Psychological Research
https://doi.org/10.1007/s00426-022-01689-5

ORIGINAL ARTICLE

Evaluating individual differences in rewarded Stroop performance: reliability and associations with self-report measures

Brent Pitchford[1] · **Karen M. Arnell**[1]

Received: 1 December 2021 / Accepted: 15 May 2022

Abstract

In three separate experiments, we examined the reliability of and relationships between self-report measures and behavioral response time measures of reward sensitivity. Using a rewarded-Stroop task we showed that reward-associated, but task-irrelevant, information interfered with task performance (MIRA) in all three experiments, but individual differences in MIRA were unreliable both within-session and over a period of approximately 4 weeks, providing clear evidence that it is not a good individual differences measure. In contrast, when the task-relevant information was rewarded, individual differences in performance benefits were remarkably reliable, even when examining performance one year later, and with a different version of a rewarded Stroop task. Despite the high reliability of the behavioral measure of reward responsiveness, behavioral reward responsiveness was not associated with self-reported reward responsiveness scores using validated questionnaires but was associated with greater self-reported self-control. Results are discussed in terms of what is actually being measured in the rewarded Stroop task.

Introduction

Only a small fraction of the vast amount of perceptual information available to us at any moment can be brought into awareness (Desimone & Duncan, 1995). What we attend to can be influenced by our goals and motivations (Kiss et al., 2009), the physical salience of the stimuli in our environment (Theeuwes, 1992, 2010), and reward history (Anderson et al., 2011; Hickey & Peelen, 2015; Hickey et al., 2010; Kim & Anderson, 2019; Raymond & O'Brien, 2009). Reward-associated information can often be involuntarily attended to (Anderson et al., 2011, 2020; Bachman et al., 2020; Hickey & Peelen, 2015; Hickey et al., 2010; Krebs et al., 2010; Raymond & O'Brien, 2009). If the reward-associated information is tied to, and therefore shifts, attention to a task-relevant dimension, then this can improve and facilitate behavioral performance (Engelmann et al., 2009; Kiss et al., 2009; Krebs et al., 2010, 2011; Raymond & O'Brien, 2009). In contrast, if the reward-associated information shifts attention to a task-irrelevant dimension, impaired performance may result (Anderson et al., 2011; Bachman et al., 2020; Hickey & Peelen, 2015; Hickey et al., 2010; Krebs et al., 2010; Wolff et al., 2016). Further, the stimulus-reward association that is processed in early sensory areas can influence higher-order cognitive functions like cognitive control (Dreisbach & Fischer, 2012; Goschke & Bolte, 2014; Krebs et al., 2011; Locke & Braver, 2008; Steenbergen et al., 2009).

One way to examine how reward associations can influence cognitive control is by using a rewarded version of the color Stroop task (Krebs et al., 2010, 2011, 2013; Stroop, 1935). In this paradigm, participants are asked to indicate the font color (task-relevant dimension) of a word while ignoring the word meaning (task-irrelevant dimension). Participants are told that fast and accurate responses to specific font colors will be rewarded whereas other font colors will not (see Fig. 1).

Behavioral performance during this task is impaired when the word meaning is different than the font color (e.g., the word RED in blue font) and this impairment is referred to as Stroop interference (Stroop, 1935). However, performance is especially impaired if the word meaning matches a color that is linked to a potential reward, suggesting that the reward-associated value of the word meaning can influence conflict processing of mismatching information in the two dimensions. This additional increase in response times

---

✉ Karen M. Arnell
karnell@brocku.ca

[1] Department of Psychology, Brock University, 500 Glenridge Ave., Saint Catharines, ON L2S 3A1, Canada

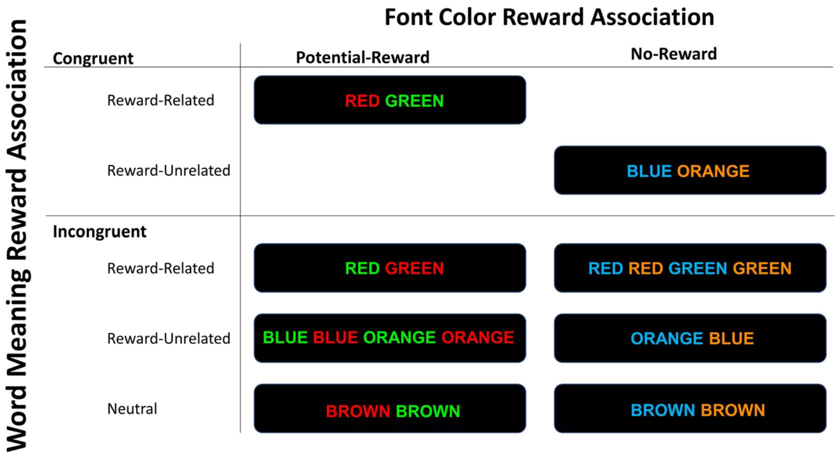

Fig. 1 Example stimuli for eight conditions of rewarded Stroop task. In this example, fast and accurate responses to red and green font colors would be rewarded (color figure online)

(RTs) when the word is associated with the reward color compared to when it is not, is referred to as the modulation of interference by reward associations (MIRA; Wolff et al., 2016). In contrast, faster RTs are observed when the font color is associated with reward than when it is not associated with reward—a finding referred to as reward responsiveness. Thus, there are three measures that can be observed in this paradigm: (1) Stroop interference that reflects the cognitive control elicited when the information in the goal-relevant dimension (the font color) does not match the goal-irrelevant information (the word meaning), (2) MIRA that represents the modulation of interference by reward associations of task-irrelevant meanings of words, and (3) reward responsiveness toward the increased value-driven saliency of font colors that are potentially rewarded. As such, MIRA appears to reflect the implicit and automatic attention capture of attention to rewarded information, whereas reward responsiveness may reflect the explicit increase in motivation to gather rewards on potential reward trials. Indeed, individual differences in these three measures have been previously associated with activation in distinct areas of the brain (Carter et al., 2000; Gruber et al., 2002; Krebs et al., 2011). Generally, activation in the Anterior Cingulate Cortex (ACC), or subdivisions of the ACC, has been associated with individual differences in Stroop interference (Carter et al., 2000; Gruber et al., 2002). The nucleus accumbens (NAcc) has been associated with individual differences in reward responsiveness, while the magnitude of interference by reward associations on conflict processing has been associated with activation in the pre-supplementary motor area (pre-SMA; Krebs et al., 2011). Despite distinct regions predicting individual differences in Stroop interference, reward responsiveness, and the interference by reward associations during conflict processing, there is also an abundance of evidence suggesting that similar brain areas are associated with both reward learning and cognitive control and these processes might be more interrelated than expected (Ridderinkhof et al., 2004a, b).

Individual variation in how reward modulates conflict processing

Individuals' ability to complete goal-directed conflict processing in laboratory tasks like the classic color Stroop task has often been related to their trait self-control (Wolff et al., 2016). Self-control is defined as "the aspect of inhibitory control that involves resisting temptations and not acting impulsively or prematurely" (Diamond, 2013, p. 137). It has previously been linked to harmful and maladaptive behaviors and disorders such as addiction (Bühringer et al., 2008; Hare et al., 2009; Moffitt et al., 2011; Tangney et al., 2004). Cognitive control is thought to play a role in self-control in that prepotent, but goal-irrelevant, impulses must be overridden to achieve successful self-control. However, the relationship between individual differences in cognitive control performance within the lab and trait self-control as assessed by self-report questionnaires is small, if not non-existent (Duckworth & Kern, 2011; Saunders et al., 2018; Wennerhold & Friese, 2020). This could be due to the low reliabilities of cognitive control measures or because of other issues that are inherent in correlating behavioral and self-report measures such as the reliability paradox whereby low between-participant variability in task performance limits the extent to which established cognitive tasks can be related to self-report measures (Dang et al., 2020; Saunders et al., 2018). Another possible explanation as to why performance on laboratory measures of cognitive control does not predict self-reported trait self-control is because these measures evaluate the ability to ignore overlearned or prepotent responses that are not associated with any reward value, whereas self-control in everyday life often includes ignoring higher reward-associated impulses (Hofmann et al., 2012). For example, refraining from eating high caloric food might be qualitatively different from ignoring an overlearned response to the semantic meaning of a word in that there is a more pleasurable association with the doughnut than with the overlearned and easy response (Hare et al., 2009;

Wennerhold & Friese, 2020). Greater self-control might rely on the ability to ignore the reward-associated but goal-irrelevant impulse, and laboratory tasks that include value-associated distracting information within conflict processing might be more ecologically valid at mimicking how self-control takes place in everyday life. Ignoring impulses that are both overlearned and value-associated may better predict individual differences in self-control (Eigsti et al., 2006; Wolff et al., 2016).

To explore this question, Wolff and colleagues (2016) included a number of Stroop task to examine how reward associations could modulate conflict processing, and whether behavioral performance would relate to individuals' trait self-control and impulsivity. In this task, a specific number of digits was reward-associated (e.g., 2 digits), whereas the value of the numbers themselves was supposed to be ignored. For example, when presented with '222', the participants would have to indicate that there were 3 digits and ignore the reward-associated value of the numbers. Individual differences in reward responsiveness (i.e., faster responses to potential reward trials versus no-reward trials) and Stroop interference did not predict individual differences in self-control or impulsivity, but the modulation of interference by reward associations (MIRA) did predict individual differences in self-control and impulsivity; individuals reporting less self-control performed worse when the information in the to-be-ignored dimension was reward associated compared to when it was not. These findings suggest that variability in how reward associations influence cognitive control are better at predicting individual differences in self-control than individual variability in either cognitive control or reward responsiveness alone.

Wolff et al.'s (2016) finding that MIRA, but not reward responsiveness, relates to self-reported self-control and impulsivity suggests that MIRA and reward responsiveness may represent independent reward-related processes. This is also suggested by Krebs et al.'s (2011) finding that the activation in the NAcc was associated with individual differences in reward responsiveness, but not MIRA, while activation in the pre-SMA was associated with MIRA, but not reward responsiveness. It is currently unclear whether the null relationship between self-control and behavioral reward responsiveness observed by Wolff et al. (2016) was due to potentially inadequate psychometric properties of the behavioral measure of reward responsiveness in the rewarded-Stroop task. The relationship between individual differences in reward responsiveness and self-control might not have been found if the behavioral measure of reward responsiveness in the rewarded Stroop task was an unreliable measure of individual differences in reward responsiveness. Indeed, reward responsiveness, MIRA and Stroop interference are all difference score measures, and the reliabilities of difference scores can often be particularly poor given that they are constrained by the reliabilities of the two measures used to create the difference (Draheimet et al., 2019). Therefore, it is particularly important to examine the reliability of difference score measures before using them as individual difference measures.

If measures like MIRA and behavioral reward responsiveness are being correlated with other measures as in Wolff et al. (2016), the reliability of these behavioral measures should be tested so their suitability as individual difference measures can be established (e.g., Dale & Arnell, 2013; Dang et al., 2020). Once higher reliabilities of the behavioral measures are confirmed, we can sufficiently examine and evaluate correlations between these measures and with previously validated self-report measures. In this study, we will examine reliability both within- and between-session to determine if the reward responsiveness and MIRA measures are reliable individual difference measures. We will also examine relationships between these behavioral measures, and relationships with self-report measures of self-control, impulsivity, and reward responsiveness. Evidence for the independence of MIRA and reward responsiveness could be strengthened if (1) there was a null relationship found between these behavioral measures, and (2) self-report measures of reward sensitivity were found to relate to either the behavioral measure of reward responsiveness in the rewarded-Stroop task, or to MIRA, but not to both. We can also learn more about the processes underlying MIRA and reward responsiveness by determining what constructs share variability with them.

There are self-report measures of reward responsiveness that might reasonably be expected to predict reward responsiveness and/or MIRA in the rewarded-Stroop task. For example, the Value-Driven Attention Questionnaire (VDAQ) was developed to measure the substantial individual differences in attention to reward (Anderson et al., 2020). Individuals' scores on this questionnaire predicted the capture of attention by task-irrelevant and previously rewarded colors in a visual search task such that greater VDAQ predicted more capture by previously-reward-associated, but task-irrelevant, information (Anderson et al., 2020). Anderson et al. (2020) also showed that higher VDAQ scores were associated with greater functional connectivity in brain areas that are implicated in reward learning (NAcc, Ventral Tegmental Area), greater trait levels of impulsivity, greater viewing of areas in images associated with higher reward probability, as well as with scores on a well-validated self-report measure of reward responsiveness (Behavioral Activation System; BAS; Carver & White, 1994). Individual differences in BAS have previously been associated with greater attention to reward and are a useful indicator of reward responsiveness and reward learning as well as a useful predictor of other cognitive measures associated with reward learning and motivation (Carver & White, 1994; Hickey & Peelen, 2015;

Hickey et al., 2010; Pitchford & Arnell, 2019; Qi et al., 2013). We might then predict that the behavioral measures of reward from individuals' performance completing a rewarded Stroop task would correlate positively with BAS and VDAQ scores as these questionnaires were ostensibly developed to measure sensitivity and/or attention-to-reward and have previously been associated with other measures of reward responsiveness and attention to reward (Anderson et al., 2020; Carver & White, 1994; Engelmann et al., 2009; Hickey et al., 2010).

The current study

The first goal of the current study was to examine whether self-reported reward responsiveness measures would relate to the behavioral measures of reward responsiveness and MIRA from both the rewarded color Stroop (Experiments 1–2) and number Stroop (Experiment 3) paradigms. To examine this question, participants completed two questionnaires that have been developed to measure self-reported attention to reward: the behavioral activation system scale (BAS; Carver & White, 1994; used here in Experiments 1–3) and the value-driven attention questionnaire (VDAQ; Anderson et al., 2020; used here in Experiment 3). If greater approach motivation to rewarding stimuli underlies MIRA, then MIRA may be positively correlated with both behavioral and self-report measures of reward responsiveness and self-reported attention to reward.

Participants also completed questionnaires to measure their trait self control (SCS; Tangney et al., 2004; Experiments 1–3) and impulsivity (BIS-11; Patton et al., 1995; Experiment 1) to examine the relationship between individual differences in MIRA, reward responsiveness, and self-control/impulsivity (Wolff et al., 2016). The SCS and BIS-11 measure how individuals generally respond to short-term immediate rewards in their everyday lives and their ability to postpone short-term impulses to achieve longer-term goals, with higher SCS and lower BIS-11 scores indicating higher self-control and lower impulsivity in everyday life.

The final goal of the current study was to examine the reliabilities of the self-report measures and the behavioral measures from the rewarded Stroop task both within the same session (split-half reliabilities) and across time (test–retest reliabilities). If MIRA and reward responsiveness in the rewarded Stroop task are good measures of individual differences in response to rewards, then we would expect these reliabilities to be high, or at least higher than reliabilities for measures that have also been used as individual measures of self-control (e.g., Dale & Arnell, 2013; Goodhew & Edwards, 2019). Split-half reliabilities were obtained for MIRA as well as other behavioral measures (reward responsiveness, Stroop interference) and self-report measures in Experiments 1–3 while test–retest reliabilities were obtained in Experiment 2.

Experiment 1

Method

Participants

The participants were 175 Brock University undergraduate students recruited through the online system for participant recruitment. Participants' ages ranged from 17 to 45 ($M = 19.66$, SD = 3.88). The data from 5 participants were removed because they had not fully completed at least one of the questionnaires and the data from one other participant was excluded due to their low accuracy (less than 60%) when completing the no-reward trials, leaving data from 169 participants available for analyses. Participants completed the study in a single in-person session in the lab in exchange for research credit towards a course. Informed consent was obtained from all participants in this experiment. GPower indicated that this sample size provided a power of more than 0.76 to detect relationships of at least $r = 0.20$ (2 -tailed).

Self-report measures

Participants completed the Tangney self-control scale (SCS; Tangney et al., 2004) to measure their trait levels of self-control. The SCS contains 36 items (e.g., "I never let myself lose control") and participants reported the degree to which the statement was true of them using a 5-point Likert scale from 1 (Not at all like me) to 5 (Very Much Like Me).

Participants also completed the Barrett's Impulsivity Scale (BIS-11; Patton et al., 1995) to measure self-reported trait impulsivity. The BIS-11contains 30 items (e.g., "I do things without thinking") and participants reported the degree to which the statement was true of them using a 4-point Likert scale from 1 (Rarely/Never) to 4 (Almost Always/Always). Both the SCS and BIS-11 have been previously associated with desirable and undesirable real-world self-control outcomes (de Ridder et al., 2012).

Participants also completed the Behavioral Inhibition System/Behavioral Activation System (BIS/BAS) scale (Carver & White, 1994). The BIS/BAS scales included 20 items (e.g., I go all out to get what I want") that were designed to measure individuals' dispositional motivational systems and their propensities to approach and avoid experiences in their everyday lives, with higher BIS scores representing greater inhibition of behavior and greater BAS scores representing tendencies of approaching exciting and rewarding experiences. Participants reported the degree to which each statement was an accurate description of their behavior on a scale of 1 (Very Inaccurate) to 4 (Very Accurate). As is typical (e.g., Carver & White, 1994), BAS was further subdivided into three BAS subscales: BAS-Fun Seeking (BAS-F), BAS-Drive (BAS-D) and BAS-Reward Responsiveness (BAS-R), where individuals' BAS-F scores represented the tendency to want to engage in rewarding experiences, BAS-D represented the tendency to pursue appetitive goals and BAS-R represented the tendency for one to feel positive emotions after receiving rewards.

Rewarded color Stroop task

Participants completed a rewarded Stroop task whereby they received rewards whenever they quickly and accurately identified reward-associated font colors. Two of the four potential font colors: blue (0, 0, 255), green (0, 128, 0), orange (255, 140, 0) and red (255, 0, 0), were randomly selected as reward-associated colors at the beginning of the experiment for each participant. For instance, red and green might be randomly selected as the two reward-associated colors (see Fig. 1). Participants were instructed that they would receive 10 points for each trial in which either of the two font colors (red, green) were presented and they responded quickly and accurately. They were told at the beginning of the task that they could choose a candy of their choice at the end of the experiment for every 400 points they had earned, and they walked by the bin with a large assortment of candies on the way to the computer. On each trial, a grey fixation cross (128, 128, 128; visual angle 0.2 °) was presented for 500–1000 ms on a black background (0, 0, 0), followed by a color word (e.g., the word BLUE, visual angle subtending a vertical angle of 0.5 ° and a horizontal visual angle ranging from 1.0 to 2.0°) in either a different font color (green, orange, red; incongruent) or in the same font color (blue; congruent) which stayed on the screen for 600 ms or until response. Neutral trials included the word BROWN since the font color could never be brown. Participants were instructed to press the corresponding color buttons as indicated by colored stickers on the A (blue), S (green), K (orange), L (red) keys on the keyboard to indicate the color of the font while ignoring the semantic meaning of the word. Participants were told to respond as quickly and accurately as possible. If the font color was a potential-reward color and they responded faster on the current trial compared to the slowest 30% of their previous accurate responses during potential-reward color trials, then they were presented with feedback telling them "you've earned 10 points!" on the screen for 1000 ms. They were told their responses were "too slow!" if they were accurate but responded slower than 70% of their previous accurate responses during potential-reward color trials, and no feedback was given following inaccurate responses or during trials with no-reward font colors. After their responses, the ITIs were randomly selected from a range of 500–1000 ms.

Following a brief practice period, there were a total of 480 trials that differed in whether the meaning of the word matched the font color (congruent; 120 trials/incongruent; 360 trials), fully crossed with whether the font color signaled a potential reward or no reward (potential-reward; 240 trials/no-reward; 240 trials). Incongruent trials in both the potential reward and no-reward conditions varied as to whether the word meaning was reward-associated or not (120 trials where the word was reward-related, 120 where the word was non-reward related, and 120 neutral "BROWN" trials. Thus, there were eight conditions with 60 trials in each condition (see Fig. 1 for example stimuli in each of the conditions). Thirty second breaks were offered after every 96 trials with the total points earned so far displayed on the screen (e.g., "You've earned 300 points so far!"). Participants earned an average of 4 (SD = 0.53) small snack size candies and chips that they chose from multiple options.

Data analysis

RTs for accurate responses that were less than 4000 ms and which did not follow inaccurate trials were subjected to a two-pass outlier removal procedure where RTs greater than 3SD away from condition means for each individual were excluded. Mean RTs and accuracies were submitted to repeated-measures ANOVAs to determine the effect of the font color association (potential-reward, no-reward) and congruency (congruent, incongruent). This was to ensure that reward facilitated Stroop performance and the replication of the typical Stroop effect. RTs and accuracies from incongruent trials were then submitted to repeated-measures ANOVAs to determine the effect of the font color association (potential-reward, no-reward) and word meaning association (reward-related, reward-unrelated). We expected to find interference by the reward association of the word meaning (task-irrelevant dimension) on both RTs and accuracies when indicating nonrewarded font colors such that the reward-related word meaning would result in increased RTs and lower accuracies compared to when the word was not reward-related (i.e., we expected to find significant MIRA). Paired-samples t-tests were then conducted to determine the magnitude of MIRA by comparing RTs for incongruent reward-related and reward-unrelated words when the font color was not reward associated (Krebs et al., 2010, 2011, 2013; Wolff et al., 2016).

Self-report measures (BIS-11, SCS, BIS/BAS) were correlated with behavioral measures from the rewarded Strop task (MIRA, Stroop Interference and Reward Responsiveness). Bonferroni corrections were applied for behaviour and self-report relationships to reduce the

Table 1 Means and standard deviations for behavioral measures

		Stroop Interference	Reward Responsiveness	MIRA
Exp. 1				
($n = 169$)	RT Mean (ms)	26**	53**	10**
	RT SD (ms)	22	48	32
Exp. 2				
($n = 223$)	RT Mean (ms)	41**	113**	11*
	RT SD (ms)	56	121	75
Exp. 3				
($n = 94$)	RT Mean (ms)	39**	69**	15**
	RT SD (ms)	35	61	37

*Significantly different from zero, $p < 0.05$

**Significantly different from zero, $p < 0.001$

likelihood of Type I errors. MIRA was calculated as the difference in RTs between the reward-related and reward-unrelated conditions when the word was incongruent and the font color was non-reward-associated (no-reward reward-related minus no-reward reward-unrelated) where higher numbers represented greater interference by the reward-associated word meanings (i.e., the task-irrelevant dimension) when reporting the font color (the task-relevant dimension). Stroop interference was calculated as the difference in RTs to incongruent and congruent trials (incongruent–congruent). Reward responsiveness was calculated as RTs to no-reward trials minus RTs to potential-reward trials, thereby reflecting how much faster they responded to potential-reward trials versus non-reward trials. (see Table 1 for means).

Experiment 1: results

The effect of reward and congruence on Stroop performance

Overall RTs were significantly faster during trials where there was a possibility of earning the reward compared to RTs during trials when there was no chance of earning the reward, $F(1, 168) = 202.65$, $p < 0.001$, $\eta_p^2 = 0.547$. The typical Stroop effect was replicated whereby RTs were slower when the meaning of the word did not match the font color (incongruent) compared to when the meaning and font color matched (congruent), $F(1, 168) = 251.31$, $p < 0.001$, $\eta_p^2 = 0.599$.The effect of faster RTs when the font color signaled potential reward was not modulated by the congruence of the stimuli, $F(1,168) = 2.36$, $p = 0.13$, $\eta_p^2 = 0.01$ (see Fig. 2).

Accuracies were greater for congruent trials relative to incongruent trials, $F(1,168) = 65.97$, $p < 0.001$, $\eta_p^2 = 0.282$. The potential-reward trials were completed more accurately than the no-reward trials, $F(1,168) = 50.20$, $p < 0.001$, $\eta_p^2 = 0.230$, and the effect of reward affected accuracies to a slightly greater extent when the stimuli were incongruent compared to when the stimuli were congruent, $F(1, 168) = 4.24$, $p = 0.04$, $\eta_p^2 = 0.025$ (see Fig. 3). Overall, these results show replication of the Stroop effect and suggest that participants successfully learned which colors were reward associated

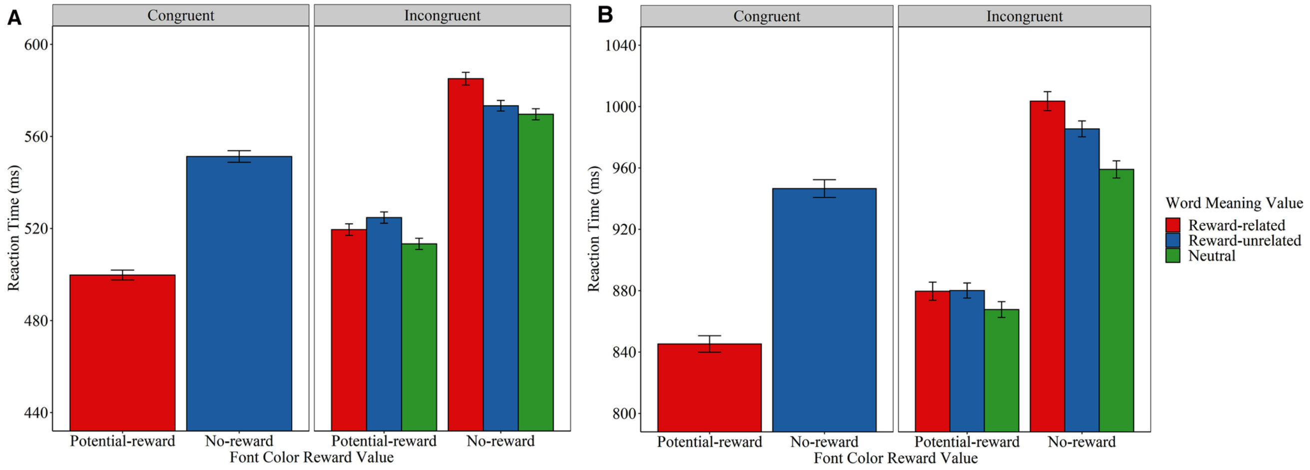

Fig. 2 Average RTs for Experiment 1 (**A**) and Experiment 2 (**B**) by the font color reward value and word meaning value of color Stroop stimuli. Error bars ± 1 within-subjects SEM

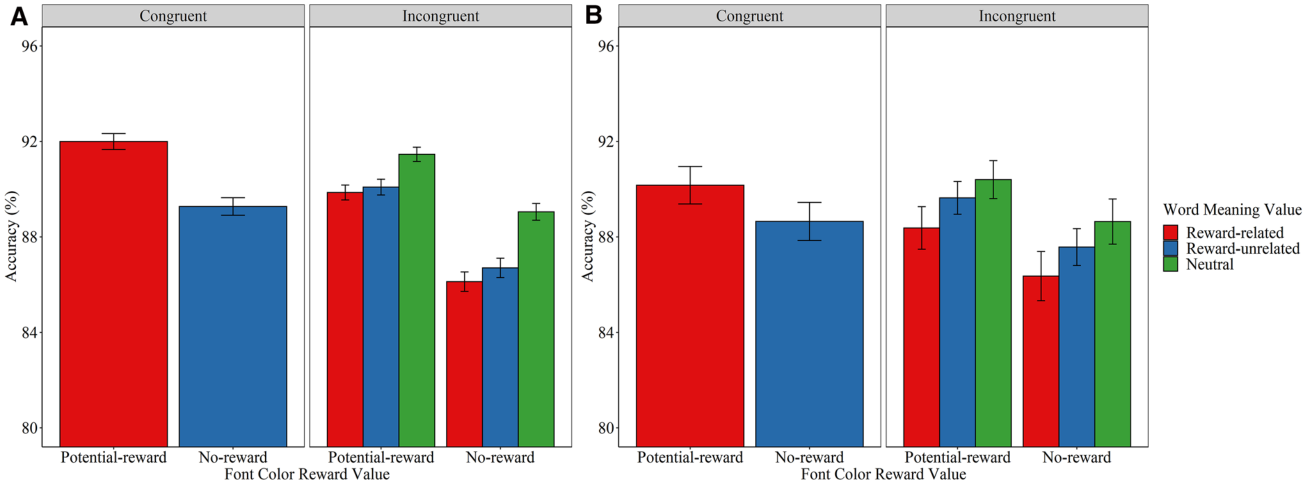

Fig. 3 Average accuracies for Experiment 1 (**A**) and Experiment 2 (**B**) by the font reward value and word meaning value of color Stroop stimuli. Error bars ± 1 within-subjects SEM

and responded faster and more accurately when words were presented in those colors.

The effect of reward-associated word meanings on Stroop performance

When examining incongruent trials only, RTs were faster for potential-reward trials compared to no-reward trials, $F(1, 168) = 170.87$, $p < 0.001$, $\eta_p^2 = 0.504$. Overall RTs were greater when the word meaning was reward-related compared to non reward related, $F(1, 168) = 7.35$, $p < 0.01$, $\eta_p^2 = 0.042$, but this effect was larger in the no-reward relative to potential reward trials as seen in Fig. 2, $F(1, 168) = 14.75$, $p < 0.001$, $\eta_p^2 = 0.081$. There was a significant MIRA effect (Krebs et al., 2010, 2011; Wolff et al., 2016) in that the RTs during no-reward trials were significantly greater when the word meaning was reward related versus non-reward related, $t(168) = 3.93$, $p < 0.001$, $d = 0.302$ (see Fig. 2), suggesting that the reward-associated word meaning interfered with the ability to indicate the correct font color.

When examining incongruent trials only, accuracies were greater for potential-reward trials compared to no-reward trials, $F(1,168) = 57.04$, $p < 0.001$, $\eta_p^2 = 0.253$. Overall accuracies were reduced when the word meaning was reward related compared to non reward related, $F(1, 168) = 4.97$, $p = 0.027$, $\eta_p^2 = 0.0287$, and this was not modulated by the font reward value, $p > 0.99$ (see Fig. 3).

Associations between measures

BAS subscales, r's > 0.28, and BIS-11 subscales, r's > 0.38, were all significantly correlated within each scale, p's < 0.001. Impulsivity (BIS-11 total) was negatively associated with self-control, $r = -0.76$, $p < 0.001$ and positively associated with approach motivation (BAS total), $r = 0.23$, $p = 0.002$. Self-control was negatively associated with higher approach motivation, $r = -0.23$, $p = 0.002$.

MIRA was significantly associated with the behavioral measure of reward responsiveness from the Stroop task, $r = 0.19$, $p = 0.02$. One possible explanation for the significant relationship is that average RTs that are included in the calculation of MIRA are also included in the calculation of reward responsiveness. When excluding the trials in these conditions, this relationship between reward responsiveness and MIRA was no longer significant, $r = 0.09$, $p = 0.25$. Thus, the significant relationship was at least partially due to the same RTs used in the calculation of both measures. MIRA was also significantly associated with Stroop interference, $r = 0.24$, $p < 0.01$. Reward responsiveness was significantly negatively associated with Stroop interference, $r = -0.18$, $p = 0.02$.

Correlations between self-report and behavioral measures are presented in Table 2. BAS total and BAS subscales did not relate to any of the behavioral measures. Neither MIRA nor the behavioral measure of reward responsiveness were related to any self-report measures, r's ranging from − 0.04 to 0.06, p's > 0.22. Greater impulsivity was associated with greater Stroop interference, $r = 0.16$, $p < 0.005$.

One possible explanation for the lack of relationships between our self-report and behavioral measures might be the low reliability of the behavioral measures. To examine split-half reliabilities, trials were first sorted based on condition (e.g., condition of potential reward font value, no reward association word meaning, incongruent), and then split into two sets by alternating trials within each condition. Correlations were then conducted between behavioral measures for each set of trials to determine split-half reliabilities. In general, the reward responsiveness measure was

Table 2 Correlations between self-report and behavioral measures, with reliabilities

		Scale	Stroop Interference	Reward Responsiveness	MIRA
Exp. 1					
(*n* = 169)		BAS-T	0.09	– 0.05	– 0.04
		SCS	– 0.09	0.04	– 0.04
		BIS-11	0.16^{*A}	0.04	0.06
Reliabilities	Split-Half		0.38**	0.88**	0.27**
Exp. 2					
(*n* = 223)		BAS-T	– 0.01	– 0.03	– 0.08
		SCS	– 0.02	0.19**	0.11
Reliabilities	Split-Half		0.52**	0.88**	0.16*
	Test- Retest		0.50**	0.66**	0.06
Exp. 1 &2					
(*n* = 392)		BAS-T	0.03	– 0.04	– 0.07
		SCS	– 0.06	0.11	0.06
Exp. 3					
(*n* = 94)		BAS-T	0.07	0.04	0.07
		SCS	– 0.35**	0.31**	– 0.17
		VDAQ	0.16	– 0.02	0.09
Reliabilities	Split-Half		0.78**	0.93**	0.26*

SCS self-control scale, *BAS* behavioral activation system scale, *BAS-T* total BAS score, *VDAQ* value-driven attention questionnaire, *BIS* Barrett's impulsivity scale, *MIRA* modulation of interference by reward associations

Bonferroni corrections were applied to behavior-self-report relationship to reduce type I error likelihood

P values for Exp. 1 and Exp. 3 were compared to a corrected alpha of 0.005, and *p* values for Exp. 2 were compared to a corrected alpha of 0.008

*$p < 0.05$, $^{*A}p < 0.005$, **$p < 0.001$

highly reliable, $r = 0.88$, $p < 0.001$, followed by the lower reliability for the Stroop interference measure, $r = 0.38$, $p < 0.001$, and MIRA measure, $r = 0.27$, $p < 0.001$.[1] The lower reliability for Stroop interference is not surprising and has been discussed previously (Hedge et al., 2017), but the low reliability for the MIRA measure might contribute to the null relationships found here.

Discussion

The results from Experiment 1 replicate previous findings that reward can facilitate RTs and accuracy during a Stroop task when it is attached to the task-relevant dimension (Krebs et al., 2010, 2011, 2013; Wolff et al., 2016). Responses were faster and more accurate when there was the potential to gain reward compared to when there was not. In contrast, reward-associated, but task-irrelevant, information impaired Stroop performance when the task-relevant information was not reward associated, replicating the MIRA effect (Krebs et al., 2010, 2011, 2013). Individual differences in the impairment of Stroop performance by the reward-associated task-irrelevant information (MIRA) had relatively poor split-half reliability, whereas there was very high split-half reliability for individual differences in reward responsiveness, suggesting that although MIRA is a robust finding, MIRA might not be a good measure for examining individual differences. Despite the potentially problematic nature of different measures in reliably capturing individual differences (Draheim et al., 2019), the reliabilities of reward responsiveness were remarkably high, while reliabilities for Stroop interference were moderate and in line with reliabilities reported in previous research (Enkavi et al., 2019;

---

[1] Only two of the conditions are included in the calculation of MIRA and the smaller number of trials per participant might have led to artificially reduced reliability for MIRA relative to the reward responsiveness measure that contains many more trials. To test this, separate correlations were conducted with a modified reward responsiveness measure that only included average RTs from two conditions: RTs to no-reward trials where the Stroop condition was incongruent, and the word was not associated with reward minus RTs to potential-reward trials where the Stroop condition was incongruent, and the word was not associated with reward. This allowed for the same number of trials to be included in the reward responsiveness and MIRA measures. In all experiments, reward responsiveness reliabilities remained remarkably high and consistent despite the reduced number of trials. Therefore, the lower reliability for the MIRA measure relative to the reward responsiveness measure is not simply a result of the fewer number of trials per participant.

Hedge et al., 2017). The difference score nature of MIRA is unlikely to explain the poor reliability of the measure.[2] The poor reliability of MIRA could explain why there were no significant relationships between the self-report measures in this experiment (BAS, SCS, BIS-11) and MIRA, as good reliabilities for behavioral measures are essential when considering them to be candidates for measuring individual differences. To examine this further, we included measures of test–retest reliability in Experiment 2 to get a better sense of how reliable these behavioral measures are in capturing individual differences.

Experiment 2

Method

Participants

Two hundred and eighty-four participants were recruited through the Amazon Mechanical Turk (MTURK) platform. The experiment was programmed in HTML and JavaScript with the help of templates provided by JsPsych (de Leeuw, 2015), which is a JavaScript library for running behavioral experiments in a web browser. The experiment was hosted on an Amazon web server for participants to access in their web browsers. Of the 284 participants, 223 (126 male) were included for further analyses because they obtained accuracies above chance in all of the 8 conditions and passed the included attention checks (e.g., "Please select Very Inaccurate for this question"), and this percentage of data kept (approximately 79%) is fairly standard when recruiting participants through MTURK (Kross et al., 2014). GPower indicated that this sample size provided a power of more than 0.91 to detect relationships of at least $r = 0.20$ (2 -tailed). The ages of the participants ranged from 18 to 53 ($M = 32$, SD = 6). Participants were told they would earn $3.50 USD for completing the experiment plus a bonus based on their behavioral performance during the task, and they were compensated an average of 6.57 USD total (SD = 0.54, range = 4.76–7.94) for participating. Participants took approximately an hour to complete the experiment ($M = 55$ min, SD = 23, median = 48 min, range = 33–175 min).

After 1 month, we invited the 223 participants to complete the experiment for a second time and stopped collecting data when we reached an n of 60. The subsample of participants was assigned the same rewarded colors as when they first completed the experiment. These participants were further compensated an average of 6.93 USD (SD = 0.47, range = 5.88–7.92) over and above the earnings they received during the first time they completed the experiment.

Self-report measures

Participants completed both the SCS and BIS/BAS questionnaires as described in Experiment 1 with the exception that there was one attention check included in each questionnaire. To ensure they were actively participating and reading the questions, they were asked to select a specific response for one of the questions in each questionnaire (e.g., "Very Inaccurate").

Rewarded color Stroop task

The rewarded Stroop task participants completed was similar to the one in Experiment 1 with some exceptions. Participants were first asked to compare one of their credit cards to a picture of a credit card on the screen and to zoom in or out in their browser until the credit card matched the size of the picture to help calibrate the screen. Two of the four potential font colors were rewarded, blue (0, 0, 255), green (0, 128, 0), yellow (255, 255, 0) and red (255, 0, 0), but rewarded colors could either be blue and red ($n = 111$) or yellow and green ($n = 112$), and these options were chosen so that the keys they would need to press for rewarded colors were never beside each other (e.g., A and K or S and L). During potential-reward trials, participants were presented the following feedback on the screen: "Congratulations! You win 2 cents!", if their response was correct and faster than 70% of their previous correct responses when shown words in reward-associated font colors. On each trial, a grey fixation cross (visual angle 0.7°) was shown for 500 ms on a white background followed by the Stroop word (visual angle subtending a vertical angle of 0.5° and a horizontal visual angle ranging from 1.4 to 3.7°) that was presented until response. Following their response, they were then shown feedback if appropriate, followed by a 2s ITI where a diagram showing which key should be pressed for each color was presented. Participants completed a practice session where they were provided with ample feedback to help them learn the task.

Data analysis

The data analysis procedures in Experiment 2 were similar to Experiment 1 with the exception that BIS-11 was not included, and a subset of the sample completed the

---

[2] One alternative method to calculating difference scores is the residualized measure approach. To calculate the residualized MIRA measure, the same conditions were included where the average RTs when the word was reward-unrelated was regressed on average RTs when the word was reward-related, and the standardized residuals were saved for further analyses. Correlations between original and residualized MIRA measures were very high (r's > .87), and reliabilities for the residualized MIRA measure remained poor in all experiments.

experiment twice which allowed us to examine test–retest reliabilities of the self-report and behavioral measures. We expected the self-report measures to be reliable over time given that their test–retest reliabilities have been previously reported elsewhere (Stanford et al., 2009; Tangney et al., 2004). The behavioral measures should have acceptable test–retest reliability to consider their use in evaluating individual differences within current and future psychological research (Goodhew & Edwards, 2019).

Experiment 2: results

The effect of reward and congruence on Stroop performance

Overall RTs were significantly faster during trials where there was a possibility of earning the reward compared to RTs during trials when there was no chance of earning the reward, $F(1{,}222) = 193.33$, $p < 0.001$, $\eta_p^2 = 0.465$. The typical Stroop effect was replicated whereby RTs were slower for incongruent trials compared to congruent trials, $F(1, 222) = 119.91$, $p < 0.001$, $\eta_p^2 = 0.351$. RTs were faster when the font color signaled potential reward and this effect was not modulated by the congruence, $F(1{,}222) = 0.52$, $p = 0.47$, $\eta_p^2 = 0.002$.

Accuracies were greater for congruent trials relative to incongruent trials, $F(1, 222) = 24.99$, $p < 0.001$, $\eta_p^2 = 0.10$. The potential-reward trials were completed more accurately than the no-reward trials, $F(1, 222) = 53.77$, $p < 0.001$, $\eta_p^2 = 0.195$, but the effect of reward affected accuracies to a slightly greater extent on incongruent trials compared to congruent trials, $F(1, 222) = 4.13$, $p = 0.04$, $\eta_p^2 = 0.018$. Overall, the results from these ANOVAs suggested that participants successfully learned which colors were reward associated and responded faster and more accurately when viewing those colors.

The effect of reward-associated word meanings on Stroop performance

When only incongruent trials were examined, RTs were faster for potential-reward compared to no-reward trials, $F(1, 222) = 184.38$, $p < 0.001$, $\eta_p^2 = 0.454$. Overall RTs were similar when the word meaning was reward-related compared to non reward related, $F(1, 222) = 0.17$, $p = 0.68$, $\eta_p^2 < 0.001$. As seen in Fig. 3, there was a significant interaction of the word meaning value and font color reward value whereby RTs were slower when the word meaning value was reward related only when the font color reward value was not reward related, $F(1, 222) = 7.09$, $p < 0.01$, $\eta_p^2 = 0.031$. There was a significant MIRA effect in that the RTs during no-reward trials were significantly greater when the word meaning was reward related versus non reward related, $t(222) = 2.26$, $p = 0.02$, $d = 0.153$, suggesting that the reward-associated word meaning interfered with the ability to quickly indicate the correct font color.

Accuracies were greater for potential-reward compared to no-reward trials, $F(1, 222) = 56.89$, $p < 0.001$, $\eta_p^2 = 0.204$. Accuracies were similar when the word meaning was reward related compared to non reward related, $F\ (1, 222) = 2.29$, $p = 0.13$, $\eta_p^2 = 0.010$, and this was seen for both the potential reward and no reward font color reward associations, $p > 0.47$ (see Fig. 3).

Associations between measures

BAS subscales were significantly positively intercorrelated, r's > 0.25, p's < 0.001. Approach motivation (BAS total) was also negatively associated with self-control, $r = -0.29$, $p < 0.001$.

MIRA was significantly positively associated with the behavioral measure of reward responsiveness, $r = 0.20$, $p < 0.01$. When excluding RTs in conditions included in the calculation of both MIRA and reward responsiveness, the relationship between reward responsiveness and MIRA remained significant, $r = 0.21$, $p = 0.002$. Stroop interference was not related to MIRA, $r = 0.06$, $p = 0.36$, or reward responsiveness, $r = -0.07$, $p = 0.30$.

Correlations between self-report and behavioral measures are presented in Table 2. Similar to Experiment 1, BAS was not associated with any of the behavioral measures. Greater self-control was associated with greater behavioral reward responsiveness, $r = 0.19$, $p < 0.001$.

Reliabilities of self-report and behavioral measures

In general, there were high split-half reliabilities for the reward responsiveness measure, $r = 0.88$, $p < 0.001$, and lower but still acceptable split-half reliability for the Stroop interference measure, $r = 0.52$, $p < 0.001$. The split-half reliability for MIRA was low, $r = 0.16$, $p = 0.02$. Test–retest reliabilities for the behavioral measures mirrored the pattern for the split-half reliabilities but were slightly lower overall; the highest test–retest reliability was found for the reward responsiveness measure, $r = 0.66$, $p < 0.001$, followed by the Stroop interference measure, $r = 0.50$, $p < 0.001$, followed by the MIRA measure, $r = 0.06$, $p = 0.64$.

All of the self-report measures were highly reliable over time with the potential exception of the BAS-R subscale (see Table 3).

Analysis using data from both experiments

To examine the associations between self-report and behavioral measures with greater power, data were examined

Table 3 Correlations between self-report measures, and test–retest reliabilities

	Scale	1	2	3	4	5	6
Exp. 1–3 (*n* = 486)	1. BAS-F	–					
	2. BAS-D	0.40**	–				
	3. BAS-R	0.27**	0.42**	–			
	4. BAS-T	0.75**	0.80**	0.73**	–		
	5. SCS	– 0.40**	– 0.18**	0.02	–0.26**	–	
Exp. 1 only (*n* = 169)	6. BIS-11	0.30**	0.12	– 0.09	0.23**	-0.76**	–
Exp. 3 only (*n* = 94)	7. VDAQ	0.54**	0.50**	0.51**	0.65**	– 0.52**	
Test–Retest (*n* = 60)		0.74**	0.79**	0.42**	0.82**	0.68**	

across both Experiments 1 and 2 with a total N of 392 participants.

MIRA was significantly associated with reward responsiveness, $r = 0.20$, $p < 0.001$, and this relationship remained when excluding RTs from conditions included in the calculation of both measures, $r = 0.20$, $p < 0.001$. Stroop interference was not significantly related to MIRA, $r = 0.08$, $p = 0.11$, or reward responsiveness, $r = -0.03$, $p = 0.58$.

After correcting for multiple corrections, there were no significant behavioural and self-report relationships, p's > 0.005.

Discussion

The results from Experiment 2 support the conclusions from Experiment 1 in that there was facilitation of RT and accuracies when the reward was tied to the task-relevant dimension but impaired performance when it was tied to the task-irrelevant dimension. Reliabilities for MIRA were poor, especially the extremely low test–retest reliability for MIRA suggesting that, although it is a robust phenomenon, MIRA is likely not a good individual differences measure.

Greater reported self-control was significantly associated with higher reward responsiveness when completing the rewarded Stroop task in Experiment 2. However, this relationship was not observed when data were combined across Experiments 1 and 2.

We did not find any support for a relationship between greater trait self-control and MIRA, as might be expected of the poor reliability of the MIRA measure. However, as discussed above, previous work by Wolff et al. (2016) showed a robust positive relationship between self-reported self-control and MIRA when participants used a number Stroop paradigm instead of a color Stroop paradigm. The final experiment included a number Stroop paradigm similar to the one used by Wolff and colleagues instead of the color-word Stroop to test whether this was responsible for the different pattern of results. Further, in Experiment 3 the new VDAQ was included as an additional measure of reward responsiveness as it has been associated with behavioral measures of attention to reward and reward learning in other reward paradigms (Anderson et al., 2020).

Experiment 3

Method

Participants

Initially, fifty-eight participants for Experiment 3 were recruited from the pool of participants that had completed Experiment 2 through the Amazon Mechanical Turk (MTURK) platform. Because a higher sample was necessary to examine correlations between behavioral and self-report measures, an additional fifty-four participants who did not complete Experiment 2 were then recruited through MTURK. The experiment was programmed in HTML and JavaScript with the help of templates provided by JsPsych and then the experiment was hosted on an Amazon web server for participants to access in their web browsers. Of the 112 participants, 94 were included for further analyses because they obtained accuracies above chance in all of the 8 conditions and passed the included attention checks (e.g., "Please select Very Inaccurate for this question"). Participants were compensated an average of 7.05 USD total (SD = 0.44, range = 2.2) for participating and took approximately an hour to complete the experiment ($M = 51$ min, SD = 1.76, median = 46.49 min, range = 35–156 min).

Self-report measures

Participants completed both the SCS and BIS/BAS questionnaires as described in Experiment 2. Participants also completed the VDAQ (Anderson et al., 2020) which contains 16 statements such as "When tasty food is placed in the open,

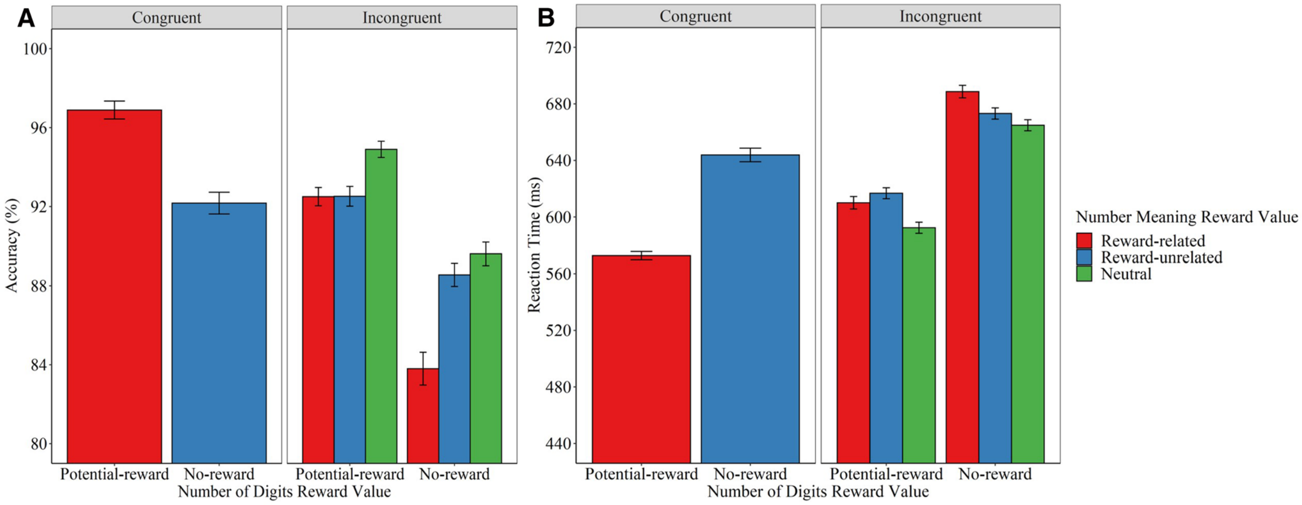

Fig. 4 Average RTs (**A**) and accuracies (**B**) in Experiment 3 by the number of digits reward value and number meaning reward value of number Stroop stimuli. Error bars ± 1 within-subjects SEM

I find it very tempting". Participants reported the degree to which the statement is true of them on a 4-point scale ranging from "the opposite of me" to "very true of me". Scores were summed across all items to create a total VDAQ score.

Rewarded number Stroop task

The rewarded Number Stroop task that participants completed was similar in structure to the rewarded color Stroop task in Experiment 2 with the exception that the stimuli were numbers (e.g., 222) where participants were asked to ignore the identity of the numbers and count the number of digits, indicating whether there were 1 (H key), 2 (J key), 3 (K key) or 4 (L key) digits. Two of the four potential number of digits were rewarded for fast (defined as in Experiment 1 and 2) and accurate responses on the counting task, and these options were chosen so that the keys they would need to press for rewarded number of digits were never beside each other (e.g., A and K or S and L).

Data analysis

The data analysis procedures in Experiment 3 were similar to Experiment 2.

Experiment 3: results

The effect of reward and congruence on Stroop performance

A repeated-measures ANOVA was conducted to examine the effect of the number of digits reward association (potential-reward, no-reward) and congruence of the stimuli (congruent, incongruent) on RTs (see Fig. 4). Overall RTs were significantly faster during trials where the number of digits signaled potential reward compared to RTs during trials when there was no chance of earning the reward, $F(1, 93) = 120.85$, $p < 0.001$, $\eta_p^2 = 0.570$. RTs were slower for incongruent trials compared to congruent trials, $F(1, 93) = 115.32$, $p < 0.001$, $\eta_p^2 = 0.554$. The faster RTs for potential reward trials was not modulated by the congruence, $F(1, 93) = 0.63$, $p = 0.43$, $\eta_p^2 < 0.01$.

A repeated-measures ANOVA was then conducted to examine the effects of reward association (potential-reward, no-reward) and congruence (congruent, incongruent) on accuracies. Accuracies were greater for congruent trials relative to incongruent trials, $F(1, 93) = 106.08$, $p < 0.001$, $\eta_p^2 = 0.533$. The potential-reward trials were completed more accurately than the no-reward trials, $F(1, 93) = 48.76$, $p < 0.001$, $\eta_p^2 = 0.344$, but the effect of reward affected accuracies to a slightly greater extent on incongruent trials compared to congruent trials, $F(1, 93) = 4.92$, $p = 0.029$, $\eta_p^2 = 0.050$. Overall, the results suggested that participants successfully learned which number of digits were reward associated and responded faster and more accurately.

The effect of reward-associated number values on Stroop performance

When examining incongruent trials only, RTs were faster and accuracies were greater for potential-reward compared to no-reward trials, $F(1, 93) = 93.29$, $p < 0.001$, $\eta_p^2 = 0.501$; $F(1, 93) = 49.38$, $p < 0.001$, $\eta_p^2 = 0.347$. Overall, RTs were similar when the number value was reward related compared to non reward related, $F(1, 93) = 3.56$, $p = 0.062$, $\eta_p^2 = 0.037$, but accuracies were lower when there was a reward-associated number value versus a non-reward associated number value, $F(1, 93) = 25.57$, $p < 0.001$, $\eta_p^2 = 0.216$. As seen in Fig. 4, there was a significant interaction of the task-irrelevant reward association and task-relevant reward association whereby RTs were slower when the number value was reward related only when the number of digits was not reward related, $F(1, 93) = 20.56$, $p < 0.001$, $\eta_p^2 = 0.181$, and for accuracies, $F(1, 93) = 29.28$, $p < 0.001$, $\eta_p^2 = 0.239$. There was a significant MIRA effect in that the RTs during no-reward trials were significantly greater when the number value was reward related versus non-reward related, $t(93) = 4.08$, $p < 0.001$, $d = 0.390$, suggesting that the reward-associated but task-irrelevant information interfered with the ability to quickly indicate the number of digits.

Associations between measures

All BAS subscales were significantly positively intercorrelated, r's > 0.26, p's < 0.02. BAS-F and BAS-D subscales, and BAS-T, were negatively associated with trait self-control, r's ranging from − 0.23 to − 0.44, p's < 0.03, but the relationship between self-control and BAS-R was found to be nonsignificant, $r = -0.01$, $p = 0.92$. VDAQ was positively associated with BAS-T, $r = 0.65$, $p < 0.001$, and all BAS subscales, r's > 0.50, p's < 0.001, and negatively associated with self-control, $r = -0.52$, $p < 0.001$.

MIRA was significantly positively associated with Stroop interference, $r = 0.35$, $p < 0.001$, but not reward responsiveness, $r = -0.13$, $p = 0.22$. When excluding RTs in conditions included in the calculation of both MIRA and reward responsiveness, the relationship between reward responsiveness and MIRA was significant, $r = -0.25$, $p = 0.02$. Note that this relationship is in the opposite direction to the significant relationship reported between reward responsiveness and MIRA in Experiment 2. Reward responsiveness was significantly negatively associated with Stroop interference, $r = -0.36$, $p < 0.001$.

Correlations between self-report and behavioral measures are presented in Table 2. BAS was not associated with any of the behavioral measures from the rewarded number Stroop task. Again, greater self-control was associated with greater behavioral reward responsiveness, $r = 0.31$, $p = 0.003$. Greater self-control was also now associated with reduced Stroop interference, $r = -0.35$, $p < 0.001$.

Consistent with the previous two experiments that included the rewarded color Stroop task, the reliabilities of the same measures with the rewarded number Stroop task followed the same pattern; the reward responsiveness measure was highly reliable, $r = 0.93$, $p < 0.001$, followed by the lower reliability for the Stroop interference measure, $r = 0.78$, $p < 0.001$, and MIRA measure, $r = 0.26$, $p = 0.01$.

Associations of color and number Stroop behavioral measures

Of the 94 participants that completed this experiment, 56 had also participated in Experiment 2 approximately one year prior to completing Experiment 3. We were able to examine whether individual differences in color Stroop performance could predict individual differences in number Stroop performance a year later. Reward responsiveness during the color Stroop task was a good predictor of reward responsiveness during the number Stroop task one year later, $r = 0.44$, $p < 0.001$. Individual differences in MIRA during the color Stroop task were not predictive of individual differences in MIRA during the number Stroop task, $r = 0.07$, $p = 0.60$. Stroop interference from the color Stroop task also did not predict Stroop interference from the number Stroop task one year later, $r = 0.08$, $p = 0.55$. These results suggest that reward responsiveness was a reliable measure across time and across the two variations of the rewarded Stroop task included in this study.

General discussion

The purpose of the current study was to examine behavioral measures of Stroop interference, reward responsiveness and MIRA from the rewarded Stroop paradigm to determine whether (1) these measures were reliable both within the same session (split-half reliabilities) and across time (test–retest reliabilities) and (2) whether the behavioral measures of reward responsiveness and MIRA would relate to two self-report measures of reward responsiveness—BAS and the VDAQ.

Reliability of the measures

The results presented here suggest that individual differences in behavioral reward responsiveness (measured as the difference in RTs on trials with and without potential reward) were highly reliable both within-session and across one month's time. The high reliability was observed both during in-lab testing when candy was used as the reward and during online testing when monetary bonus payment was used as the reward, and for both the color Stroop task in Experiments 1 and 2, and the number Stroop task in Experiment 3. Indeed, for the subset of participants who completed both Experiments 2 and 3, reward responsiveness was even reliable across the color and number versions of the Stroop task one full year later. The faster RTs on potential reward trials were not the result of a speed-accuracy trade-off as participants were also more accurate on potential reward trials relative to trials without a potential reward. These results provide strong evidence that the improved performance when a reward is associated with the task-relevant dimension of the stimulus is a reliable individual difference measure that is suitable for examining relationships with other individual difference measures.

Individual differences in Stroop interference (measured as incongruent minus congruent trial RTs) were less reliable compared to individual differences in behavioral reward responsiveness but were still adequate. This moderate reliability was found within-session and across sessions spanning approximately one month. However, Stroop interference did not generalize across tasks as the interference observed in the color Stroop task was not significantly related to the interference observed in the number Stroop task performed one year later. The modest reliability found here for Stroop interference is consistent with previous work showing that the Stroop effect is remarkably robust when examining group means but is not particularly suitable as an individual difference measure (Enkavi et al., 2019; Hedge et al., 2017).

In contrast, individual differences in MIRA showed low within-session reliability and were unreliable across time, both with the identical Stroop task or different tasks, suggesting that MIRA is not a suitable measure for examining individual differences in reward behavior. One possibility was that the low reliability of MIRA was due to the smaller number of trials included when computing MIRA compared to the number of trials used to calculate reward responsiveness. However, calculating reward responsiveness with the same number of trials that were included when calculating MIRA did not significantly influence the reliability of reward responsiveness, suggesting that the relative unreliability of MIRA was not due to the smaller number of trials alone. Although MIRA is a difference score measure, and difference scores are particularly prone to low reliabilities (e.g., Draheim et al., 2019), this is also unlikely to be the explanation for the low MIRA reliability here. First, Stroop interference and the behavioral measure of reward responsiveness are also both difference score measures that included some of the same RT conditions, yet these were found to be moderately (Stroop) or highly (reward responsiveness) reliable in the present study. Second, when residuals were used instead of difference scores, MIRA reliability was still low, and Stroop and reward responsiveness reliabilities were still high. Third, the RTs that make up the MIRA difference score both had test–retest reliabilities above 0.80 suggesting that they did not constrain the reliability of the MIRA difference scores which approximated 0 across sessions.

Although individual differences in MIRA were unreliable across time, the support for the MIRA phenomenon occurring at the group level was substantial whereby impaired performance was consistently demonstrated when the task-irrelevant dimension included reward-associated information and the task-relevant dimension did not (Krebs et al., 2010, 2011, 2013). The consistent replication of this effect despite differences in the samples recruited for the in-lab and online experiments, the task being completed online for two of the experiments and within the lab for the other, and differences in the type of rewards given (i.e., money, candy) suggests that MIRA is a robust and replicable effect and that the rewarded Stroop paradigm included in this study was able to elicit MIRA. Reward can have a detrimental effect on cognitive performance when it is tied to task-irrelevant information that is supposed to be ignored (Anderson et al., 2011; Bachman et al., 2020; Hickey & Peelen, 2015; Hickey et al., 2010; Krebs et al., 2010; Wolff et al., 2016). However, the robustness of an effect does not require that individual differences in that effect are reliable (Goodhew & Edwards, 2019; Hedge et al., 2017). MIRA appears to be an example of a robust effect that is not a reliable individual differences measure. Indeed, it is possible that trial-to-trial variation is a hallmark of MIRA such that font naming impairment with a reward-related color word may be greater immediately after a trial where the color of the word was just rewarded at the font level. Perhaps trial-to-trial variation can explain the unreliability of MIRA within the same session. In an exploratory analysis, we examined whether the impairment of font color naming in Experiments 1 and 2 combined was influenced by the characteristics of the previous trial (font color reward association, word meaning reward association, Stroop congruence). Potential reward trials were followed by faster responses on the subsequent trials, perhaps due to increased motivation elicited by the potential reward. Contrary to the hypothesis above, font naming impairment with a reward-related color word was greater immediately after a trial where the color of the word was not rewarded at the font level. Despite the influence of the previous trial on MIRA overall, split-half reliability of MIRA remained unchanged when excluding post-non-rewarded trials or when excluding post-rewarded trials, suggesting that the influence of N-1 trial font color reward association did not explain the low reliability of MIRA reported here.

Relationships with self-report measures

Reward responsiveness

Individual differences in behavioral reward responsiveness were highly reliable in comparison to individual differences in Stroop interference and MIRA. Despite this finding, there was no evidence to suggest that individual differences in behavioral reward responsiveness were associated with individual differences in self-reported reward responsiveness (i.e., BAS), even though BAS measures showed high test–retest reliability themselves, or with VDAQ scores even though VDAQ scores have been shown to predict behavioral measures of attention to reward in other paradigms (Anderson et al., 2020). Further, individual differences in self-reported trait self-control were negatively associated with self-reported reward responsiveness but positively associated with behavioral reward responsiveness. These findings suggest that the behavioral reward responsiveness measure cannot be used interchangeably with self-report measures of reward responsiveness, even though they are thought to measure similar constructs. One possible explanation is that the self-report measures examining individuals' subjective evaluations of their own reward responsiveness in everyday life might differ from variability in performance as measured by RTs and accuracies (Dang et al., 2020). However, individual differences in self-reported reward responsiveness have been previously associated with behavioral measures of reward responsiveness while completing different cognitive tasks (Anderson et al., 2020; Carver & White, 1994). Anderson et al. (2020) showed that the VDAQ scores were significantly and positively related to attention capture by previously rewarded items in a visual search task and to the preference to fixate a previously rewarded quadrant of the screen while viewing scenes. These results are consistent with Anderson et al.'s conceptualization of the VDAQ score as representing a largely implicit attraction to items potentially associated with reward. Indeed, the majority of VDAQ questions ask about the ability to ignore potentially rewarding stimuli in a variety of contexts (incoming text message, background TV, tempting food, billboards). Thus, the lack of relationship observed here between the VDAQ and the behavioral measure of reward responsiveness in the Stroop task may result from the VDAQ measuring the automatic capturing power of stimuli associated with reward, even when they are task irrelevant, in contrast to the reward responsiveness measure from the Stroop task which may measure the explicit seeking of existing rewards. However, the BAS-D and BAS-R scales do appear to measure explicit reward-seeking behavior. The BAS scale includes questions like "When I want something, I usually go all-out to get it." and have an absence of attention capture questions of the sort seen in the VDAQ. BAS-D and BAS-R scores have also been shown to predict increases in positive mood in response to reward cues (Carver & White, 1994). However, BAS scales have also been shown to be moderately positively correlated with VDAQ scores (Anderson et al., 2020) showing that even if they measure different aspects of sensitivity to reward (i.e., implicit capture versus explicit seeking), these aspects are related. Interestingly, the behavioral reward responsiveness measure from the Stroop tasks was found to be positively related to self-control in both Experiment 2 and Experiment 3, but not the combined analyses from Experiments 1 and 2, suggesting that although it is not related to BAS or VDAQ scores, this measure of reward responsiveness is related to self-control abilities. Indeed, such self-control abilities may be essential to stay on-task and discriminate potential reward from non-reward trials when they are mixed within blocks. Overall, it appears that further validation of this behavioral measure of reward responsiveness in the rewarded Stroop task is necessary to determine what aspect(s) of reward responsiveness are being measured, and whether it is possible that reward responsiveness in the context of the rewarded Stroop paradigm is similar to reward responsiveness in other contexts such as a less demanding reward-training paradigm (e.g., responding to rewarded versus non-rewarded colored shapes without to-be-ignored information).

Another key goal of this study was to examine whether behavioral and self-report measures of reward responsiveness would associate with individual differences in MIRA. The independence of reward responsiveness and MIRA has been implied but not explicitly examined in previous research (Krebs et al., 2011; Wolff et al., 2016). We did not find evidence suggesting that individual differences in self-reported reward responsiveness using the BAS scales or the VDAQ were related to individual differences in MIRA. Initially, we might have expected to find a positive relationship between MIRA and VDAQ scores in particular as MIRA appears to represent the capture of attention by rewarding stimuli, which is exactly what the VDAQ purports to measure. However, no relationship between MIRA and VDAQ would be expected given the poor reliability of MIRA and evidence that it does not make a good trait measure.

Within the Stroop task, reward responsiveness and MIRA were positively related in Experiment 2, negatively related in Experiment 3, and not related in Experiment 1. The lack of consistency when comparing the reward responsiveness—MIRA relationships in the three experiments suggests that there is no relationship. Individuals who were highly responsive to reward did not show greater impairment to their Stroop performance by the reward associations relative to individuals who were less responsive to reward. The perceived value of the reward does not explain the extent to which the reward-associated but task-irrelevant information impairs conflict processing. It is presently unclear whether MIRA is even specific to reward sensitivity per se or whether it might be part of a more general arousal-based emotional Stroop effect such as in the taboo Stroop paradigm where color naming RTs are slowed when the task-irrelevant word is associated with non-rewarded but highly arousing information (e.g., MacKay et al., 2004).

It has previously been suggested that individuals with lower self-control are more influenced by reward-associated but task-irrelevant information as based on a significant positive correlation between self-reported self-control and MIRA (Wolff et al., 2016). This relationship was not replicated here in any of the three studies and even with our large sample size in the combined Experiment 1 and 2 analyses where the N approximated 400. The failure to replicate was found both for the rewarded color Stroop task used by Krebs and colleagues

(2011, 2013) and a number Stroop task like that used by Wolff et al. (2016), so the nature of the Stroop task cannot be an explanation for the different patterns of results. Indeed, the null relationship observed here between trait measures and MIRA would be expected given that MIRA showed no evidence of being a reliable measure within or across sessions and a measure cannot be expected to relate to other measures if it does not relate to itself. In fact, compared to MIRA, higher reliabilities were found for the Stroop interference, which has elicited criticism in the past for being an unreliable measure (Duckworth & Kern, 2011; Wolff et al., 2016). Further, in two of three experiments, there were correlations found between impulsivity/self-control and Stroop interference such that lesser self-control was associated with greater Stroop interference, although previous work has found evidence suggesting there is no relationship between Stroop interference and trait self-control (Saunders et al., 2018). This implies that the correlation reported by Wolff et al. (2016) may be spurious, especially given that it was based on a small sample size of approximately 40 participants and relationships between self-report and behavioral measures are generally much lower than the relationship observed by Wolff et al. (e.g., Dang et al., 2020). Indeed, large spurious correlations are more likely to result from small sample sizes (Hedge et al., 2017; Schönbrodt & Perugini, 2013).

The self-report measures of self-control/impulsivity (SCS and BIS-11) and reward responsiveness (BAS) have been previously validated and were found to be highly reliable over time in this study. Further, we found that higher self-control was associated with lower self-reported reward responsiveness, which has also been reported by others (e.g., Hamilton et al., 2014; O'Gorman & Baxter, 2002). In contrast, the behavioral RT measure of reward responsiveness was positively related to self-control. In the rewarded Stroop task, self-control may aid in keeping participants on-task to discriminate reward from non-reward trials, but in the everyday world, high sensitivity to reward may be associated with self-control failures (e.g., I really found the doughnut tempting so I ate it; Jimura et al., 2013). No relationships were observed between MIRA and self-control. Although MIRA seemed like a good candidate for examining how reward can influence cognitive processing for individuals based on their levels of trait self-control, the unreliability of MIRA should direct our attention to other behavioral measures that are more reliable over time to better explore the relationship between individual differences in reward processing and trait self-control.

Conclusion

Reward associations that were tied to task-irrelevant information impaired Stroop performance (MIRA). This was consistently found across three experiments and two different versions of rewarded Stroop. However, the extent to which reward associations could impair Stroop performance for individuals was not reliable within or across sessions and was therefore not associated with individual differences in self-reported reward responsiveness or trait self-control. More reliable individual differences measures than MIRA could be used to better understand the link between attention to reward and self-control. How much faster individuals responded when there was a potential reward versus no reward (i.e., behavioral reward responsiveness) was very reliable over time, but was not associated with validated and reliable self-report measures of reward responsiveness like the VDAQ and BAS. In contrast, self-reported self-control and self-reported reward responsiveness were strongly related whereby individuals with higher self-control reported less reward responsiveness. This work suggests that future work is necessary to understand the behavioral measures of reward responsiveness and MIRA in the rewarded Stroop paradigm.

Acknowledgements The work was supported by a Canadian Graduate Scholarship from the Natural Sciences and Engineering Research Council of Canada (NSERC) to the first author, and by a grant from NSERC to the second author. We thank Carly Lundale, Daniella Zambito and Lauren Kremble for their assistance with data collection.

Funding The Natural Sciences and Engineering Research Council (NSERC) provided Discovery grant funding to KA and scholarship funding to BP.

Declarations

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

Ethical approval All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

Informed consent Informed consent was obtained from all individual participants included in the study.

References

Anderson, B. A., Kim, H., Britton, M. K., & Kim, A. J. (2020). Measuring attention to reward as an individual trait: The value-driven attention questionnaire (VDAQ). *Psychological Research Psychologische Forschung, 84*(8), 2122–2137. https://doi.org/10.1007/s00426-019-01212-3.

Anderson, B. A., Laurent, P. A., & Yantis, S. (2011). Value-driven attentional capture. *Proceedings of the National Academy of Sciences, 108*(25), 10367–10371. https://doi.org/10.1073/pnas.1104047108.

Bachman, M. D., Wang, L., Gamble, M. L., & Woldorff, M. G. (2020). Physical salience and value-driven salience operate through different neural mechanisms to enhance attentional selection. *The*

Journal of Neuroscience, 40(28), 5455–5464. https://doi.org/10.1523/JNEUROSCI.1198-19.2020.

Bühringer, G., Wittchen, H.-U., Gottlebe, K., Kufeld, C., & Goschke, T. (2008). Why people change? The role of cognitive-control processes in the onset and cessation of substance abuse disorders. *International Journal of Methods in Psychiatric Research, 17*(S1), S4–S15. https://doi.org/10.1002/mpr.246.

Carter, C. S., Macdonald, A. M., Botvinick, M., Ross, L. L., Stenger, V. A., Noll, D., & Cohen, J. D. (2000). Parsing executive processes: Strategic vs. evaluative functions of the anterior cingulate cortex. *Proceedings of the National Academy of Sciences, 97*(4), 1944–1948. https://doi.org/10.1073/pnas.97.4.1944.

Carver, C. S., & White, T. L. (1994). Behavioral inhibition, behavioral activation, and affective responses to impending reward and punishment: The BIS/BAS Scales. *Journal of Personality and Social Psychology, 67*(2), 319–333. https://doi.org/10.1037/0022-3514.67.2.319.

Dale, G., & Arnell, K. M. (2013). Investigating the stability of and relationships among global/local processing measures. *Attention, Perception, & Psychophysics, 75*, 394–406. https://doi.org/10.3758/s13414-012-0416-7.

Dang, J., King, K. M., & Inzlicht, M. (2020). Why are self-report and behavioral measures weakly correlated? *Trends in Cognitive Sciences, 24*(4), 267–269. https://doi.org/10.1016/j.tics.2020.01.007.

de Leeuw, J. R. (2015). jsPsych: A JavaScript library for creating behavioral experiments in a Web browser. *Behavior Research Methods, 47*(1), 1–12. https://doi.org/10.3758/s13428-014-0458-y.

de Ridder, D. T., Lensvelt-Mulders, G., Finkenauer, C., Stok, F. M., & Baumeister, R. F. (2012). Taking stock of self-control: A meta-analysis of how trait self-control relates to a wide range of behaviors. *Personality and Social Psychology Review, 16*(1), 76–99. https://doi.org/10.1177/1088868311418749.

Desimone, R., & Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annual Review of Neuroscience, 18*(1), 193–222. https://doi.org/10.1146/annurev.ne.18.030195.001205.

Diamond, A. (2013). Executive functions. *Annual Review of Psychology, 64*(1), 135–168. https://doi.org/10.1146/annurev-psych-113011-143750.

Draheim, C., Mashburn, C. A., Martin, J. D., & Engle, R. W. (2019). Reaction time in differential and developmental research: A review and commentary on the problems and alternatives. *Psychological Bulletin, 145*(5), 508. https://doi.org/10.1037/BUL0000192.

Dreisbach, G., & Fischer, R. (2012). The role of affect and reward in the conflict-triggered adjustment of cognitive control. *Frontiers in Human Neuroscience, 6*, 342–342. https://doi.org/10.3389/fnhum.2012.00342.

Duckworth, A. L., & Kern, M. L. (2011). A meta-analysis of the convergent validity of self-control measures. *Journal of Research in Personality, 45*(3), 259–268. https://doi.org/10.1016/j.jrp.2011.02.004.A.

Eigsti, I. M., Zayas, V., Mischel, W., Shoda, Y., Ayduk, O., Dadlani, M. B., & Casey, B. J. (2006). Predicting cognitive control from preschool to late adolescence and young adulthood. *Psychological Science, 17*(6), 478–484. https://doi.org/10.1111/J.1467-9280.2006.01732.X.

Engelmann, J. B., Damaraju, E., Padmala, S., & Pessoa, L. (2009). Combined effects of attention and motivation on visual task performance: Transient and sustained motivational effects. *Frontiers in Human Neuroscience, 3*, 4. https://doi.org/10.3389/neuro.09.004.2009.

Enkavi, A. Z., Eisenberg, I. W., Bissett, P. G., Mazza, G. L., MacKinnon, D. P., Marsch, L. A., & Poldrack, R. A. (2019). Large-scale analysis of test–retest reliabilities of self-regulation measures. *Proceedings of the National Academy of Sciences, 116*(12), 5472–5477. https://doi.org/10.1073/pnas.1818430116.

Goodhew, S. C., & Edwards, M. (2019). Translating experimental paradigms into individual-differences research: Contributions, challenges, and practical recommendations. *Consciousness and Cognition, 69*, 14–25. https://doi.org/10.1016/j.concog.2019.01.008.

Goschke, T., & Bolte, A. (2014). Emotional modulation of control dilemmas: The role of positive affect, reward, and dopamine in cognitive stability and flexibility. *Neuropsychologia, 62*, 403–423. https://doi.org/10.1016/j.neuropsychologia.2014.07.015.

Gruber, S. A., Rogowska, J., Holcomb, P., Soraci, S., & Yurgelun-Todd, D. (2002). Stroop performance in normal control subjects: An fMRI study. *NeuroImage, 16*(2), 349–360. https://doi.org/10.1006/nimg.2002.1089.

Hamilton, K. R., Sinha, R., & Potenza, M. N. (2014). Self-reported impulsivity, but not behavioral approach or inhibition, mediates the relationship between stress and self-control. *Addictive Behaviors, 39*(11), 1557–1564. https://doi.org/10.1016/j.addbeh.2014.01.003.

Hare, T. A., Camerer, C. F., & Rangel, A. (2009). Self-control in decision-making involves modulation of the vmPFC valuation system. *Science, 324*, 646–648. https://doi.org/10.1126/science.1168450.

Hedge, C., Powell, G., & Sumner, P. (2017). The reliability paradox: Why robust cognitive tasks do not produce reliable individual differences. *Behavior Research Methods, 50*, 1166–1186. https://doi.org/10.3758/s13428-017-0935-1.

Hickey, C., Chelazzi, L., & Theeuwes, J. (2010). Reward guides vision when it's your thing: Trait reward-seeking in reward-mediated visual priming. *PLoS ONE, 5*(11), e14087–e14087. https://doi.org/10.1371/journal.pone.0014087.

Hickey, C., & Peelen, M. V. (2015). Neural mechanisms of incentive salience in naturalistic human vision. *Neuron, 85*(3), 512–518. https://doi.org/10.1016/j.neuron.2014.12.049.

Hofmann, W., Vohs, K. D., & Baumeister, R. F. (2012). What people desire, feel conflicted about, and try to resist in everyday life. *Psychological Science, 23*(6), 582–588. https://doi.org/10.1177/0956797612437426.

Jimura, K., Chushak, M. S., & Braver, T. S. (2013). Impulsivity and self-control during intertemporal decision making linked to the neural dynamics of reward value representation. *Journal of Neuroscience, 33*(1), 344–357. https://doi.org/10.1523/JNEUROSCI.0919-12.2013.

Kim, H., & Anderson, B. A. (2019). Dissociable neural mechanisms underlie value-driven and selection-driven attentional capture. *Brain Research, 1708*, 109–115. https://doi.org/10.1016/j.brainres.2018.11.026.

Kiss, M., Driver, J., & Eimer, M. (2009). Reward priority of visual target singletons modulates event-related potential signatures of attentional selection. *Psychological Science, 20*(2), 245–251. https://doi.org/10.1111/j.1467-9280.2009.02281.x.

Krebs, R. M., Boehler, C. N., Appelbaum, L. G., & Woldorff, M. G. (2013). Reward associations reduce behavioral interference by changing the temporal dynamics of conflict processing. *PLoS ONE*. https://doi.org/10.1371/journal.pone.0053894.

Krebs, R. M., Boehler, C. N., Egner, T., & Woldorff, M. G. (2011). The neural underpinnings of how reward associations can both guide and misguide attention. *The Journal of Neuroscience, 31*(26), 9752–9759. https://doi.org/10.1523/JNEUROSCI.0732-11.2011.

Krebs, R. M., Boehler, C. N., & Woldorff, M. G. (2010). The influence of reward associations on conflict processing in the Stroop task. *Cognition, 117*(3), 341–347. https://doi.org/10.1016/j.cognition.2010.08.018.

Kross, E., Bruehlman-Senecal, E., Park, J., Burson, A., Dougherty, A., Shablack, H., Bremner, R., Moser, J., & Ayduk, O. (2014). Self-talk as a regulatory mechanism: How you do it matters. *Journal of Personality and Social Psychology, 106*(2), 304–324. https://doi.org/10.1037/a0035173.

Locke, H. S., & Braver, T. S. (2008). Motivational influences on cognitive control: Behavior, brain activation, and individual differences. *Cognitive, Affective, & Behavioral Neuroscience, 8*(1), 99–112. https://doi.org/10.3758/CABN.8.1.99.

MacKay, D. G., Shafto, M., Taylor, J. K., Marian, D. E., Abrams, L., & Dyer, J. R. (2004). Relations between emotion, memory, and attention: Evidence from taboo Stroop, lexical decision, and immediate memory tasks. *Memory & Cognition, 32*(3), 474–488. https://doi.org/10.3758/BF03195840.

Moffitt, T. E., Arseneault, L., Belsky, D., Dickson, N., Hancox, R. J., Harrington, H., Houts, R., Poulton, R., Roberts, B. W., Ross, S., Sears, M. R., Thomson, W. M., & Caspi, A. (2011). A gradient of childhood self-control predicts health, wealth, and public safety. *Proceedings of the National Academy of Sciences, 108*(7), 2693–2698. https://doi.org/10.1073/pnas.1010076108.

O'Gorman, J. G., & Baxter, E. (2002). Self-control as a personality measure. *Personality and Individual Differences, 32*(3), 533–539. https://doi.org/10.1016/S0191-8869(01)00055-1.

Patton, J. H., Stanford, M. S., & Barratt, E. S. (1995). Factor structure of the barratt impulsiveness scale. *Journal of Clinical Psychology, 51*(6), 768–774. https://doi.org/10.1002/1097-4679(199511)51:63.0.CO;2-1.

Pitchford, B., & Arnell, K. M. (2019). Self-control and its influence on global/local processing: An investigation of the role of frontal alpha asymmetry and dispositional approach tendencies. *Attention, Perception, and Psychophysics, 81*(1), 173–187. https://doi.org/10.3758/s13414-018-1610-z.

Qi, S., Zeng, Q., Ding, C., & Li, H. (2013). Neural correlates of reward-driven attentional capture in visual search. *Brain Research, 1532*, 32–43. https://doi.org/10.1016/j.brainres.2013.07.044.

Raymond, J. E., & O'Brien, J. L. (2009). Selective visual attention and motivation: The consequences of value learning in an attentional blink task. *Psychological Science, 20*(8), 981–988. https://doi.org/10.1111/j.1467-9280.2009.02391.x.

Ridderinkhof, K., Ulldsperger, M., Crone, E. A., & Nieuwenhuis, S. (2004a). The role of the medial frontal cortex in cognitive control. *Science, 306*(5695), 443–447. https://doi.org/10.1126/science.1100301.

Ridderinkhof, K. R., van Den Wildenberg, W. P. M., Segalowitz, S. J., & Carter, C. S. (2004b). Neurocognitive mechanisms of cognitive control: The role of prefrontal cortex in action selection, response inhibition, performance monitoring, and reward-based learning. *Brain and Cognition, 56*(2), 129–140. https://doi.org/10.1016/j.bandc.2004.09.016.

Saunders, B., Milyavskaya, M., Etz, A., Randles, D., & Inzlicht, M. (2018). Reported self-control is not meaningfully associated with inhibition-related executive function: A Bayesian analysis. *Collabra, 4*(1), 9. https://doi.org/10.1525/COLLABRA.134.

Schönbrodt, F. D., & Perugini, M. (2013). At what sample size do correlations stabilize? *Journal of Research in Personality, 47*(5), 609–612. https://doi.org/10.1016/j.jrp.2013.05.009.

Stanford, M. S., Mathias, C. W., Dougherty, D. M., Lake, S. L., Anderson, N. E., & Patton, J. H. (2009). Fifty years of the Barratt Impulsiveness Scale: An update and review. *Personality and Individual Differences, 47*(5), 385–395. https://doi.org/10.1016/j.paid.2009.04.008.

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology, 18*(6), 643–662. https://doi.org/10.1037/h0054651.

Tangney, J. P., Baumeister, R. F., & Boone, A. L. (2004). High self-control predicts good adjustment, less pathology, better grades, and interpersonal success. *Journal of Personality, 72*(2), 271–324. https://doi.org/10.1111/j.0022-3506.2004.00263.x.

Theeuwes, J. (1992). Perceptual selectivity for color and form. *Perception & Psychophysics, 51*(6), 599–606. https://doi.org/10.3758/BF03211656.

Theeuwes, J. (2010). Top-down and bottom-up control of visual selection. *Acta Psychologica, 135*(2), 77–99. https://doi.org/10.1016/J.ACTPSY.2010.02.006.

van Steenbergen, H., Band, G. P., & Hommel, B. (2009). Reward counteracts conflict adaptation: Evidence for a role of affect in executive control. *Psychological Science, 20*(12), 1473–1477. https://doi.org/10.1111/j.1467-9280.2009.02470.x.

Wennerhold, L., & Friese, M. (2020). Why self-report measures of self-control and inhibition tasks do not substantially correlate. *Collabra, 6*(1), 9. https://doi.org/10.1525/collabra.276.

Wolff, M., Krönke, K.-M., & Goschke, T. (2016). Trait self-control is predicted by how reward associations modulate Stroop interference. *Psychological Research Psychologische Forschung, 80*(6), 944–951. https://doi.org/10.1007/s00426-015-0707-4.