Cognition 117 (2010) 341–347

Contents lists available at ScienceDirect

Cognition

journal homepage: www.elsevier.com/locate/COGNIT

Brief article

The influence of reward associations on conflict processing in the Stroop task

Ruth M. Krebs *, Carsten N. Boehler, Marty G. Woldorff

Center for Cognitive Neuroscience, Duke University, Durham, NC 27708, USA

ARTICLE INFO

Article history:
Received 8 April 2010
Revised 26 August 2010
Accepted 30 August 2010

Keywords:
Reward
Conflict
Stroop
Interference

ABSTRACT

Performance in a behavioral task can be facilitated by associating stimulus properties with reward. In contrast, conflicting information is known to impede task performance. Here we investigated how reward associations influence the within-trial processing of conflicting information using a color-naming Stroop task in which a subset of ink colors (task-relevant dimension) was associated with monetary incentives. We found that color-naming performance was enhanced on trials with *potential-reward* versus those without. Moreover, in *potential-reward* trials, typical conflict-induced performance decrements were attenuated if the incongruent word (task-irrelevant dimension) was unrelated to reward. In contrast, incongruent words that were semantically related to reward-predicting ink colors interfered with performance in *potential-reward* trials and even more so in no-reward trials, despite the semantic meaning being entirely task-irrelevant. These observations imply that the prospect of reward enhances the processing of task-relevant stimulus information, whereas incongruent *reward-related* information in a task-irrelevant dimension can impede task performance.

1. Introduction

Reward is known to be an effective motivator of behavior and a driving force for learning (for a review see Wise, 2004). Numerous studies in humans have demonstrated that reward anticipation is associated with performance improvement in diverse behavioral domains, including response speed and accuracy (e.g., Bijleveld, Custers, & Aarts, 2010; Knutson, Adams, Fong, & Hommer, 2001), visual discrimination and visual search (e.g., Engelmann & Pessoa, 2007; Kristjansson, Sigurjonsdottir, & Driver, 2010), cognitive control (e.g., Locke & Braver, 2008), negative priming (e.g., Della Libera & Chelazzi, 2006), and memory processes (e.g., Adcock, Thangavel, Whitfield-Gabrieli, Knutson, & Gabrieli, 2006; Krebs, Schott, Schutze, & Duzel, 2009; Wittmann et al., 2005).

While reward generally exerts enhancing effects on behavior, the presence of conflicting information is known to disrupt performance, as commonly demonstrated by conflict paradigms such as the Stroop color-naming task (Stroop, 1935). In this task, subjects respond to the ink color of a color word (e.g., "RED") while ignoring its semantic meaning. Typically, subjects' performance is facilitated in trials in which the information in the task-relevant (ink color) and task-irrelevant (word meaning) dimensions are congruent, and impeded if they are incongruent (MacLeod, 1991). According to influential parallel distributed processing models of the Stroop effect, information from both input dimensions is conveyed in parallel, and the ultimate response depends on the relative activation of the two pathways (Carter & van Veen, 2007; Cohen, Dunbar, & McClelland, 1990). In the color-naming Stroop task, it has been proposed that automatic reading of the irrelevant word meaning strongly co-activates the

* Corresponding author. Address: Center for Cognitive Neuroscience, Duke University, B203 Levine Science Research Center, Box 90999, Durham, NC 27708, USA. Tel.: +1 919 668 1334; fax: +1 919 681 0815.
E-mail address: ruth.krebs@duke.edu (R.M. Krebs).

0010-0277/$ - see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.cognition.2010.08.018

corresponding pathway in parallel to the processing of the relevant ink color, and, if incongruent, interferes with performance.

More recently, observations that brain regions implicated in human cognitive control are also critically involved in reward-based learning (Miller & Cohen, 2001; Ridderinkhof, Ullsperger, Crone, & Nieuwenhuis, 2004; Ridderinkhof, van den Wildenberg, Segalowitz, & Carter, 2004; Ullsperger & von Cramon, 2003) have given rise to the question of how far processes related to reward and conflict interact (Holroyd & Coles, 2002; Jocham & Ullsperger, 2009). Supporting such an interaction, it has been demonstrated that reward information has the potential to disrupt the behavioral adjustments that are typically observed subsequent to incongruent trials in a flanker task (van Steenbergen, Band, & Hommel, 2009). According to this study, the commonly observed behavioral adjustments (for a review see Egner, 2007) might be counteracted by the receipt of reward, thereby suggesting a shared mechanism (van Steenbergen et al., 2009). However, these observations were limited to sequential effects, and reward was delivered incidentally (i.e., subjects' responses were not instrumental to obtaining rewards) and thus it remains unknown how conflict processing would be modulated if reward is associated with components of the task itself.

We sought to investigate this question by associating reward with two of the four ink colors in a Stroop task. While subjects responded to the ink color, the irrelevant semantic meaning of the word could be congruent, incongruent, or neutral with regard to the ink color. In addition to these typical Stroop-paradigm categories, the irrelevant word could be semantically linked to a color that was either part of the *potential-reward* ink-color subset or not. However, the semantic information was entirely task-irrelevant and thus never associated with obtaining reward.

Based on the notion that cognitive control in concert with attention can differentially emphasize the pathways of potential competing inputs we hypothesized that reward associations in the relevant dimension would further promote effective stimulus processing. Specifically, we predicted general response facilitation and reduced interference in *potential-reward* as compared to no-reward trials. Additionally, we hypothesized that reward associations with an ink color would generalize to its semantic representation (i.e., word meaning). Consequently, incongruent word meanings that are implicitly linked to reward, might cause greater interference by emphasizing the incongruent information.

2. Experiment 1

2.1. Methods

2.1.1. Participants

Twenty healthy right-handed subjects participated (mean age ± SD: 22.5 ± 3.2, 14 female) and gave written informed consent in accordance with the Duke Medical Center Institutional Review Board for human subjects. Subjects were paid a basic amount of $15 plus an average reward bonus of $15.

2.1.2. Paradigm and procedure

Subjects performed a version of the classic color-naming Stroop task in which they responded to the ink color of words while ignoring their semantic meaning. A small gray fixation square (visual angle 0.3°) was maintained in the center of a black screen (Fig. 1A). In each trial a colored capitalized word was presented above fixation for 600 ms, randomly chosen from the following set: "RED", "YELLOW", "BLUE", "GREEN", or "BROWN" (vertical 0.8°, horizontal ranging from 2.1° to 4.6°). The words were separated by a variable stimulus onset asynchrony (SOA) of 1800–2200 ms and were written in one of four ink colors (red, yellow, blue, or green). Subjects were instructed to respond as quickly as possible by pressing the button associated with the current ink color (*Color*; task-relevant dimension) while ignoring the semantic meaning (*Word*; task-irrelevant dimension; Fig. 1B). Responses were given with the index and middle fingers of the left and right hands, with color-button assignments and color-reward associations counterbalanced across subjects. The semantic meaning of a given word could be congruent (e.g., "GREEN" written in green) or incongruent (e.g., "RED" written in green) with regard to the ink color. Furthermore, trials consisting of words with no conflicting response mapping (e.g., "BROWN" written in green) were intermixed to provide a neutral category.

While responses to two of the four possible ink colors were associated with the potential for monetary reward (*potential-reward*), the remaining two colors represented standard Stroop trials (*no-reward*; Fig. 1B). Accordingly, a fast and correct response in *potential-reward* trials resulted in a 10-cent gain, while an incorrect or slow response resulted in a 10-cent penalty. In order to keep all subjects in a similar reward range, the response time-out was adjusted dynamically based on individual performance. This procedure led to an average gain of $2.50 per run for each subject (70:30 gain-to-loss ratio). Following a short training session, subjects completed six experimental 6-min runs, yielding a total of 480 *potential-reward* and 480 *no-reward* trials. During four 20-s breaks within each run, the updated dollar amount was displayed, serving as performance feedback.

The information conveyed by the irrelevant semantic meaning of the word resulted in equally distributed congruency conditions for both *potential-reward* and *no-reward* trials (Fig. 1B): *congruent, incongruent reward-unrelated, incongruent reward-related*, and *neutral*. It should be emphasized that, although the irrelevant incongruent word could be implicitly "*related*" to either the *potential-reward* or no-reward ink-color subset, the monetary incentives were exclusively dependent on the ink-color dimension. This manipulation allowed us to investigate the explicit effects of reward in the relevant dimension (*potential-reward* versus *no-reward*), as well as indirect effects of reward associations that were entirely irrelevant to the task (*incongruent reward-related* versus *reward-unrelated*).

The averaged response times (RT) and error rates were submitted to repeated-measures analyses of variance (rANOVAs) to verify the overall main effects of the relevant dimension (*Color*: *potential-reward*, *no-reward*) and the

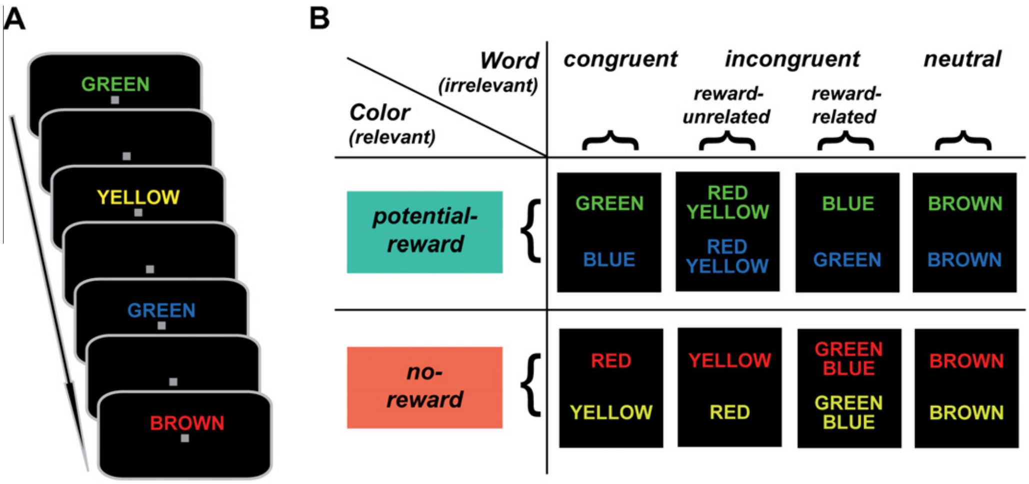

Fig. 1. Stimuli and experimental conditions. (A) Subjects responded to the ink color (relevant dimension) of presented words (SOA: stimulus onset asynchrony). (B) Counterbalanced across subjects, a subset of ink colors was associated with the potential of reward (*potential-reward*; e.g., green and blue), while the remaining ink colors were not (*no-reward*; e.g., red and yellow). The word meaning (irrelevant dimension) could be *congruent, incongruent reward-unrelated, incongruent reward-related*, or *neutral* with regard to the ink color.

irrelevant dimension (*Word*: *congruent, incongruent, neutral*). In order to investigate differential effects of *reward-related* and *reward-unrelated* irrelevant information, additional 2by2-rANOVAs were conducted focusing on the two types of incongruent trials.

2.1.3. Results

Subjects responded faster in *potential-reward* as compared to *no-reward* trials ($F_{(1,19)} = 78.28$, $p < .001$; $\eta_p^2 = .805$; Fig. 2A). In addition, in keeping with common findings in the Stroop task, RTs were significantly modulated by the presence of irrelevant semantic information, with fastest responses when the task-irrelevant word meaning was congruent, intermediate when it was neutral, and slowest when it was incongruent ($F_{(1,18)} = 50.6$, $p < .001$; $\eta_p^2 = .727$; Table 1). When confining the rANOVA to incongruent trials (Fig. 2A; dotted box), we again observed generally shorter RTs in *potential-reward* versus *no-reward* trials ($F_{(1,19)} = 65.96$, $p < .001$; $\eta_p^2 = .776$). In contrast, *reward-related* information in the irrelevant dimension resulted in the opposite pattern, with slower responses for incongruent words semantically related to *potential-reward* colors (*reward-related* > *reward-unrelated*; $F_{(1,19)} = 13.08$, $p = .002$; $\eta_p^2 = .408$). No interaction effects were observed ($p > .1$).

Considering *no-reward* trials alone, planned post hoc *t*-tests comparing incongruent to neutral trials confirmed the typical incongruency Stroop effects, with significant differences between *neutral* trials and both incongruent trial types (*neutral* < *reward-unrelated*: $t_{(19)} = 4.35$, $p < .001$; *neutral* < *reward-related*: $t_{(19)} = 3.13$, $p = .006$). In *potential-reward* trials, conflicting information still led to significantly slower responses relative to neutral trials if the incongruent word was related to *potential-reward* colors (*neutral* < *reward-related*: $t_{(19)} = 3.39$, $p = .003$). However, this effect was absent for *reward-unrelated* words (*neutral* < *reward-unrelated*: $t_{(19)} = 1.30$, $p = .211$). The direct comparison between incongruent *reward-related* and *reward-unrelated* words confirmed the relative RT-slowing for the former within both *potential-reward* ($t_{(19)} = 2.30$, $p = .033$) and *no-reward* trials ($t_{(19)} = 2.61$, $p = .017$).

Participants also committed less errors in *potential-reward* compared to *no-reward* trials ($F_{(1,19)} = 22.87$, $p < .001$; $\eta_p^2 = .546$). We furthermore found a main effect of the irrelevant word-meaning dimension ($F_{(1,18)} = 21.9$, $p < .001$; $\eta_p^2 = .535$), with highest error rates for incongruent trials, intermediate for neutral, and least for congruent. Focusing on incongruent trials only, we observed lower error rates for *potential-reward* versus *no-reward* trials ($F_{(1,19)} = 20.85$, $p < .001$; $\eta_p^2 = .523$), accompanied by a significant interaction with the word dimension ($F_{(1,19)} = 8.44$, $p = .009$; $\eta_p^2 = .308$), reflecting greater interference from incongruent *reward-unrelated* words in *no-reward* trials. No main effect of the irrelevant dimension alone was observed ($p > .2$). Planned post hoc *t*-tests revealed that in *no-reward* trials, error rates were significantly higher for both types of incongruent information compared to neutral trials (*neutral* < *reward-related*: $t_{(19)} = 2.89$, $p = .009$; *neutral* < *reward-unrelated*: $t_{(19)} = 4.75$, $p < .001$). In *potential-reward* trials, however, a robust increase in error rates was only observed for incongruent *reward-related* words relative to neutral ones (*neutral* < *reward-related*: $t_{(19)} = 4.27$, $p < .001$; *neutral* < *reward-unrelated*: $t_{(19)} = 2.04$, $p = .056$).

2.1.4. Discussion

In summary, the observed differential pattern in experiment 1 indicates a beneficial influence of reward associations in the relevant dimension, including a performance-enhancing suppression of incongruent irrelevant information when it was semantically unrelated to reward. At the same time, conflicting irrelevant information that was implicitly linked to reward led to substantially stronger interference as compared to incongruent information that was entirely unrelated to reward.

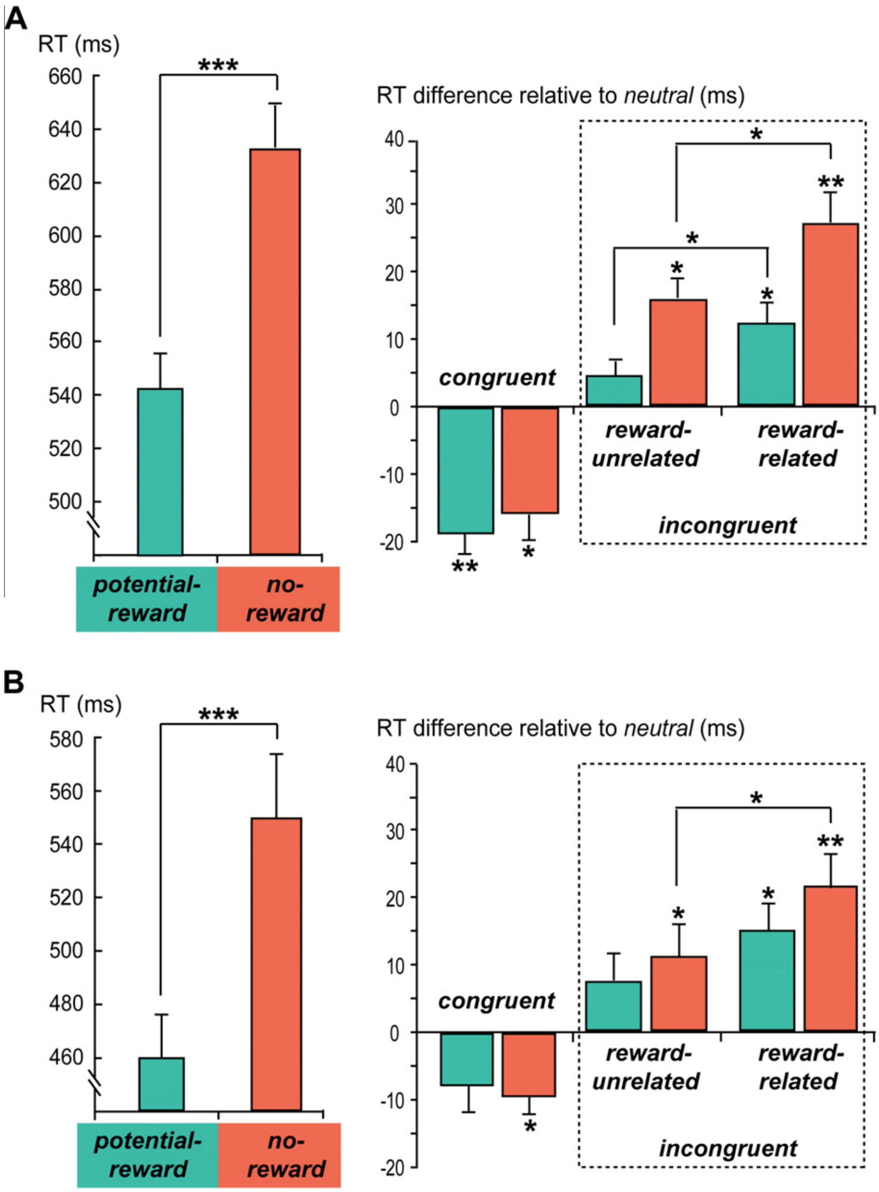

Fig. 2. RT results from experiment 1 (A) and experiment 2 (B). Differences between *potential-reward* (turquoise bars) and *no-reward* (orange bars) trials averaged across all word-meaning categories are displayed on the left. Effects of the irrelevant word-meaning dimension are displayed separately for the respective ink-color subsets on the right. RT values are depicted as the difference relative to neutral words. Error bars represent standard errors of the means (significance level $^{***}p < .001$; $^{**}p < .005$; $^{*}p < .05$).

In order to verify the observed effects and to replicate the results in an independent subject sample, a second experiment was conducted. First, in order to ensure that the observed effects can be exclusively evoked by the prospect of reward rather than a mixture of "maximizing wins" and "minimizing losses," we excluded the concept of punishment from experiment 2. While fast and accurate responses were still rewarded, subjects were no longer penalized for being slow or incorrect. Second, in order to reduce the potential confusion about the task-relevant and therefore reward-predictive stimulus dimension, we included a *color-reward training phase* preceding the rewarded Stroop task. In this training phase, subjects responded to colored rectangles in the absence of semantic information and received visual feedback to strengthen the reward associations. Third, following a rewarded Stroop phase similar to experiment 1, subjects performed an unrewarded Stroop task (termed "extinction phase") in which they were instructed that none of the colors was associated with reward any longer. This phase was included in order to investigate to what extent the behavioral influences of established reward associations would persist after removing the prospect of reward.

Table 1
Effects of relevant and irrelevant reward associations on performance.

Color (relevant)	Word (irrelevant)				
		Congruent	Incongruent		Neutral
			Reward-unrelated	*Reward-related*	
Exp1: Rewarded Stroop					
Potential-reward	RT ms (SD)	521 (71.7)	544 (77.0)	552 (79.9)	540 (76.6)
	Errors % (SD)	5.7 (2.6)	8.8 (5.1)	10.5 (4.3)	6.7 (3.4)
No-reward	RT ms (SD)	607 (87.0)	637 (89.7)	650 (93.7)	623 (80.7)
	Errors % (SD)	10.2 (5.1)	19.7 (10.4)	15.6 (8.0)	12.5 (7.5)
Exp2: Rewarded Stroop					
Potential-reward	RT ms (SD)	450 (35.5)	466 (44.9)	473 (41.9)	458 (43.6)
	Errors % (SD)	8.6 (6.4)	10.5 (6.7)	14.1 (9.4)	11.1 (8.3)
No-reward	RT ms (SD)	542 (45.5)	561 (46.9)	572 (45.7)	550 (45.3)
	Errors % (SD)	16 (8.4)	22.8 (11.2)	23.2 (10.1)	17.9 (8.7)
Exp2: Extinction phase					
Former potential-reward	RT ms (SD)	516 (60.4)	535 (66.7)	553 (65.7)	529 (62.1)
	Errors % (SD)	11.4 (7.4)	16.4 (10)	17.2 (9.6)	12.5 (6.6)
Former no-reward	RT ms (SD)	560 (71.1)	577 (81.5)	581 (67.7)	565 (73.4)
	Errors % (SD)	17 (11.8)	17.2 (10.1)	20.8 (10.4)	15.8 (9.9)

Exp1: Experiment 1; Exp2: Experiment2; RT: response time; SD: standard deviation.

3. Experiment 2

3.1. Methods

3.1.1. Participants

Sixteen healthy right-handed subjects (mean age ± SD: 22.6 ± 3.2, 9 female) participated after giving written informed consent in accordance with the Duke Medical Center Institutional Review Board for human subjects. Subjects were paid a basic amount of $15 plus an average reward bonus of $16.

3.1.2. Paradigm and procedure

Experiment 2 consisted of three successive phases: a color-reward training phase, a rewarded Stroop phase, and an extinction phase in which rewards were no longer given. The color-button mappings, as well as the color-reward associations, remained constant throughout the three phases but were counterbalanced across subjects.

Color-reward training phase: Colored rectangles (red, yellow, blue, green; 0.8° by 2°) were presented right above fixation for 600 ms, and subjects were asked to indicate the current color as fast as possible by pressing one of four buttons. Similar to experiment 1, fast and accurate responses to two of the four colors were associated with obtaining incentives (*potential-reward*) while the remaining colors were not (*no-reward*). Additionally, responses were followed by visual feedback (500 ms) indicating if it was fast/accurate ("+10ct") or slow/incorrect ("±0ct"). Thus, fast and correct responses were rewarded, while slow or incorrect responses did not affect the total gain. Subjects performed one 6-min run consisting of 56 trials (SOA = 1800–2200 ms) and the total gain was displayed in the end.

Rewarded Stroop phase: The task was identical to the one in experiment 1 with two exceptions: Subjects performed four instead of six 6-min runs, resulting in 320 *potential-reward* and 320 *no-reward* trials. Moreover, subjects were no longer penalized for incorrect or slow responses. A staircase procedure analogous to experiment 1 guaranteed a winning rate of 70%, which translated to an average win of $4 per run.

Extinction phase: Prior to performing two additional runs of the Stroop task, subjects were explicitly instructed that none of the colors would be associated with reward anymore. We refer to colors that were formerly associated with *potential-reward* and *no-reward* as *former potential-reward* and *former no-reward*, respectively. The stimuli and timing parameters were identical to the rewarded Stroop task, except for the interim feedback which was dropped.

3.2. Results

3.2.1. Color-reward training phase

Subjects responded significantly faster to *potential-reward* colors (mean ± SD: *potential-reward* 467 ± 57 ms; *no-reward* 530 ± 69; $t_{(15)} = 4.3$, $p = .001$) while maintaining similar error rates (mean ± SD: *potential-reward* 17 ± 13%; *no-reward* 20 ± 16%; $p > .5$), indicating that subjects learned to associate specific colors with the prospect of reward.

3.2.2. Rewarded Stroop phase

Similar to experiment 1, responses were faster in *potential-reward* compared to *no-reward* trials ($F_{(1,15)} = 51.29$, $p < .001$; $\eta_p^2 = .774$; Fig. 2B and 1). Also as in experiment 1, RTs were longest for incongruent trials, intermediate for neutral, and fastest for congruent ones ($F_{(1,14)} = 16.76$, $p < .001$; $\eta_p^2 = .528$). Focusing on incongruent trials only, the observed main effects were also comparable to experiment 1 (*potential-reward* < *no-reward*: $F_{(1,15)} = 48.73$, $p < .001$; $\eta_p^2 = .765$; *reward-related* > *reward-unrelated*: $F_{(1,15)} = 7.55$, $p = .015$; $\eta_p^2 = .335$).

Planned post hoc *t*-tests revealed significantly longer RTs for incongruent compared to neutral words in *no-reward* trials (*neutral* < *reward-related*: $t_{(15)} = 4.1$, $p = .001$;

neutral < *reward-unrelated*: $t_{(15)} = 2.3$, $p = .034$; *reward-related* > *reward-unrelated*: $t_{(15)} = 2.1$, $p = .045$). In contrast, in *potential-reward* trials, only incongruent *reward-related* words led to a significant RT difference relative to neutral ones (*neutral* < *reward-related*: $t_{(15)} = 3.3$, $p = .005$; *neutral* < *reward-unrelated*: $t_{(15)} = 1.9$, $p = .072$; *reward-related* > *reward-unrelated*: $t_{(15)} = 2.0$, $p = .065$).

Error rates were significantly reduced in *potential-reward* compared to *no-reward* trials ($F_{(1,15)} = 10.85$, $p = .005$; $\eta_p^2 = .420$) and again highest for incongruent, intermediate for neutral, and lowest for congruent words ($F_{(1,14)} = 10.34$, $p < .001$; $\eta_p^2 = .565$). Within incongruent trial types, error rates were again lower in *potential-reward* as compared to *no-reward* trials ($F_{(1,15)} = 10.94$, $p = .005$; $\eta_p^2 = .422$), but there was no significant difference between incongruent *reward-related* and *reward-unrelated* words ($p > .1$).

3.2.3. Extinction phase

Subjects still responded significantly faster to colors that formerly indicated potential reward (*former potential-reward* < *no-reward*: $F_{(1,15)} = 14.75$, $p = .002$; $\eta_p^2 = .496$), despite the explicit removal of actual reward. However, the reward-driven effect on accuracy did no longer reach significance ($F_{(1,15)} = 3.5$, $p > .08$). In line with typical Stroop effects, responses were slower ($F_{(1,14)} = 10.86$, $p = .001$; $\eta_p^2 = .420$) and less accurate ($F_{(1,14)} = 7.63$, $p = .004$; $\eta_p^2 = .337$) on incongruent as compared to neutral and congruent trials.

Focusing on the incongruent trials, responses were still faster in trials consisting of former *potential-reward* colors ($F_{(1,15)} = 12.1$, $p = .003$; $\eta_p^2 = .447$), but there were no significant differences in accuracy ($p > .4$). Furthermore, there were no significant differences between incongruent *reward-related* and *reward-unrelated* words in either RT or accuracy (all $p > .2$).

4. General discussion

In the present study, using a version of the Stroop task with reward associations for a subset of the ink colors, we systematically manipulated reward anticipation in the task-relevant dimension (ink color), thereby also implicitly imparting reward associations in the task-irrelevant dimension (word meaning). In line with previous research demonstrating the performance-enhancing effect of reward (e.g., Bijleveld et al., 2010; Engelmann & Pessoa, 2007), we found that reward anticipation linked to the word's ink color generally led to performance facilitation, as reflected by faster responses and lower error rates. In addition, although the typical conflict-induced slowing of responses was observed in all no-reward trials, this effect was partially suppressed in *potential-reward* trials. In contrast, reward associations in the irrelevant word-meaning dimension inflicted costs on performance if they were incongruent to the relevant ink color. More specifically, incongruent words semantically related to *potential-reward* colors interfered more strongly with performance than words semantically unrelated to reward, despite the word meaning always being entirely task-irrelevant.

The observation of response facilitation for the *potential-reward* ink colors is consistent with the notion that reward anticipation has an overall enhancing effect on task performance. In terms of parallel distributed processing models of conflict processing (MacLeod, 1991), reward in the relevant dimension seems to selectively enhance the processing of the currently relevant stimulus property, thereby reducing influence from interfering information and increasing the probability and speed of the correct response. In experiment 2, the stronger reduction of interference, as well as the reduction of facilitation in *potential-reward* trials, might be related to the additional practice in the color-naming task.

Interestingly, the general response facilitation observed in *potential-reward* trials persisted in the subsequent extinction phase, despite the explicit instruction that none of the colors was any longer predictive of reward. The latter observation further supports the idea that reward selectively increases the processing of behaviorally relevant stimuli. Once established, this preferential stimulus processing might persist in a relatively automatized fashion even if the original higher cognitive goal is no longer explicitly reinforced.

Beyond the performance-improving effects of relevant reward associations (*ink color*) in both experiments, we observed an opposite effect of reward when presented in the irrelevant dimension (*word meaning*), with slower responses to incongruent *reward-related* words. This pattern suggests that the strong reward association with the relevant stimulus dimension generalized to the irrelevant one, thereby inducing greater interference in incongruent trials (i.e., when the prepotent response to the highly salient word is incongruent with the required response). This *reward-related* increase in interference was very robust in no-reward trials in both experiments. However, in *potential-reward* trials, interference from incongruent *reward-related* words appeared to be slightly smaller in the second compared to the first experiment, possibly due to practice effects. Notably, the increased interference from *reward-related* words reported here relied on a newly learned stimulus-reward association rather than highly overlearned relationships or even automatized processes such as word reading itself. More generally, the observation of a performance decrement by reward information in the irrelevant dimension supports the notion that such associations might not always be beneficial for behavioral performance (Padmala & Pessoa, 2010; Pessoa, 2009).

With respect to the neural processes that shape these associations, the differential processing patterns might rely on dopaminergic pathways that are known to be involved in both reward (Schott et al., 2008; Schultz, 1997; Wise, 2004; Zellner & Ranaldi, 2009) and conflict processing (Holroyd & Coles, 2002; Jocham & Ullsperger, 2009). By enhancing the color-reward associations, dopamine may support cognitive control processes that facilitate the processing of the relevant stimulus property and thereby help reduce conflict-induced distraction (Locke & Braver, 2008; Montague, Hyman, & Cohen, 2004; Ridderinkhof, van den Wildenberg et al., 2004). At the same time, implicit reward associations with task-irrelevant stimulus properties – here, the word meaning – may induce an increase in salience of these properties. This increased salience may disrupt task performance by enhancing the incorrect stimulus-response mapping or by drawing some attentional resources away from the processing of the relevant dimension (Pessoa, 2009). Importantly, at least in the present study, the mechanism that brings about these *reward-related* modulations with regard to the irrelevant dimension must involve a relatively abstract color representation that is shared by processes extracting information about the word's ink color and its semantic meaning, thereby leading to a transfer of reward association from the relevant to the irrelevant stimulus dimension.

In summary, the current observations support the idea that task-relevant reward acts as an overall enhancer of behavioral performance. Specifically, the anticipation of reward seems to promote effective stimulus processing, including a reduction of interference from conflicting information. However, these beneficial effects can come at a cost: If irrelevant *reward-related* information is incongruent with the relevant dimension, it appears to disrupt task performance, presumably by emphasizing an already prepotent but incorrect response to the highly salient word.

In this regard, the color-naming version of the Stroop task we used here appears to be especially well suited to provoke such robust interference effects due to word reading being a highly trained skill as compared to color-naming. In contrast, in the *reverse* Stroop task, which requires responses to the word instead of to the ink color, the interference effects from the task-irrelevant color dimension tend to be relatively weak or even absent (MacLeod, 1991). Accordingly, future studies will have to verify how this commonly observed processing asymmetry interacts with *reward-related* stimulus saliency. Reward associations in the *reverse Stroop* task may facilitate responses to *reward-related* words. In turn, the concomitantly increased saliency of irrelevant *reward-related* colors might boost the typically weak interference effects in this task.

Acknowledgement

This work was funded by NIH Grant R01-MH060415 to M.G.W. and by DFG Grant BO 3345/1-1 to C.N.B.

References

Adcock, R. A., Thangavel, A., Whitfield-Gabrieli, S., Knutson, B., & Gabrieli, J. D. (2006). Reward-motivated learning: Mesolimbic activation precedes memory formation. *Neuron, 50*(3), 507–517.

Bijleveld, E., Custers, R., & Aarts, H. (2010). Unconscious reward cues increase invested effort, but do not change speed-accuracy tradeoffs. *Cognition, 115*(2), 330–335.

Carter, C. S., & van Veen, V. (2007). Anterior cingulate cortex and conflict detection: An update of theory and data. *Cognitive, Affective, &Behavioral Neuroscience, 7*(4), 367–379.

Cohen, J. D., Dunbar, K., & McClelland, J. L. (1990). On the control of automatic processes: A parallel distributed processing account of the Stroop effect. *Psychological Review, 97*(3), 332–361.

Della Libera, C., & Chelazzi, L. (2006). Visual selective attention and the effects of monetary rewards. *Psychological Science, 17*(3), 222–227.

Egner, T. (2007). Congruency sequence effects and cognitive control. *Cognitive, Affective, & Behavioral Neuroscience, 7*(4), 380–390.

Engelmann, J. B., & Pessoa, L. (2007). Motivation sharpens exogenous spatial attention. *Emotion, 7*(3), 668–674.

Holroyd, C. B., & Coles, M. G. (2002). The neural basis of human error processing: Reinforcement learning, dopamine, and the error-related negativity. *Psychological Review, 109*(4), 679–709.

Jocham, G., & Ullsperger, M. (2009). Neuropharmacology of performance monitoring. *Neuroscience and Biobehavioral Reviews, 33*(1), 48–60.

Knutson, B., Adams, C. M., Fong, G. W., & Hommer, D. (2001). Anticipation of increasing monetary reward selectively recruits nucleus accumbens. *Journal of Neuroscience, 21*(16), RC159.

Krebs, R. M., Schott, B. H., Schutze, H., & Duzel, E. (2009). The novelty exploration bonus and its attentional modulation. *Neuropsychologia, 47*(11), 2272–2281.

Kristjansson, A., Sigurjonsdottir, O., & Driver, J. (2010). Fortune and reversals of fortune in visual search: Reward contingencies for pop-out targets affect search efficiency and target repetition effects. *Attention Perception & Psychophysics, 72*(5), 1229–1236.

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.

MacLeod, C. M. (1991). Half a century of research on the Stroop effect: An integrative review. *Psychological Bulletin, 109*(2), 163–203.

Miller, E. K., & Cohen, J. D. (2001). An integrative theory of prefrontal cortex function. *Annual Review of Neuroscience, 24*, 167–202.

Montague, P. R., Hyman, S. E., & Cohen, J. D. (2004). Computational roles for dopamine in behavioural control. *Nature, 431*(7010), 760–767.

Padmala, S., & Pessoa, L. (2010). Interactions between cognition and motivation during response inhibition. *Neuropsychologia, 48*(2), 558–565.

Pessoa, L. (2009). How do emotion and motivation direct executive control? *Trends in Cognitive Sciences, 13*(4), 160–166.

Ridderinkhof, K. R., Ullsperger, M., Crone, E. A., & Nieuwenhuis, S. (2004). The role of the medial frontal cortex in cognitive control. *Science, 306*(5695), 443–447.

Ridderinkhof, K. R., van den Wildenberg, W. P., Segalowitz, S. J., & Carter, C. S. (2004). 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.

Schott, B. H., Minuzzi, L., Krebs, R. M., Elmenhorst, D., Lang, M., Winz, O. H., et al. (2008). Mesolimbic functional magnetic resonance imaging activations during reward anticipation correlate with reward-related ventral striatal dopamine release. *Journal of Neuroscience, 28*(52), 14311–14319.

Schultz, W. (1997). Dopamine neurons and their role in reward mechanisms. *Current Opinion in Neurobiology, 7*(2), 191–197.

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology, 18*, 643–662.

Ullsperger, M., & von Cramon, D. Y. (2003). Error monitoring using external feedback: Specific roles of the habenular complex, the reward system, and the cingulate motor area revealed by functional magnetic resonance imaging. *Journal of Neuroscience, 23*(10), 4308–4314.

van Steenbergen, H., Band, G. P., & Hommel, B. (2009). Reward counteracts conflict adaptation: Evidence for a role of affect in executive control. *Psychological Science*.

Wise, R. A. (2004). Dopamine, learning and motivation. *Nature Reviews Neuroscience, 5*(6), 483–494.

Wittmann, B. C., Schott, B. H., Guderian, S., Frey, J. U., Heinze, H. J., & Duzel, E. (2005). Reward-related FMRI activation of dopaminergic midbrain is associated with enhanced hippocampus-dependent long-term memory formation. *Neuron, 45*(3), 459–467.

Zellner, M. R., & Ranaldi, R. (2009). How conditioned stimuli acquire the ability to activate VTA dopamine cells: A proposed neurobiological component of reward-related learning. *Neuroscience and Biobehavioral Reviews*.