Psychostimulants such as methylphenidate increase the activity of the central nervous system through inhibiting the reuptake of dopamine and norepinephrine which is thought to exert a positive effect on the decision-making process (Solanto 1998). Note, in our study only 38% of youths with CD + ADHD were treated with psychostimulants compared to almost 80% of youths with ADHD. Therefore, future studies need to investigate whether psychostimulants can have a positive effect on reinforcement-based decision-making in CD (vs. ADHD). Moreover, it should be investigated to what extent impaired decision-making in CD is associated with general impairments in the decision-making process or related to deficiencies in specific decision-making phases (i.e. punishment anticipation vs. choice selection vs. outcome processing). This knowledge could inform effective treatments tailored to the specific needs of the affected individuals.

Statement of interest

We have no potential conflicts of interests to declare.

Funding

This study was funded by the European Commission's Seventh Framework Programme [FP7/2007-2013] under Grant Agreement no. 602407 (FemNAT-CD, coordinator: Christine M. Freitag).

References

Alegria AA, Radua J, Rubia K. 2016. Meta-analysis of fMRI studies of disruptive behavior disorders. Am J Psychiatry. 173(11):1119–1130.

American Psychiatric Association. 2000. Diagnostic and statistical manual of mental disorders: DSM-IV-TR. 4th, rev. ed. Washington DC: Author.

American Psychiatric Association. 2013. Diagnostic and statistical manual of mental disorders: DSM-V. 5th ed. Washington DC: Author.

Banaschewski T, Hollis C, Oosterlaan J, Roeyers H, Rubia K, Willcutt E, Taylor E. 2005. Towards an understanding of unique and shared pathways in the psychopathophysiology of ADHD. Dev Sci. 8(2):132–140.

Blair RJR. 2016. The neurobiology of impulsive aggression. J Child Adolesc Psychopharmacol. 26(1):4–9.

Blair RJ, Veroude K, Buitelaar J. 2018. Neuro-cognitive system dysfunction and symptom sets: a review of fMRI studies in youth with conduct problems. Neurosci Biobehav Rev. 91: 69–90.

Cohen J. 1988. Statistical power analysis for the behavioral sciences (2nd ed.). Hillsdale (NJ): Lawrance Eribaum Association.

Connor DF, Doerfler LA. 2008. ADHD with comorbid oppositional defiant disorder or conduct disorder: discrete or nondistinct disruptive behavior disorders? J Atten Disord. 12(2):126–134.

Delgado MR, Nystrom LE, Fissell C, Noll DC, Fiez JA. 2000. Tracking the hemodynamic responses to reward and punishment in the striatum. J Neurophysiol. 84(6):3072–3077.

Desikan RS, Ségonne F, Fischl B, Quinn BT, Dickerson BC, Blacker D, Buckner RL, Dale AM, Maguire RP, Hyman BT, Albert MS, Killiany RJ. 2006. An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. Neuroimage. 31(3):968–980.

Eklund A, Nichols TE, Knutsson H. 2016. Cluster failure: why fMRI inferences for spatial extent have inflated false-positive rates. Proc Natl Acad Sci USA. 113(28):7900–7905.

Ernst M, Paulus MP. 2005. Neurobiology of decision making: a selective review from a neurocognitive and clinical perspective. Biol Psychiatry. 58 (8):597–604.

Essau CA, Sasagawa S, Frick PJ. 2006. Psychometric properties of the Alabama parenting questionnaire. J Child Fam Stud. 15(5):595–614.

Fairchild G, Hagan CC, Passamonti L, Walsh ND, Goodyer IM, Calder AJ. 2014. Atypical neural responses during face processing in female adolescents with conduct disorder. J Am Acad Child Adolesc Psychiatry. 53(6):677–687.

Fairchild G, Hawes DJ, Frick PJ, Copeland WE, Odgers CL, Franke B, Freitag CM, De Brito SA. 2019. Conduct disorder. Nat Rev Dis Primers. 5(1):1–25.

Fairchild G, van Goozen SHM, Calder AJ, Goodyer IM. 2013. Research review: evaluating and reformulating the developmental taxonomic theory of antisocial behaviour. J Child Psychol Psychiatry. 54(9):924–940.

Field A. 2009. Discovering statistics using SPSS. Book plus code for E version of text. Thousand Oaks: SAGE Publications Limited.

Finger EC, Marsh AA, Blair KS, Reid ME, Sims C, Ng P, Pine DS, Blair RJR. 2011. Disrupted reinforcement signaling in the orbitofrontal cortex and caudate in youths with conduct disorder or oppositional defiant disorder and a high level of psychopathic traits. Am J Psychiatry. 168(2): 152–162.

Finger EC, Marsh AA, Mitchell DG, Reid ME, Sims C, Budhani S, Kosson DS, Chen G, Towbin KE, Leibenluft E, et al. 2008. Abnormal ventromedial prefrontal cortex function in children with psychopathic traits during reversal learning. Arch Gen Psychiatry. 65(5):586–594.

Galvan A, Hare TA, Parra CE, Penn J, Voss H, Glover G, Casey BJ. 2006. Earlier development of the accumbens relative to orbitofrontal cortex might underlie risk-taking behavior in adolescents. J Neurosci. 26(25):6885–6892.

Grady CL, Rieck JR, Nichol D, Rodrigue KM, Kennedy KM. 2021. Influence of sample size and analytic approach on stability and interpretation of brain-behavior correlations in task-related fMRI data. Hum Brain Mapp. 42(1):204–219.

Grossberg S, Gutowski WE. 1987. Neural dynamics of decision making under risk: affective balance and cognitive-emotional interactions. Psychol Rev. 94(3):300–318.

Kaufman J, Birmaher B, Brent D, Rao U, Flynn C, Moreci P, Williamson D, Ryan N. 1997. 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. 36(7): 980–988.

Kim H, Shimojo S, O'Doherty JP. 2006. Is avoiding an aversive outcome rewarding? Neural substrates of avoidance learning in the human brain. PLoS Biol. 4(8):e233–1461.

Knutson B, Wimmer GE. 2007. Splitting the difference: How does the brain code reward episodes? Ann NY Acad Sci. 1104(1):54–69.

Murray J, Farrington DP. 2010. Risk factors for conduct disorder and delinquency: key findings from longitudinal studies. Can J Psychiatry. 55(10):633–642.

Plichta MM, Scheres A. 2014. Ventral-striatal responsiveness during reward anticipation in ADHD and its relation to trait impulsivity in the healthy population: a meta-analytic review of the fMRI literature. Neurosci Biobehav Rev. 38: 125–134.

Polanczyk G. v, Salum GA, Sugaya LS, Caye A, Rohde LA. 2015. Annual research review: a meta-analysis of the worldwide prevalence of mental disorders in children and adolescents. J Child Psychol Psychiatry. 56(3):345–365.

Pringsheim T, Hirsch L, Gardner D, Gorman DA. 2015. The pharmacological management of oppositional behaviour, conduct problems, and aggression in children and adolescents with attention-deficit hyperactivity disorder, oppositional defiant disorder, and conduct disorder: a systematic review and meta-analysis. Part 2: antipsychotics and traditional mood stabilizers. Can J Psychiatry. 60(2): 52–61.

Puiu AA, Wudarczyk O, Goerlich KS, Votinov M, Herpertz-Dahlmann B, Turetsky B, Konrad K. 2018. Impulsive aggression and response inhibition in attention-deficit/hyperactivity disorder and disruptive behavioral disorders: findings from a systematic review. Neurosci Biobehav Rev. 90:231–246.

Rubia K. 2011. "Cool" inferior frontostriatal dysfunction in attention-deficit/hyperactivity disorder versus "hot" ventromedial orbitofrontal-limbic dysfunction in conduct disorder: a review". Biol Psychiatry. 69(12):e69–e87.

Rubia K. 2018. Cognitive neuroscience of attention deficit hyperactivity disorder (ADHD) and its clinical translation. Front Hum Neurosci. 12(March):100–123.

Rubia K, Halari R, Cubillo A, Mohammad AM, Brammer M, Taylor E. 2009. Methylphenidate normalises activation and functional connectivity deficits in attention and motivation networks in medication-naïve children with ADHD during a rewarded continuous performance task. Neuropharmacology. 57(7–8):640–652.

Rubia K, Halari R, Cubillo A, Mohammad AM, Scott S, Brammer M. 2010. Disorder-specific inferior prefrontal hypofunction in boys with pure attention-deficit/hyperactivity disorder compared to boys with pure conduct disorder during cognitive flexibility. Hum Brain Mapp. 31(12):1823–1833.

Rubia K, Smith AB, Halari R, Matsukura F, Mohammad M, Taylor E, Brammer MJ. 2009. Disorder-specific dissociation of orbitofrontal dysfunction in boys with pure conduct disorder during reward and ventrolateral prefrontal dysfunction in boys with pure ADHD during sustained attention. Am J Psychiatry. 166(1):83–94.

Schachar R, Tannock R. 1995. Test of four hypotheses for the comorbidity of attention-deficit hyperactivity disorder and conduct disorder. J Am Acad Child Adolesc Psychiatry. 34(5):639–648.

Scholl J, Klein-Flügge M. 2018. Understanding psychiatric disorder by capturing ecologically relevant features of learning and decision-making. Behav Brain Res. 355:56–75.

Solanto M. v. 1998. Neuropsychopharmacological mechanisms of stimulant drug action in attention-deficit hyperactivity disorder: a review and integration. Behav Brain Res. 94(1):127–154.

Sonuga-Barke EJS, Cortese S, Fairchild G, Stringaris A. 2016. Annual research review: Transdiagnostic neuroscience of child and adolescent mental disorders - differentiating decision making in attention-deficit/hyperactivity disorder, conduct disorder, depression, and anxiety. J Child Psychol Psychiatr. 57(3):321–349.

The MathWorks Inc. 2014. MATLAB 2014a.

Turgay A. 2004. Aggression and disruptive behavior disorders in children and adolescents. Expert Rev Neurother. 4(4):623–632.

Wechsler D. 2011. Wechsler intelligence scale for children – Fourth Edition. New York: Pearson.

Wechsler D. 2012. Wechsler adult intelligence scale – Fourth Edition. New York: Pearson.

White SF, Fowler K. a, Sinclair S, Schechter JC, Majestic CM, Pine DS, Blair RJ. 2014. Disrupted expected value signaling in youth with disruptive behavior disorders to environmental reinforcers. J Am Acad Child Adolesc Psychiatry. 53(5):579–588.

White SF, Pope K, Sinclair S, Fowler KA, Brislin SJ, Williams WC, Pine DS, Blair RJR. 2013. Disrupted expected value and prediction error signaling in youths with disruptive behavior disorders during a passive avoidance task. Am J Psychiatry. 170(3):315–323.

World Health Organization. 1992. The ICD-10 classification of mental and behavioural disorders: clinical descriptions and diagnostic guidelines. Geneva: World Health Organization.