



OPEN ACCESS

EDITED BY
Vincent D. Campese,
University of Evansville, United States

REVIEWED BY
Bernard W. Balleine,
University of New South Wales,
Australia
Yingqi Gu,
Zhejiang Sci-Tech University, China

*CORRESPONDENCE
Dirk E. M. Geurts
Dirk.Geurts@radboudumc.nl

SPECIALTY SECTION
This article was submitted to
Learning and Memory,
a section of the journal
Frontiers in Behavioral Neuroscience

RECEIVED 07 June 2022
ACCEPTED 21 September 2022
PUBLISHED 14 October 2022

CITATION
Geurts DEM, von Borries K, Huys QJM,
Bulten BH, Verkes R-J and Cools R
(2022) Psychopathic tendency
in violent offenders is associated with
reduced aversive Pavlovian inhibition
of behavior and associated striatal
BOLD signal.
Front. Behav. Neurosci. 16:963776.
doi: 10.3389/fnbeh.2022.963776

COPYRIGHT
© 2022 Geurts, von Borries, Huys,
Bulten, Verkes and Cools. This is an
open-access article distributed under
the terms of the [Creative Commons
Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use,
distribution or reproduction in other
forums is permitted, provided the
original author(s) and the copyright
owner(s) are credited and that the
original publication in this journal is
cited, in accordance with accepted
academic practice. No use, distribution
or reproduction is permitted which
does not comply with these terms.

Psychopathic tendency in violent offenders is associated with reduced aversive Pavlovian inhibition of behavior and associated striatal BOLD signal

Dirk E. M. Geurts^{1,2*}, Katinka von Borries³,
Quentin J. M. Huys⁴, Berend H. Bulten³,
Robbert-Jan Verkes^{1,3} and Roshan Cools^{1,2}

¹Centre for Cognitive Neuroimaging, Donders Institute for Brain, Cognition and Behaviour, Radboud University, Nijmegen, Netherlands, ²Department of Psychiatry, Radboud University Medical Center, Nijmegen, Netherlands, ³Pompestichting Center for Forensic Psychiatry, Pro Persona Mental Health, Nijmegen, Netherlands, ⁴Division of Psychiatry and Max Planck UCL Centre for Computational Psychiatry and Ageing Research, Mental Health Neuroscience Department, Institute of Neurology, University College London, London, United Kingdom

Background: Violent offenders with psychopathic tendencies are characterized by instrumental, i.e., planned, callous, and unemotional (aggressive) behavior and have been shown to exhibit abnormal aversive processing. However, the consequences of abnormal aversive processing for instrumental action and associated neural mechanisms are unclear.

Materials and methods: Here we address this issue by using event-related functional magnetic resonance imaging (fMRI) in 15 violent offenders with high psychopathic tendencies and 18 matched controls during the performance of an aversive Pavlovian-to-instrumental transfer paradigm. This paradigm allowed us to assess the degree to which aversive Pavlovian cues affect instrumental action and associated neural signaling.

Results: Psychopathic tendency scores were associated with an attenuation of aversive Pavlovian inhibition of instrumental action. Moreover, exploratory analyses revealed an anomalous positive association between aversive inhibition of action and aversive inhibition of BOLD signal in the caudate nucleus of violent offenders with psychopathic tendencies. In addition, psychopathic tendency also correlated positively with amygdala reactivity during aversive versus neutral cues in Pavlovian training.

Conclusion: These findings strengthen the hypothesis that psychopathic tendencies in violent offenders are related to abnormal impact of aversive

processing on instrumental behavior. The neural effects raise the possibility that this reflects deficient transfer of aversive Pavlovian inhibitory biases onto neural systems that implement instrumental action, including the caudate nucleus.

KEYWORDS

psychopathy, Pavlovian-to-instrumental transfer, inhibition, fMRI, amygdala, caudate, putamen

Introduction

Instrumental decision making is susceptible to emotional/affective influences (Estes and Skinner, 1941; Damasio, 1997). Evidence suggests that this affective biasing of action selection can reflect an interaction between distinct behavioral control systems (Cardinal et al., 2002; Dayan et al., 2006; Kahneman and Frederick, 2007). For example, instrumentally controlled action selection is well established to be sensitive to biasing by a Pavlovian or hardwired “affective” system that regulates innately specified responses to aversive stimuli (Dayan and Seymour, 2013; Guitart-Masip et al., 2014). This Pavlovian system allows agents to control behavior through strategies that have been learnt across a lifetime and/or generations to be adaptive and thus to be generalizable to novel situations. Examples of such strategies are our tendencies to promote approach (and suppress withdrawal) actions when facing reward or to suppress approach (and potentiate withdrawal) actions when facing punishment. These strategies allow us to circumvent more expensive, rational instrumental (context-appropriate) calculations and to make judgments quickly and efficiently. However, they can also contribute to maladaptive behavior.

Anomalies in the interaction between these Pavlovian and instrumental control systems have been proposed to account for behavioral impairments seen in a wide variety of neuropsychiatric disorders (Dayan et al., 2006; Heinz et al., 2016; Huys et al., 2016; Hallquist et al., 2018; Nord et al., 2018; Chen et al., 2021). Here we focus on the high end of a psychiatric dimension (Patrick, 2022) that imposes a large burden on individual victims and society as a whole: psychopathic tendency. Specifically we study the interaction between Pavlovian and instrumental control systems in a group of violent offenders with high degrees of psychopathic tendency and a group of matched healthy controls. Psychopathic tendency is characterized by affective and behavioral anomalies (De Brito et al., 2021) and has been associated, in violent offenders, with “instrumental aggression” (i.e., callous and unemotional aggressive behavior) and high rates of recidivism even after prison sentences (Hare, 2003; Leistico et al., 2008; Warren and Burnette, 2013). Given our interest in psychopathic tendency,

we employed the Psychopathy Checklist—revised (PCL-R) (Hare, 2003) as a psychological assessment tool to quantify, in each violent offenders, the degree of psychopathic tendency. The degree to which variation in Pavlovian-instrumental interaction varies with regard to individual differences in the PCL-R score were then assessed using correlational analyses.

A core feature of the crimes committed by violent offenders with psychopathic tendency is their “instrumentality,” i.e., their planned, callous, and unemotional nature (Blair, 2001; De Brito et al., 2021). These crimes are premeditated and committed to achieve a desired goal at the expense of others. Despite the centrality of such callous and unemotional action in clinical observations and in elaborate cognitive models of psychopathic tendency [e.g., the violence inhibition model (Blair, 2005)], neuroscientific research on the mechanisms of instrumental action (i.e., actions planned to obtain a certain outcome) in the face of aversive cues is scarce. So far, the neuroscience of psychopathic tendency has focused mainly on reduced affective (primarily aversive) processing *per se* and associated neural signals, for example, in limbic circuitry (Brook et al., 2013). There is evidence (albeit in small samples) that people with psychopathic tendency respond normally to unconditioned aversive Pavlovian stimuli (US), but that their psychophysiological responses to conditioned aversive stimuli (CS) are compromised (Flor et al., 2002; Veit et al., 2002; Birbaumer et al., 2005; Rothmund et al., 2012; Schultz et al., 2016). However, it is unclear how such a deficiency in aversive information processing is related to the behavioral abnormalities of psychopathic tendency. Studies focusing on affective anomalies *per se* do not provide insight in the behavioral deficits that might stem from these affective anomalies. We set up the current study to test directly the hypothesis that psychopathic tendency is associated not only with abnormal aversive processing *per se*, but rather also with reduced transfer of aversive Pavlovian biases to instrumental behavior. We thus addressed one instance of the more general proposal that neuropsychiatric abnormality is associated with an absence of Pavlovian solutions to behavioral control.

The present study was conducted around the same time as another study we performed with violent offenders

(Ly et al., 2016) to test this same hypothesis. This prior study indeed demonstrated reduced potentiation of instrumental avoidance (versus approach) actions by aversive angry (versus appetitive happy) faces in a group of violent offenders compared with a group of matched controls. The added value of the present study is threefold. First, we provide a conceptual replication, thus increasing the construct validity of this prior finding by showing reduced impact of aversive Pavlovian cues on instrumental action in a different group of violent offenders with high levels of psychopathic tendencies, using a different paradigm. Notably, by including a neutral Pavlovian cue, this paradigm allowed us to establish that the altered impact of Pavlovian cues was due to reductions in aversive bias instead of increases in appetitive bias. Second, the present study addresses neural BOLD responses associated with aversive Pavlovian conditioning and the influence of aversive Pavlovian cues on instrumental behavior in violent offenders, showing a key role for the striatum in abnormal Pavlovian control of behavior. Finally, we demonstrate that the behavioral and neural changes are a function of individual differences in psychopathic tendency.

In our previous study, affective biases by facial cues were indexed during one and the same instrumental learning phase (Ly et al., 2016). By contrast, the paradigm employed here comprised three separate phases, allowing us to disentangle (i) instrumental action learning impairment, indexed during a first phase, from (ii) changes in the learning of, and responsiveness to Pavlovian cues themselves, indexed during a second Pavlovian conditioning phase, and (iii) changes in the key process of interest: Pavlovian-to-instrumental transfer, indexed during a final task phase. This key PIT process of interest was anticipated, based on prior work (Huys et al., 2011, 2016; Geurts et al., 2013a), to surface, across all participants, as potentiation of aversive instrumental withdrawal actions, but suppression of instrumental approach actions in the context of aversive Pavlovian cues, i.e., stimuli that predict aversive outcomes. Following our prior observation (Ly et al., 2016), violent offenders were expected to exhibit reduced impact of aversive cues on both types of instrumental action and we assess specifically whether this surfaces in a psychopathic tendency-dependent manner. Thus, we predicted that they exhibit reduced aversive inhibition of approach as well as reduced aversive potentiation of withdrawal actions.

Next, we assessed the neural mechanisms underlying the aversive PIT effects. Animal and human studies consistently implicate frontostriatal brain regions in instrumental action, especially the dorsomedial (caudate nucleus) and dorsolateral (putamen) parts of the striatum and the ventromedial regions of the prefrontal cortex (Tricomi et al., 2004; Valentin et al., 2007; Balleine and O'Doherty, 2009; Wunderlich et al., 2009; Dolan and Dayan, 2013). In addition, affective information is known to influence instrumental actions *via* the amygdala (Cardinal et al., 2002; Talmi et al., 2008; Balleine and O'Doherty, 2009;

Prevost et al., 2012; Geurts et al., 2013a; Ly et al., 2014), and extensive evidence implicates dysfunction of the amygdala in people with psychopathic tendency (Veit et al., 2002; Birbaumer et al., 2005; Blair, 2008; Glenn and Raine, 2009; Moul et al., 2012). Thus, we anticipated, that violent offenders with psychopathic tendency would exhibit changes in aversive cue-related BOLD signal in the amygdala as well as differential aversive modulation of instrumental action-related signals in frontal and striatal brain regions. To this end, we focused our primary analyses on the striatum, ventromedial prefrontal cortex, and the amygdala.

Materials and methods

Subjects

Eighteen male violent offenders with psychopathic tendency (three left-handed) volunteered and were selected based on available information about clinical status and history from an in-patient population of a forensic hospital (**Supplementary material** and methods). All received a court-imposed placement under a hospital order with imprisonment for committing violence offenses repeatedly, including murder, slaughter, battery, rape, while suffering from psychiatric illness or disorder. The violent offenders all had a score of ≥ 26 on the Hare Psychopathy Check List-Revised (PCL-R) (Hare, 2003; **Table 1**). Additionally, twenty healthy men matched for age and IQ without criminal records or a history of psychiatric disorders were recruited from among the employees of the same hospital by advertisement. Participants in both groups were screened for drug use and for medical/neurological history (**Supplementary material** and methods and **Table 1**). Considering the particularities of the population, the testing environment, and the time period when testing was possible, these were the maximum numbers of inclusion.

Following previous studies (Brazil et al., 2009; von Borries et al., 2009), exclusion criteria were all major Axis-I and Axis-II disorders except for cluster B personality disorders in violent offenders, psychotropic medication, cannabis or other drug use 1 week before, alcohol or oxazepam use within 24 h before experiment, visual disorder, and neurological disorder. Furthermore, individuals not eligible for MRI scanning were excluded.

All participants received oral and written information about the experiment and gave written informed consent. They received payment as a reimbursement for participation. The study was performed in accordance with the Declaration of Helsinki and approved by the local ethical committee (NL30545.091.09).

Two violent offenders withdrew from participation and one violent offender was excluded because of excessive head

TABLE 1 Group characteristics (mean, standard deviation) of the group of violent offenders with psychopathic tendency and healthy matched control subjects.

	Violent offenders with psychopathic tendency (<i>n</i> = 15)	Healthy controls (<i>n</i> = 18)	Statistics (<i>P</i>-value)
Age	40.2 (9.1)	41.2 (10.4)	0.78
IQ (NLV)	101.7 (8.8)	101.5 (8.7)	0.96
PCL-R total	30.7 (4.0)	–	–
PCL-R factor 1	11.9 (2.9)	–	–
PCL-R factor 2	13.9 (2.1)	–	–

Exclusion criteria for both groups were: (i) Use of alcohol more than 3 units/day during the week preceding the experimental measure and use of alcohol within 24 h of the measurement.

(ii) Use of cannabis or other illicit drugs within the week before measurement and use of psychotropic medication other than oxazepam during the 5 days before measurement.

(iii) Use of oxazepam within 12 h before measurement.

(iv) Smoking within 3 h before measurement.

(v) History of trauma capitis, visual and auditory disorders, neurological disorders, first degree relative with any relevant neurological disorders.

movement (more than twice the voxel size). Two non-criminal healthy controls were excluded because their behavioral data suggested they did not follow the instructions during the PIT stage [despite instructions to play the instrumental game (see paradigm) these participants determined their actions solely on the Pavlovian CS, but never on the instrumental stimuli in more than half of the trials: 58 and 83% resp., compared with on average 1%, range 0–17%, for all other participants]. Moreover, due to technical issues with the scanner and excessive head movement only one of two runs could be analyzed for one healthy control and two violent offenders. Thus, we analyzed datasets of 15 violent offenders and 18 healthy controls.

Pavlovian-instrumental transfer paradigm

Subjects performed a computerized task to assess aversive PIT (Geurts et al., 2013a; [Supplementary material](#); [Figure 1](#)). The experiment consisted of three stages: (1) instrumental, (2) Pavlovian, and (3) PIT stage. The instrumental stage contained two Action Contexts: (i) a context in which the active response led to an approach action and (ii) another in which the active response led to a withdrawal action ([Figure 1A](#)). In the approach Action Context subjects learned through monetary feedback (wins and losses) whether to “collect” the instrumental stimulus (approach-go) or not (approach-no-go). In the withdrawal Action Context they learned to avoid collecting instrumental stimuli (withdrawal-go) or not (withdrawal-no-go). Instrumental stimuli were randomly assigned to one of the four trial types. Thus, in both the approach and withdrawal Action Contexts, there were two go-stimuli, which yielded reward more often (i.e., ~85% of the cases) after active

responses (and punishment after not responding), and two nogo-stimuli, which yielded reward more often (i.e., also ~85% of the cases) after not responding (and punishment after go-responding).

The second, Pavlovian stage consisted of repeated presentation of three audiovisual stimuli ([Figure 1B](#)): The appetitive and aversive conditioned stimuli (CS) were followed, respectively, by appetitive or aversive juice (i.e., the unconditioned stimuli USs) on 50% of trials. The neutral CS resulted in no outcome. The appetitive juice was based on subjective preference for apple, orange, or strawberry lemonade. The aversive juice was a bitter magnesium sulfate solution (0.3M). Conditioning was assessed in two ways: (1) subjects indicated the degree to which they liked each of the CSs (and USs) by use of visual analog scales (VAS), before and after the experiment; (2) subjects chose one of the two presented Pavlovian stimuli (presented for 2 s; ITI 0.5 s) in extinction on 12 interspersed query trials.

In the third (PIT) stage stimulus presentation was the same as in the instrumental stage, except that Pavlovian stimuli tiled the background from 250 ms before ([Larson et al., 2013](#)) and no outcomes were presented ([Figure 1C](#)). Subjects were instructed that their choices counted toward the final monetary total, and that the juices associated with the Pavlovian outcomes were collected outside the scanner for them to drink afterward. There were two independent runs which each comprised different stimuli/CSs. Both runs included all three stages, and were separated by a 2-min break.

Image acquisition

Whole-brain imaging was performed on a 3 Tesla MR scanner (Magnetom Trio Tim, Siemens Medical Systems, Erlangen, Germany). Functional data were obtained using a multi-echo gradient T2*-weighted echo-planar scanning sequence ([Poser et al., 2006](#); [Supplementary material](#)).

Analyses

Behavioral data analysis

In keeping with our research aims we assessed group differences as well as parametric associations with PCL-R score. These scores were only available for the violent offenders. The behavioral data were analyzed using the statistic software SPSS 16.0 and Matlab® 2009b.

Instrumental training

First, the proportion of correct responses was calculated for the first ten and last ten trials for each of the four trial types (covering all 80 instrumental trials). To assess whether subjects learned to make the correct choice, data were

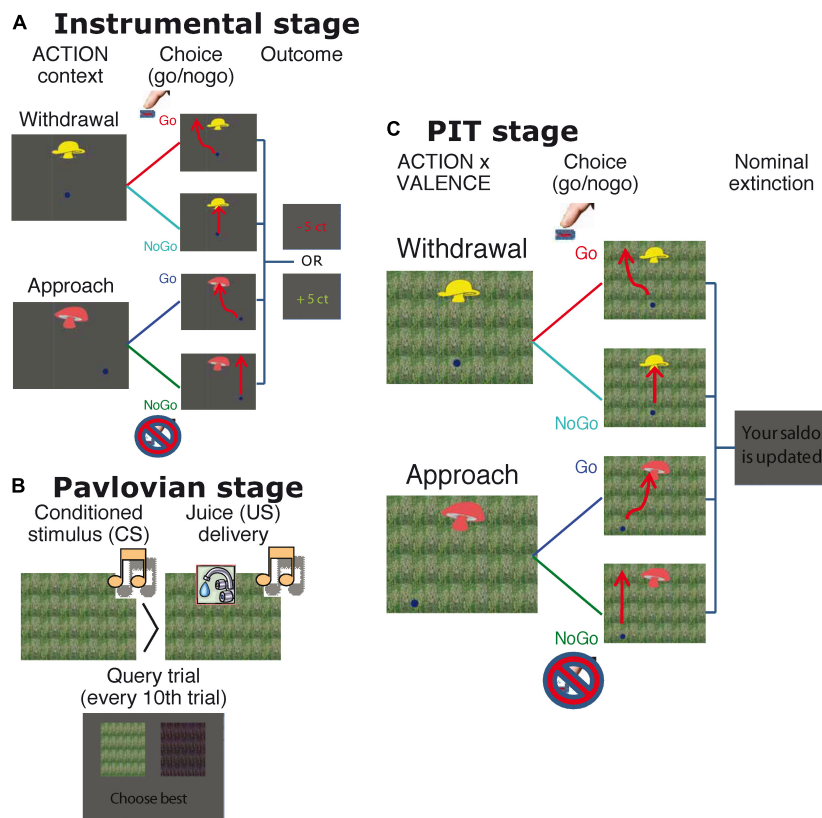


FIGURE 1

Task details. **(A)** Instrumental stage. Trials started with the appearance of the instrumental stimulus at the top center of the screen and of a dot at the bottom of the screen. In approach trials, the dot started either on the left or on the right bottom side of the screen. Participants could choose to do nothing (approach-no-go), in which case the dot would wiggle past the instrumental stimulus. Alternatively, they could push the button repeatedly to steer the dot through the instrumental stimulus (approach-go). In withdrawal trials, the dot started centrally at the bottom beneath the instrumental stimulus. Participants could choose to push the button repeatedly to avoid moving through instrumental stimulus (withdrawal-go) or to do nothing (withdrawal-no-go). The four possible trajectories are drawn in the figure (red lines). If the dot entered the goal region, then the instrumental stimulus was collected. After the dot moved outside the window feedback was provided. Thus, there were 2 ACTION contexts (approach and withdrawal), with each 4 different instrumental stimuli, with 2 stimuli resulting more often in reward after a go-action and 2 resulting more often in reward after a no-go. Each stimulus was presented 10 times, resulting in $(2 \times 4 \times 10=)$ 80 instrumental trials (divided in mini blocks of 8 withdrawal or approach trials). The straight line just to one side of the instrumental stimulus was a reflecting boundary that the dot could not cross. Timings were as follows: Instrumental stimuli were presented for 2.5 s, during which responses were collected. After 2.5 s, feedback was presented for 1 s. The intertrial interval (ITI) was 1 s (blank screen). **(B)** Pavlovian stage. Each Pavlovian conditioned stimuli (CS) was presented 20 times, and for each session there was a separate set of three stimuli. Stimulus presentation order was fully randomized across participants. Stimulus duration was 4.5 s, and juice delivery (2 ml) occurred between 0 and 1.5 s after stimulus onset. The ITI was 1 s. Query trials were presented after every 10 Pavlovian trials. On these trials, participants chose one of the two presented Pavlovian (audiovisual) stimuli (presented for 2 s; ITI 0.5 s) without any feedback. **(C)** PIT stage. The PIT stage paralleled the instrumental training, except that Pavlovian CSs tiled the background. Each instrumental stimulus was presented 12 times and each Pavlovian CS was presented 32 times, counterbalanced across the different instrumental stimuli. No outcomes were presented, but participants were instructed that their choices counted toward the final total. Participants were explicitly instructed that the juices were collected outside the scanner, and they agreed before the start of the experiment to drink them afterward. Timing of one trial was as follows: 250 ms after the onset of the Pavlovian stimulus, the instrumental stimulus (and dot) was overlaid on top of this Pavlovian stimulus. Duration of the instrumental stimulus was 2.5 s; duration of the Pavlovian stimulus was 2.75 s. Upon offset of both stimuli, feedback was presented, which consisted only of the words "Balance is updated" (duration = 1 s, ITI = 1 s). Note that there were two runs in which all three stages (with new independent Pavlovian and instrumental stimuli) were assessed.

averaged across sessions and submitted to a repeated measures analysis of variance (rmANOVA) with Time (beginning/end of instrumental training), Action Context (approach/withdrawal) and Response (go/nogo) as within-subject and Group (healthy controls/violent offenders) as between-subject factor. Second, we assessed whether the learned behavior generalized to the PIT stage. Therefore, the factor Time was changed to include three

levels: the end of the instrumental training and the beginning and the end of the PIT stage.

Pavlovian conditioning

Non-parametric tests were used to assess the proportion of correct responses on Pavlovian query trials and pre- and

post-conditioning VAS ratings of the CS, because data were not distributed normally.

Pavlovian-instrumental transfer

The behavioral outcome measure was proportion of go actions, $p(\text{go})$, as a function of trial type (i.e., Action Context and CS Valence). Effects of CS Valence and Action on $p(\text{go})$ reflect PIT effects on choice. This dependent variable was first averaged across runs before it were submitted to an rmANOVA with Action Context (approach/withdrawal), and CS Valence(neutral/aversive) as within-subject factors and Group (healthy controls versus violent offenders) as a between-subject factor. Note that we focused our analyses on aversive PIT, based on our hypothesis (see Section “Introduction”) and on our previous work ($n = 33$) showing that the current paradigm was not sensitive to (and therefore not valid to assess) appetitive PIT [neutral vs. appetitive (Geurts et al., 2013a), [Supplementary results](#)]. The PCL-R-score was added as a covariate to assess its association with aversive PIT.

Functional magnetic resonance imaging analysis

Functional magnetic resonance imaging analysis was performed with SPM5 software (Wellcome Trust Centre for Cognitive Neuroimaging, London, UK). Pre-processing steps and first-level fMRI analysis were exactly as described by Geurts et al. (2013a): Pre-processing steps included applying a PAID-weight algorithm (Poser et al., 2006) to combine the different echoes, slice-time correction, coregistration, normalization based on parameters estimated through segmentation of the structural images, and smoothing.

The primary analysis was restricted to the PIT-stage. At the subject level a general linear model (GLM) was specified with 6 main regressors (4 of interest) representing the onset of the six different PIT trials of this paradigm [Action Context (approach/withdrawal) – Valence(appetitive/neutral/aversive)]. For each main regressor two additional parametric regressors were added (Büchel et al., 1996): The PIT-regressor (Talmi et al., 2008) was a parametric modulator of BOLD responses by the number of button presses per trial. A further parametric regressor contained the expectation associated with each instrumental stimulus (the Q -value) per trial as estimated from a model-based analysis of behavior (Huys et al., 2011). This was done based on prior data showing that BOLD signal in the prefrontal cortex and striatum, our regions of interest, covary with instrumental action value (e.g., Valentin et al., 2007; Wunderlich et al., 2009). As such, this approach maximized the degree to which our GLM captured variability in relevant BOLD signal. Furthermore, realignment parameters were added, high-pass filtering (128 s) was applied and parameter estimates were obtained by maximum-likelihood estimation (AR1).

The parameter estimates for the 4 parametric PIT-regressors were used in a $2 \times 2 \times 2$ rmANOVA at the group-level (with

random effects) with Action Context(approach/withdrawal) and Valence (neutral/aversive) as within-subject factors and GROUP (healthy controls/Vos) as between-subjects factor. Planned contrasts were the same as in Geurts et al. (2013a), but now assessed as a function of Group: [aversive-neutral] to reveal regions involved in aversive PIT across Action Contexts, and [(approach aversive-approach neutral)–(withdrawal aversive-withdrawal neutral)] to reveal regions involved in action-specific aversive PIT, and [approach–withdrawal] to reveal action-specific regions.

To capture group differences in brain-behavior associations, beyond those related to trial-by-trial variation, we contrasted the main regressors (Talmi et al., 2008; Geurts et al., 2013a) at the subject-level to calculate the main effect of Valence [(approach aversive + withdrawal aversive)–(approach neutral + withdrawal neutral)] and an interaction between Valence and Action Context [(approach aversive–approach neutral)–(withdrawal aversive–withdrawal neutral)]. The resulting individual contrasts were then used in a two-sample t -test at the group-level with behavioural aversive PIT-effects [$p(\text{go})$] as a covariate for each group separately enabling comparison between groups. Thus, these analyses reveal regions, on a subject-by-subject basis, in which CS Valence-dependent BOLD signal change during the PIT stage was associated with aversive PIT. This association was assessed as a function of Action Context and Group. These analyses were repeated with PCL-R score (instead of behavioral PIT) as a covariate to assess whether CS Valence dependent BOLD signal change during the PIT stage was associated with psychopathy severity.

Next, additional analyses were performed to assess whether positive behavioral and fMRI findings from the PIT stage could be explained by BOLD signal change in the Pavlovian conditioning stage. Thus, we analyzed CS Valence-dependent BOLD signal change during the Pavlovian training phase as a function of individual differences in PCL-R score, aversive PIT and neural signaling in the caudate nucleus (see Section “Results”) during the PIT stage (each inserted as a covariate in three separate whole-brain analyses of CS Valence-related signals during the Pavlovian training phase).

First, at the subject level, a GLM was specified with six main regressors of interest representing the onset of the CS trials (during which no US was presented) in the beginning and the end of the conditioning stage: Valence (appetitive/neutral/aversive) \times Time (early/late). This latter distinction between early and late acquisition, was based on evidence of rapid habituation of the responses in the amygdala during conditioning (Birbaumer et al., 2005). Early trials were the first three trials following the first US presentation for aversive and appetitive CS trials. For the neutral CS, the early trials were the first three presentations and the late trials were all the remaining CS presentations thereafter. To capture the other parts of the Pavlovian training, four regressors were added: for appetitive and aversive US onsets, for juice delivery

onset (duration 2 s) and for the query trial onset (duration 2 s). Realignment and high-pass filtering was applied as before. Parameter estimates were obtained by maximum-likelihood estimation (AR1). We calculated the main effect of CS Valence (i.e., aversive-neutral) at the subject level for early, late and overall [early + late] conditioning and correlated the effects at the group-level (one sample *t*-test with covariate) with the PCL-R score; behavioral aversive PIT; and the extracted betas of the caudate nucleus as covariates of interest.

Regions of interest analysis

We report those effects that survive family wise error (FWE) correction for multiple comparisons across the whole brain ($P_{WB} < 0.05$, voxel-level) or regions of interest (ROIs, see Section “Introduction” for rationale): Following exactly the same procedure as in our previous work (Geurts et al., 2013a) the bilateral amygdala, caudate nucleus and putamen were defined using the automated anatomical labeling atlas in the WFU PickAtlas toolbox (Tzourio-Mazoyer et al., 2002). The bilateral nucleus accumbens was segmented for each participant using the FSL FIRST segmentation tool (Patenaude et al., 2011). These individual segments were then overlaid onto each other, generating one nucleus accumbens for the group (cf. Geurts et al., 2013a). Again following this prior work, we used the action-specific (approach > withdrawal) activation cluster [$p < 0.001$ uncorrected; peak voxel MNI-coordinates: $(-8, 36, -8)$] revealed by this previous PIT study to assess action specific signal in the vmPFC. The left and right elements of the bilateral volumes of interest were combined using Marsbar™ (Brett et al., 2002). Masks of the ROIs can be found in the [Supplementary material](#).

Results

Behavioral data

Instrumental and Pavlovian stage

Healthy controls tended to learn faster than the violent offenders during instrumental training {Group \times Time $F_{(1,31)} = 4.3, p = 0.072$; note, however, that they did not differ from violent offenders in terms of instrumental performance during the PIT stage [main effect of Group: $F_{(1,31)} = 2.0, P = 0.17$]. There were no other relevant group differences in the instrumental and Pavlovian stages of the PIT paradigm ([Supplementary material](#)).

Pavlovian-instrumental transfer

Data revealed the expected action-specific PIT effects across groups (Huys et al., 2011; cf. Geurts et al., 2013a): Aversive stimuli inhibited approach, but promoted withdrawal [[Figure 2A](#); interaction Action Context (approach/withdrawal) \times Valence (neutral/aversive):

$F_{(1,31)} = 8.6, p = 0.006$; Valence during approach: $F_{(1,31)} = 4.7, p = 0.037$; Valence during withdrawal: $F_{(1,31)} = 4.3, p = 0.046$]. Subjects made more go-responses in the approach than in the withdrawal context overall [main Action Context effect: $F_{(1,31)} = 9.9, p = 0.004$]. However, in contrast to our expectation no significant main effect of or interaction with Group was found ($F < 2.4, p > 0.161$, [Supplementary Table 2](#)).

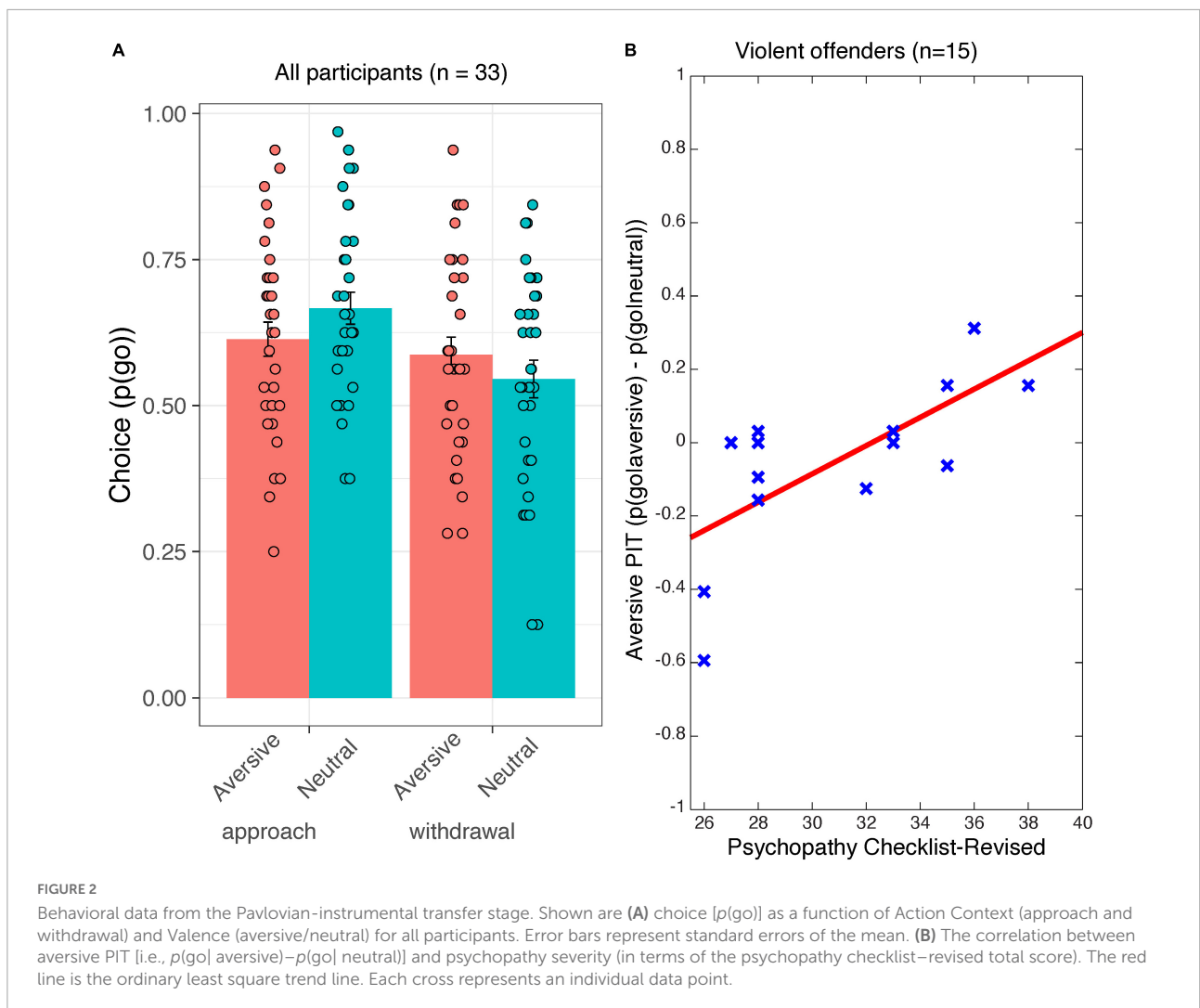
Regarding individual differences in psychopathic tendencies, we found that higher PCL-R scores were associated with less aversive inhibition, in a manner that was action-non-specific [interaction PCL-R score \times Valence (neutral/aversive): $F_{(1,13)} = 12.6, p = 0.004$; [Figure 2B](#)]. Thus, higher PCL-R scores were associated with reduced aversive inhibition of approach actions [simple effects for approach block only: $(p(\text{go}|\text{aversive}\&\text{approach}) - p(\text{go}|\text{neutral}\&\text{approach})) \times \text{PCL-R score}$: $F_{(1,13)} = 5.6, p = 0.035$] and a tendency toward increased aversively motivated withdrawal actions [simple effects for withdrawal block only: $(p(\text{go}|\text{aversive}\&\text{withdrawal}) - p(\text{go}|\text{neutral}\&\text{withdrawal})) \times \text{PCL-R score}$: $F_{(1,13)} = 4.3, p = 0.059$].

Imaging data

In contrast to our hypothesis, the neural responses during aversive PIT did not differ between groups [no significant Action Context (approach/withdrawal) \times Valence (neutral/aversive) \times Group (violent offenders/healthy controls) interactions]. Action-specific signals (approach vs. withdrawal) across CS Valence were found in the precuneus [(12, -78, 6), $k = 5453, Z = 7.02, p_{FWE} < 0.001$, whole brain corrected], lingual [(10, -52, 52), $k = 310, Z = 5.50, p_{FWE} = 0.001$, whole brain corrected] and middle occipital gyrus [(34, -88, 2), $k = 159, Z = 4.98, p_{FWE} = 0.016$, whole brain corrected].

Moreover, within the group of violent offenders there was no significant effect of psychopathic tendency on BOLD signal during aversive PIT at the whole brain or in the pre-specified regions of interest.

Individual differences in behavioral aversive PIT [$p(\text{go}|\text{aversive}) - p(\text{go}|\text{neutral})$] correlated positively with BOLD signal (aversive main regressor - neutral main regressor) in the putamen across both groups: Greater aversive inhibition of behavior was associated with reduced BOLD signal during aversive PIT trials (versus neutral) in the left putamen [[Figure 3](#), peakvoxel MNI-coordinates $(-26, 2, 6), k = 205, Z = 4.59, p_{FWE} = 0.004$, small volume corrected]. In other words, higher BOLD signal in the putamen during the presentation of the aversive (versus neutral) Pavlovian cue was accompanied by greater disinhibition of go-actions, in line with the hypothesis that putamen signal reflects motor execution. Conversely, there was a significant difference between groups in the caudate nucleus [peakvoxel MNI-coordinates $(14, 20, 10), k = 98, Z = 4.26, p_{FWE} = 0.006$, small volume corrected, [Figure 3](#)]: By contrast to the



putamen, signal in the caudate nucleus of healthy controls (aversive versus neutral) correlated negatively with aversive PIT [$p_{\text{uncorrected}} = 0.012$ at peak voxel (14, 18, 10), Figure 3], so that greater aversive inhibition of behavior was associated with greater aversive BOLD signal during aversive (versus neutral) PIT trials. This concurs generally with the hypothesis that caudate nucleus signal reflects the operation of a more complex computational operation related to motor planning (Alexander et al., 1986; Herz et al., 2013; Provost et al., 2015), such as the transfer of affective information onto action. Critically, this positive relation between BOLD signal in the caudate during aversive trials and behavioral inhibition was completely reversed in the violent offenders so that greater aversive inhibition of behavior in the violent offenders was associated with reduced BOLD signal during aversive (versus neutral) PIT trials in the caudate nucleus [peak voxel MNI-coordinates (14, 18, 10), $k = 225$, $Z = 4.85$, $p_{\text{FWE}} = 0.039$, corrected for the whole-brain]. Thus, the caudate nucleus of violent offenders acted like their putamen,

perhaps reflecting a failure to engage the computation that is required for translating aversive information into action suppression. Note, that we repeated this correlation analysis within the violent offender group by means of Kendall's tau analyses, because Kendall's tau analysis is preferred over parametric methods for small samples and is robust to outliers (Croux and Dehon, 2010). Correlation of the mean beta estimates from the caudate nucleus with behavior remained significant [$\tau_{(15)} = 0.520$, $p = 0.008$]. Moreover, based on reviewer comments we also conducted this analyses without the two violent offenders that visually might seem to drive these findings and found that the correlation is robust to excluding these data points [$\tau_{(13)} = 0.450$, $p = 0.047$].

Next, we asked whether the psychopathic tendency-related behavioral PIT effects (Figure 2B) were accompanied by psychopathic tendency-related differences in Pavlovian conditioning. Psychopathic tendency in terms of PCL-R score was indeed related to CS-dependent BOLD signal change

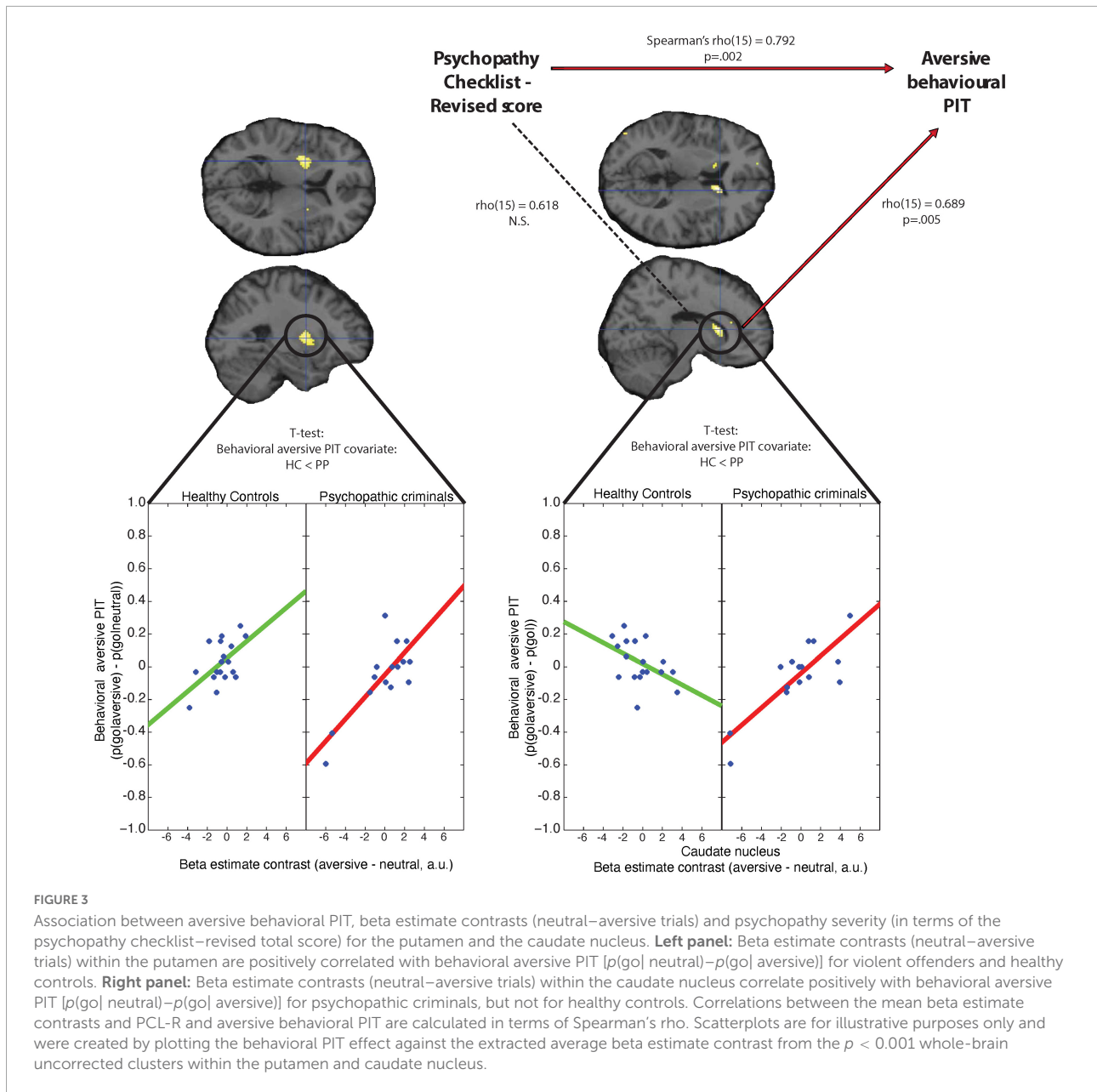


FIGURE 3

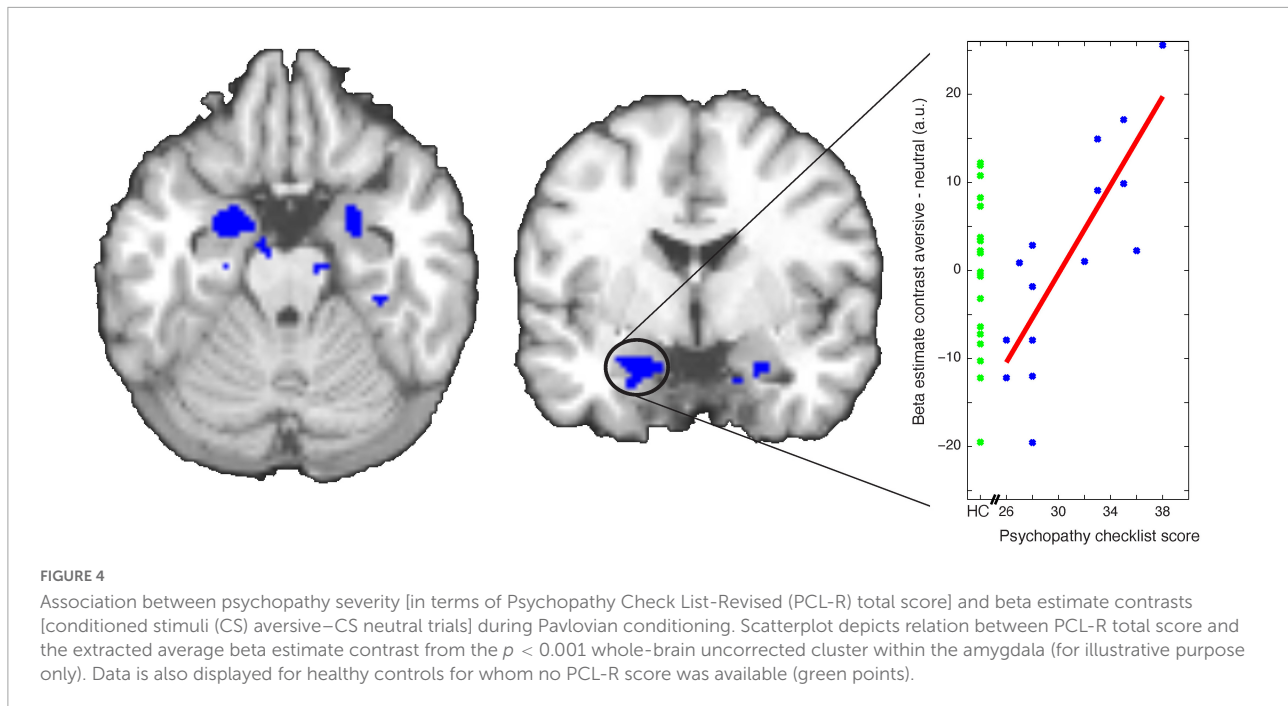
Association between aversive behavioral PIT, beta estimate contrasts (neutral–aversive trials) and psychopathy severity (in terms of the psychopathy checklist–revised total score) for the putamen and the caudate nucleus. **Left panel:** Beta estimate contrasts (neutral–aversive trials) within the putamen are positively correlated with behavioral aversive PIT [$p(\text{go}|\text{neutral}) - p(\text{go}|\text{aversive})$] for violent offenders and healthy controls. **Right panel:** Beta estimate contrasts (neutral–aversive trials) within the caudate nucleus correlate positively with behavioral aversive PIT [$p(\text{go}|\text{neutral}) - p(\text{go}|\text{aversive})$] for psychopathic criminals, but not for healthy controls. Correlations between the mean beta estimate contrasts and PCL-R and aversive behavioral PIT are calculated in terms of Spearman’s rho. Scatterplots are for illustrative purposes only and were created by plotting the behavioral PIT effect against the extracted average beta estimate contrast from the $p < 0.001$ whole-brain uncorrected clusters within the putamen and caudate nucleus.

(aversive versus neutral) over the whole conditioning stage) in the bilateral amygdala [right amygdala: peakvoxel MNI-coordinates (24 2 -24), $k = 22$, $Z = 3.51$, $p_{FWE} = 0.033$, left amygdala: peakvoxel MNI-coordinates (-18 -2 -22), $k = 37$, $Z = 3.48$, $p_{FWE} = 0.036$, small volume corrected for the bilateral amygdala; **Figure 4**]. This psychopathy severity dependent effect was also accompanied by group differences in the left amygdala [left amygdala: peakvoxel MNI-coordinates (-12 -2 -22), $k = 5$, $Z = 3.65$, $p = 0.010$, small volume corrected for the bilateral amygdala], so that violent offenders exhibited greater amygdala signal during aversive versus neutral cues compared with healthy controls. Note, that we did not find a significant association between CS-dependent amygdala

signal during Pavlovian conditioning and aversive PIT behavior ($p_{\text{uncorrected}} > 0.001$).

Discussion

The present study shows that, although we did not find behavioral group differences between healthy controls and violent offenders, higher psychopathic tendency within the violent offender sample was accompanied by attenuated inhibition of instrumental behavior in the presence of (non-consequential) Pavlovian aversive cues. In addition, while aversive PIT and caudate BOLD signal correlated negatively



in healthy volunteers, this correlation was completely reversed in the violent offenders with psychopathy (compared with non-criminal healthy controls). Moreover, within the group of violent offenders, we established a positive association between psychopathic tendency and amygdala reactivity to aversive cues during Pavlovian conditioning. Together, these data suggest that psychopathic tendency is associated with enhanced aversive Pavlovian cue reactivity (cf. Schultz et al., 2016), yet reduced aversive inhibition of instrumental behavior. In addition, these data suggest a link between psychopathic tendency, Pavlovian aversive inhibition, and striatal action selection.

The finding that increased psychopathic tendency was associated with reduced aversive inhibition is reminiscent of findings from our previous work with a comparable experimental task, showing that central serotonin depletion reduced aversive inhibition in healthy volunteers (Geurts et al., 2013b). This observation is remarkable considering the prior finding that psychopathic tendency in a violent offender sample is accompanied by reduced serotonin metabolites in the cerebrospinal fluid (Soderstrom et al., 2001, 2003) and that callous-unemotional traits in boys was related to lower serum levels of serotonin (Moul et al., 2013). In line with these observations, we found a strong correlation between the PCL-R-score and aversive inhibition in a PIT task. The next step will be to assess whether aversive Pavlovian disinhibition in psychopathy can be countered by serotonergic drugs, such as selective serotonin reuptake inhibitors, consistent with findings that provoked aggression in primary psychopathy can be reduced by serotonin augmentation by paroxetine (Fanning et al., 2014; e.g., Butler et al., 2021).

Our finding that the amygdala of violent offenders with high psychopathic tendency was more responsive to aversive (versus neutral) CSs than those with lower psychopathic tendency (Figure 4) is inconsistent with the hypothesis that violent offenders with high psychopathy severity scores would be insensitive to aversive cues (Flor et al., 2002; Birbaumer et al., 2005). This generally concurs with other prior findings that challenge the view that violent offenders with psychopathic tendency are insensitive to punishment and/or lack fear (e.g., Gregory et al., 2015; Hoppenbrouwers et al., 2016). More specifically our findings are in line with the largest study on Pavlovian conditioning in psychopathy that shows increased rather than reduced amygdala BOLD signal (Schultz et al., 2016) as well as recent meta-analyses of task-based activation studies in psychopathy that challenge the predominant view of amygdala hypo-reactivity in psychopathy (Deming and Koenigs, 2020). Remarkably, as was the case for central serotonin depletion (Geurts et al., 2013a) increased psychopathic tendency was associated with reduced aversive inhibition of both the approach and withdrawal actions. This suggests that the aversive transfer computation that is disrupted operates at the level of action intensity rather than of action valence.

Both healthy volunteers and violent offenders exhibited a negative between-participant correlation between the degree to which instrumental actions were inhibited by aversive cues and the degree to which BOLD signal in the putamen was activated during aversive cues. This observation is generally in line with putamen signal reflecting motor execution (Alexander et al., 1986; Herz et al., 2013; Provost et al., 2015). By contrast, signal in the caudate nucleus of healthy volunteers correlated

positively with the degree to which instrumental actions were inhibited by aversive cues. Thus, greater aversive inhibition was associated with greater caudate nucleus signal during aversive cues. This raises the possibility that signal in the caudate nucleus of healthy controls does not reflect motor execution *per se*, but rather reflects the operation of the aversive Pavlovian inhibition computation of interest. The key finding is that this association was completely reversed in the violent offenders, so that the across-participant pattern of caudate nucleus signal resembled that in their putamen. In line with evidence from work with experimental rodents, evidence in humans (Balleine and O'Doherty, 2009) implicates the caudate nucleus [together with the ventromedial prefrontal cortex (Valentin et al., 2007)] in the instrumental control of behavior (Tanaka et al., 2008; de Wit et al., 2012) as well as response inhibition (Watanabe and Munoz, 2010; Schmidt et al., 2020). Accordingly, one might speculate, based on our neural findings, that aversive Pavlovian disinhibition in psychopathy is accompanied by a failure of the caudate nucleus to exhibit the Pavlovian inhibition computations that it exhibits normally, as suggested by the negative relation in healthy controls. Instead, the caudate nucleus of violent offenders exhibited the across-participant pattern of effects seen in their putamen, which instead reflects increases in motor execution: greater signal with more Go actions. We note that the group differences in the continuous brain-behavior associations were not accompanied by group differences in average BOLD signal. This suggests that the behavioral deficits in violent offenders do not reflect a failure to recruit the caudate nucleus *per se*, but rather that they reflect a failure to recruit the caudate nucleus as a function of the relevant aversive inhibition computation.

One puzzle is that the range of behavioral PIT scores in the violent offenders was comparable with that in the controls. Thus, we did not provide a conceptual replication of the expected group effect on behavior found in our previous study (Ly et al., 2016): While we did find a relation between psychopathic tendency and attenuated aversive inhibition within the group of violent offenders, we did not observe the impact of aversive cues on instrumental action to be altered in the group of violent offenders compared with the group of non-violent healthy controls. This complicates the interpretation of our results. One implication is that abnormal aversive PIT *per se* is not a sufficient prerequisite for developing criminal psychopathic tendency. Criminal psychopathy might surface only if abnormal aversive PIT is accompanied, for example, by excessive impact of reward on behavior and cognition (Buckholz et al., 2010; Bjork et al., 2012; Yildirim and Derksen, 2015; Geurts et al., 2016). This would be in line with the literature on increased reward seeking and decreased sensitivity to punishment in psychopathic individuals (e.g., Newman et al., 1990). Thus, it might be that one factor or an interaction between multiple factors moderates the impact of differences in aversive PIT on clinical symptomatology (cf. Plichta and Scheres, 2014 for

explanation of the moderator model). PCL-R-scores from the healthy controls were not available and therefore we cannot exclude that a similar association exists in healthy controls. However, we think this is unlikely, because there were no correlations between scores on the Psychopathic Personality Inventory and aversive PIT (Supplementary results).

We consider several explanations for this lack of a group effect. First, we have a relatively small sample size to detect such a group difference, which might have led to false negative findings. Second, our paradigm is not optimized for indexing appetitive PIT. As the main finding of Ly et al. (2016) is based on a contrast between aversive and appetitive cues, this effect might be due to aberrant effects of violent offenders to either appetitive, aversive or both cues. In the current study, focusing on aversive versus neutral cues, we did not find this group difference. This raises the possibility that the findings of Ly et al. (2016) reflect combined modulation of appetitive biasing as well as aversive biasing. Unfortunately, the insensitivity of our paradigm to appetitive PIT (cf. Geurts et al., 2013a; see Supplementary results) precludes strong conclusion about the valence-specificity of the effects. Third, where Ly et al. (2016) use happy and sad faces as affective cues that are presented during instrumental learning, we use neutral stimuli associated with appetitive and aversive juices that we present during already learned instrumental trials presented in nominal extinction. It might be that these latter Pavlovian CS sort more effect in violent offenders than (sad) faces (cf. Brennan and Baskin-Sommers, 2021). Future studies comparing effects of facial expressions with Pavlovian CS on instrumental behavior might elucidate these discrepancies.

Finally, we highlight the following limitations of the current study: First, although our sample size is comparable to that of other neuroimaging studies focusing on psychopathic criminals (cf. Brazil et al., 2009; von Borries et al., 2009; Rothmund et al., 2012; Contreras-Rodríguez et al., 2014), we recognize the limitation of such a small sample size especially when considering our parametric analysis (Button et al., 2013). We did find statistically significant, robust effects, but future studies with larger sample sizes are necessary to confirm the relationships uncovered in this study. Second, we were not able to replicate the strong action-specific BOLD signal found in the ventromedial prefrontal cortex as observed in our previous fMRI study with the same paradigm in healthy young volunteers (Geurts et al., 2013a). We have recently replicated this ventromedial effect in young women (both healthy and with borderline personality disorder) in another independent dataset (submitted to this issue, D. E. M. Geurts, T. J. van den Heuvel, R. Cools). One factor that might account for this discrepancy is that the average performance at the end of the instrumental task was significantly better in the latter studies (with mainly graduate students) compared with that of the healthy controls in the current study [mean accuracy (SEM): 2013 study = 0.76 (0.023), current = 0.64 (0.027), t -test: $t_{36} = 3.6$, $p = 0.001$]. Thus, it

is possible that the healthy controls and patients in the current study relied to a lesser degree on a goal-directed control strategy and to a greater degree on a habitual control strategy than did the subjects in our previous study. Although speculative, this might explain why the putamen was recruited as a function of aversive PIT in the current study in both the healthy control group and the psychopathy group, which was not the case in our previous study. Relevant in this context might also be the fact that the current study included only men, whereas the other studies mainly included women. Third, we should note that our group comparison is necessarily confounded by overt criminal history. As such, we cannot and do not claim specificity of our findings to violent offenders with psychopathic tendency compared with non-psychopathic violent offenders or psychopathic non-offenders (cf. “successful psychopaths”).

In sum, our results strengthen the hypothesis that psychopathic tendency in violent offenders is associated with abnormal impact of aversive Pavlovian suppression of instrumental behavior. The neural results raise the possibility that this reflects deficits in neural computations involving the caudate nucleus.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

The studies involving human participants were reviewed and approved by Medisch Ethische Toetsing Commissie (METC) Oost-Nederland. The patients/participants provided their written informed consent to participate in this study.

Author contributions

DG and KB collected the data. DG performed the statistical analyses under supervision of RC. DG and RC wrote the first

draft of the manuscript. All authors contributed to conception and design of the study and manuscript revision, read, and approved the submitted version.

Funding

DG acknowledged the Netherlands Organisation for Health Research and Development (AGIKO grant 92003576). RC was supported by a Human Frontiers Science Program grant to Kae Nakamura, Nathaniel Daw, and RC, as well as a VIDI grant from the Innovational Research Incentives Scheme of the Netherlands Organisation for Scientific Research (VIDI grant NWO-452-08-009) and a James McDonnell scholar award. QH was supported by the German Research Foundation (Deutsche Forschungsgemeinschaft, DFG, FOR 1617: grant RA1047/2-1). R-JV was supported by grant from the National Initiative Brain and Cognition, NWO: 056-24-011.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fnbeh.2022.963776/full#supplementary-material>

References

- Alexander, G. E., DeLong, M. R., and Strick, P. L. (1986). Parallel Organization of Functionally Segregated Circuits Linking Basal Ganglia and Cortex. *Annu. Rev. Neurosci.* 9, 357–381. doi: 10.1146/annurev.ne.09.030186.002041
- Balleine, B. W., and O'Doherty, J. P. (2009). Human and Rodent Homologies in Action Control: Corticostriatal Determinants of Goal-Directed and Habitual Action. *Neuropsychopharmacology* 35, 48–69. doi: 10.1038/npp.2009.131

- Birbaumer, N., Veit, R., Lotze, M., Erb, M., Hermann, C., Grodd, W., et al. (2005). Deficient fear conditioning in psychopathy: A functional magnetic resonance imaging study. *Arch. Gen. Psychiatry* 62, 799–805. doi: 10.1001/archpsyc.62.7.799
- Bjork, J. M., Chen, G., and Hommer, D. W. (2012). Psychopathic tendencies and mesolimbic recruitment by cues for instrumental and passively obtained rewards. *Biol. Psychol.* 89, 408–415. doi: 10.1016/j.biopsycho.2011.12.003

- Blair, R. J. (2001). Neurocognitive models of aggression, the antisocial personality disorders, and psychopathy. *J. Neurol. Neurosurg. Psychiatry* 71, 727–731.
- Blair, R. J. R. (2005). Applying a cognitive neuroscience perspective to the disorder of psychopathy. *Dev. Psychopathol.* 17, 865–891.
- Blair, R. J. R. (2008). The amygdala and ventromedial prefrontal cortex: Functional contributions and dysfunction in psychopathy. *Philos. Trans. R Soc. Lond. B Biol. Sci.* 363, 2557–2565. doi: 10.1098/rstb.2008.0027
- Brazil, I. A., de Bruijn, E. R. A., Bulten, B. H., von Borries, A. K. L., van Lankveld, J. J. D. M., Buitelaar, J. K., et al. (2009). Early and Late Components of Error Monitoring in Violent Offenders with Psychopathy. *BPS* 65, 137–143. doi: 10.1016/j.biopsych.2008.08.011
- Brennan, G. M., and Baskin-Sommers, A. R. (2021). Cognitive mechanisms influencing facial emotion processing in psychopathy and externalizing. *Personal. Disord.* 12, 581–593. doi: 10.1037/per0000473
- Brett, M., Anton, J.-L., Valabregue, R., and Poline, J.-B. (2002). Region of interest analysis using an SPM toolbox. *NeuroImage* 16:497.
- Brook, M., Brieman, C. L., and Kosson, D. S. (2013). Emotion processing in Psychopathy Checklist assessed psychopathy: A review of the literature. *Clin. Psychol. Rev.* 33, 979–995. doi: 10.1016/j.cpr.2013.07.008
- Büchel, C., Wise, R. J., Mummary, C. J., Poline, J. B., and Friston, K. J. (1996). Nonlinear regression in parametric activation studies. *NeuroImage* 4, 60–66. doi: 10.1006/nimg.1996.0029
- Buckholtz, J. W., Treadway, M. T., Cowan, R. L., Woodward, N. D., Benning, S. D., Li, R., et al. (2010). Mesolimbic dopamine reward system hypersensitivity in individuals with psychopathic traits. *Nat. Neurosci.* 13, 419–421. doi: 10.1038/nn.2510
- Butler, T., Schofield, P. W., Knight, L., Ton, B., Greenberg, D., Scott, R. J., et al. (2021). Sertraline hydrochloride for reducing impulsive behaviour in male, repeat-violent offenders (ReINVEST): Protocol for a phase IV, double-blind, placebo-controlled, randomised clinical trial. *BMJ Open* 11:e044656. doi: 10.1136/bmjopen-2020-044656
- Button, K. S., Ioannidis, J. P. A., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S. J., et al. (2013). Power failure: Why small sample size undermines the reliability of neuroscience. *Nat. Rev. Neurosci.* 14, 365–376. doi: 10.1038/nrn3475
- Cardinal, R. N., Parkinson, J. A., Hall, J., and Everitt, B. J. (2002). Emotion and motivation: The role of the amygdala, ventral striatum, and prefrontal cortex. *Neurosci. Biobehav. Rev.* 26, 321–352. doi: 10.1016/S0149-7634(02)00007-6
- Chen, H., Mojtahedzadeh, N., Belanger, M. J., Nebe, S., Kuitunen-Paul, S., Sebold, M., et al. (2021). Model-Based and Model-Free Control Predicts Alcohol Consumption Developmental Trajectory in Young Adults: A 3-Year Prospective Study. *Biol. Psychiatry* 89, 980–989. doi: 10.1016/j.biopsych.2021.01.009
- Contreras-Rodríguez, O., Pujol, J., Batalla, I., Harrison, B. J., Soriano-Mas, C., Deus, J., et al. (2014). Functional Connectivity Bias in the Prefrontal Cortex of Psychopaths. *Biol. Psychiatry* 78, 647–655. doi: 10.1016/j.biopsych.2014.03.007
- Croux, C., and Dehon, C. (2010). Influence Functions of the Spearman and Kendall Correlation Measures. *Stat. Methods Appl.* 19, 497–515. doi: 10.1007/s10260-010-0142-z
- Damasio, A. R. (1997). Neuropsychology Towards a neuropathology of emotion and mood. *Nature* 386, 769–770. doi: 10.1038/386769a0
- Dayan, P., Niv, Y., Seymour, B., and Daw, N. (2006). The misbehavior of value and the discipline of the will. *Neural Netw.* 19, 1153–1160. doi: 10.1016/j.neunet.2006.03.002
- Dayan, P., and Seymour, B. (2013). “Values and actions in aversion,” in *Neuroeconomics: Decision Making and the Brain*, ed. P. W. Glimcher (Cambridge: Academic Press).
- De Brito, S. A., Forth, A. E., Baskin-Sommers, A. R., Brazil, I. A., Kimonis, E. R., Pardini, D., et al. (2021). Psychopathy. *Nat. Rev. Dis. Primers* 7:49. doi: 10.1038/s41572-021-00282-1
- de Wit, S., Watson, P., Harsay, H. A., Cohen, M. X., van de Vijver, I., and Ridderinkhof, K. R. (2012). Corticostriatal connectivity underlies individual differences in the balance between habitual and goal-directed action control. *J. Neurosci.* 32, 12066–12075. doi: 10.1523/JNEUROSCI.1088-12.2012
- Deming, P., and Koenigs, M. (2020). Functional neural correlates of psychopathy: A meta-analysis of MRI data. *Transl. Psychiatry* 10:133. doi: 10.1038/s41398-020-0816-8
- Dolan, R. J., and Dayan, P. (2013). Goals and Habits in the Brain. *Neuron* 80, 312–325. doi: 10.1016/j.neuron.2013.09.007
- Estes, W. K., and Skinner, B. F. (1941). Some quantitative properties of anxiety. *J. Exp. Soc. Psychol.* 29:390.
- Fanning, J. R., Berman, M. E., Guillot, C. R., Marsic, A., and McCloskey, M. S. (2014). Serotonin (5-HT) augmentation reduces provoked aggression associated with primary psychopathy traits. *J. Pers. Disord.* 28, 449–461. doi: 10.1521/pedi_2012_26_065
- Flor, H., Birbaumer, N., Hermann, C., Ziegler, S., and Patrick, C. J. (2002). Aversive Pavlovian conditioning in psychopaths: Peripheral and central correlates. *Psychophysiology* 39, 505–518.
- Geurts, D. E. M., Huys, Q. J. M., den Ouden, H. E. M., and Cools, R. (2013a). Aversive Pavlovian Control of Instrumental Behavior in Humans. *J. Cogn. Neurosci.* 25, 1428–1441. doi: 10.1162/jocn_a_00425
- Geurts, D. E. M., Huys, Q. J. M., den Ouden, H. E. M., and Cools, R. (2013b). Serotonin and Aversive Pavlovian Control of Instrumental Behavior in Humans. *J. Neurosci.* 33, 18932–18939. doi: 10.1523/JNEUROSCI.2749-13.2013
- Geurts, D. E. M., von Borries, K., Volman, I., Bulten, B. H., Cools, R., and Verkes, R. J. (2016). Neural connectivity during reward expectation dissociates psychopathic criminals from non-criminal individuals with high impulsive/antisocial psychopathic traits. *Soc. Cogn. Affect. Neurosci.* 11, 1326–1334. doi: 10.1093/scan/nsw040
- Glenn, A. L., and Raine, A. (2009). Psychopathy and instrumental aggression: Evolutionary, neurobiological, and legal perspectives. *Int. J. Law Psychiatry* 32, 253–258. doi: 10.1016/j.ijlp.2009.04.002
- Gregory, S., Blair, R. J., Simmons, A., and Kumari, V. (2015). Punishment and psychopathy: A case-control functional MRI investigation of reinforcement learning in violent antisocial personality disordered men. *Lancet Psychiatry* 2, 153–160. doi: 10.1016/S2215-0366(14)00071-6
- Guitart-Masip, M., Duzel, E., Dolan, R., and Dayan, P. (2014). Action versus valence in decision making. *Trends Cogn. Sci.* 18, 194–202. doi: 10.1016/j.tics.2014.01.003
- Hallquist, M. N., Hall, N. T., Schreiber, A. M., and Dombrowski, A. Y. (2018). Interpersonal dysfunction in borderline personality: A decision neuroscience perspective. *Curr. Opin. Psychol.* 21, 94–104. doi: 10.1016/j.copsyc.2017.09.011
- Hare, R. D. (2003). *Manual for the Hare Psychopathy Checklist-Revised*, 2nd Edn. Toronto: Multi Health Systems.
- Heinz, A., Schlagenhauf, F., Beck, A., and Wackerhagen, C. (2016). Dimensional psychiatry: Mental disorders as dysfunctions of basic learning mechanisms. *J. Neural Transm. Suppl.* 123, 809–821. doi: 10.1007/s00702-016-1561-2
- Herz, D. M., Eickhoff, S. B., Løkkegaard, A., and Siebner, H. R. (2013). Functional neuroimaging of motor control in parkinson's disease: A meta-analysis. *Hum. Brain Mapp.* 35, 3227–3237. doi: 10.1002/hbm.22397
- Hoppenbrouwers, S. S., Bulten, B. H., and Brazil, I. A. (2016). Parsing fear: A reassessment of the evidence for fear deficits in psychopathy. *Psychol. Bull.* 142, 573–600. doi: 10.1037/bul0000040
- Huys, Q. J. M., Cools, R., Gölzer, M., Friedel, E., Heinz, A., Dolan, R. J., et al. (2011). Disentangling the Roles of Approach, Activation and Valence in Instrumental and Pavlovian Responding. *PLoS Comput. Biol.* 7:e1002028. doi: 10.1371/journal.pcbi.1002028.t002
- Huys, Q. J. M., Gölzer, M., Friedel, E., Heinz, A., Cools, R., Dayan, P., et al. (2016). The specificity of Pavlovian regulation is associated with recovery from depression. *Psychol. Med.* 46, 1027–1035. doi: 10.1017/S0033291715002597
- Kahneman, D., and Frederick, S. (2007). Frames and brains: Elicitation and control of response tendencies. *Trends Cogn. Sci.* 11, 45–46. doi: 10.1016/j.tics.2006.11.007
- Larson, C. L., Baskin-Sommers, A. R., Stout, D. M., Balderston, N. L., Curtin, J. J., Schultz, D. H., et al. (2013). The interplay of attention and emotion: Top-down attention modulates amygdala activation in psychopathy. *Cogn. Affect. Behav. Neurosci.* 13, 757–770. doi: 10.3758/s13415-013-0172-8
- Leistico, A.-M. R., Salekin, R. T., DeCoster, J., and Rogers, R. (2008). A Large-Scale Meta-Analysis Relating the Hare Measures of Psychopathy to Antisocial Conduct. *Law Hum. Behav.* 32, 28–45. doi: 10.1007/s10979-007-9096-6
- Ly, V., Cools, R., and Roelofs, K. (2014). Aversive disinhibition of behavior and striatal signaling in social avoidance. *Soc. Cogn. Affect. Neurosci.* 9, 1530–1536. doi: 10.1093/scan/nst145
- Ly, V., von Borries, A. K. L., Brazil, I. A., Bulten, B. H., Cools, R., and Roelofs, K. (2016). Reduced transfer of affective value to instrumental behavior in violent offenders. *J. Abnorm. Psychol.* 125, 657–663. doi: 10.1037/abn0000166
- Moul, C., Dobson-Stone, C., Brennan, J., Hawes, D., and Dadds, M. (2013). An Exploration of the Serotonin System in Antisocial Boys with High Levels of Callous-Unemotional Traits. *PLoS One* 8:e56619. doi: 10.1371/journal.pone.0056619

- Moul, C., Killcross, S., and Dadds, M. R. (2012). A model of differential amygdala activation in psychopathy. *Psychol. Rev.* 119, 789–806. doi: 10.1037/a0029342
- Newman, J. P., Patterson, C. M., Howland, E. W., and Nichols, S. L. (1990). Passive avoidance in psychopaths: The effects of reward. *Pers. Individual Differ.* 11, 1101–1114.
- Nord, C. L., Lawson, R. P., Huys, Q. J. M., Pilling, S., and Roiser, J. P. (2018). Depression is associated with enhanced aversive Pavlovian control over instrumental behaviour. *Sci. Rep.* 8:12582. doi: 10.1038/s41598-018-30828-5
- Patenaude, B., Smith, S. M., Kennedy, D. N., and Jenkinson, M. (2011). A Bayesian model of shape and appearance for subcortical brain segmentation. *NeuroImage* 56, 907–922. doi: 10.1016/j.neuroimage.2011.02.046
- Patrick, C. J. (2022). Psychopathy: Current Knowledge and Future Directions. *Annu. Rev. Clin. Psychol.* 18, 387–415.
- Plichta, M. M., and 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. doi: 10.1016/j.neubiorev.2013.07.012
- Poser, B. A., Versluis, M. J., Hoogduin, J. M., and Norris, D. G. (2006). BOLD contrast sensitivity enhancement and artifact reduction with multiecho EPI: Parallel-acquired inhomogeneity-desensitized fMRI. *Magn. Reson. Med.* 55, 1227–1235. doi: 10.1002/mrm.20900
- Provost, C., Liljeholm, M., Tyszka, J. M., and O'Doherty, J. P. (2012). Neural Correlates of Specific and General Pavlovian-to-Instrumental Transfer within Human Amygdalar Subregions: A High-Resolution fMRI Study. *J. Neurosci.* 32, 8383–8390. doi: 10.1523/JNEUROSCI.6237-11.2012
- Provost, J.-S., Hanganu, A., and Monchi, O. (2015). Neuroimaging studies of the striatum in cognition Part I: Healthy individuals. *Front. Syst. Neurosci.* 9:140. doi: 10.3389/fnsys.2015.00140
- Rothmund, Y., Ziegler, S., Hermann, C., Gruesser, S. M., Foell, J., Patrick, C. J., et al. (2012). Fear conditioning in psychopaths: Event-related potentials and peripheral measures. *Biol. Psychol.* 90, 50–59. doi: 10.1016/j.biopsycho.2012.02.011
- Schmidt, C. C., Timpert, D. C., Arend, I., Vossel, S., Fink, G. R., Henik, A., et al. (2020). Control of response interference: Caudate nucleus contributes to selective inhibition. *Sci. Rep.* 10:20977. doi: 10.1038/s41598-020-77744-1
- Schultz, D. H., Balderston, N. L., Baskin-Sommers, A. R., Larson, C. L., and Helmstetter, F. J. (2016). Psychopaths Show Enhanced Amygdala Activation during Fear Conditioning. *Front. Psychol.* 7:903. doi: 10.1037/a0016480
- Soderstrom, H., Blennow, K., Manhem, A., and Forsman, A. (2001). CSF studies in violent offenders I. 5-HIAA as a negative and HVA as a positive predictor of psychopathy. *J. Neural Transm.* 108, 869–878. doi: 10.1007/s007020170036
- Soderstrom, H., Blennow, K., Sjodin, A. K., and Forsman, A. (2003). New evidence for an association between the CSF HVA: 5-HIAA ratio and psychopathic traits. *J. Neurol. Neurosurg. Psychiatry* 74, 918–921. doi: 10.1136/jnnp.74.7.918
- Talmi, D., Seymour, B., Dayan, P., and Dolan, R. J. (2008). Human Pavlovian Instrumental Transfer. *J. Neurosci.* 28, 360–368. doi: 10.1523/JNEUROSCI.4028-07.2008
- Tanaka, S. C., Balleine, B. W., and O'Doherty, J. P. (2008). Calculating Consequences: Brain Systems That Encode the Causal Effects of Actions. *J. Neurosci.* 28, 6750–6755. doi: 10.1523/JNEUROSCI.1808-08.2008
- Tricomi, E. M., Delgado, M. R., and Fiez, J. A. (2004). Modulation of caudate activity by action contingency. *Neuron* 41, 281–292.
- Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., et al. (2002). Automated Anatomical Labeling of Activations in SPM Using a Macroscopic Anatomical Parcellation of the MNI MRI Single-Subject Brain. *NeuroImage* 15, 273–289. doi: 10.1006/nimg.2001.0978
- Valentin, V. V., Dickinson, A., and O'Doherty, J. P. (2007). Determining the Neural Substrates of Goal-Directed Learning in the Human Brain. *J. Neurosci.* 27, 4019–4026. doi: 10.1523/JNEUROSCI.0564-07.2007
- Veit, R., Flor, H., Erb, M., Hermann, C., Lotze, M., Grodd, W., et al. (2002). Brain circuits involved in emotional learning in antisocial behavior and social phobia in humans. *Neurosci. Lett.* 328, 233–236. doi: 10.1016/s0304-3940(02)00519-0
- von Borries, A. K. L., Brazil, I. A., Bulten, B. H., Buitelaar, J. K., Verkes, R. J., and de Bruijn, E. R. A. (2009). Neural correlates of error-related learning deficits in individuals with psychopathy. *Psychol. Med.* 40, 1559–1568. doi: 10.1017/S0033291709992017
- Warren, J. I., and Burnette, M. (2013). The Multifaceted Construct of Psychopathy: Association with APD, Clinical, and Criminal Characteristics among Male and Female Inmates. *Int. J. Forensic Ment. Health* 12, 265–273. doi: 10.1080/14999013.2013.857739
- Watanabe, M., and Munoz, D. P. (2010). Presetting Basal Ganglia for Volitional Actions. *J. Neurosci.* 30, 10144–10157. doi: 10.1523/JNEUROSCI.1738-10.2010
- Wunderlich, K., Rangel, A., and O'Doherty, J. P. (2009). Neural computations underlying action-based decision making in the human brain. *Proc. Natl. Acad. Sci. U S A.* 106, 17199–17204. doi: 10.1073/pnas.0901077106
- Yildirim, B. O., and Derksen, J. J. L. (2015). Mesocorticolimbic dopamine functioning in primary psychopathy: A source of within-group heterogeneity. *Psychiatry Res.* 229, 633–677. doi: 10.1016/j.psychres.2015.07.005