medRxiv preprint doi: https://doi.org/10.1101/2023.12.16.23300091; this version posted December 18, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Reward processing in young people with self-harm behaviour: an absence of motivational bias towards SH-related cues and naturalistic rewards

Emre Yavuz, MSc*

Division of Psychiatry, Department of Brain Sciences, Imperial College London
Institute of Behavioural Neuroscience, Department of Experimental Psychology, Division of Psychology and Language Sciences, University College London, London, UK

Rachel Rodrigues, PhD*
Division of Psychiatry, Department of Brain Sciences, Imperial College London

Ana Pascual Sanchez, PhD
Child and Adolescent Mental Health Service (CAMHS) Enhanced Treatment Service, South London and Maudsley NHS Foundation Trust, London, UK

Anne Lingford-Hughes, PhD
Division of Psychiatry, Department of Brain Sciences, Imperial College London, London, UK

Martina Di Simplicio, PhD
Division of Psychiatry, Department of Brain Sciences, Imperial College London, London, UK

**joint first authors*

Corresponding author:

Dr Martina Di Simplicio
Division of Psychiatry, Department of Brain Sciences
Imperial College London
2nd Floor Commonwealth Building
Du Cane Road
London
W12 0NN
UK
Tel - 0208 383 41071
Email – m.di-simplicio@imperial.ac.uk

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/2023.12.16.23300091; this version posted December 18, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Abstract

Twenty percent of young people report a lifetime presence of self-harm (SH) behaviour, associated with negative health and functional outcomes. Understanding the underlying cognitive mechanisms is needed to develop targeted early interventions. Reward processing biases may underlie SH, aligning with accounts of the behaviour acquiring “addictive” characteristics. However, the specific nature of such biases remains unclear, particularly its relationship with negative affect (NA) that frequently triggers SH. In Study 1, we compared young people (aged 16-25) with SH to a group with NA but no SH history and a healthy control group on performance of a novel Incentive Delay Task (IDT), with SH-related (SH trials), positive social (social trials) or monetary images (money trials) as stimuli. In Study 2, a different sample of SH and HC participants completed the same IDT following NA induction via an online Trier Social Stress Test. For both studies, we hypothesised faster and more correct responses in the SH group than control groups on SH trials. Contradicting our hypothesis, there were no significant between-group differences in IDT performance on SH, social and money trials in either study. Certain SH characteristics (positive reinforcement, SH mental imagery, urge) were significantly correlated with better performance on SH trials in SH participants. Thus, broadly SH behaviour may not be underpinned by motivational biases towards SH-related cues or naturalistic rewards. Future studies should clarify whether incentivisation of SH-related cues instead explains individual differences in SH behaviour and its relation with treatment and prognosis.

medRxiv preprint doi: https://doi.org/10.1101/2023.12.16.23300091; this version posted December 18, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Introduction

Self-harm (SH) is defined as "any act of self-injury or self-poisoning carried out by a person, irrespective of their motivation"[1]. SH most commonly occurs in young people by self-cutting, occurring in about 18% of adolescents[2-7]. Engaging in SH increases risk of future suicide attempts and other adverse outcomes including substance abuse and impaired psychosocial functioning[3], thus early intervention is important.

SH is understood as a manifestation of emotional distress and is often used to cope with difficult emotions[2]. As the severity and frequency of SH varies widely[2-7], it is essential to understand the motivational processes that underlie self-harm and may differentiate those that self-harm repeteadly. In keeping with personal accounts of self-harm[7], this has also been conceptualised as an addictive behaviour that can be learned and repeated as a result of reinforcement from reducing negative and/or enhancing positive emotions[4-6]. Hence, an understanding of the underlying mechanisms of this addictive component could aid the development of preventative strategies for repeated SH.

One of the theoretical frameworks for understanding addictive behaviours is the Incentive Sensitisation Theory[8]. Incentive sensitisation is thought to occur due to reward system hypersensitisation to a substance or behaviour and its associated cues, and helps explain how the occasional use of substances can lead to addiction or relapse[8]. According to this theory, substance-related cues become conditioned and predictive of reward, resulting in attentional resources being preferentially allocated to them[8]. Exposure to these cues can elicit reward anticipation and motivation to obtain the reward[9]. Initial evidence supports the idea that SH-related cues may also become sensitised and elicit reward anticipation[10,11]. These SH cues may

medRxiv preprint doi: https://doi.org/10.1101/2023.12.16.23300091; this version posted December 18, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

include external cues such as objects to self-harm or internal representations, such as mental imagery of SH[12], which is highly prevalent in those who SH[13] and could play a role similar to mental imagery of substances, a key feature of craving in addiction[14].

Past research suggests these biases in reward processing systems may be present in SH[15-17]. For example, a genetic predisposition to lower dopamine transmission is correlated with greater SH in individuals with bulimia[15] and in individuals who experienced child abuse[16]. Using fMRI, neural activity in reward-related regions has been shown to be both higher[18.19] and lower[17] when anticipating monetary rewards in SH groups compared with healthy controls, which may be associated with predisposition to SH like in other addictive behaviours[20]. However, studies also report no differences in sensitivity to social rewards in SH groups compared with controls[21,22]. The term 'waiting impulsivity' has also been used to characterise other addictive behaviours aside SH, describing the ability to refrain from responding until signalled to do so by explicit cues[23-25]. These inconsistent findings highlight the need to clarify the presence and direction of motivational biases for natural rewards in SH. Given that greater self-reported impulsivity has also been associated with a greater likelihood of SH behaviour[23], this could be explored in the context of different reward stimuli.

Past research also suggests that individuals engage in SH to relieve negative affect (NA)[6]. Greater NA often preceeds SH enactment[26,27] In line with this, NA has been reported in individuals with SH compared with healthy controls in cross-sectional and longitudinal behavioural studies[28-30]. In young people with SH, BOLD activity in reward-related regions correlates with self-reported deficits in emotion regulation[31], and is modulated by stimuli's valence[32]. NA is also associated with reduced reward signalling for non-salient rewards in addiction, often leading to relapse and increasing motivation for the addictive behaviour[33]. For

medRxiv preprint doi: https://doi.org/10.1101/2023.12.16.23300091; this version posted December 18, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

example, some young people frequently report that they self-harm because that provides more intense stimulation than other alternative strategies. However, the interaction between NA and reward processing remains unexplored in those with SH.

We conducted two studies to investigate motivational biases for SH-related stimuli and natural rewards in young people with SH (Study 1). In a separate sample of participants, we investigated how these biases are influenced by NA using a NA induction (Study 2). We used a novel behavioural Incentive Delay Task (IDT[34]) with money, positive social and SH cues, measuring the number of correct responses and reaction latency across trials (motivational bias). We also measured the number of premature responses across trials as an indicator of waiting impulsivity.

In Study 1 we compared individuals with SH to both a HC group, and a group with equivalent levels of NA but no SH history, to control for the effect of low mood and anxiety on reward processing and explore why some individuals do not SH despite experiencing high NA[2,35]. As motivational biases may only be elicited in the presence of NA, a different SH and HC group repeated the same experiment in Study 2 after undergoing the Trier Social Stress Test. Our primary hypothesis for both studies was that if SH participants were biased to salient SH-related stimuli, the SH group would respond faster to- and make more correct or premature responses to SH stimuli compared with NA and HC groups in the IDT. As findings on motivational biases for natural stimuli in SH are inconsistent, we explored differences in RT, accuracy and premature responses to social or monetary rewards in the IDT. We also explored the association between responses to SH stimuli and self-reported measures of SH frequency, duration and recency, endorsed motivation for SH, urge to SH, SH imagery characteristics (Study 1 only) in the SH group, and state depression, anxiety and stress levels in SH and NA groups.

medRxiv preprint doi: https://doi.org/10.1101/2023.12.16.23300091; this version posted December 18, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Methods

We collaborated with a Young Person's Research Group (YPRG) formed of young people aged 16 - 25 with lived experience of self-harm and other mental health difficulties to deliver these studies. The YPRG contributed to the experimental design, development of the IDT, recruitment methods and interpretation of results. The study was approved by the NRES Committee South Central Oxford-C (REF: 19/SC/0275). All materials and data are reported on Open Science Framework (https://osf.io/8kcg3/).

Participants

Participants were recruited using social media advertisements on Instagram and Facebook (see Supplementary Material for details) and reimbursed with a £50 voucher for completing the testing. Inclusion criteria for all 3 groups included being 16-25 years old, having adequate English to provide informed consent and understand the study and the tasks. Exclusion criteria included having severe neurological impairment, learning disability, or current, acute psychotic symptoms that could interfere with task performance. The group-specific inclusion criteria were: 1) SH group having at least one episode of SH within the past year; 2) NA group having depression and/or anxiety scores on the Depression, Anxiety and Stress Scale (DASS-21)[36] above the normal range (depression score >= 10, anxiety score >= 8, stress score >= 15); healthy controls (HC) group having DASS-21 depression, anxiety and stress scores within the normal range. Participants were excluded from the NA group if they had a history of SH or suicide, current or past substance abuse or dependence, as well as if they had smoked regularly or vaped nicotine (self-reported), assessed using the MINI International Neuropsychiatric Interview (MINI)[37] and smoking assessment. Participants were excluded from

medRxiv preprint doi: https://doi.org/10.1101/2023.12.16.23300091; this version posted December 18, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

the HC group if they had presented a history of mental health difficulties, SH, suicide or substance misuse or dependence, as well as if they had smoked regularly or vaped nicotine (self-reported), also assessed using the MINI and smoking assessment.

Experimental Procedure

The testing session for studies 1 and 2 was conducted online over MS Teams by members of the Mood Instability Research Group. Both studies took place online given the COVID pandemic, with study 1 taking place from January 2019-2021 and study 2 from January 2021-2023.

Self-report questionnaires

Participants answered self-report questionnaires via Qualtrics links shared by the experimenter over MS Teams. Clinical characteristics were assessed via the MINI[37] and the McLean Screening Instrument for Borderline Personality Disorder (MSI-BPD)[38]. Questionnaires relating to SH included the Self-Injurious Thoughts and Behaviour Interview (SITBI)[39], the Self-Harm Imagery Interview (SHII)[40] (Study 1 only), the Alexian Brothers Urge to Self-Injure Scale (ABUSI)[41] and the Craving Experiences Questionnaire adapted for SH (CEQ-SH)[14] (Study 2 only). State depression, anxiety and stress was assessed using the DASS-21[42]. The Positive and Negative Affect Schedule (PANAS) was administered before and after the Trier Social Stress Test in Study 2 (see below). See Questionnaires details in the Supplementary Materials.

Incentive Delay Task

medRxiv preprint doi: https://doi.org/10.1101/2023.12.16.23300091; this version posted December 18, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

This study used a novel IDT developed by our team and adapted from the original fMRI Monetary Incentive Delay paradigm[17,34], Participants completed the task on the online platform Pavlovia. The task had 3 reward types: SH, positive social and money, with two blocks per reward type and 50-60 trials/block. For each reward type, half the trials were a neutral condition and half were a 'win' condition with reward-based stimuli (SH, positive social, money). For the SH reward type, half the SH-related stimuli were standardised across tasks and half were tailored to each participant's SH methods. NA and HC participants tasks were matched to SH participants. The order of the reward types was counterbalanced across participants.

During the task, a cue appeared which predicted the type of feedback image (reward) participants would receive, followed by a target (black square) which then disappeared. Participants had to respond correctly by pressing a key before the target disappeared to obtain the feedback images. If the target was correctly responded to, a neutral image (either an image of a neutral object, a neutral social scene or an image of an empty container; neutral condition) or reward image (either an image of act of SH, a positive social scene or an image of money in a container; win condition) appeared (Figure 1). See Supplementary Figure S1 for details of the IDT stimuli. Participants also won 20p per correct response in win money trials, as informed in the pre-task instructions. If participants responded to the cue, the feedback 'Too early ' appeared. If they responded after the target disappeared, the feedback 'Try to be faster ' appeared. The target's minimum and maximum duration of presentation was calculated based on the standard deviation of the participant's baseline RTs from the fastest half of trials during a 20-trial practice block (based on a previous study[43]). This ensured sufficient task difficulty.

Accuracy and mean RTs on the IDT were calculated for correct responses to the target. Accuracy was calculated by summing the number of correct target responses occurring after

medRxiv preprint doi: https://doi.org/10.1101/2023.12.16.23300091; this version posted December 18, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

0.1s from the target's appearance for reward and neutral conditions, for each reward type separately (log transformed). Premature responses, measuring waiting impulsivity, were calculated by summing the number of responses to the cue and target that occurred within 0.1s of the target's appearance for reward and neutral conditions, for each reward type separately (log-transformed).

In Study 1, after IDT completion, participants also rated each of the SH stimuli from 0 (not at all) to 10 (extremely) on five scales: calming, exciting, pleasant, unpleasant and distressing, creating one score for each scale. These 5 scores were also summed to create a total score given previous findings suggesting that SH elicits mixed emotions[44,45]. These scores were used to explore whether perception of SH stimuli influences behavioural responses during the IDT.

medRxiv preprint doi: https://doi.org/10.1101/2023.12.16.23300091; this version posted December 18, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

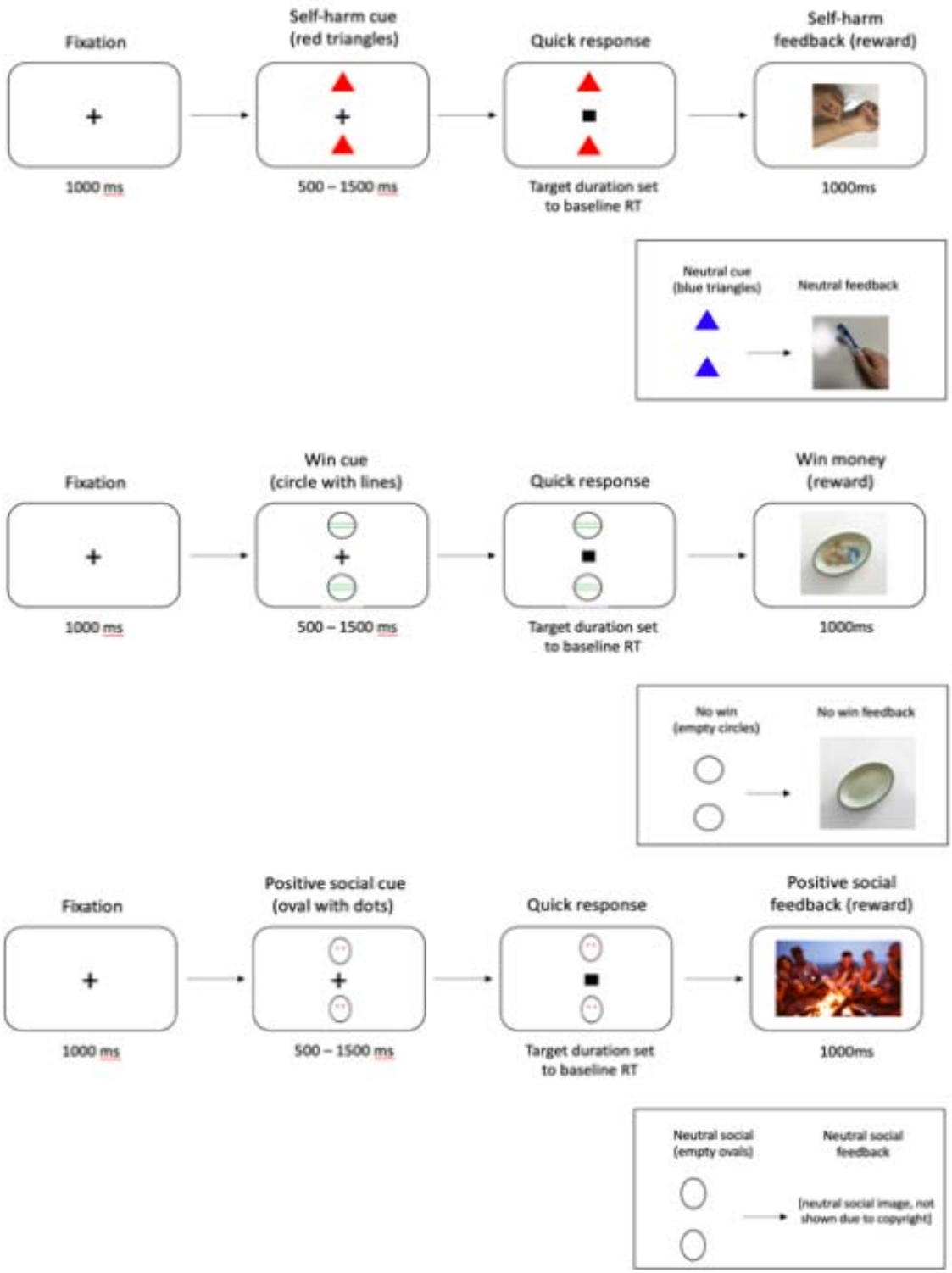

Figure 1: Incentive Delay Task, showing the self-harm reward type (top), money reward type (middle) and social reward type (bottom).

medRxiv preprint doi: https://doi.org/10.1101/2023.12.16.23300091; this version posted December 18, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Trier Social Stress Test

In Study 2, negative affect was induced by administering the Trier Social Stress Test (TSST). The TSST is the most frequently used social evaluative stress task in children, adolescents and adults[46,47], and effectively induces self-reported NA and physiological stress responses in HC and SH groups[48,49]. An online TSST[50] was used based on the original in-person TSST[51] and shown to be effective in inducing self-reported and physiological stress responses in healthy adolescents[50].The TSST has 3 sections: a 5-minute speech preparation period; a 5-minute speech presentation; and a 5-minute arithmetic task. The speech and arithmetic task are completed in front of 2 judges (Figure 2). See Supplementary Materials for details.

Statistical analysis

A power analysis (based on an Incentive Delay Task used by a previous study comparing reaction times in SH compared HC participants[17]) indicated that N=48 participants were required per group for an effect size of d = 0.580 at 80% power, with an alpha threshold of 0.05 (see Supplementary Materials for details).

In Study 1, correct and premature responses were analysed using a 3 x 3 x 2 ANOVA, with group (SH, NA and HC) as a between-group factor and reward (SH, social and money) and condition (win, neutral) as within-group factors. Significant main effects of group and reward were examined using Bonferroni-corrected post-hoc t-tests or Games-Howell tests where homoscedasticity was not met, and significant interactions were analysed using simple main effects analyses. Total money won was compared between groups using a one-way ANOVA, with significant main effects of group examined with Bonferroni-corrected post-hoc t-tests. A reward-neutral trials contrast was calculated for the percentage of correct trials, premature

medRxiv preprint doi: https://doi.org/10.1101/2023.12.16.23300091; this version posted December 18, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

responses and RTs for correct responses in the IDT to run correlational analysis. Each of these 3 contrasts were correlated with past-year, lifetime frequency and duration of SH and endorsement of automatic positive or negative reinforcement as SH motivator (SITBI); urge to SH (ABUSI and CEQ); characteristics of self-harm-related mental imagery (Self-Harm Images Interview)(Study 1 only); and state depression, anxiety and stress (DASS-21) scores using Spearman-rank correlations. As these correlations were exploratory, we did not correct for multiple comparisons.

The analysis procedure for the IDT in Study 2 was identical to part 1, except only 2 groups (SH vs HCs) were used for analysis and SH-related imagery scores were not available, as mentioned above. Additionally, two separate contrasts for positive and negative PANAS scores respectively (PANAS scores after - before TSST) were correlated with contrasts for each of the IDT outcome measures (reward - neutral trials) for each group and for each condition separately using Spearman-rank correlations. This was to assess whether TSST-induced changes in negative or positive affect were associated with a motivational bias in the IDT.

Results

	Study 1		**Study 2**	
Self-harm Characteristic	**Mean (± SD)**	**Median (IQR)**	**Mean (± SD)**	**Median (IQR)**
Recency: No. days since last self-harm episode	72.6 (± 90.7)	30 (143)		
Lifetime frequency	396.5 (± 829.5)	100 (217)	283.6	46.5 (234.5)
Past-year frequency	52.6 (± 829.5)	100 (217)	41.2 (± 72.0)	10 (31.5)
Past-month frequency	3.2 (± 6.0)	1 (4)		

medRxiv preprint doi: https://doi.org/10.1101/2023.12.16.23300091; this version posted December 18, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Age onset, yrs	13.8 (± 3.5)	13 (4)	13.5 (± 2.4)	13 (3)
Duration, yrs	5.5 (± 3.6)	5 (5)	4.8 (± 3.2)	4 (3)
Number of self-harm methods	2.4 (± 1.2)	2 (1)	1.5 (± 0.8)	1 (1)
Ever needed medical treatment (yes:no:unknown)	15:39:0	-	15:61:4	-
Future likelihood of self-harm (0-4 scale)	2.7 (± 1.2)			
Positive reinforcement endorsement score	2.33 (± 1.2)	3(2)	2.39 (± 1.3)	3(1)
Negative reinforcement endorsement score	2.67 (±0.9)	3(1)	2.95 (± 0.9)	3(2)
Urge to self-harm (ABUSI, max score 35)	16.7 (± 6.8)	15 (10)	12.0 (± 7.2)	12 (11.5)

Table 1. Self-harm characteristics of SH group in Study 1 and 2. ABUSI: Alexian Brothers Urge to Self-harm scale

Study 1

Fifty-four SH, 56 NA and 51 HC participants initially completed the IDT task. After removal of outliers (see Supplementary Materials), a final sample of 49 SH, 54 NA and 49 HC was left. Demographic and SH characteristics are reported in Table 1 and 3 respectively. Clinical characteristics are reported in Supplementary Table S4.

medRxiv preprint doi: https://doi.org/10.1101/2023.12.16.23300091; this version posted December 18, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

	SH (N = 54)	NA (N = 57)	HC (N = 51)	F/$\chi 2$	*p*
Age, yrs (mean ± SD)	19.4 (± 2.9)	18.6 (± 2.4)	19.0 (± 2.5)	1.1	NS
Gender (f:m:nb:o)	38:12:3:1	52:5:0:0	40:11:0:0	5.0	NS
Ethnicity:				5.0	0.004
White	44 (81.5%)	32 (56.1%)	27 (52.9%)		
Black/African/Carribean	2 (3.7%)	5 (9.6%)	4 (7.8%)		
Asian	6 (11.1%)	16 (28.0%)	13 (25.5%)		
Arab	1 (1.9%)	3 (5.2%)	4 (7.8%)		
Mixed/Multiple Ethnic Groups	1 (1.9%)	1 (1.7%)	2 (3.9%)		
Not specified	0 (0%)	0 (0%)	1 (2.0%)		
Yrs education (mean ± SD)	14.0 (± 2.0)	13.5 (± 1.6)	13.8 (± 1.9)	1.1	NS
IQ (mean ± SD)	13.5 (± 7.7)	13.5 (± 7.1)	11.6 (± 6.4)	0.5	NS

Table 2. Demographic characteristics of the Study 1 sample. IQ was calculated based on National Adult Reading Test (NART) score.*White vs non-White.

Incentive Delay Task performance

There was a significant main effect of group on correct responses (F(2, 148) = 4.98, p = 0.008, $\eta p2$ = 0.06), where post-hoc t tests showed that the SH group made more correct responses to the target across the entire task than the NA group (p = 0.006) but not HCs (p = 0.215) (Figure 3). This significant difference was present for all 3 reward types (group x reward type x condition interaction: (F(4, 296) = 2.37, p = 0.053, $\eta p2$ = 0.03). There was a significant main effect of reward type for correct responses (F(2, 296) = 44.33, p < 0.001, $\eta p2$ = 0.23), where all participants had fewer correct responses in win and neutral money trials than in positive and

medRxiv preprint doi: https://doi.org/10.1101/2023.12.16.23300091; this version posted December 18, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

neutral social trials and SH and neutral SH trials ($p < 0.001$). There was no significant main effect of condition (win vs neutral) for correct responses across all reward types ($F(1,148) = 0.16$, $p = 0.691$, $\eta p2 = <0.01$). However, there was a significant reward type x condition interaction ($F(2, 296) = 6.76$, $p = 0.001$, $\eta p2 = 0.04$), with a higher number of correct responses in the win money trials compared with the neutral money trials ($p = 0.001$), but not in SH 'win' trials compared with neutral SH trials ($p = 0.423$) and 'win' (positive) social trials compared with neutral social trials ($p = 0.072$).

There was no significant main effect of group on correct RTs to the target ($F(2, 149) = 1.63$, $p = 0.199$, $\eta p2 = 0.02$)(Figure 4). There was a significant main effect of reward type for correct RTs ($F(2, 298) = 5.49$, $p = 0.005$, $\eta p2 = 0.04$), with significantly shorter RTs for the money reward type than the SH reward type ($p < 0.05$). There was a significant main effect of condition ($F(2, 149) = 23.79$, $p < 0.001$, $\eta p2 = 0.14$), with significantly shorter correct RTs in win trials compared with neutral trials across all reward types ($p < 0.005$). There was a significant reward type x condition interaction for correct RTs ($F(2, 298) = 16.91$, $p <0.001$), with significantly shorter RTs in win money compared with neutral money trials ($p < 0.001$), but not in SH 'win' trials compared with neutral SH trials ($p = 0.78$) and 'win' (positive) social compared with neutral social trials ($p = 0.248$). There was no significant group x reward type x condition interaction for correct RTs ($p > 0.05$).

There was a significant main effect of group for premature responses ($F(2, 147) = 3.57$, $p = 0.031$, $\eta p2 = 0.05$), although between-group differences did not reach significance ($p > 0.05$)(Figure 5). There was a significant main effect of reward type ($F(2, 294) = 23.25$, $p < 0.001$, $\eta p2 = 0.14$), where all groups made significantly fewer premature responses for the money reward type than the SH and social rewards ($p < 0.001$). There was a significant reward type x condition interaction ($F(2, 294)) = 19.46$, $p < 0.001$, $\eta p2 = 0.12$), with all groups making more

medRxiv preprint doi: https://doi.org/10.1101/2023.12.16.23300091; this version posted December 18, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

premature responses on SH 'win' trials relative to neutral SH trials and on win money trials relative to neutral money trials ($p < 0.001$), but not on 'win '(positive) social compared with neutral social trials (p=0.841). There was no significant group x reward type x condition interaction for premature responses ($F(4, 294) = 1.42$, $p = 0.229$, $\eta p2 = 0.02$).

There was a significant main effect of group for money won ($F(2, 149) = 10.03$, $p = 0.025$, $\eta p2 = 0.05$), with the SH group winning significantly more money than the HC group ($p = 0.041$), but not the NA group ($p = 0.073$)(Supplementary materials).

Associations between Incentive Delay Task performance and self-report questionnaires

Participants in the SH group who had been self-harming for fewer years had significantly more correct responses on SH trials ($rs = 0.33$, $p = 0.024$). Participants in the SH group who reported greater automatic positive reinforcement motivations for SH on the SITBI responded faster to the target on SH trials (rs = -0.316, p = 0.027). SH participants with higher DASS-21 depression subscale scores had significantly more premature responses in SH trials ($rs = 0.31$, $p = 0.032$). Participants in the SH group whose SH-related mental imagery generated more positive emotions (on PANAS score embedded in the SHII) made more correct responses on SH trials (rs = 0.348, p = 0.017). See Supplementary Figure S4.

Associations between Incentive Delay Task performance and appraisal of SH stimuli

Ratings for SH stimuli in the three groups are reported in Supplementary Table S1. In the SH group, those who rated SH stimuli as more pleasant and salient (total rating score) had a greater number of correct responses on SH trials (pleasant $rs = 0.29$, $p = 0.032$; total rating

medRxiv preprint doi: https://doi.org/10.1101/2023.12.16.23300091; this version posted December 18, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

score *rs* = 0.29, *p* = 0.034). Distressing and unpleasant ratings were not associated with performance in the SH group. In the NA group, those rating SH stimuli as more unpleasant, distressing and salient had shorter mean RTs on SH trials (unpleasant *rs* = -0.32, *p* = 0.015; distressing *rs* = -0.27, *p* = 0.049; total rating score *rs* = -0.31, *p* = 0.021). There were no significant associations between SH stimuli ratings and IDT performance in the HC group (*p* > 0.05).

Study 2

Seventy-five SH and 70 HC participants initially completed the Trier Social Stress Test and IDT task. After removal of outliers (10 SH, 7 HC) (see Supplementary Material), a final sample of 65 SH and 63 HC participants was left. Demographic characteristics and scores on self-report questionnaires are reported in Table 2 and 3 respectively. Clinical characteristics are reported in Supplementary Table S4.

	SH (N = 80)	**HC (N = 76)**	**F/$\chi 2$**	***p***
Age, yrs (mean ± SD)	18.2 (± 2.4)	20.4 (± 2.6)	-5.58	<0.001
Gender (f:m:nb)	71:5:4	56:20:0	14.68	<0.001
Ethnicity:				NS
White	41 (51%)	31		
Black/African/Carribean	8 (10%)	7		
Asian	24 (30%)	34		
Arab	0 (0%)	2		
Mixed/Multiple Ethnic Groups	4 (5%)	2		

medRxiv preprint doi: https://doi.org/10.1101/2023.12.16.23300091; this version posted December 18, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Not specified	3 (4%)	0		
Yrs education (mean ± SD)	12.7 (± 5.8)	15.3 (± 5.6)	-2.81	0.006
IQ (mean ± SD)	10.7 (± 8.6)	11.7 (± 5.6)	-0.84	NS

Table 3. Demographic characteristics for sample in Study 2. IQ was calculated based on National Adult Reading Test (NART) scores.

Trier Social Stress Test (TSST)

There was a significant main effect of time on negative affect (PANAS scores) ($F(1, 236) = 53.69$, $p < 0.001$, $\eta p2 = 0.19$), where negative affect scores were significantly higher following the TSST than before across both SH and HC groups ($p < 0.0001$). There was also a significant main effect of group on negative affect ($F(1, 236) = 111.23$, $p < 0.001$, $\eta p2 = 0.32$) and positive affect (PANAS scores) ($F(1, 236) = 29.96$, $p < 0.001$, $\eta p2 = 0.11$), where SH participants had significantly higher negative affect scores and lower positive affect scores than HC participants both before and after the TSST ($p < 0.0001$)(Figure 2).

Incentive Delay Task performance

There was a significant main effect of reward type on the number of correct responses ($F(1, 810) = 47.65$, $p < 0.001$, $\eta p2 = <0.01$), with significantly greater number of correct responses for the money reward than the SH or social rewards across both groups ($p < 0.0001$)(Figure 3). All other main effects and interaction effects were non-significant ($p > 0.05$).

medRxiv preprint doi: https://doi.org/10.1101/2023.12.16.23300091; this version posted December 18, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

There was a significant main effect of group on RT ($F(1, 786) = 8.68$, $p < 0.05$, $\eta p2 = 0.01$), where those in the SH group had significantly faster RTs than the HC group, regardless of the reward type and condition ($p < 0.05$)(Figure 4). All other main effects and interaction effects were non-significant ($p > 0.05$).

There was a significant main effect of condition on the number of premature responses ($F(1, 375) = 19.43$, $p < 0.001$, $\eta p2 = 0.02$), with a significantly higher number of premature responses in 'win' trials compared with neutral trials across both groups ($p < 0.0001$)(Figure 5). There was a significant main effect of reward type on the number of premature responses ($F(2, 461) = 11.93$, $p < 0.001$, $\eta p2 = 0.03$), where participants across both groups had a significantly greater number of premature responses for the money rewards than the SH and social rewards ($p < 0.05$). There was also a significant reward type x condition interaction ($F(2, 242) = 6.27$, $p < 0.05$, $\eta p2 = 0.02$), where there were a significantly greater number of premature responses in 'win' trials compared with neutral trials for the SH and money rewards across both groups ($p < 0.05$), but not for 'win' (positive) social trials compared with neutrals social trials. All other main effects and interaction effects were non-significant ($p > 0.05$).

There was no significant difference in the total amount of money won between the SH and HC groups ($p > 0.05$)(Supplementary Materials).

<u>Association between Incentive Delay Task performance and affect change after TSST</u>

There was no significant association between any of the IDT outcome measures and change in positive or negative affect (PANAS scores after TSST - PANAS scores before TSST) in either the SH or HC group ($p > 0.05$).

medRxiv preprint doi: https://doi.org/10.1101/2023.12.16.23300091; this version posted December 18, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Associations between IDT performance and self-report questionnaires

Total CEQ-SH score was significantly associated with premature responding on SH trials, where a higher urgency to engage in SH was significantly associated with fewer premature responses in the SH group ($rs = -0.29$, $p = 0.04$). All other associations were non-significant ($p > 0.05$) (Supplementary Figure S5).

medRxiv preprint doi: https://doi.org/10.1101/2023.12.16.23300091; this version posted December 18, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

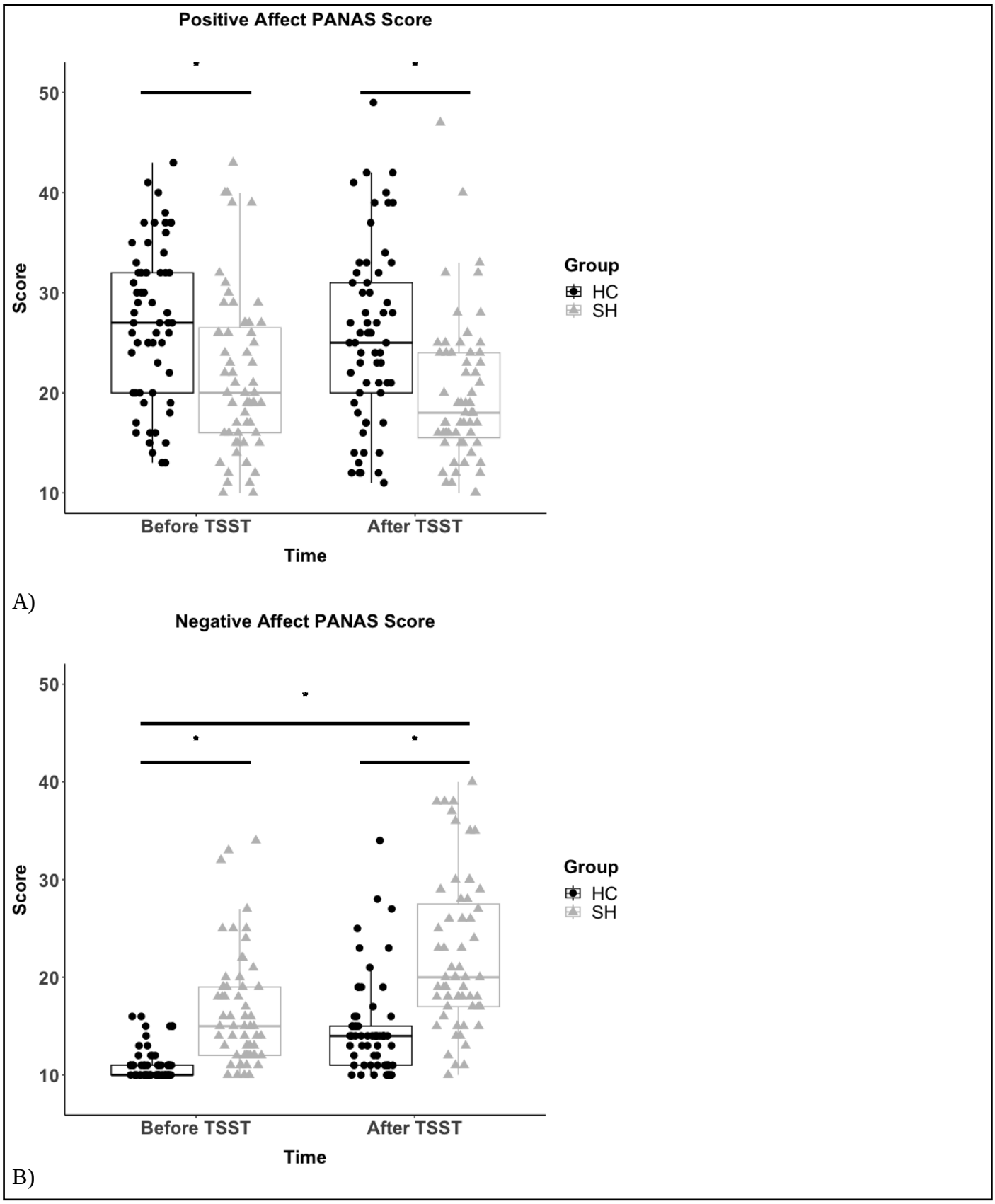

Figure 2. Change in Positive Affect PANAS scores (A) and Negative Affect PANAS scores (B) from before to after the TSST in Study 2. * = $p < 0.05$.

medRxiv preprint doi: https://doi.org/10.1101/2023.12.16.23300091; this version posted December 18, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

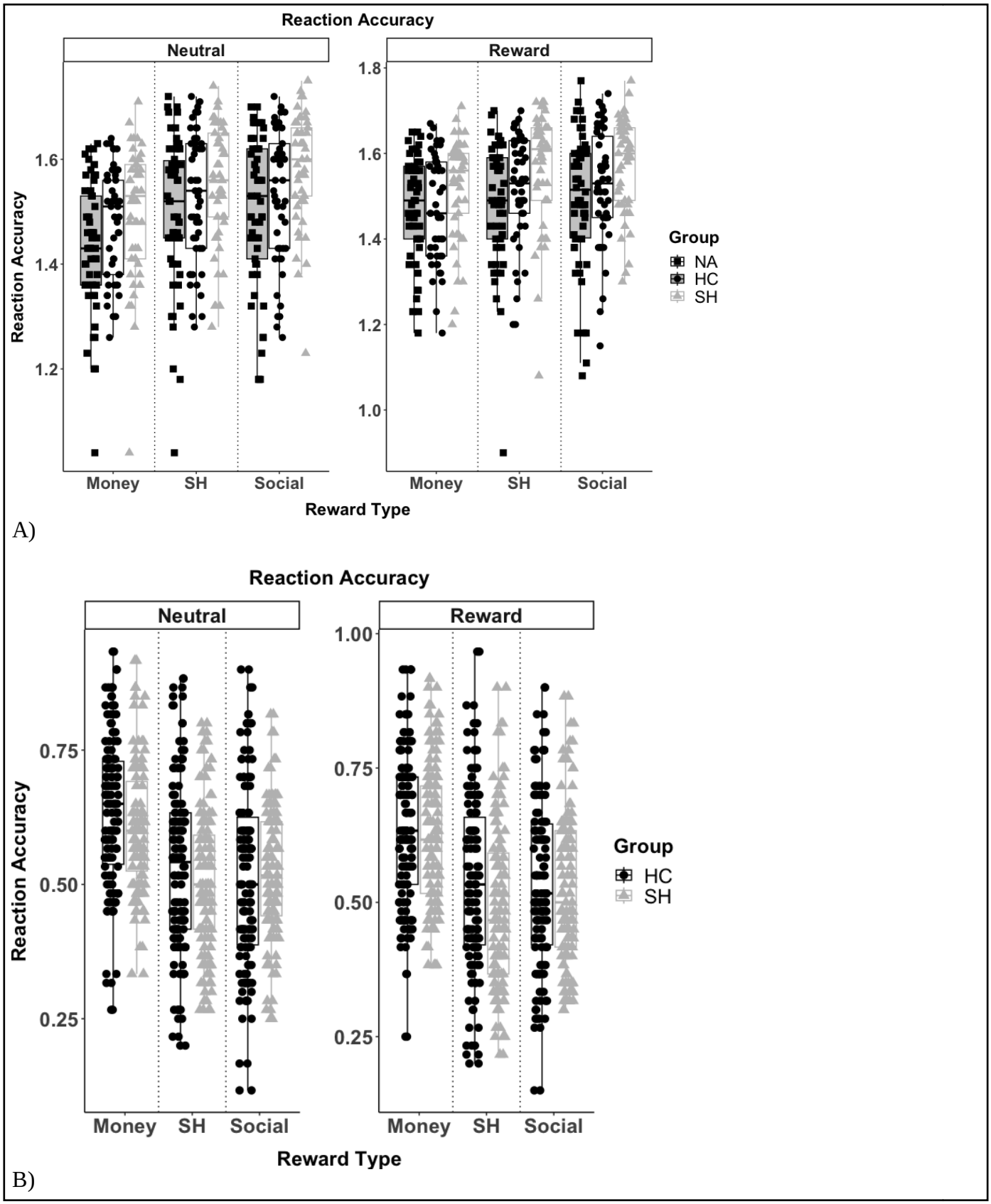

Figure 3. Correct responses for the Incentive Delay Task in Study 1 (A) and Study 2 (B) of the study.

medRxiv preprint doi: https://doi.org/10.1101/2023.12.16.23300091; this version posted December 18, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

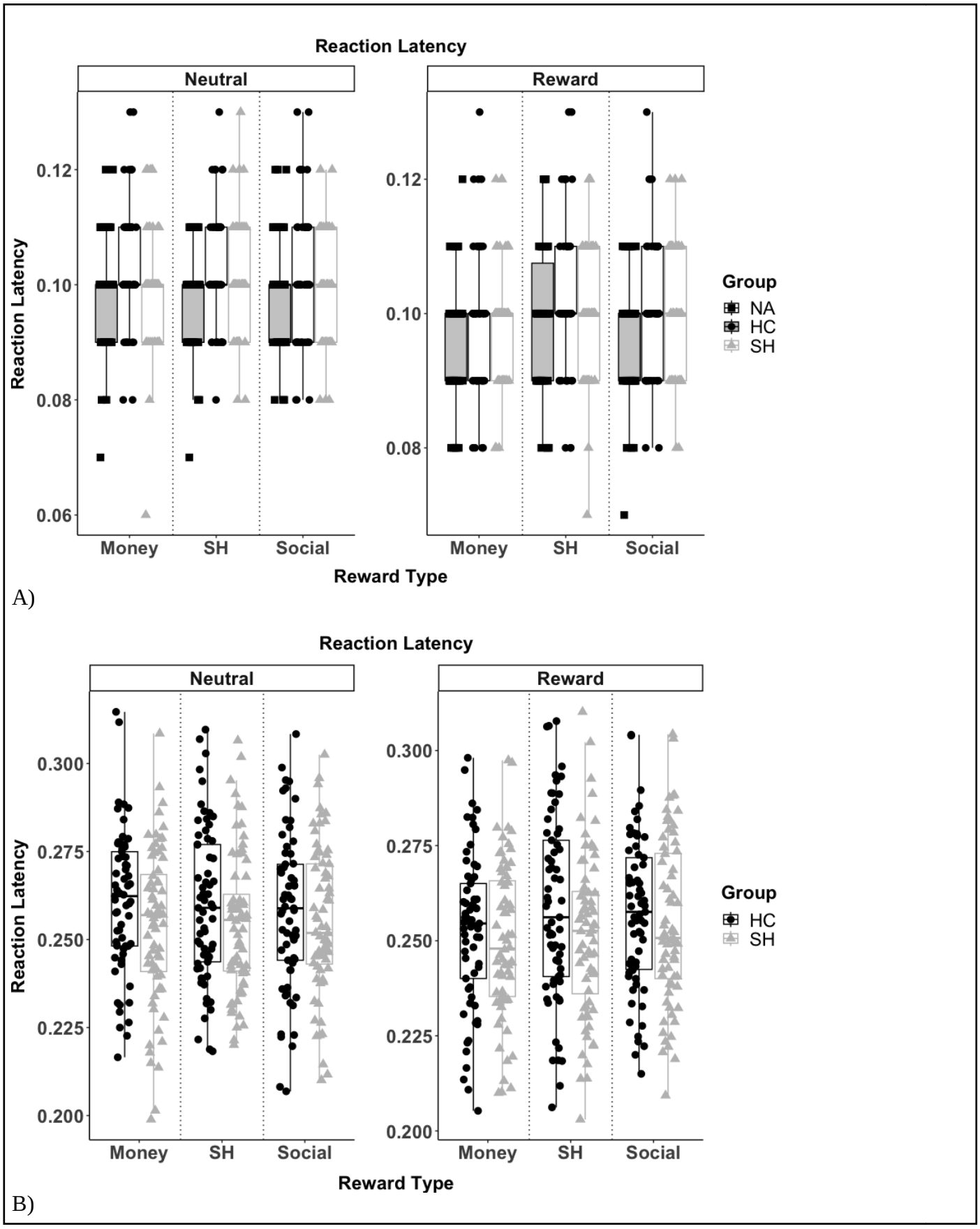

Figure 4. Reaction latency for the Incentive Delay Task for Study 1 (A) and Study 2 (B).

medRxiv preprint doi: https://doi.org/10.1101/2023.12.16.23300091; this version posted December 18, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

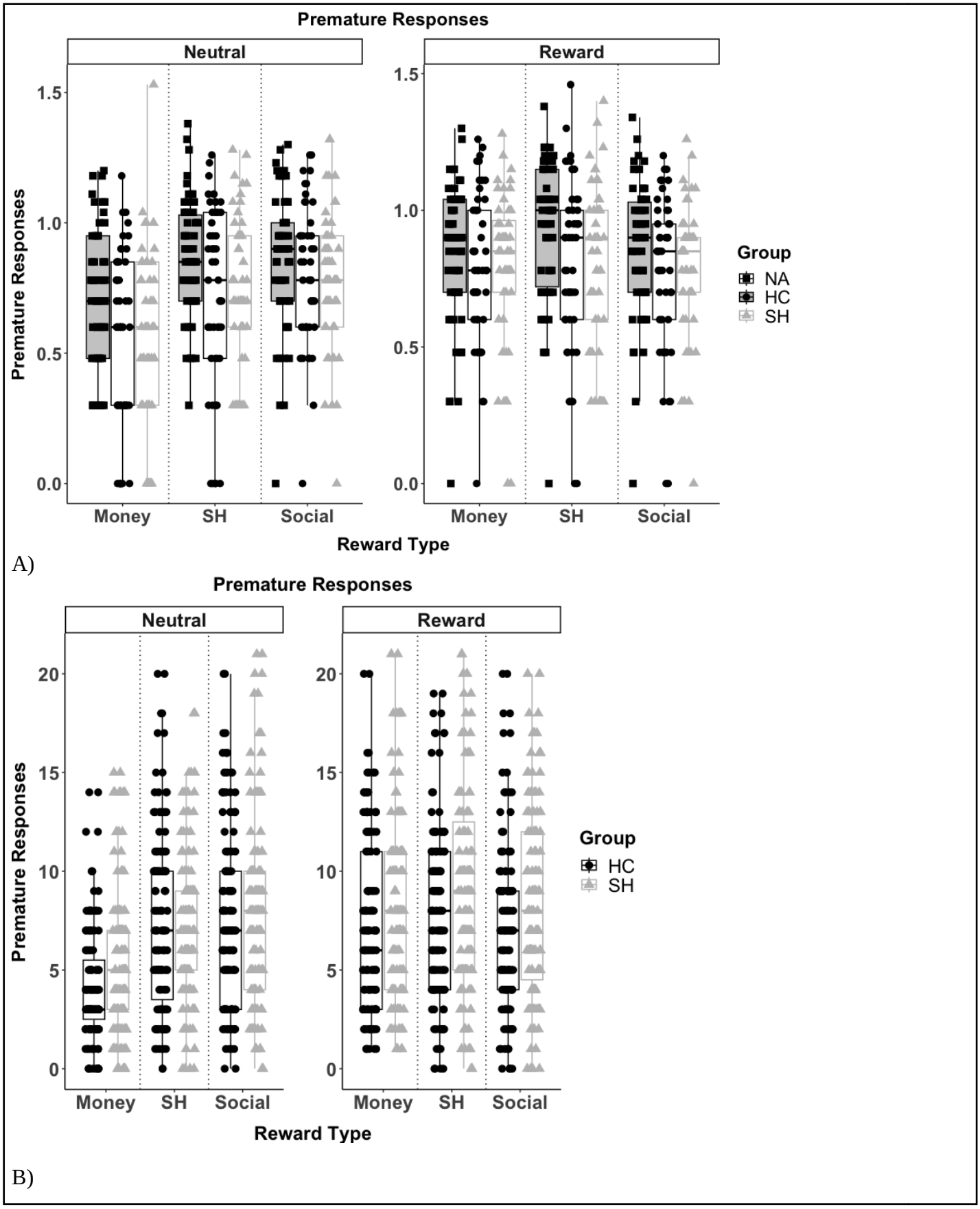

Figure 5. Premature responses for the Incentive Delay Task for Study 1 (A) and Study 2 (B).

medRxiv preprint doi: https://doi.org/10.1101/2023.12.16.23300091; this version posted December 18, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Discussion

Contrary to our hypothesis, young people with SH behaviour did not show a motivational bias towards salient SH-related stimuli on a novel behavioural Incentive Delay Task compared to controls. Moreover, we did not find any biases in behavioural measures of reward anticipation for monetary or social rewards in individuals with SH compared to controls. Negative affect induction via a social stress task was effective at inducing negative affect, but did not modulate the response to either SH-related cues or natural rewards. When exploring SH characteristics, those with fewer years of-, a greater urge to- and a higher automatic positive reinforcement motivation for SH, as well as those with higher positive affect following SH-related mental imagery, made more correct, faster and fewer premature responses on reward trials in the SH condition of the IDT. However, these associations were not replicated across our two studies.

Our results suggest that SH-related stimuli do not acquire an incentive value in young people with a history of SH. However, it is possible that SH stimuli may have triggered bi-directional biases at the individual-level in the SH group, in the form of both approach and avoidance behaviours, thus minimizing any group-level directionality of bias. This would be consistent with ambivalence models reported in substance use[52-54], other emerging evidence in SH[45,55] and in those with suicidal ideation[56]. Results from an attentional Dot Probe task also conducted in Study 1, support an avoidance of SH-related cues in the same SH group[55]. As reported in this latter study[55], members of our YPRG described having both a desire to look at the SH images and to avoid them, and suggested that this may be influenced by the motivation to stop self-harming in keeping with previous literature[57]. It is unclear if and how this is associated with individual differences in SH characteristics such as duration of SH history, given that a shorter duration of SH was associated with faster responses on SH trials in the Incentive Delay Task, but also greater attentional avoidance of SH cues in the Dot Probe Task[55] in Study 1.

medRxiv preprint doi: https://doi.org/10.1101/2023.12.16.23300091; this version posted December 18, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Exploratory analyses tentatively point to other factors that may enhance or contribute towards motivation for self-harm. Specifically, young people who more strongly endorsed self-harming for automatic positive reinforcement, for example “to feel something” or increase positive affect, and rated the SH stimuli as more pleasant had faster and more correct responses on SH trials in Study 1. This finding is consistent with greater attention towards SH cues on the Dot Probe task in the same study[55]. Additionally, young people whose mental imagery of self-harm generated more positive emotions made more correct responses to SH trials. Based on these exploratory findings, future work could investigate if acquisition of an incentive bias towards SH-related stimuli might be present in a sub-group of young people for whom positive reinforcement drives the behaviour, and if emotional response to SH-related imagery may capture this mechanism[12].

We found a significant correlation between negative affect DASS-21 scores and faster responding to SH images in the SH but not the NA group in Study 1. However, our NA induction in Study 2 refutes the hypothesis that negative affect would elicit sensitisation to SH cues in individuals with SH. This is contrary to findings in reported in addiction[58-60]. A possible explanation could be that TSST-related effects on affect may have been too brief to influence IDT performance, as subjective stress levels have been shown to decrease significantly from the beginning of the arithmetic part of the online TSST to the end[51]. Whilst we found a significant increase in PANAS negative affect after the TSST, this may not have been sufficient to effectively influence IDT task performance, given the difference between our procedure and past studies, including the number of judges and length per section[49,50]. Moreover, TSST effects may be mediated by stress hormones such as cortisol, which in turn modulate reward processing, and stress-response dysfunction has been associated with SH[48,49]. Alternatively, NA

medRxiv preprint doi: https://doi.org/10.1101/2023.12.16.23300091; this version posted December 18, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

may drive SH behaviour via other mechanisms than modulating approach towards SH cues. Previous findings showed that acute stress does not impact responses to rewarding stimuli in patients with Borderline Personality Disorder[61] and Major Depressive Disorder[62], which commonly co-occur with SH[2].

Another explanation for the inability of our NA induction to elicit sensitization to SH cues may be due to heterogeneity in our SH samples in terms of SH characteristics. For example, SH frequency spanned from only once to greater than 400 times in a lifetime (see Table 1). This may explain why some individual differences in specific factors including duration of SH were tentatively associated with IDT performance in the absence of significant between-group differences, but equally why these associations don't always replicate across our two studies. Future studies should clarify whether SH participants can be clustered based on variables such as SH frequency and whether such clusters may be associated with unique cognitive mechanisms underpinning SH behaviour.

Although previous studies using fMRI have shown reduced neural activity during reward anticipation in SH participants compared with HCs, our findings do not support the presence of biases in processing monetary rewards in young people with SH[17]. Rather than showing reduced anticipation to monetary reward, the SH group performed with better accuracy across all trials and won more money than the control groups in Study 1. It is possible that the feedback "try to be faster" after incorrect responses was perceived as particularly stressful by this population, and motivated them to do better, as reported by our YPRG members. This may have skewed results towards maximised performance across all reward type trials. Interestingly, our NA group in Study 1 also appeared as sensitive to monetary reward as the HC group despite depressive symptoms. This is in line with previous monetary IDT findings showing that

medRxiv preprint doi: https://doi.org/10.1101/2023.12.16.23300091; this version posted December 18, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

low mood per se is not associated with reduced reward anticipation but anhedonia is[63]. Of note, medication use influenced IDT performance, but did not affect our results (Supplementary Tables S6-S8).

Limitations

A possible limitation of our study might be that as participants received a tangible reward (money) only in the monetary condition, this could have masked more subtle motivational biases towards image-based cues for SH and social rewards. Further, our SH images may not have been salient enough, for example they did not show blood or wounds, while seeing blood during SH is thought to contribute to its reinforcing nature[64] . More ecologically-valid stimuli, such as social media 'likes'[21], reactions to text messages or interpersonal situations could have been used in the IDT social condition. It is also possible that simple behavioural markers as our modified IDT are unable to detect reward system biases underpinning SH behaviour. Finally, we did not ask participants about their motivation to stop self-harming or engagement in therapy, which may impact on response to SH-related stimuli; and we did not collect data on SH-related mental imagery in Study 2. Our YPRG also suggested that our sample may have been biased towards individuals trying to stop self-harming, as they would be more likely to take part in research.

Conclusion

Overall, young people who SH did not show incentive sensitisation to SH stimuli, nor biases in processing social or monetary rewards. This also remained true also following a NA induction, strengthening the phenomenological validity of our findings. As such, from a translational

medRxiv preprint doi: https://doi.org/10.1101/2023.12.16.23300091; this version posted December 18, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

perspective, anticipation of SH-related cues or rewards do not appear to have potential as markers for targeting SH interventions. However, we found an indication that motivational biases may be present in a sub-group of young people with SH. Future studies should replicate our findings either by recruiting large samples powered to robustly capture individual differences with small effect size or by focusing on those who describe their SH as 'addictive 'and perceive SH-related cues, including mental imagery, as motivating. Furthermore, understanding how 'addictive' properties of SH behaviour may relate to greater suicide risk may also help with the development of targeted preventative strategies for SH. The addition of physiological measures such as salivary cortisol may also better elucidate SH profiles for the purpose of risk, trajectory and intervention stratification.

Acknowledgements

We thank the Young Person's Research Group: Catherine Thomas, Caroline Crandell, Maya Kamall, Martin Sinclair, Katy McSweeny, Tabitha A B Podger and Abby Martin for their valuable insight and work on this research**.** We would also like to thank members of the Department of Psychiatry at Imperial College London for acting as judges during the TSST.

Financial support

This work was supported by the Imperial NIHR BRC (Award P78558 to Dr Di Simplicio). This paper presents independent research funded by Imperial NIHR BRC and supported by the NIHR CRF at Imperial College Healthcare NHS Trust. The views expressed are those of the authors and not necessarily those of Imperial NIHR BRC, the NHS, the NIHR or the Department of Health.

medRxiv preprint doi: https://doi.org/10.1101/2023.12.16.23300091; this version posted December 18, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Conflict of interest

The authors have no conflict of interest to declare.

References

1. NICE. (2022, September 7). *Overview | Self-harm: Assessment, management and preventing recurrence | Guidance | NICE*. Www.nice.org.uk. https://www.nice.org.uk/guidance/NG225
2. Wolff, J. C., Thompson, E., Thomas, S. A., Nesi, J., Bettis, A. H., Ransford, B., Scopelliti, K., Frazier, E. A., & Liu, R. T. (2019). Emotion dysregulation and non-suicidal self-injury: A systematic review and meta-analysis. *European Psychiatry*, *59*, 25–36. https://doi.org/10.1016/j.eurpsy.2019.03.004
3. Titelius, E. N., Cook, E., Spas, J., Orchowski, L., Kivisto, K., O'Brien, K., Frazier, E., Wolff, J. C., Dickstein, D. P., Kim, K. L., & Seymour, K. E. (2017). Emotion Dysregulation Mediates the Relationship Between Child Maltreatment and Non-Suicidal Self-Injury. *Journal of Aggression, Maltreatment & Trauma*, *27*(3), 323–331. https://doi.org/10.1080/10926771.2017.133881
4. Blasco-Fontecilla, H., Fernández-Fernández, R., Colino, L., Fajardo, L., Perteguer-Barrio, R., & de Leon, J. (2016). The Addictive Model of Self-Harming (Non-suicidal and Suicidal) Behavior. *Frontiers in Psychiatry*, *7*(8). https://doi.org/10.3389/fpsyt.2016.00008

medRxiv preprint doi: https://doi.org/10.1101/2023.12.16.23300091; this version posted December 18, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

5. Gardner, K. J., Paul, E., Selby, E. A., Klonsky, E. D., & Mars, B. (2021). Intrapersonal and Interpersonal Functions as Pathways to Future Self-Harm Repetition and Suicide Attempts. *Frontiers in Psychology*, *12*. https://doi.org/10.3389/fpsyg.2021.688472
6. Nock, M. K., & Prinstein, M. J. (2004). A Functional Approach to the Assessment of Self-Mutilative Behavior. *Journal of Consulting and Clinical Psychology*, *72*(5), 885–890. https://doi.org/10.1037/0022-006x.72.5.885
7. Davis, S., & Lewis, C. A. (2019). Addiction to Self-harm? The Case of Online Postings on Self-harm Message Boards. *International Journal of Mental Health and Addiction*, *17*(4), 1020–1035. https://doi.org/10.1007/s11469-018-9975-8
8. Berridge, K. C., & Robinson, T. E. (2016). Liking, wanting, and the incentive-sensitization theory of addiction. *American Psychologist*, *71*(8), 670–679. https://doi.org/10.1037/amp0000059
9. Rømer Thomsen, K., Fjorback, L. O., Møller, A., & Lou, H. C. (2014). Applying incentive sensitization models to behavioral addiction. *Neuroscience & Biobehavioral Reviews*, *45*, 343–349. https://doi.org/10.1016/j.neubiorev.2014.07.009
10. Osuch, E., Ford, K., Wrath, A., Bartha, R., & Neufeld, R. (2014). Functional MRI of pain application in youth who engaged in repetitive non-suicidal self-injury vs. psychiatric controls. *Psychiatry Research: Neuroimaging*, *223*(2), 104–112. https://doi.org/10.1016/j.pscychresns.2014.05.003
11. Hayashi, T., Ko, J. H., Strafella, A. P., & Dagher, A. (2013). Dorsolateral prefrontal and orbitofrontal cortex interactions during self-control of cigarette craving. *Proceedings of the National Academy of Sciences*, *110*(11), 4422–4427. https://doi.org/10.1073/pnas.1212185110

medRxiv preprint doi: https://doi.org/10.1101/2023.12.16.23300091; this version posted December 18, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

12. Dargan, P., Reid, G., & Hodge, S. (2016). Exploring the role of mental imagery in the experience of self-injury: an interpretative phenomenological analysis. *Behavioural and cognitive psychotherapy, 44*(1), 92-103.

13. Lawrence, H. R., Balkind, E. G., Ji, J. L., Burke, T. A., & Liu, R. T. (2023). Mental imagery of suicide and non-suicidal self-injury: A meta-analysis and systematic review. *Clinical Psychology Review* , *103*, 102302–102302. https://doi.org/10.1016/j.cpr.2023.102302

14. May, J., Andrade, J., Kavanagh, D. J., Feeney, G. F. X., Gullo, M. J., Statham, D. J., Skorka-Brown, J., Connolly, J. M., Cassimatis, M., Young, R. McD., & Connor, J. P. (2014). The Craving Experience Questionnaire: a brief, theory-based measure of consummatory desire and craving. *Addiction*, *109*(5), 728–735. https://doi.org/10.1111/add.12472

15. Thaler, L., Groleau, P., Badawi, G., Sycz, L., Zeramdini, N., Too, A., Israel, M., Joober, R., Bruce, K. R., & Steiger, H. (2012). Epistatic interactions implicating dopaminergic genes in bulimia nervosa (BN): Relationships to eating- and personality-related psychopathology. *Progress in Neuro-Psychopharmacology and Biological Psychiatry*, *39*(1), 120–128. https://doi.org/10.1016/j.pnpbp.2012.05.019

16. Gao, Y., Xiong, Y., Liu, X., & Wang, H. (2021). The Effects of Childhood Maltreatment on Non-Suicidal Self-Injury in Male Adolescents: The Moderating Roles of the Monoamine Oxidase A (MAOA) Gene and the Catechol-O-Methyltransferase (COMT) Gene. *International Journal of Environmental Research and Public Health*, *18*(5), 2598. https://doi.org/10.3390/ijerph18052598

17. Sauder, C. L., Derbidge, C. M., & Beauchaine, T. P. (2015). Neural responses to monetary incentives among self-injuring adolescent girls. *Development and Psychopathology*, *28*(1), 277–291. https://doi.org/10.1017/s0954579415000449

medRxiv preprint doi: https://doi.org/10.1101/2023.12.16.23300091; this version posted December 18, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

18. Poon, J. A., Thompson, J. C., Forbes, E. E., & Chaplin, T. M. (2018). Adolescents ' Reward-related Neural Activation: Links to Thoughts of Nonsuicidal Self-Injury. *Suicide and Life-Threatening Behavior*, *49*(1), 76–89. https://doi.org/10.1111/sltb.12418
19. Vega, D., Ripollés, P., Soto, À., Torrubia, R., Ribas, J., Monreal, J. A., Pascual, J. C., Salvador, R., Pomarol-Clotet, E., Rodríguez-Fornells, A., & Marco-Pallarés, J. (2017). Orbitofrontal overactivation in reward processing in borderline personality disorder: the role of non-suicidal self-injury. *Brain Imaging and Behavior*, *12*(1), 217–228. https://doi.org/10.1007/s11682-017-9687-x
20. Luijten, M., Schellekens, A. F., Kühn, S., Machielse, M. W. J., & Sescousse, G. (2017). Disruption of Reward Processing in Addiction : An Image-Based Meta-analysis of Functional Magnetic Resonance Imaging Studies. *JAMA Psychiatry*, *74*(4), 387–398. https://doi.org/10.1001/jamapsychiatry.2016.3084
21. Perini, I., Gustafsson, P. A., Hamilton, J. P., Kämpe, R., Mayo, L. M., Heilig, M., & Zetterqvist, M. (2019). Brain-based Classification of Negative Social Bias in Adolescents With Nonsuicidal Self-injury: Findings From Simulated Online Social Interaction. *EClinicalMedicine*, *13*, 81–90. https://doi.org/10.1016/j.eclinm.2019.06.016
22. Olié, E., Doell, K. C., Corradi-Dell'Acqua, C., Courtet, P., Perroud, N., & Schwartz, S. (2018). Physical pain recruits the nucleus accumbens during social distress in borderline personality disorder. *Social Cognitive and Affective Neuroscience*, *13*(10), 1071–1080. https://doi.org/10.1093/scan/nsy078
23. Cassels, M., Neufeld, S., van Harmelen, A.-L., Goodyer, I., & Wilkinson, P. (2020). Prospective Pathways From Impulsivity to Non-Suicidal Self-Injury Among Youth. *Archives of Suicide Research*, 1–14. https://doi.org/10.1080/13811118.2020.1811180
24. McHugh, C. M., Chun Lee, R. S., Hermens, D. F., Corderoy, A., Large, M., & Hickie, I. B. (2019). Impulsivity in the self-harm and suicidal behavior of young people: A systematic

medRxiv preprint doi: https://doi.org/10.1101/2023.12.16.23300091; this version posted December 18, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

review and meta-analysis. *Journal of Psychiatric Research*. https://doi.org/10.1016/j.jpsychires.2019.05.012

25. Dalley, J. W., & Ersche, K. D. (2018). Neural circuitry and mechanisms of waiting impulsivity: relevance to addiction. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *374*(1766), 20180145. https://doi.org/10.1098/rstb.2018.0145
26. Miller, M., Redley, M., & Wilkinson, P. O. (2021). A Qualitative Study of Understanding Reasons for Self-Harm in Adolescent Girls. *International Journal of Environmental Research and Public Health*, *18*(7), 3361. https://doi.org/10.3390/ijerph18073361
27. Muehlenkamp, J. J., Engel, S. G., Wadeson, A., Crosby, R. D., Wonderlich, S. A., Simonich, H., & Mitchell, J. E. (2009). Emotional states preceding and following acts of non-suicidal self-injury in bulimia nervosa patients. *Behaviour Research and Therapy*, *47*(1), 83–87. https://doi.org/10.1016/j.brat.2008.10.011
28. Chen, W.-L., & Chun, C.-C. (2019). Association between Emotion Dysregulation and Distinct Groups of Non-Suicidal Self-Injury in Taiwanese Female Adolescents. *International Journal of Environmental Research and Public Health*, *16*(18), 3361. https://doi.org/10.3390/ijerph16183361
29. Tatnell, R., Kelada, L., Hasking, P., & Martin, G. (2013). Longitudinal Analysis of Adolescent NSSI: The Role of Intrapersonal and Interpersonal Factors. *Journal of Abnormal Child Psychology*, *42*(6), 885–896. https://doi.org/10.1007/s10802-013-9837-6
30. Zanus, C., Battistutta, S., Aliverti, R., Monasta, L., Montico, M., Ronfani, L., & Carrozzi, M. (2021). High-school students and self-injurious thoughts and behaviours: clues of emotion dysregulation. *Italian Journal of Pediatrics*, *47*(1). https://doi.org/10.1186/s13052-021-00958-0
31. Miller, M., Redley, M., & Wilkinson, P. O. (2021). A Qualitative Study of Understanding Reasons for Self-Harm in Adolescent Girls. *International Journal of Environmental Research and Public Health*, *18*(7), 3361. https://doi.org/10.3390/ijerph18073361

medRxiv preprint doi: https://doi.org/10.1101/2023.12.16.23300091; this version posted December 18, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

32. Muehlenkamp, J. J., Engel, S. G., Wadeson, A., Crosby, R. D., Wonderlich, S. A., Simonich, H., & Mitchell, J. E. (2009). Emotional states preceding and following acts of non-suicidal self-injury in bulimia nervosa patients. *Behaviour Research and Therapy*, *47*(1), 83–87. https://doi.org/10.1016/j.brat.2008.10.011

33. Niedtfeld, I., Schulze, L., Kirsch, P., Herpertz, S. C., Bohus, M., & Schmahl, C. (2010). Affect Regulation and Pain in Borderline Personality Disorder: A Possible Link to the Understanding of Self-Injury. *Biological Psychiatry*, *68*(4), 383–391. https://doi.org/10.1016/j.biopsych.2010.04.015

34. Hooley, J. M., Dahlgren, M. K., Best, S. G., Gonenc, A., & Gruber, S. A. (2020). Decreased Amygdalar Activation to NSSI-Stimuli in People Who Engage in NSSI: A Neuroimaging Pilot Study. *Frontiers in Psychiatry*, *11*. https://doi.org/10.3389/fpsyt.2020.00238

35. Volkow, N. D., Michaelides, M., & Baler, R. (2019). The neuroscience of drug reward and addiction. *Physiological Reviews*, *99*(4), 2115–2140. https://doi.org/10.1152/physrev.00014.2018

36. Knutson, B., Bhanji, J. P., Cooney, R. E., Atlas, L. Y., & Gotlib, I. H. (2008). Neural Responses to Monetary Incentives in Major Depression. *Biological Psychiatry*, *63*(7), 686–692. https://doi.org/10.1016/j.biopsych.2007.07.023

37. Ho, T. C., Walker, J. C., Teresi, G. I., Kulla, A., Kirshenbaum, J. S., Gifuni, A. J., Singh, M. K., & Gotlib, I. H. (2021). Default mode and salience network alterations in suicidal and non-suicidal self-injurious thoughts and behaviors in adolescents with depression. *Translational Psychiatry*, *11*(1). https://doi.org/10.1038/s41398-020-01103-x

38. Antony, M. M., Bieling, P. J., Cox, B. J., Enns, M. W., & Swinson, R. P. (1998). Psychometric properties of the 42-item and 21-item versions of the Depression Anxiety Stress Scales in clinical groups and a community sample. *Psychological Assessment*, *10*(2), 176–181. https://doi.org/10.1037/1040-3590.10.2.176

medRxiv preprint doi: https://doi.org/10.1101/2023.12.16.23300091; this version posted December 18, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

39. Sheehan, D. V., Lecrubier, Y., Sheehan, K. H., Amorim, P., Janavs, J., Weiller, E., Hergueta, T., Baker, R., & Dunbar, G. C. (1998). The Mini-International Neuropsychiatric Interview (M.I.N.I.): The Development and Validation of a Structured Diagnostic Psychiatric Interview for DSM-IV and ICD-10. *The Journal of Clinical Psychiatry*, *59*(suppl 20). https://www.psychiatrist.com/jcp/neurologic/neurology/mini-international-neuropsychiatric-interview-mini/
40. Zanarini, M. C., Vujanovic, A. A., Parachini, E. A., Boulanger, J. L., Frankenburg, F. R., & Hennen, J. (2003). A Screening Measure for BPD: The McLean Screening Instrument for Borderline Personality Disorder (MSI-BPD). *Journal of Personality Disorders*, *17*(6), 568–573. https://doi.org/10.1521/pedi.17.6.568.25355
41. Nock, M. K., Holmberg, E. B., Photos, V. I., & Michel, B. D. (2007). Self-Injurious Thoughts and Behaviors Interview: Development, reliability, and validity in an adolescent sample. *Psychological Assessment*, *19*(3), 309–317. https://doi.org/10.1037/1040-3590.19.3.309
42. Di Simplicio, M., Appiah-Kusi, E., Wilkinson, P., Watson, P., Meiser-Stedman, C., Kavanagh, D. J., & Holmes, E. A. (2020). Imaginator : A Proof-of-Concept Feasibility Trial of a Brief Imagery-Based Psychological Intervention for Young People Who Self-Harm. *Suicide and Life-Threatening Behavior*, *50*(3), 724–740. https://doi.org/10.1111/sltb.12620
43. Washburn, J. J., Juzwin, K. R., Styer, D. M., & Aldridge, D. (2010). Measuring the urge to self-injure: Preliminary data from a clinical sample. *Psychiatry Research*, *178*(3), 540–544. https://doi.org/10.1016/j.psychres.2010.05.018
44. Antony, M. M., Bieling, P. J., Cox, B. J., Enns, M. W., & Swinson, R. P. (1998). Psychometric properties of the 42-item and 21-item versions of the Depression Anxiety Stress Scales in clinical groups and a community sample. *Psychological Assessment*, *10*(2), 176–181. https://doi.org/10.1037/1040-3590.10.2.176

medRxiv preprint doi: https://doi.org/10.1101/2023.12.16.23300091; this version posted December 18, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

45. Roiser, J. P., Stephan, K. E., den Ouden, H. E. M., Barnes, T. R. E., Friston, K. J., & Joyce, E. M. (2008). Do patients with schizophrenia exhibit aberrant salience? *Psychological Medicine*, *39*(2), 199–209. https://doi.org/10.1017/s0033291708003863

46. Edmondson, A. J., Brennan, C. A., & House, A. O. (2016). Non-suicidal reasons for self-harm: A systematic review of self-reported accounts. *Journal of Affective Disorders*, *191*, 109–117. https://doi.org/10.1016/j.jad.2015.11.043

47. McEvoy, P. M., Hayes, S., Hasking, P. A., & Rees, C. S. (2017). Thoughts, images, and appraisals associated with acting and not acting on the urge to self-injure. *Journal of Behavior Therapy and Experimental Psychiatry*, *57*, 163–171. https://doi.org/10.1016/j.jbtep.2017.05.010

48. Goodman, W. K., Janson, J., & Wolf, J. M. (2017). Meta-analytical assessment of the effects of protocol variations on cortisol responses to the Trier Social Stress Test. *Psychoneuroendocrinology*, *80*, 26–35. https://doi.org/10.1016/j.psyneuen.2017.02.030

49. Seddon, J. A., Rodriguez, V. J., Provencher, Y., Raftery-Helmer, J., Hersh, J., Labelle, P. R., & Thomassin, K. (2020). Meta-analysis of the effectiveness of the Trier Social Stress Test in eliciting physiological stress responses in children and adolescents. *Psychoneuroendocrinology*, *116*, 104582. https://doi.org/10.1016/j.psyneuen.2020.104582

50. Kaess, M., Hille, M., Parzer, P., Maser-Gluth, C., Resch, F., & Brunner, R. (2012). Alterations in the neuroendocrinological stress response to acute psychosocial stress in adolescents engaging in nonsuicidal self-injury. *Psychoneuroendocrinology*, *37*(1), 157–161. https://doi.org/10.1016/j.psyneuen.2011.05.009

51. Klimes-Dougan, B., Begnel, E., Almy, B., Thai, M., Schreiner, M. W., & Cullen, K. R. (2019). Hypothalamic-pituitary-adrenal axis dysregulation in depressed adolescents with non-suicidal self-injury. *Psychoneuroendocrinology*, *102*, 216–224. https://doi.org/10.1016/j.psyneuen.2018.11.004

medRxiv preprint doi: https://doi.org/10.1101/2023.12.16.23300091; this version posted December 18, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

52. Kirschbaum, C., Pirke, K.-M., & Hellhammer, D. H. (1993). The "Trier Social Stress Test" – A Tool for Investigating Psychobiological Stress Responses in a Laboratory Setting. *Neuropsychobiology*, *28*(1-2), 76–81. https://doi.org/10.1159/000119004
53. Gunnar, M. R., Reid, B. M., Donzella, B., Miller, Z. R., Gardow, S., Tsakonas, N. C., Thomas, K. M., DeJoseph, M., & Bendezu, J. J. (2021). Validation of an online version of the Trier Social Stress Test in a study of adolescents. *Psychoneuroendocrinology*, *125*, 105111. https://doi.org/10.1016/j.psyneuen.2020.105111
54. Breiner, M. J., Stritzke, W. G., & Lang, A. R. (1999). Approaching avoidance. A step essential to the understanding of craving. *Alcohol Research & Health : The Journal of the National Institute on Alcohol Abuse and Alcoholism*, *23*(3), 197–206. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6760377/
55. Feldstein Ewing, S. W., Tapert, S. F., & Molina, B. S. G. (2016). Uniting adolescent neuroimaging and treatment research: Recommendations in pursuit of improved integration. *Neuroscience & Biobehavioral Reviews*, *62*, 109–114. https://doi.org/10.1016/j.neubiorev.2015.12.011
56. Stritzke, W. G., McEvoy, P. M., Wheat, L. R., Dyer, K. R., & French, D. J. (2007). The yin and yang of indulgence and restraint: The ambivalence model of craving. Motivation of health behavior, 31-47.
57. Rodrigues, R., Erzsebet Mehesz, Lingford-Hughes, A., & Martina Di Simplicio. (2023). Approach-avoidance biases to self-harm cues in young people with self-harm. *Journal of Affective Disorders*. https://doi.org/10.1016/j.jad.2023.08.021
58. Naherkiak, B., Bhaskaran, J., Sareen, J., Wang, Y., & Bolton, J. M. (2019). Ambivalence About Living and the Risk for Future Suicide Attempts. *The Primary Care Companion for CNS Disorders*, *21*(2). https://doi.org/10.4088/pcc.18m02361
59. Sandel et al., 2020 (?)

medRxiv preprint doi: https://doi.org/10.1101/2023.12.16.23300091; this version posted December 18, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

60. Bresin, K., & Gordon, K. H. (2013). Endogenous opioids and nonsuicidal self-injury: A mechanism of affect regulation. *Neuroscience & Biobehavioral Reviews*, *37*(3), 374–383. https://doi.org/10.1016/j.neubiorev.2013.01.020

61. H. Johnston, J., E.J. Linden, D., & B.M. van den Bree, M. (2016). Combining Stress and Dopamine Based Models of Addiction: Towards a Psycho-Neuro-Endocrinological Theory of Addiction. *Current Drug Abuse Reviews*, *9*(1), 61–74. https://doi.org/10.2174/1874473709666151209113913

62. Hogarth, L. (2020). Addiction is driven by excessive goal-directed drug choice under negative affect: translational critique of habit and compulsion theory. *Neuropsychopharmacology*, *45*(5), 720–735. https://doi.org/10.1038/s41386-020-0600-8

63. Berghorst, L. H., Kumar, P., Greve, D. N., Deckersbach, T., Ongur, D., Dutra, S. J., & Pizzagalli, D. A. (2016). Stress and reward processing in bipolar disorder: a functional magnetic resonance imaging study. *Bipolar Disorders*, *18*(7), 602–611. https://doi.org/10.1111/bdi.12444

64. Kumar, P., Slavich, G. M., Berghorst, L. H., Treadway, M. T., Brooks, N. H., Dutra, S. J., Greve, D. N., O'Donovan A., Bleil, M. E., Maninger, N., & Pizzagalli, D. A. (2015). Perceived life stress exposure modulates reward-related medial prefrontal cortex responses to acute stress in depression. *Journal of Affective Disorders*, *180*, 104–111. https://doi.org/10.1016/j.jad.2015.03.035

65. Stringaris, A., Vidal-Ribas Belil, P., Artiges, E., Lemaitre, H., Gollier-Briant, F., Wolke, S., Vulser, H., Miranda, R., Penttilä, J., Struve, M., Fadai, T., Kappel, V., Grimmer, Y., Goodman, R., Poustka, L., Conrod, P., Cattrell, A., Banaschewski, T., Bokde, A. L. W., & Bromberg, U. (2015). The Brain's Response to Reward Anticipation and Depression in Adolescence: Dimensionality, Specificity, and Longitudinal Predictions in a Community-

medRxiv preprint doi: https://doi.org/10.1101/2023.12.16.23300091; this version posted December 18, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Based Sample. *American Journal of Psychiatry*, *172*(12), 1215–1223. https://doi.org/10.1176/appi.ajp.2015.14101298

66. Glenn, C. R., & Klonsky, E. D. (2010). The Role of seeing blood in non-suicidal self-injury. *Journal of Clinical Psychology*, n/a-n/a. https://doi.org/10.1002/jclp.20661
67. Yim, I. S., Quas, J. A., Cahill, L., & Hayakawa, C. M. (2010). Children's and adults' salivary cortisol responses to an identical psychosocial laboratory stressor. *Psychoneuroendocrinology*, *35*(2), 241–248. https://doi.org/10.1016/j.psyneuen.2009.06.014
68. Saunders, J. B., Aasland, O. G., Babor, T. F., de la Fuente, J. R., & Grant, M. (1993). Development of the Alcohol Use Disorders Identification Test (AUDIT): WHO Collaborative Project on Early Detection of Persons with Harmful Alcohol Consumption--II. *Addiction (Abingdon, England)*, *88*(6), 791–804. https://doi.org/10.1111/j.1360-0443.1993.tb02093.x
69. Adamson, S. J., Kay-Lambkin, F. J., Baker, A. L., Lewin, T. J., Thornton, L., Kelly, B. J., & Sellman, J. D. (2010). An improved brief measure of cannabis misuse: The Cannabis Use Disorders Identification Test-Revised (CUDIT-R)□. *Drug and Alcohol Dependence*, *110*(1-2), 137–143. https://doi.org/10.1016/j.drugalcdep.2010.02.017
70. Nelson, H. E., & Wilson, J. R. (1991). National Adult Reading Test (NART). 2nd Edn Windsor: NFER-Nelson.
71. Hales, S., Blackwell, S. E., Di Simplicio, M., Iyadurai, L., Young, K., & Holmes, E. A. (2015). *Imagery-Based Cognitive-Behavioral Assessment* (G. P. Brown & D. A. Clark, Eds.). PubMed; Guilford Press. https://pubmed.ncbi.nlm.nih.gov/26334081/
72. Delmonte, S., Balsters, J. H., McGrath, J., Fitzgerald, J., Brennan, S., Fagan, A. J., & Gallagher, L. (2012). Social and monetary reward processing in autism spectrum disorders. *Molecular autism, 3*(1), 1-13.

medRxiv preprint doi: https://doi.org/10.1101/2023.12.16.23300091; this version posted December 18, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

73. Dichter, G. S., Richey, J. A., Rittenberg, A. M., Sabatino, A., & Bodfish, J. W. (2012). Reward circuitry function in autism during face anticipation and outcomes. *Journal of autism and developmental disorders, 42*(2), 147-160.