Hao Chen | Stephan Nebe | Negin Mojtahedzadeh | Soren Kuitunen-Paul | Maria Garbusow | Daniel J. Schad | Michael Armin Rapp | Quentin J. M. Huys | Andreas Heinz | Michael N. Smolka

Susceptibility to interference between Pavlovian and instrumental control is associated with early hazardous alcohol use

Suggested citation referring to the original publication:
Addiction biology 26 (2020) 4, Art. e12983 pp. 1 - 14
DOI https://doi.org/10.1111/adb.12983
ISSN 1355-6215, 1369-1600

Journal article | Version of record

Secondary publication archived on the Publication Server of the University of Potsdam:
Zweitveröffentlichungen der Universität Potsdam : Humanwissenschaftliche Reihe 855
ISSN: 1866-8364
https://nbn-resolving.org/urn:nbn:de:kobv:517-opus4-569609
DOI: https://doi.org/10.25932/publishup-56960

Received: 4 March 2020 | Revised: 31 August 2020 | Accepted: 9 October 2020

DOI: 10.1111/adb.12983

ORIGINAL ARTICLE

Addiction Biology SSA WILEY

Susceptibility to interference between Pavlovian and instrumental control is associated with early hazardous alcohol use

Hao Chen[1] | Stephan Nebe[1,2] | Negin Mojtahedzadeh[1] | Sören Kuitunen-Paul[3,4] | Maria Garbusow[5] | Daniel J. Schad[5,6] | Michael A. Rapp[6] | Quentin J.M. Huys[7] | Andreas Heinz[5] | Michael N. Smolka[1]

[1]Department of Psychiatry and Neuroimaging Center, Technische Universität Dresden, Dresden, Germany

[2]Zurich Center for Neuroeconomics, Department of Economics, University of Zurich, Zurich, Switzerland

[3]Institute of Clinical Psychology and Psychotherapy, Technische Universität Dresden, Dresden, Germany

[4]Research Group Stress & Addiction, Department of Child and Adolescent Psychiatry, Faculty of Medicine, Technische Universität Dresden, Dresden, Germany

[5]Department of Psychiatry and Psychotherapy, Charité – Universitätsmedizin Berlin, Campus Charité Mitte, Berlin, Germany

[6]Area of Excellence Cognitive Sciences, University of Potsdam, Potsdam, Germany

[7]Division of Psychiatry and Max Planck UCL Centre for Computational Psychiatry and Ageing Research, University College London, London, UK

Correspondence
Michael N. Smolka, Section of Systems Neuroscience, Department of Psychiatry and Psychotherapy, Technische Universität Dresden, Würzburger Str. 35, 01187 Dresden, Germany.
Email: michael.smolka@tu-dresden.de

Funding information
Deutsche Forschungsgemeinschaft, Grant/Award Numbers: 186318919, 402170461, 178833530; UZH Grants Office, Grant/Award Number: FK-19-020

Abstract

Pavlovian-to-instrumental transfer (PIT) tasks examine the influence of Pavlovian stimuli on ongoing instrumental behaviour. Previous studies reported associations between a strong PIT effect, high-risk drinking and alcohol use disorder. This study investigated whether susceptibility to interference between Pavlovian and instrumental control is linked to risky alcohol use in a community sample of 18-year-old male adults. Participants (N = 191) were instructed to 'collect good shells' and 'leave bad shells' during the presentation of appetitive (monetary reward), aversive (monetary loss) or neutral Pavlovian stimuli. We compared instrumental error rates (ER) and functional magnetic resonance imaging (fMRI) brain responses between the congruent and incongruent conditions, as well as among high-risk and low-risk drinking groups. On average, individuals showed a substantial PIT effect, that is, increased ER when Pavlovian cues and instrumental stimuli were in conflict compared with congruent trials. Neural PIT correlates were found in the ventral striatum and the dorsomedial and lateral prefrontal cortices (lPFC). Importantly, high-risk drinking was associated with a stronger behavioural PIT effect, a decreased lPFC response and an increased neural response in the ventral striatum on the trend level. Moreover, high-risk drinkers showed weaker connectivity from the ventral striatum to the lPFC during incongruent trials. Our study links interference during PIT to drinking behaviour in healthy, young adults. High-risk drinkers showed higher susceptibility to Pavlovian cues, especially when they conflicted with instrumental behaviour, indicating lower interference control abilities. Increased activity in the ventral striatum (bottom-up), decreased lPFC response (top-down), and their altered interplay may contribute to poor interference control in the high-risk drinkers.

KEYWORDS

high-risk drinking, interference control, Pavlovian-to-instrumental transfer

1 | INTRODUCTION

To behave efficiently in one's daily life and to adapt one's actions to a dynamic environment, a response selection system is frequently engaged. Critical control components involved when making such choices include Pavlovian and instrumental control. Through Pavlovian conditioning, inborn and hard-wired responses (e.g., approach or avoidance) to biologically potent (unconditioned) stimuli can be associated with neutral stimuli. Thereafter, such conditioned responses to Pavlovian cues are independent of their outcomes. Conversely, instrumental behaviour, more specifically, goal-directed instrumental behaviour, is controlled by the contingencies between actions and outcomes. Pavlovian cues can influence ongoing instrumental behaviour, even though the responses to the Pavlovian cues were acquired separately from the instrumental responses—this process is called Pavlovian-to-instrumental transfer (PIT). To elaborate, a food's enticing scent (Pavlovian) may encourage people to partake in eating behaviour (Instrumental), whereas an unpleasant scent may hinder eating behaviour. In a typical human PIT task,[1,2] participants need to perform learned instrumental responses (press a button for approach or avoidance) in the presence of previously and independently trained Pavlovian cues (appetitive or aversive).

Most previous human PIT studies investigated how Pavlovian cues influence instrumental approach behaviour. Accordingly, appetitive Pavlovian cues were found to promote instrumental approach responses compared with the neutral cues,[3–9] whereas aversive Pavlovian cues were found to reduce instrumental approach behaviour.[10,11] Additionally, some studies have examined PIT effects in the avoidance context by rewarding successful instrumental avoidance behaviour, in which aversive Pavlovian cues were shown to promote instrumental avoidance behaviours.[12–14]

Moreover, in an orthogonal experimental design with the appetitive–aversive Pavlovian axis and the approach–avoidance instrumental axis, instrumental behaviour was impaired by incongruent Pavlovian cues (instrumental approach behaviour by aversive Pavlovian cues or instrumental avoidance behaviour by appetitive Pavlovian cues) but was promoted by congruent Pavlovian cues.[10,15] Freeman, et al.[16] used a go–no-go/PIT task, which resembles a classical go–no-go task. In this task, participants learned to respond to one stimulus in the go trials while withholding their responses to another stimulus in no-go trials. The authors modified the proportion of no-go trials where appetitive Pavlovian cues were presented. It was then found that when the proportion of incongruent no-go trials out of all no-go trials was higher, the provocation of the appetitive cues on instrumental approach behaviour (go trials) in the subsequent trials was reduced. Additionally, in one EEG study, Cavanagh et al.[17] used another variant of a go–no-go task to investigate how Pavlovian biases influence instrumental learning during the conflict between both systems. It was found that midfrontal theta power, sensitive to conflict and the following adaptive control, was associated with the ability to overcome Pavlovian biases when they interfered with the instrumental behaviour. Taken together, these four studies imply that cognitive control is to be allocated to overcome the conflict between Pavlovian and instrumental control.

Linked to alcohol drinking behaviour, previous studies from our group have found associations between the stronger motivational effect of Pavlovian cues on instrumental behaviour and alcohol dependence,[18–20] as well as high-risk drinking during young adulthood.[21] In addition to the enhanced behavioural effect, the neural correlates of the motivational PIT effect in the nucleus accumbens[19,20] and the amygdala[21] were also associated with alcohol dependence and high-risk drinking during young adulthood, respectively. Notably, when whether the Pavlovian cue interferes with the instrumental behaviour was taken into account, alcohol-dependent patients committed more errors compared with healthy controls when Pavlovian stimuli and instrumental responses were in conflict, especially when participants needed to inhibit instrumental approach responses during the presence of appetitive Pavlovian cues[15]; this behavioural impairment was also stronger for future relapsers.[22] As of yet, whether this interference effect along with its neural correlates was associated with high-risk drinking during young adulthood is not clear.

We thus investigated interference control during a PIT task in a group of healthy, young men at age 18, who were drinking occasionally but did not fulfil the criteria for Diagnostic and Statistical Manual of Mental Disorders (DSM-IV) alcohol dependence. The rationale behind this is that social drinking behaviour is influenced by numerous environmental cues during social occasions, which reflects the PIT task in the experimental settings to some extent. A reduction in the ability to allocate cognitive resources in order to control the response to cues that look tempting but violate the long-term goals may contribute to hazardous drinking development. From this perspective, we assumed that the ability to allocate interference control when the Pavlovian cues conflict with the instrumental behaviour, along with its associated neural responses, could be potential (bio)markers of hazardous drinking behaviour during early adulthood. More specifically, on the behavioural level, it was hypothesised that error rates (ERs) would increase in the incongruent condition, that is, when Pavlovian cues and instrumental stimuli are incongruent, as compared with the congruent condition. Importantly, individuals with higher levels of risk in drinking should show more susceptibility to this effect, that is, show lower interference control.

On the neural level, previous literature has found neural correlates of motivational effects of Pavlovian cues in the amygdala,[11,23–25] the ventral striatum (VS),[11,23,25] and the dorsal striatum.[12,26] Accordingly, the VS and amygdala were expected to show responses during the PIT task. Importantly, referring to the meta-analysis of tasks that require different dimensions of inhibitory or interference control,[27] we also hypothesised that conflict between Pavlovian cues and required instrumental behaviour would elicit responses in cognitive control areas—the lateral prefrontal cortex (lPFC) and the dorsomedial prefrontal cortex (dmPFC). Further, low-risk drinkers were hypothesised to allocate more top-down interference control as compared with high-risk drinkers. If this were to be the case, we would expect the effective connectivity between the aforementioned brain regions

to be altered in the high-risk drinkers, which we would explore with dynamic causal models.

2 | MATERIALS AND METHODS

2.1 | Participants and general procedure

Invitation letters were first sent to 1937 males at age 18 who were randomly sampled from the local registration offices in Dresden and Berlin, Germany. At the baseline of the longitudinal study, only males were recruited because of the higher prevalence of risky drinking behaviour. After screening 445 respondents, those with the inclusion criteria of right-handedness, no history of major mental disorders including substance dependence (except for nicotine dependence), eligibility for magnetic resonance imaging (MRI) and having had at least two drinking occasions in the past 3 months were further invited. Of those who met the inclusion criteria, 201 participants completed the behavioural and MRI assessment. After excluding participants with incomplete behavioural data because of technical issues, 191 participants were included for the final analysis.

Participants went through the experimental procedure with two appointments. During the first appointment, participants finished the Munich Composite International Diagnostic Interview (M-CIDI[28,29]) according to the DSM-IV[30], along with cognitive ability assessment (details in Supporting Information S2). The risk status of our subjects was defined according to their binge drinking behaviour based on World Health Organization (WHO) standards[31]: as recommended, an average intake of more than 60 g of ethanol on a given drinking occasion was used as a cut-off for high-risk and low-risk drinkers. According to the self-reported alcohol intake per occasion during the last year reported in the M-CIDI, 97 participants were classified as low-risk drinkers and the other 94 as high-risk drinkers (drinking behaviours of the two groups shown in Table 1).

During the second appointment, approximately 9 days (SD = 16 days) later, participants performed the PIT task consisting of four phases. The Pavlovian phase and the PIT phase were done within the MRI scanner, whereas the instrumental phase and the forced-choice phase were conducted outside the scanner. As briefly mentioned above, participants were presented with images of various shells whose quality (good or bad) was randomly assigned. During the instrumental training, participants were asked to learn the quality of each shell through trial-and-error instrumental responses. When collecting or leaving the shells, the participants received probabilistic feedback that dictated whether their action resulted in a monetary gain or loss. To collect a shell, the participants were required to press the left mouse button five or more times. Each button press resulted in a visual cue (a small dot) moving closer to the image of the shell. To leave a shell, there was no action required. A shell was only considered 'collected' if the threshold of five button presses was reached or surpassed. During the Pavlovian conditioning, participants passively learned the association between five types of compound conditional stimuli (CSs, consisting of fractal-like images and pure tones) and positive (€1, €2), negative (€−1, €−2) or neutral (€0) unconditioned stimuli (USs). Following this, participants performed the instrumental task again (90 trials) with the fractal images of the CSs tiled in the background. This phase, referred to as the PIT phase, was performed under nominal extinction to avoid further learning. Additionally, there were 72 trials with pictures of alcoholic/water beverages presented in the background in combination with the two instrumental stimuli (details about the alcohol/water PIT trials shown in Supporting Information S1). In the last phase, participants were presented with two CSs within 2 s and were required to choose one. A more detailed PIT task description is provided in Figure 1

TABLE 1 Drinking behaviour of the sample

	High-risk drinkers			Low-risk drinkers		
	N	Min–max	Mean (SD)	*N*	Min–max	Mean (SD)
General description of the sample						
Age	94	18.1–18.9	18.4 (0.2)	97	18.1–18.8	18.4 (0.2)
Years of education	94	11–13.5	11.6 (0.6)	96	4–14.5	11.6 (1.1)
Drinking behaviour						
Age first drinking	94	10–16	14.1 (1.4)	97	9–18	14.4 (1.3)
Age first drunk	94	12–18	15.5 (1.1)	89	10–18	16.0 (1.1)
Alcohol consumption last year (g/day)	94	3.2–112.5	19.4 (16.8)	97	0.6–22.5	5.1 (4.6)
Alcohol consumption (g/occasion)	94	63–225	104.2 (40.4)	97	18–54	39.2 (11.5)
Age first bingeing	86	14–18	16.5 (0.8)	52	14–18	16.5 (0.9)
Frequency bingeing (lifetime)	86	1–150	26.1 (29.7)	97	0–100	5.3 (14.3)
Alcohol consumption per bingeing (g/occasion)	94	63–450	130.9 (52.5)	97	0–225	57.2 (59.5)
Generic drink score[a]	94	−4.5-19.2	3.0 (4.2)	97	−8.4-8.5	−2.8 (3.2)

[a]Detailed information about how the Generic Drink Score was computed and the statistical analysis regarding this variable are shown in Supporting Information S3.

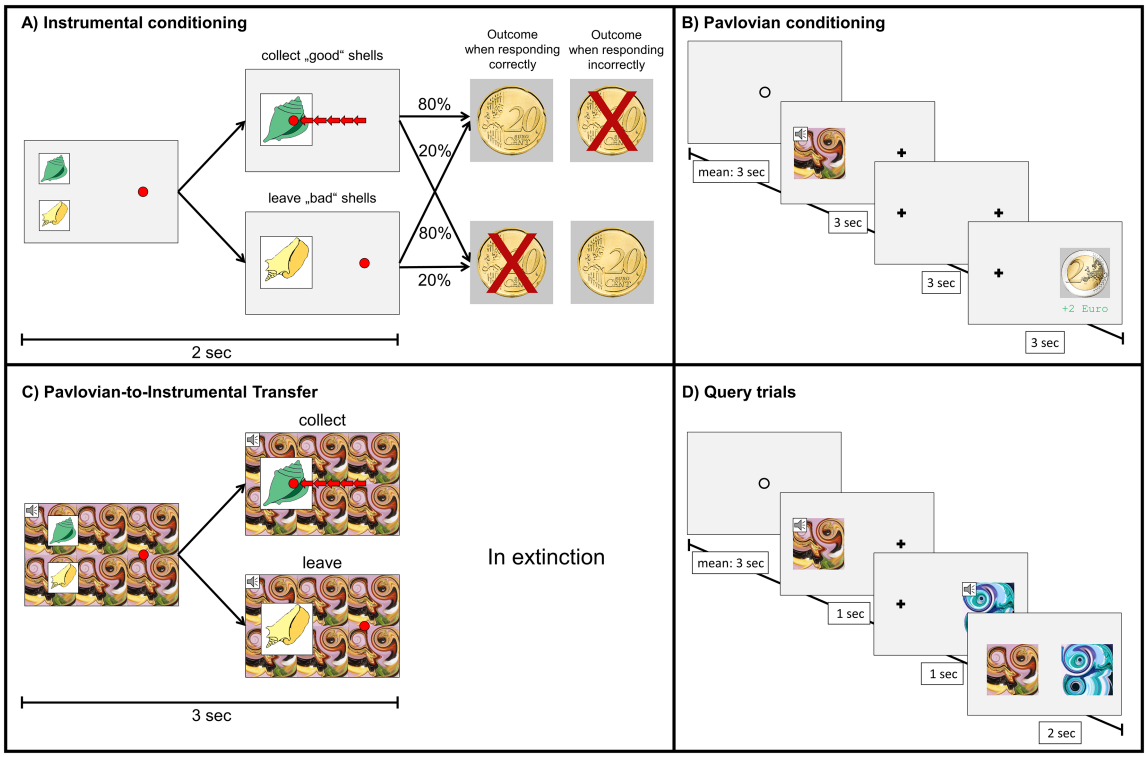


FIGURE 1 Pavlovian-to-instrumental transfer (PIT) experiment procedure (also see Garbusow et al.[18,19]). (A) Instrumental phase: participants learned to collect the good shells (press the button more than five times to move the dot towards the shell) and leave the bad shells (no action was required) according to the probabilistic feedback. After 60 trials, instrumental training ended once participants reached the learning criterion (80% correct choices over the last 16 consecutive trials) or at a maximum of 120 trials. (B) Pavlovian phase: participants passively learned the association between five types of compound conditional stimuli (CSs, consisting of fractal-like images and pure tones) and positive (€1, €2), negative (€−1, €−2) or neutral (€0) unconditioned stimuli (USs). There were 80 trials in total with 16 trials of each type. (C) PIT phase: Participants performed the instrumental task again with the tiled fractal images of the CSs in the background. Each trial lasted 3 s, with the fractal images shown 0.6 s before the instrumental shells. Therefore, participants had a response window of 2.4 s. There were 90 trials in total. This phase was done under nominal extinction to avoid further learning. Additionally, there were 72 trials with alcohol/water pictures presented in the background in combination with the two instrumental stimuli (details about the alcohol/water PIT trials shown in Supporting Information S6). (D) Query trials: in order to verify the acquisition of Pavlovian expectations, participants needed to make forced choices between two CSs within 2 s. Each possible pair of the CSs was presented three times in a randomised order

(also see Garbusow et al.[19]). Participants also rated their subjective experience with the five Pavlovian fractals after the experiment. The analyses for the subjective ratings and forced-choice query trials are presented in Supporting Information S6.

2.2 | Behavioural analysis

It is important to note that the same dataset was used in a previous study from within our group[21]; however, the analysis of the current study uses these data for a different purpose: to investigate the interference of Pavlovian cues on the ongoing instrumental behaviour. A detailed discussion about the difference between the analyses of the current study and Garbusow et al.[21] is provided in Supporting Information S7.

The analysis for this study was restricted to PIT trials that could either be categorised as 'congruent' or 'incongruent'. In the congruent condition, the Pavlovian background value and the instrumental stimulus were positively or negatively concordant, meaning the Pavlovian fractal images corresponding to the monetary gains of 1 or 2€ were paired with the 'good' shells. Additionally, the congruent condition consisted of trials in which the Pavlovian images corresponding to monetary losses of 1 or 2€ were paired with the 'bad' shells. For the incongruent condition, the opposite is true; this condition consisted of trials that were paired discordantly. To keep the analysis parsimonious, trials with neutral Pavlovian stimuli in the background were disregarded for the analysis. Moreover, trials with or alcoholic/water beverages in the background were also disregarded because it is not clear how healthy young adults would perceive the valence of these backgrounds. Thus, classifying these trials a priori as either congruent or incongruent would not have been viable.

The behavioural data were analysed with R 3.4.0 (R Core Team, Vienna, Austria). ER was used as a primary measurement of

task performance in the PIT phase. Correct responses were defined as at least five button presses in collect trials, or less than five button presses in the leave condition, regardless of the background stimuli.

To check whether our approach for PIT data analysis is suitable, we first compared the ER across the 14 conditions (7 Pavlovian cues × 2 instrumental behaviours), which confirmed that the main difference in ER arises from the incongruent versus congruent contrast (Figure S1). Within the incongruent condition, the ER showed a symmetric pattern: collecting a good shell with a negative Pavlovian background did not differ from leaving a bad shell with a positive background. This symmetric pattern held true when assessing the association between the ER and the drinking behaviour; a detailed description and the exploratory analyses of alcoholic/water beverage background trials are displayed in Supporting Information S1.

The interference PIT effect score was calculated by subtracting the ER in the congruent condition from the ER in the incongruent condition for each individual. To test whether the participants make more errors in the incongruent condition compared with the congruent condition, a one-tailed, one-sample *t* test was conducted on the interference PIT effect score. The one-tailed test was used on the basis of our a priori hypothesis that the ER is higher in the incongruent compared with the congruent condition.

The association between performance during the PIT task and the alcohol drinking behaviour was then tested, particularly binge drinking behaviour. Again, on the basis of our hypothesis, a one-tailed two-sample *t* test was performed accordingly to test whether the interference PIT effect was higher in the high-risk compared with the low-risk drinking group.

2.3 | fMRI data acquisition and analysis

2.3.1 | fMRI data acquisition and preprocessing

The imaging data (echo-planar imaging [EPI]) sequence and structural T1 weighted image were acquired using a Siemens 3-Tesla MRI scanner (Magnetom Trio, Siemens, Erlangen, Germany). Preprocessing of the fMRI data was performed with Nipype.[32] The 480 EPI images were slice time corrected, realigned to the first image of the sequence, coregistered to the individual segmented and normalised structural image and then smoothed with a Gaussian Kernel of full width at half maximum of 8 mm (see Supporting Information S4 for detailed information).

After the preprocessing, 139 subjects were included in the fMRI analysis. Among the 52 subjects who were excluded from the fMRI analysis, there were four participants with incidental findings, 22 participants with more than 3 mm volume-to-volume movement or 3° rotation and 26 participants without valid data for the first-level model as they did not press a button at least once for some stimuli, thus having an empty regressor in the first-level model preventing model estimation.

2.3.2 | fMRI data analysis

Statistical analyses of the fMRI data during the PIT phase were performed by the general linear model (GLM) in SPM12 (Wellcome Trust Centre for Neuroimaging, London, UK). On the first level, a model that consisted of 10 onset regressors of our main interest was used: five Pavlovian CS values (€−2, −1, 0, 1 and 2 monetary loss or reward) × two instrumental conditions (collect or leave). Additionally, four onset regressors for the alcohol/water trials (collect/leave × alcohol/water) were also included in the first level model. The onset of each registered button press was entered into a regressor of no interest. Finally, six nuisance (motion) regressors were also included in the model.

On the first level, the incongruent versus congruent contrast was defined as follows: the four types of incongruent trials were collapsed (CSs paired with €−1 or €−2 in the collect trials or CSs paired with €1 or €2 in the leave trials) together and then the four types of congruent trials were subtracted, thus mirroring the behavioural analysis. These individual contrast images were then entered into second-level SPM analysis (one-sample *t* test). To associate the neural incongruency effect (i.e., brain response to interference) with the behavioural performance at the group level, the individual behavioural interference PIT effect was included as a covariate in the second-level model. Additionally, a covariate of no interest was also included to specify the site information (the experiment was performed in either Berlin or Dresden). To test the hypotheses, brain responses in four regions of interest (ROIs) were analysed. The dmPFC, lPFC and VS masks were defined on the basis of the 12 mm spheres around the peaks from previous review papers.[27,33] The amygdala mask was defined anatomically (details in Figure 2). The mean individual parameter estimates were then extracted within the four ROIs from the first-level incongruent versus congruent contrast. To examine the neural incongruency effect on the group level, the mean parameter estimates from the four ROIs were tested in 4 one-sample *t* tests. Following this, the association between the brain response to interference and the behavioural interference PIT effect (ΔER) was tested with Pearson correlation tests for the four ROIs separately. These results were corrected for four comparisons with Bonferroni correction, with $p_{corr.} = 0.05$ ($p_{uncorr.} = 0.0125$) as the threshold.

These ROI analyses were followed by an exploratory whole-brain analysis of the incongruent versus congruent contrast and its association with the behavioural interference PIT effect (i.e., covariate effect on the second level) at an uncorrected threshold of $p < 0.001$, cluster size $k \geq 50$. Whether or not the association between behavioural and neural incongruency effect differs from risk status was also explored. The detailed description for this analysis is shown in Supporting Information S5.

To further explore how the effective connectivity modulated by the incongruent condition differs between the two groups, especially regarding the interplay between the VS and the dmPFC and lPFC areas, dynamic causal modelling (DCM) analyses were applied to the data.[34] The time series were extracted from the

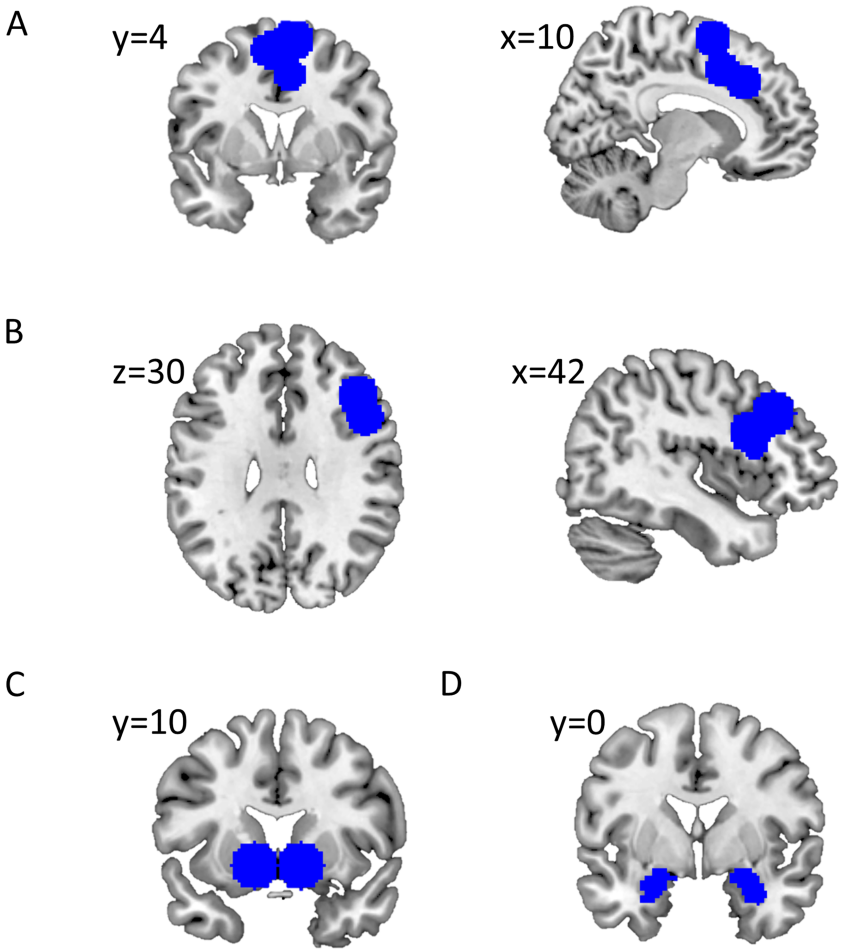


FIGURE 2 Regions of interest (ROI) masks. (A) dorsomedial prefrontal cortex (dmPFC) mask: generated according to the review paper by Hung et al.[27] In the cognitive inhibition meta-analysis, there were four peaks located in the dmPFC regions (Talairach coordinate: 6/14/40; 6/26/32; 8/8/58; −6/0/54). Four 12 mm spheres were generated around each peak, and the conjunctions of these spheres were used as the dmPFC masks. (B) lateral prefrontal cortex (lPFC) mask: conjunction of the three 12 mm spheres generated around the three peaks in the lPFC according to the same meta-analysis (Talairach coordinates: 42/26/30; 46/14/22; 52/16/14). (C) ventral striatum (VS) mask: defined on the basis of the peak of a previous meta-analysis on functional magnetic resonance imaging (fMRI) reward-related tasks.[33] The conjunction of the two 12 mm spheres around the peak Montreal Neurological Institute (MNI) coordinates: −12/10/−6 and 12/10/−6 were defined as the VS mask. (D) Amygdala mask: the bilateral amygdala mask was defined anatomically on the basis of the AAL atlas in the WFU PickAtlas toolbox[53]

peak voxels within the VS, lPFC and dmPFC that showed more activation during the conflict (i.e., incongruent–congruent contrast) because no regions were less activated during the conflict. Accordingly, for each individual, the time series of the three regions were extracted from 8 mm spheres centred on the individual local maxima, which were allowed to vary within 5 mm spheres around the group peak voxels during the conflict (incongruent-congruent contrast). The amygdala was excluded for this exploratory analysis, as there was no neural response in the amygdala within our contrast of interest; detailed information about this can be found in Section 3. In the model space, full intrinsic connections were assumed among the three regions, including self-connections. All PIT trials were used as driving inputs to enter VS, and the incongruent condition was used as modulatory input. Three possible modulatory effects were assumed on the connections between each pair of the three regions: forward, backward or bidirectional. With three possible connection structures between each pair, our model space consisted of 27 models in total (three possible structures × three pairs between the three regions; Figure 3).

Following this, Bayesian model selection (BMS) was conducted in combination with family-level inference.[35] The aim of the family-level inference, in this case, was to compare the models with different types of interplay between the VS and the two prefrontal areas during the incongruent condition. Six families (Figure 4) were defined accordingly: (1) incongruent condition only modulates the top-down connections; (2) incongruent condition only modulates the bottom-up connections; (3) incongruent condition modulates top-down connections between the VS and the dmPFC but bottom-up connections between the VS and the lPFC, or vice versa; (4) incongruent condition modulates both top-down and bottom-up connections only for the lPFC; (5) incongruent condition modulates both top-down and bottom-up connections only for the dmPFC; and (6) incongruent condition modulates both top-down and bottom-up connections for both the lPFC and the dmPFC. The BMS was done separately for the two groups. Given that fixed optimal model structures were not assumed among individuals, a random-effects analysis was used on the group level. This method takes into account the individual differences in model structures.[36] Following the BMS, Bayesian model averaging (BMA) was performed across the entire model space to further obtain parameter estimates of the effective connectivity. Finally, two-sample *t* tests were done to compare the connectivity between the two groups. The results were corrected for six comparisons with Bonferroni correction, with $p_{corr.} = 0.05$ as the threshold.

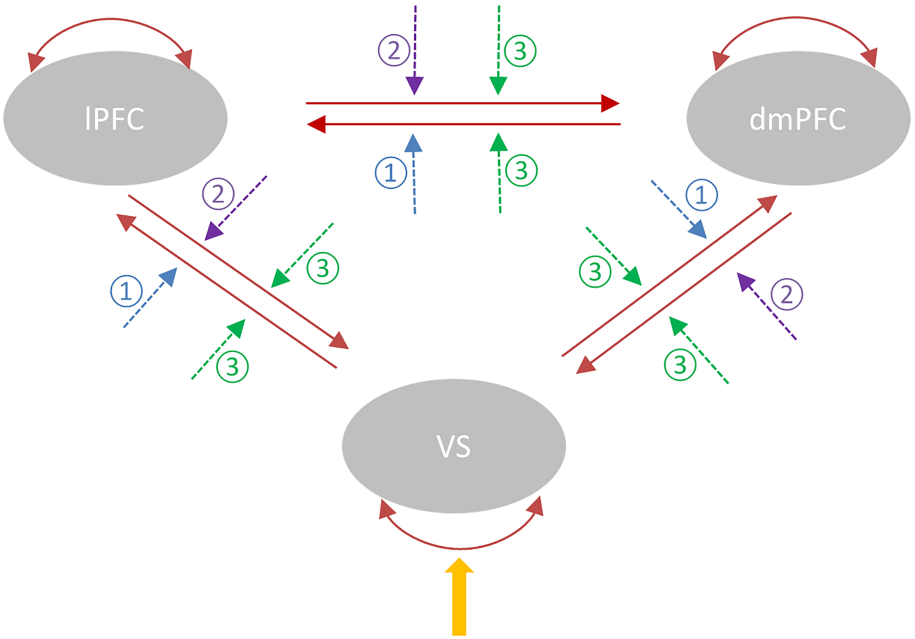
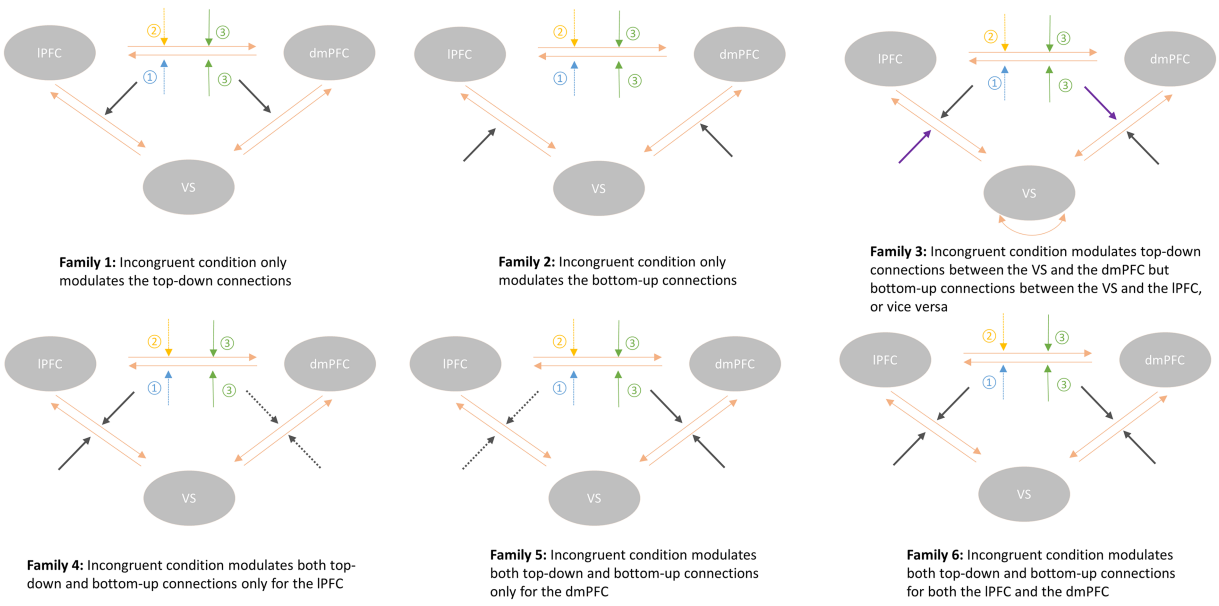


FIGURE 3 Dynamic causal modelling (DCM) model space. There were 27 models in the DCM model space. The driving input consisted of all Pavlovian-to-instrumental transfer (PIT) trials that entered the ventral striatum. The red arrows specify the intrinsic connectivity: all three regions were assumed to be intrinsically connected to each other and to themselves. The incongruent condition was assumed to modulate the connectivity between each pair of regions in three ways: forward, backward or bidirectional, which resulted in 27 modulatory structures in total



Family 1: Incongruent condition only modulates the top-down connections

Family 2: Incongruent condition only modulates the bottom-up connections

Family 3: Incongruent condition modulates top-down connections between the VS and the dmPFC but bottom-up connections between the VS and the lPFC, or vice versa

Family 4: Incongruent condition modulates both top-down and bottom-up connections only for the lPFC

Family 5: Incongruent condition modulates both top-down and bottom-up connections only for the dmPFC

Family 6: Incongruent condition modulates both top-down and bottom-up connections for both the lPFC and the dmPFC

FIGURE 4 Dynamic causal modelling (DCM) model families. The 27 DCM models were divided into six model families on the basis of the modulatory effect of the incongruent condition on the connectivity between the ventral striatum (VS) and the two prefrontal regions. Within each model family, there were three possible types of modulatory effects of the incongruent condition on the lateral prefrontal cortex (lPFC)–dorsomedial prefrontal cortex (dmPFC) connection: forward, backward and bidirectional

2.4 | Association between risk status and PIT effect

To further examine whether the PIT effects were associated with risk status, logistic regression was employed with risk status as the dependent variable. Possible predictors included the behavioural interference PIT effect and parameter estimates from the neural activated clusters in the incongruent condition (after adjusting for the behavioural interference PIT effect to avoid collinearity in predicting). In a stepwise backward selection process, the best combination of predictors was examined. Data-driven clusters were again used for this analysis because it was expected that these regions would reflect the neural responses within the PIT task more precisely compared with the ROIs.

3 | RESULTS

3.1 | Behavioural results

The ER was found to be, on average, approximately twice as high in the incongruent condition (30.8%) as compared with the congruent condition (15.6%, Figure 5A). This increase of ER was highly significant (T = 7.23; df = 190; $p = 5.47 \times 10^{-12}$; d = 0.52), indicating a substantial interference PIT effect in the whole sample. As hypothesised, the PIT effect was substantially stronger in the high-risk compared with the low-risk drinking group ($\Delta ER_{high\text{-}risk}$ = 21.3%, $\Delta ER_{low\text{-}risk}$ = 9.2%, T = 2.96; df = 189; $p = 1.74 \times 10^{-3}$; d = 0.43). The results are displayed in Figure 5B. t tests on working memory, processing speed and crystallised intelligence revealed no significant differences between the two groups (for details, see Supporting Information S2).

3.2 | fMRI results

3.2.1 | Neural incongruency effect—ROI analysis

In the ROI analyses, the four one-sample t tests of the parameter estimates within the four ROIs did not survive the correction for multiple comparisons, thus indicating no significant difference in the congruent condition compared with the incongruent condition on the group level.

3.2.2 | Neural correlates of the behavioural interference PIT effect—ROI analysis

When correlating the behavioural interference PIT effect and neural responses (incongruent–congruent condition) in the four ROIs, positive correlations were found between the behavioural interference PIT effect (ΔER) and the neural responses in the lPFC ($r(137)$ = 0.23; $p_{one\text{-}tailed;corr.}$ = 0.012) as well as in the dmPFC ($r(137)$ = 0.25; $p_{one\text{-}tailed;corr.}$ = 0.007). The correlation between neural responses in the VS and the behavioural interference PIT effect was also positive, but it did not survive the control for multiple comparisons ($r(137)$ = 0.16; $p_{one\text{-}tailed}$ = 0.080 without the Bonferroni correction). However, correlations were not seen between the behavioural interference PIT effect and responses in the amygdala ($r(137)$ = −0.02; $p_{one\text{-}tailed}$ = 0.790 without the Bonferroni correction).

3.2.3 | Neural incongruency effect—Whole-brain analysis

With respect to the explorative whole-brain analysis, the second-level t-contrast of the incongruent versus congruent PIT condition was first investigated; this included the individual behavioural interference PIT effect as a covariate. Increased brain responses during the incongruent compared with the congruent PIT trials (neural incongruency effect) were found in the ventral tegmental area (VTA; k = 50, t = 4.01, peak Montreal Neurological Institute [MNI] templates coordinates: −10/−16/−22) at a whole-brain uncorrected threshold of $p < 0.001$, cluster size $k \geq 50$ (Figure 6A). As an additional sanity check, at a lower threshold ($p < 0.01$, cluster size $k \geq 50$), the BOLD response of parietal top-down control regions (BA40, peak MNI coordinates: −34/−48/50, k = 265, t = 2.93) were also more pronounced during the incongruent condition. In contrast, no brain region showed higher activity during the congruent compared with the incongruent PIT trials at the same statistical threshold (whole-brain $p < 0.001$, cluster size $k \geq 50$).

3.2.4 | Neural correlates of the behavioural interference PIT effect—Whole-brain analysis

In the next step of the whole-brain analyses, whether or not the neural response to interference was associated with the behavioural interference PIT effect was investigated by conducting a one-sample t test on the behavioural interference PIT effect covariate. Neural correlates of the behavioural interference PIT effect were seen in the VS

FIGURE 5 Behavioural interference Pavlovian-to-instrumental transfer (PIT) effect. (A) Error rate (ER) increased on average by 15.2% in the incongruent condition compared with the congruent condition ($p = 5.47 \times 10^{-12}$). (B) High-risk drinkers (n = 94), in contrast to the low-risk drinkers (n = 97), reflected increased ER in the incongruent condition compared with congruent condition ($p = 1.74 \times 10^{-3}$). (C) Individual ER change in the incongruent compared with the congruent condition, separated between high- and low-risk drinkers

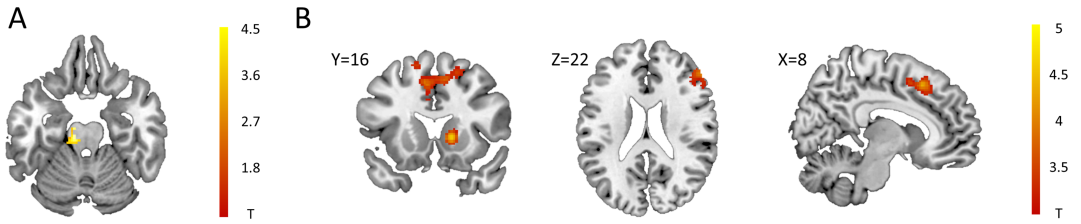


FIGURE 6 Neural incongruency effect and neural correlates of behavioural interference Pavlovian-to-instrumental transfer (PIT) effect ($p_{uncorrected}$ < 0.001, cluster size$k$ ≥ 50). (A) Interference (incongruent–congruent trials) elicited activation in ventral tegmental areas (VTA) (t = 4.01, k = 50, peak Montreal Neurological Institute (MNI) coordinates: −10/−16/−22). (B) A neural PIT effect (brain response to interference correlated with behavioural interference PIT effect) was found in the ventral striatum (VS) (t = 4.58, k = 168, peak MNI coordinates: 14/16/0), lateral prefrontal cortex (lPFC) (t = 3.97, k = 235, peak MNI coordinates: 50/38/22) and dorsomedial prefrontal cortex (dmPFC) (t = 4.35, k = 955, peak MNI coordinates: 8/20/48)

(k = 168, t = 4.58, peak MNI coordinate: 14/16/0), lPFC (k = 235, t = 3.97, peak MNI coordinate: 50/38/22) and dmPFC (k = 955, t = 4.35, peak MNI coordinate: 8/20/48) at a whole-brain uncorrected threshold of p < 0.001, k ≥ 50 (Figure 6B; detailed results displayed in Table 2). To illustrate the brain correlates of the behavioural interference PIT effect (ΔER), the neural activation within the three activated clusters was plotted in response to incongruent over congruent trials (neural incongruency effect) against the behavioural interference PIT effect (Figure 7). As can be seen, the neural response to incongruency in the VS, lPFC and dmPFC was higher in subjects with a stronger behavioural interference PIT effect. However, not all the individuals showed responses to incongruency—this effect was driven by around half of the individuals who committed more errors in the incongruent condition as compared with the congruent condition. The association between the behavioural interference PIT effect and the neural incongruency effect was stronger for low-risk drinkers compared with high-risk drinkers in the VS and the lPFC, but the difference was marginal in the dmPFC (detailed result in Figures S3 and S4).

3.2.5 | Effective connectivity difference between high- and low-risk drinkers

The model selection was first performed in order to select an optimal family of models among the six families in Figure 4. The selection was performed separately for the high- and low-risk drinking groups to test whether the winning family of models was different for the two groups. The selection was based on the exceedance probability: a higher exceedance probability suggests one family of models has more evidence compared with other specified families of models. According to the family exceedance probability, the winning family for the high-risk drinking group was the family in which the incongruent condition only modulated the bottom-up but not the top-down connections between the VS, lPFC or the dmPFC. The winning family had an exceedance probability of 0.32 (compared with the second-best family with an exceedance probability of 0.17). In contrast, for the low-risk drinkers, the model family in which the incongruent condition fully modulated all the connections between the VS and both the lPFC

TABLE 2 fMRI results table

Whole-brain results ($p_{uncorrected.}$ < 0.001, cluster size ≥ 50)						
		Peak MNI				
Region	Side	x	y	z	Peak-level t score	Cluster size
Neural incongruency effect (incongruent–congruent)						
Brain-stem (midbrain)	L	−10	−22	−22	4.19	50
Inferior temporal gyrus	R	58	−42	−16	3.76	157
Neural correlates of the behavioural interference PIT effect						
Right ventral striatum (extended to caudate)	R	14	16	0	4.58	168
SMA (BA32, extended to BA8 and BA6)	R	8	20	48	4.35	955
Middle frontal gyrus (SMA; BA6)	L	−28	2	58	4.03	226
Middle frontal gyrus (DLPFC/VLPFC; BA 45)	R	50	38	22	3.97	235
Middle frontal gyrus (IFG; BA 44)	L	−36	22	34	3.84	69

Abbreviations: DLPFC, dorsal lateral prefrontal cortex; fMRI, functional magnetic resonance imaging; IFG, inferior frontal gyrus; MNI, Montreal Neurological Institute; PIT, Pavlovian-to-instrumental transfer; SMA, supplementary motor area; VLPFC, ventral lateral prefrontal cortex.

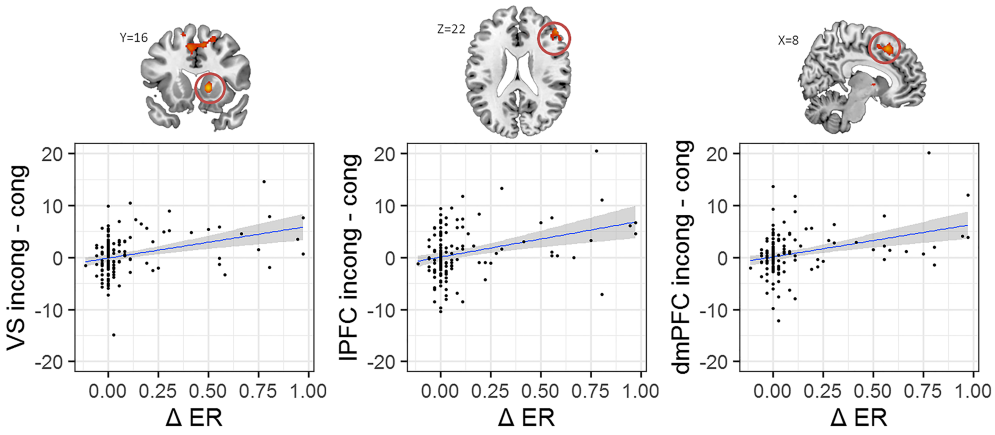


FIGURE 7 Neural correlates of behavioural interference PIT effect. Illustration of the positive association between neural activation in the ventral striatum (VS), lateral prefrontal cortex (lPFC), dorsomedial prefrontal cortex (dmPFC) and the behavioural interference Pavlovian-to-instrumental transfer (PIT) effect

and the dmPFC had the highest exceedance probability of 0.38 (the second-best family had an exceedance probability of 0.19). Generally speaking, with around twice the exceedance probability of the winning family compared with the second-best family, it was concluded that there was only weak support for the two different winning families for the two groups (plotted in Figure 8). Because of the different winning families, the strength of the connectivity was further obtained through BMA across the entire model space for both groups; this ensured the parameter estimates were comparable. The BMA does not make inferences about the model structure, but it rather computes a weighted average of the effective connectivity parameters from all the specified models. The weights are given by the posterior probabilities of different models.[36] On the basis of the BMA results, one can directly compare whether the effective connectivity parameters between certain brain regions are different for the two groups. According to the criteria that the posterior mean is larger than zero at a probability threshold of 95%, the incongruent condition significantly modulated the connection from the VS to the lPFC and the bidirectional connection between the lPFC and the dmPFC for the low-risk but not the high-risk drinkers (Table 3). By comparing the modulatory parameters between the two groups, significantly higher effective connectivity was found from the VS to the lPFC modulated by the incongruent condition in the low-risk compared with the high-risk drinking group (p = 0.004 after Bonferroni correction for six comparisons) (Table 3).

3.3 | Association between risk status and PIT effects

In the backward stepwise logistic regression with risk status as the dependent variable, the best model ($\chi^2(3, N = 139) = 8.966$, $p = 0.030$) included three of the four predictors: the behavioural interference PIT effect ($\beta = 2.073$; $p = 0.014$), the neural activation in the incongruent condition in the VS ($\beta = 0.298$; $p = 0.091$) and the lPFC ($\beta = -0.391$; $p = 0.042$), but not in the dmPFC. The logistic regression thus indicated a positive association between risk status and behavioural

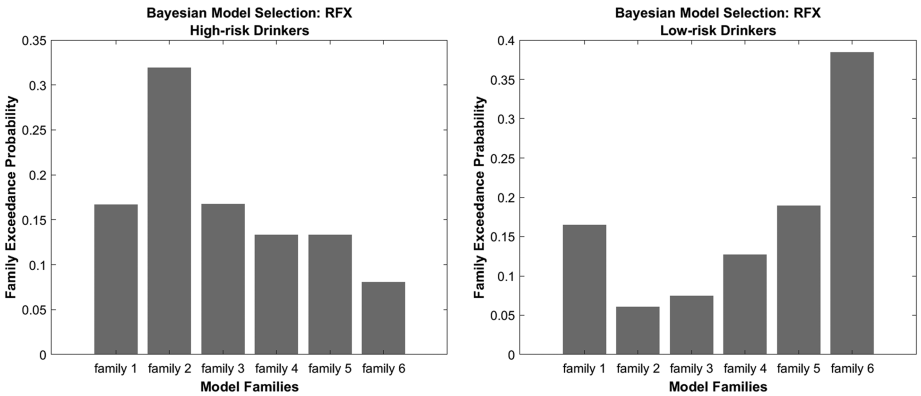


FIGURE 8 Bayesian model selection (random-effects analysis; RFX) results for the high-risk and low-risk drinkers. According to the family exceedance probability, the winning family for the high-risk drinking group was the family where incongruent condition only modulates the bottom-up but not the top-down connections between the ventral striatum (VS) and the lateral prefrontal cortex (lPFC) as well as the dorsomedial prefrontal cortex (dmPFC) (Family 2). In contrast, for the low-risk drinkers, the model family where the incongruent condition fully modulates all the connections between the VS and both the lPFC and dmPFC had the highest exceedance probability (Family 6)

TABLE 3 DCM results

Modulatory effects of the incongruent condition				
			Two-sample *t* test	
	Low-risk drinkers	High-risk drinkers	*t* value	*p* value
VS → lPFC	0.056 (0.099)**	−0.002 (0.095)	3.52	0.001 **
VS → dmPFC	0.021 (0.097)	0.017 (0.093)	0.22	0.829
lPFC → VS	0.001 (0.098)	0.004 (0.097)	−0.22	0.828
lPFC → dmPFC	0.049 (0.100)**	0.013 (0.097)	2.13	0.035 *
dmPFC → VS	−0.010 (0.098)	0.006 (0.097)	−1.00	0.317
dmPFC → lPFC	0.045 (0.099)**	0.020 (0.097)	1.52	0.132
Driving input from all PIT trials				
→VS	0.011 (0.008)**	0.005 (0.009)	3.98	0.001 **

Abbreviations: DCM, dynamic causal modelling; dmPFC, dorsomedial prefrontal cortex; lPFC, lateral prefrontal cortex; PIT, Pavlovian-to-instrumental transfer; VS, ventral striatum.
*Significant at uncorrected threshold $p < 0.05$.**Survives Bonferroni correction for multiple comparisons (six comparisons).

interference PIT effect and the VS (trend-wise), whereas the risk status was negatively associated with the neural responses in the lPFC.

4 | DISCUSSION

In this study, we investigated whether interference between Pavlovian and instrumental control, assessed with a PIT task, is associated with risky alcohol use in a cohort of healthy males aged 18 years. As expected, participants committed substantially more errors in the incongruent compared with the congruent condition, which suggests that interference by incongruent Pavlovian cues impairs instrumental performance. Importantly, the instrumental performance was substantially more impaired by Pavlovian interference in high-risk compared with low-risk drinkers, indicating better interference control abilities in the latter. At the neural level, participants with a stronger behavioural instrumental impairment showed higher activation in the VS, the dmPFC and the lPFC during incongruent PIT trials. Furthermore, the neural responses also differed with risk status: high-risk drinkers showed a blunted top-down control response of the lPFC, as well as reduced effective connectivity from the VS to the lPFC during the incongruent (i.e., conflict) condition. Taken together, these findings indicate that individuals who can allocate top-down control to overcome conflict, that is, interference between Pavlovian and instrumental cues, are less likely to show risky alcohol consumption.

At the behavioural level, the effect of interference was very pronounced; however, at the neural level, interference was not detected in the a priori ROIs. The subsequent explorative whole-brain analysis revealed that incongruence was reflected by stronger activation in the VTA and parietal areas, but these activations would not have survived correction for multiple comparisons. Thus, for the entire sample of young males, the neural effect of interference between Pavlovian and instrumental control was rather modest. Regarding brain regions, this finding is in line with previous animal studies, which showed that inactivation of the VTA reduced the PIT effect.[37,38] Additionally, activation of the parietal areas, which has been suggested to be part of the inhibitory brain network,[27] may indicate the conflict participants experienced in the incongruent condition. The modest effect on the group level might be due to the fact that only about half of the sample showed impaired performance during interference between instrumental and Pavlovian control.

In contrast, when the interindividual differences in interference were considered, it was found that the VS, lPFC and dmPFC activation correlated positively with the behavioural interference PIT effect. Previous literature repeatedly reported the VS to reflect the influence of the Pavlovian cue on instrumental behaviour.[11,23,25] The VS cluster that was found also extended to the dorsal striatum; this has also been shown by two previous studies.[12,26] In contrast to previous studies, we did not find amygdala activation.[11,21,23–25] As suggested by these studies, the amygdala may compute the affective valence of Pavlovian cues in the PIT task. Notably, one difference between the previously mentioned studies and the current study involves the valence signal. In the aforementioned PIT studies, when comparing the positive/negative Pavlovian cue condition with the neutral condition, the finding reflected a mixture of salience and valence signal. Conversely, in the current analysis, the valence signal was averaged out when pooling the different combinations of Pavlovian cues and instrumental stimuli into incongruent and congruent conditions. This may begin to explain why activation in the amygdala was not found. Taken together, the signal seen in the VS may reflect a salience signal indicating that the Pavlovian cue is at odds with the required instrumental behaviour.

The response elicited by incongruent trials was also found in the dmPFC. This region has been extensively linked to conflict-related performance monitoring, in which it plays an important role in deciding the subsequent adjustments in performance.[39,40] Additionally, incongruent trials also evoked a response of the lPFC, which is a critical structure that gathers task-related information and exhibits top-down cognitive control[41,42] in relation to conflict monitoring, error monitoring and response selection.[43] To summarise, the activation

found in the VS, lPFC and dmPFC is part of a corticostriatal circuit that is critical for response selection and cognitive control through the extensive communication between the subcortical and cortical parts[44,45]—which makes it essential for overcoming interference during incongruent task trials.

Compared with low-risk drinkers, the high-risk drinkers showed a stronger association between the behavioural and the neural PIT effect. This effect may be related to the findings from the DCM analysis, which suggested that the incongruent stimuli tended not to modulate the effective connectivity from the dmPFC and lPFC to the VS for the high-risk drinkers. Parameter estimates further indicated that the effective connectivity from the VS to the lPFC was higher in response to the incongruent stimuli in the low-risk compared with the high-risk drinking group. It is also worth mentioning that the VS mask for the DCM analysis was generated around the peak activation from the analysis—this mask also partly consisted of the dorsal striatum. Therefore, the interplay between the VS and the lPFC may have also involved the dorsal striatum to some extent. Taken together, the neural response in this network may explain why low-risk drinkers showed better interference control (i.e., were less susceptible to response conflicts induced by incongruent stimuli) when the Pavlovian cue conflicts with the instrumental behaviour. It is plausible that the VS of low-risk drinkers sends a salience signal that helps allocate cognitive top-down control to resolve the response conflict.

It is worth noting that a previous paper from our group found that the association between the valence of the Pavlovian cues and response rates (indicating response vigour) was stronger for high-risk than low-risk drinkers.[21] However, in this study, the main focus was to investigate the motivational effect of Pavlovian cues on the ongoing instrumental behaviours, regardless of whether they promote (congruent condition) or hinder (incongruent condition) the required instrumental response. Despite using the same dataset, the main focus of the current study was to examine the interference effect of Pavlovian cues when they are in conflict with the necessary instrumental behaviour. By doing this, the motivational and cognitive control perspectives were able to be examined simultaneously, as both perspectives were present during trials with interference from Pavlovian cues. Therefore, these results connect previous research in the fields of cognitive control and motivated behaviour. Even though the interplay of cognitive control and motivated behaviour is essential to understand addictive behaviour, most experimental approaches either focus on one or the other. An exception would be the go–no-go/PIT task,[16,46] which assesses the influence of non-drug Pavlovian cues on response inhibition. So far, go–no-go/PIT tasks have not been used to study substance use or dependence. These results, therefore, complement previous studies that reported an association between binge drinking and impaired interference control in young adults.[47]

Importantly, the conflict between Pavlovian and instrumental control substantially differs from conflict seen in traditional interference tasks such as the classical colour-word Stroop task (conflict at stimulus level)[48,49] or the Simon task (conflict at response level).[50,51] In these 'cold' interference tasks, responses are instructed and are not the result of learning based on rewards or punishments. Interference in these tasks mainly results from automated response tendencies (i.e., neither the colour representation in the Stroop task nor the location cue representation in the Simon task triggers motivational responses). In contrast, in our 'hot' interference task, Pavlovian cues trigger a motivational response, that is, approach or avoidance behaviour and interfere with motivated instrumental behaviour. On the basis of the hypothesis about the difference between the 'cold' and 'hot' interference task, future studies could investigate whether the PIT effect we found could (to some extent) be explained by these 'cold' interference tasks or it involves fundamentally different mechanisms.

To conclude, the results of the current study show that the susceptibility to Pavlovian interference during a PIT task is linked to hazardous drinking behaviours at age 18. Although the imbalance between the top-down and bottom-up systems has been suggested to be associated with addictive behaviour, previous studies have tended to consider either the perspective of cognitive control or motivated behaviour but not both at the same time. Using a PIT task, we assessed the top-down control and its interaction with bottom-up Pavlovian and instrumental processes. Our experimental data indicate that a poor interplay between top-down and bottom-up processes may contribute to early hazardous alcohol use.

5 | LIMITATIONS

We investigated a sample of 18-year-old social drinkers. In this sample, some participants did not commit any errors during the PIT task. It is thus unclear whether these participants experienced no interference at all or they had better interference control. Another explanation could be that the PIT task was not sensitive enough to capture the very subtle effects that may have been present in these participants. Therefore, a possible solution to this issue could be found in further refinement of the PIT task to increase the sensitivity to more subtle effects. Additionally, the classification of high- and low-risk drinkers based on the self-reported alcohol consumption data during the past year may not be entirely accurate because of the possible memory bias; future studies may improve this by using more frequently assessed electronic diary data. Another limitation of the current study is that these results cannot be generalised to nonmale populations.

ACKNOWLEDGEMENTS

This study was supported by the Deutsche Forschungsgemeinschaft (DFG; grants for FOR 1617 [Project number 186318919], TRR 265 [Project number 402170461][52] and SFB 940 [Project number 178833530]). S.N. received funding from the UZH Grants Office (FK-19-020). MR-imaging for this study was, in part, performed at the Berlin Center for Advanced Neuroimaging (BCAN). We thank Matthew Belanger for his proofreading. Open access funding enabled and organized by Projekt DEAL.

AUTHORS CONTRIBUTION

MNS, AH, QJMH and MAR were responsible for the study concept and design. SN, MG and SKP contributed to data acquisition. HC, SN and MNS analysed the data. HC, MNS and SN interpreted the findings. HC drafted the manuscript. MNS, SN, QJMH, SKP, MG, AH, DJS and NM provided critical revision of the manuscript for important intellectual content. All authors critically reviewed content and approved the final version for publication.

ORCID

Hao Chen https://orcid.org/0000-0002-0026-3623
Stephan Nebe https://orcid.org/0000-0003-3968-9557
Sören Kuitunen-Paul https://orcid.org/0000-0001-8224-6490
Maria Garbusow https://orcid.org/0000-0001-7247-6700
Daniel J. Schad https://orcid.org/0000-0003-2586-6823
Michael A. Rapp https://orcid.org/0000-0003-0106-966X
Quentin J.M. Huys https://orcid.org/0000-0002-8999-574X
Andreas Heinz https://orcid.org/0000-0001-5405-9065
Michael N. Smolka https://orcid.org/0000-0001-5398-5569

REFERENCES

1. Cartoni E, Balleine B, Baldassarre G. Appetitive Pavlovian-instrumental transfer: a review. *Neurosci Biobehav Rev*. 2016;71:829-848.
2. Holmes NM, Marchand AR, Coutureau E. Pavlovian to instrumental transfer: a neurobehavioural perspective. *Neurosci Biobehav Rev*. 2010;34(8):1277-1295.
3. Allman MJ, DeLeon IG, Cataldo MF, Holland PC, Johnson AW. Learning processes affecting human decision making: an assessment of reinforcer-selective Pavlovian-to-instrumental transfer following reinforcer devaluation. *J Exp Psychol Anim Behav Process*. 2010;36(3):402-408.
4. Eder AB, Dignath D. Asymmetrical effects of posttraining outcome revaluation on outcome-selective Pavlovian-to-instrumental transfer of control in human adults. *Learn Motiv*. 2016;54:12-21.
5. Paredes-Olay C, Abad MJ, Gamez M, Rosas JM. Transfer of control between causal predictive judgments and instrumental responding. *Anim Learn Behav*. 2002;30(3):239-248.
6. Quail SL, Morris RW, Balleine BW. Stress associated changes in Pavlovian-instrumental transfer in humans. *Q J Exp Psychol (Hove)*. 2017;70(4):675-685.
7. Rosas JM, Paredes-Olay MC, García-Gutiérrez A, Espinosa JJ, Abad MJ. Outcome-specific transfer between predictive and instrumental learning is unaffected by extinction but reversed by counter-conditioning in human participants. *Learn Motiv*. 2010;41:48-66.
8. Watson P, Wiers RW, Hommel B, de Wit S. Working for food you don't desire. Cues interfere with goal-directed food-seeking. *Appetite*. 2014;79:139-148.
9. Eder AB, Dignath D. Cue-elicited food seeking is eliminated with aversive outcomes following outcome devaluation. *Q J Exp Psychol (Hove)*. 2016;69(3):574-588.
10. Huys QJM, Cools R, Golzer M, et al. Disentangling the roles of approach, activation and valence in instrumental and Pavlovian responding. *PLoS Comput Biol*. 2011;7(4):e1002028.
11. Geurts DE, Huys QJ, den Ouden HE, Cools R. Aversive Pavlovian control of instrumental behavior in humans. *J Cogn Neurosci*. 2013;25(9):1428-1441.
12. Lewis AH, Niznikiewicz MA, Delamater AR, Delgado MR. Avoidance-based human Pavlovian-to-instrumental transfer. *Eur J Neurosci*. 2013;38(12):3740-3748.
13. Nadler N, Delgado MR, Delamater AR. Pavlovian to instrumental transfer of control in a human learning task. *Emotion*. 2011;11(5):1112-1123.
14. Garofalo S, Robbins TW. Triggering Avoidance: Dissociable Influences of Aversive Pavlovian Conditioned Stimuli on Human Instrumental Behavior. *Front Behav Neurosci*. 2017;11:63.
15. Sommer C, Garbusow M, Junger E, et al. Strong seduction: impulsivity and the impact of contextual cues on instrumental behavior in alcohol dependence. *Transl Psychiatry*. 2017;7(8):e1183.
16. Freeman SM, Alvernaz D, Tonnesen A, Linderman D, Aron AR. Suppressing a motivationally-triggered action tendency engages a response control mechanism that prevents future provocation. *Neuropsychologia*. 2015;68:218-231.
17. Cavanagh JF, Eisenberg I, Guitart-Masip M, Huys Q, Frank MJ. Frontal theta overrides Pavlovian learning biases. *J Neurosci*. 2013;33(19):8541-8548.
18. Garbusow M, Schad DJ, Sebold M, et al. Pavlovian-to-instrumental transfer effects in the nucleus accumbens relate to relapse in alcohol dependence. *Addict Biol*. 2016;21(3):719-731.
19. Garbusow M, Schad DJ, Sommer C, et al. Pavlovian-to-instrumental transfer in alcohol dependence: a pilot study. *Neuropsychobiology*. 2014;70(2):111-121.
20. Schad DJ, Garbusow M, Friedel E. Neural correlates of instrumental responding in the context of alcohol-related cues index disorder severity and relapse risk. *Eur Arch Psychiatry Clin Neurosci*. 2019;269(3):295-308.
21. Garbusow M, Nebe S, Sommer C, et al. Pavlovian-to-instrumental transfer and alcohol consumption in young male social drinkers: behavioral, neural and polygenic correlates. *J Clin Med*. 2019;8(8):1188.
22. Sommer C, Birkenstock J, Garbusow M, et al. Dysfunctional approach behavior triggered by alcohol-unrelated Pavlovian cues predicts long-term relapse in alcohol dependence. *Addict Biol*. 2018;25(1):e12703.
23. Mendelsohn A, Pine A, Schiller D. Between thoughts and actions: motivationally salient cues invigorate mental action in the human brain. *Neuron*. 2014;81(1):207-217.
24. Prevost C, Liljeholm M, Tyszka JM, O'Doherty JP. Neural correlates of specific and general Pavlovian-to-instrumental transfer within human amygdalar subregions: a high-resolution fMRI study. *J Neurosci*. 2012;32(24):8383-8390.
25. Talmi D, Seymour B, Dayan P, Dolan RJ. Human pavlovian-instrumental transfer. *J Neurosci*. 2008;28(2):360-368.
26. Bray S, Rangel A, Shimojo S, Balleine B, O'Doherty JP. The neural mechanisms underlying the influence of Pavlovian cues on human decision making. *J Neurosci*. 2008;28(22):5861-5866.
27. Hung Y, Gaillard SL, Yarmak P, Arsalidou M. Dissociations of cognitive inhibition, response inhibition, and emotional interference: Voxelwise ALE meta-analyses of fMRI studies. *Hum Brain Mapp*. 2018;39(10):4065-4082.
28. Jacobi F, Mack S, Gerschler A, et al. The design and methods of the mental health module in the German Health Interview and Examination Survey for Adults (DEGS1-MH). *Int J Methods Psychiatr Res*. 2013;22(2):83-99.
29. Wittchen H-U, Pfister H. *DIA-X-Interviews: Manual für Screening-Verfahren und Interview; Interviewheft*. Frankfurt: Swets Test Services; 1997.
30. Saß H, Wittchen H-U, Zaudig M, Houben I. *DSM-IV-TR—Diagnostisches und Statistisches Manual Psychischer Störungen–Textrevision*. Göttingen: Hogrefe; 2003.
31. Stockwell T, Chikritzhs T, Dawson D. *International guide for monitoring alcohol consumption and related harm*. Geneva, Switzerland: World Health Organization; 2000.
32. Gorgolewski K, Burns CD, Madison C, et al. Nipype: a flexible, lightweight and extensible neuroimaging data processing framework in python. *Front Neuroinform*. 2011;5:13.

33. Liu X, Hairston J, Schrier M, Fan J. Common and distinct networks underlying reward valence and processing stages: a meta-analysis of functional neuroimaging studies. *Neurosci Biobehav Rev*. 2011;35(5):1219-1236.
34. Friston KJ, Harrison L, Penny W. Dynamic causal modelling. *Neuroimage*. 2003;19(4):1273-1302.
35. Penny WD, Stephan KE, Daunizeau J. Comparing Families of Dynamic Causal Models. *PLoS Comput Biol*. 2010;6(3):e1000709.
36. Stephan KE, Penny WD, Moran RJ, den Ouden HE, Daunizeau J, Friston KJ. Ten simple rules for dynamic causal modeling. *Neuroimage*. 2010;49(4):3099-3109.
37. Corbit LH, Janak PH, Balleine BW. General and outcome-specific forms of Pavlovian-instrumental transfer: the effect of shifts in motivational state and inactivation of the ventral tegmental area. *Eur J Neurosci*. 2007;26(11):3141-3149.
38. Murschall A, Hauber W. Inactivation of the ventral tegmental area abolished the general excitatory influence of Pavlovian cues on instrumental performance. *Learn Mem*. 2006;13(2):123-126.
39. Domenech P, Koechlin E. Executive control and decision-making in the prefrontal cortex. *Curr Opin Behav Sci*. 2015;1:101-106.
40. Ridderinkhof KR, Ullsperger M, Crone EA, Nieuwenhuis S. The role of the medial frontal cortex in cognitive control. *Science*. 2004;306(5695):443-447.
41. Egner T, Hirsch J. Cognitive control mechanisms resolve conflict through cortical amplification of task-relevant information. *Nat Neurosci*. 2005;8(12):1784-1790.
42. Kouneiher F, Charron S, Koechlin E. Motivation and cognitive control in the human prefrontal cortex. *Nat Neurosci*. 2009;12(7):939-945.
43. Amodio DM, Frith CD. Meeting of minds: the medial frontal cortex and social cognition. *Nat Rev Neurosci*. 2006;7(4):268-277.
44. Haber SN. Corticostriatal circuitry. Neuroscience in the 21st Century 2016:1–21.
45. Peters SK, Dunlop K, Downar J. Cortico-striatal-thalamic loop circuits of the salience network: a central pathway in psychiatric disease and treatment. *Front Syst Neurosci*. 2016;10:104.
46. Freeman SM, Razhas I, Aron AR. Top-down response suppression mitigates action tendencies triggered by a motivating stimulus. *Curr Biol*. 2014;24(2):212-216.
47. Carbia C, Lopez-Caneda E, Corral M, Cadaveira F. A systematic review of neuropsychological studies involving young binge drinkers. *Neurosci Biobehav Rev*. 2018;90:332-349.
48. MacLeod CM. The Stroop task: the "gold standard" of attentional measures. *J Exp Psychol Gen*. 1992;121(1):12-14.
49. Stroop JR. Studies of interference in serial verbal reactions. *J Exp Psychol*. 1935;18(6):643-662.
50. Simon JR, Rudell AP. Auditory SR compatibility: the effect of an irrelevant cue on information processing. *J Appl Psychol*. 1967;51(3):300-304.
51. Hommel B. The Simon effect as tool and heuristic. *Acta Psychol (Amst)*. 2011;136(2):189-202.
52. Heinz A, Kiefer F, Smolka MN, et al. Addiction research consortium: losing and regaining control over drug intake (ReCoDe)—from trajectories to mechanisms and interventions. *Addict Biol*. 2020;25(2):e12866.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.

How to cite this article: Chen H, Nebe S, Mojtahedzadeh N, et al. Susceptibility to interference between Pavlovian and instrumental control is associated with early hazardous alcohol use. *Addiction Biology*. 2021;26:e12983. https://doi.org/10.1111/adb.12983