Best, M., Lawrence, N. S., Logan, G. D., McLaren, I. P. L., & Verbruggen, F. (2015). Should I stop or should I go? The role of associations and expectancies. ***Journal of Experimental Psychology: Human Perception and Performance, in press.***

Should I stop or should I go? The role of associations and expectancies

Best, M. (University of Exeter)

Lawrence, N.S. (University of Exeter)

Logan, G. D. (Vanderbilt University)

McLaren, I. P. L. (University of Exeter)

Verbruggen, F. (University of Exeter)

Correspondence concerning this article should be addressed to Maisy Best or Frederick Verbruggen, School of Psychology, University of Exeter, Exeter, EX4 4QG. UK. E-mail: mjb246@exeter.ac.uk or f.l.j.verbruggen@exeter.ac.uk. We thank Robert Gaschler and two anonymous reviewers for their for their insightful and constructive feedback. This work was supported by a studentship from the Economic and Social Research Council to MB, an Economic and Social Research Council Grant (ES/J00815X/1) to FV and IPLM, and a starting grant to FV from the European Research Council (ERC) under the European Union's Seventh Framework Program (FP7/2007-2013)/ ERC Grant Agreement No. 312445.

Abstract

Following exposure to consistent stimulus-stop mappings, response inhibition can become automatized with practice. What is learned is less clear, even though this has important theoretical and practical implications. A recent analysis indicates that stimuli can become associated with a stop signal or with a stop 'goal'. Furthermore, expectancy may play an important role. Previous studies that have used stop or no-go signals to manipulate stimulus-stop learning cannot distinguish between stimulus-signal and stimulus-goal associations, and expectancy has not been measured properly. In the present study, participants performed a task that combined features of the go/no-go task and the stop-signal task in which the stop-signal rule changed at the beginning of each block. The go and stop signals were superimposed over forty task-irrelevant images. Our results show that participants can learn direct associations between images and the stop goal without mediation via the stop signal. Exposure to the image-stop associations influenced task performance during training, and the expectancies measured following task completion or measured within the task. But, despite this, we found an effect of stimulus-stop associations on test performance only when the task increased the task-relevance of the images. This could indicate that the influence of stimulus-stop learning on go performance is strongly influenced by attention to both task-relevant and task-irrelevant stimulus features. More generally, our findings suggest a strong interplay between 'automatic' and 'controlled' processes.

Response inhibition is often considered to be a deliberate act of top-down cognitive control. It allows people to quickly stop and replace actions that are no longer relevant or that are inappropriate in the current task environment. Longitudinal studies have shown that response inhibition, and self-control more generally, in childhood and adolescence correlates with a variety of life outcomes in adulthood, including personal finances and engagement in healthy behaviors (Diamond, 2013; Moffitt et al., 2011; Nigg et al., 2006). Furthermore, clinical research suggests that impairments in response inhibition may contribute to the development of a range of psychopathological and impulse-control disorders, such as attention deficit/hyperactivity disorder, obsessive-compulsive disorder, substance abuse, pathological gambling, and eating disorders (Bechara, Noel & Crone, 2006; Crews & Boettiger, 2009; de Wit, 2009; Fernie et al., 2013; Garavan & Stout, 2005; Nigg, 2001; Noël, Brevers & Bechara, 2013). Response inhibition efficiency also correlates with the treatment outcome in people with such disorders (e.g. Nederkoorn, Jansen, Mulkens & Jansen, 2007). Thus, the ability to stop actions seems very important for adaptive and goal-directed behavior. However, in the past few years, research has demonstrated that response inhibition may not always be the executive, deliberate, act of control that it is typically assumed to be. In the present study, we will further explore the interplay between 'bottom-up' and 'top-down' control processes when stopping a response.

Popular paradigms used to study response inhibition in healthy and clinical populations are the go/no-go task (Donders, 1868/1969) and the stop-signal task (Logan & Cowan, 1984; Verbruggen & Logan, 2008c). Research using these tasks has demonstrated both short-term and long-term after-effects of stopping (e.g. Bissett & Logan, 2011; Enticott, Bradshaw, Bellgrove, Upton & Ogloff, 2009; Giesen & Rothermund, 2014; Rieger & Gauggel, 1999; Verbruggen, Logan, Liefooghe & Vandierendonck, 2008; Verbruggen & Logan, 2008a). For example, responding to a stimulus is typically slowed after a stop-signal trial. This slowing is more pronounced when the primary-task stimulus of the previous stop-signal trial is repeated, which has led to the suggestion that people can learn associations between specific stimuli and stopping (Rieger & Gauggel, 1999; Verbruggen et al., 2008; Verbruggen & Logan, 2008a). The idea that a specific stimulus can become associated with stopping is consistent with studies that have highlighted the role of stimulus-response (S-R) bindings in other cognitive control paradigms, such as the negative priming paradigm (cf. the 'do-not-respond' tag account; Neill, Valdes, Terry & Gorfein, 1992; Neill & Valdes, 1992), the task-switching paradigm (e.g. Koch & Allport, 2006; Waszak, Hommel & Allport, 2003, 2004, 2005), and interference control tasks (e.g. Anderson & Folk, 2013; Hommel, Proctor & Vu, 2004). The formation of stimulus-stop bindings may be the first step towards automaticity (Logan, 1990). Memory-retrieval accounts of automatization assume that every time people respond to a stimulus, processing episodes are stored as

'instances' (Logan, 1988) or 'event files' (Hommel, 1998, 2004) in memory. These instances or event files may contain information about the stimulus (e.g. the word), the interpretation given to the stimulus (e.g. 'natural'), the task goal (e.g. 'go')[1] and the response (e.g. 'left key press'). These episodes are retrieved when a stimulus is repeated, and will influence responding. For example, the Instance Theory (Logan, 1988) postulates that action selection can be construed as a race between an algorithmic response-selection process and a memory-retrieval process; the process that finishes first determines which action is selected. When the memory-retrieval process wins the race, the decision is said to be automatic, whereas decisions based on algorithmic processing are deliberate or intentional (Logan, 1988). Therefore, when the stimulus-response or stimulus-stop mapping is the same throughout practice, multiple instances are formed and automatic processing can develop (Logan, 1988; Shiffrin & Schneider, 1977).

Some of us examined the idea that inhibitory control in go/no-go and stop-signal tasks can be triggered automatically via the retrieval of stimulus-stop associations from memory (Verbruggen & Logan, 2008b). For example, in a series of go/no-go experiments, a stimulus category determined if a participant should respond or not (e.g. living word referents = go; non-living word referents = no-go). After a training phase, the go/no-go mapping was reversed in a test phase. We found that responding to the old stop stimuli was slowed compared with new stimuli that were not previously presented during training (consequently, these new stimuli were not associated with going or stopping) or old stimuli that were associated with going. This response slowing was also found in modified versions of the stop-signal task in which the contingencies between specific go stimuli and stopping were manipulated, such that certain items were consistently presented on stop-signal trials, whereas other items were presented on both go and stop-signal trials. Consistent with the go/no-go results, we found that responding was slowed for old stop items compared with inconsistent items that were not particularly associated with going or stopping (Lenartowicz, Verbruggen, Logan & Poldrack, 2011; Verbruggen & Logan, 2008b). Furthermore, the Lenartowicz et al. (2011) study demonstrated that old stop items activated the neural stopping network. Thus, response inhibition may become automatized after sufficient practice with consistent stimulus-stop mappings (Jasinska, 2013; Lenartowicz et al., 2011; Spierer, Chavan & Manuel, 2013; Verbruggen, Best, Bowditch, Stevens & McLaren, 2014; Verbruggen & Logan, 2008b). These findings may have important implications for our current theories of response inhibition and executive control. Furthermore, they could also have practical applications. Recent studies suggest that the acquisition of stimulus-stop associations could be an effective way to reduce engagement in impulsive behaviors, such as excessive food (e.g. Houben & Jansen, 2011) and alcohol (e.g. Jones & Field, 2013) consumption. These studies used paradigms in

which no-go or stop signals were superimposed over, or presented around, images of unhealthy foods or alcohol (e.g. Bowley, Faricy, Johnstone & Smith, 2013; Houben & Jansen, 2011; Houben, Nederkoorn, Wiers & Jansen, 2011; Houben, 2011; Jones & Field, 2013; Lawrence, Verbruggen, Morrison, Adams & Chambers, 2015; Veling, Aarts & Papies, 2011; Veling, Aarts & Stroebe, 2013; Veling, van Koningsbruggen, Aarts & Stroebe, 2014). Pairing these images with stopping reduced subsequent consumption of unhealthy foods and alcohol. Therefore, this research suggests that automatic inhibition could be useful in the treatment of a variety of impulse-control disorders (for a recent meta-analysis, see Jones et al., 2015).

Current research in the stop-learning literature appears to provide strong support for the 'automatic inhibition' account that postulates that stimuli can become associated with the act of stopping. However, a recent review indicates that it is still unclear exactly what is learned in these tasks and how this influences performance (Verbruggen, Best et al., 2014). The present study was designed to address two of the main outstanding issues that we highlighted in our review (similar issues were also recently raised in the context of S-R bindings; Henson, Eckstein, Waszak, Frings & Horner, 2014): (1) are associations between stimuli and stopping direct, and (2) to what extent does expectancy play a role?

Are associations between stimuli and stopping direct?

The automatic inhibition account assumes that people learn direct associations between a stimulus and the act of stopping in go/no-go tasks and modified versions of the stop-signal task. However, the results of a recent experiment are inconsistent with this account (Verbruggen, Best et al., 2014). In that experiment, participants made speeded semantic categorizations (living/non-living) of a series of words. On some trials (stop-signal trials) an additional visual signal was presented below the word, instructing participants to withhold their planned response. Certain words were consistently presented on stop-signal trials, whereas other words were presented on go and stop-signal trials with equal probability. We found that the probability of responding on stop-signal trials was lower for the consistent words than for the inconsistent words in the training phase, indicating that learning had occurred. However, we found no go reaction time (RT) difference between the old stop words and the inconsistent words when the stimulus-stop mapping was subsequently reversed in the test phase. In other words, learning influenced stop performance on signal trials in the training phase, but it did not influence go performance on no-signal trials in the test phase. We proposed that this pattern of results indicates that participants learned stimulus-signal associations rather than stimulus-stop associations. Such associations between the stop words and the stop signal (i.e. the line turning bold) will prime the

representation of the stop signal rather than the stop goal. Signal detection plays a critical role in successful stopping (e.g. Verbruggen, Stevens & Chambers, 2014), and computational work suggests that a considerable proportion of the stopping latency is occupied by perceptual or afferent processes (Boucher, Palmeri, Logan & Schall, 2007; Logan, Van Zandt, Verbruggen & Wagenmakers, 2014; Logan, Yamaguchi, Schall, & Palmeri, 2015; Salinas & Stanford, 2013). Thus, by priming the representation of the stop signal, learning could lead to improvements in stopping performance on stop-signal trials without influencing responding on go trials in the test phase.

The idea that participants could learn stimulus-signal associations is also consistent with a range of research on learning and conditioning in humans and other animals that indicates that stimulus detection can itself become conditioned (McLaren, Wills & Graham, 2010) and, of course, that links between perceptual stimuli can be established. As an illustrative (and rather basic) example, in a classic autoshaping paradigm with pigeons, the presentation of a conditioned stimulus (e.g. a keylight) and an unconditioned stimulus (e.g. the delivery of food) usually co-occur. With practice, the presentation of the conditioned stimulus alone can come to elicit the conditioned response (e.g. pecking at this key). The conditioned stimulus can activate this response via two routes; either indirectly via the CS-US link, or more directly, via a CS-R link (Hall, 2002). Thus, it seems plausible that learning can also influence perception of the no-go or stop signal in response-inhibition paradigms.

The potential for stimulus-signal associations has important implications for the interpretation of previously reported behavioral effects in the stop-learning literature. Previous studies that have used no-go or stop signals to manipulate stimulus-stop learning cannot distinguish between stimulus-goal and stimulus-signal learning. It is therefore possible that previously observed RT effects and neural activations (Lenartowicz et al., 2011; Manuel, Bernasconi & Spierer, 2013; Manuel, Grivel, Bernasconi, Murray & Spierer, 2010) could be mediated by a link between the stimulus, the stop signal, and stopping (see Figure 1). Similarly, in go/no-go experiments in which the go/no-go rules are explicit (e.g. living = go, non-living is no-go), the stimulus-stop association could be mediated via the go/no-go category (e.g. 'desk = non-living -> non-living = no-go', instead of 'desk = no-go'). In addition to being of theoretical interest, the idea of stimulus-stop associations also has implications for applied stop-training research (see above). Therefore, in the present study, we investigated whether there is any evidence for the original idea (i.e. as suggested by Verbruggen & Logan, 2008b) that direct associations can be acquired between a stimulus and the stop goal, without mediation via a representation of the stop signal (or no-go category). To discourage the formation of stimulus-signal associations, we changed the stop signal and the task rules at the beginning of each block. The

demonstration of response slowing for consistent stop items in the present experiment would provide the strongest evidence to date for the direct stimulus-stop hypothesis.

What is the role of expectancy in stimulus-stop learning?

In the associative-learning literature, there is an on-going debate surrounding the involvement of explicit and implicit processes in the acquisition of stimulus-action associations (Mitchell, De Houwer & Lovibond, 2009). To make a broad distinction, 'explicit' processes are assumed to be controlled, intentional, effortful and rule-based; by contrast, 'implicit' processes are assumed to be automatic, effortless, and associative (e.g. McLaren, Green & Mackintosh, 1994; for a recent discussion of the distinction between associative and propositional processes, see McLaren et al., 2014). Expectancy ratings have been used to dissociate between the two processes (e.g. McLaren et al., 2014; Newell & Shanks, 2014). In the context of stop-learning, this dissociation between rule-based processes and associative (S-S or S-R) processes has important theoretical implications. After all, expectancy of a stop signal for old stop items could indicate that the response slowing observed for old stop items is due to proactive inhibitory control, rather than 'automatic inhibition'. When a cue indicates that a stop signal is likely to occur on the following trial(s), participants proactively increase response thresholds or suppress motor activation (e.g. Jahfari et al., 2012; Ramautar, Kok & Ridderinkhof, 2004; Verbruggen & Logan, 2009; Zandbelt, Bloemendaal, Neggers, Kahn & Vink, 2013). Stimuli associated with stopping could act as such cues (e.g. 'if stimulus X then *p*(stop) is high'), and participants would adjust their response strategies accordingly. In other words, slowing for old stop items could be due to proactive control (which may be conceived as another 'algorithmic' process; cf. Logan, 1988), rather than to the direct activation of the stop response via memory retrieval. The role of expectancy-driven processes is also relevant for the applied stop-training research. Indeed, the extent to which training effects like these reflect implicit or associative effects has been called into question. For example, Boot, Simons, Stothart and Stutts (2013) argued that many 'control' training effects could be due to changes in expectations and demand characteristics. The involvement of expectancies would have implications for the longevity of these inhibitory control training effects and the variability of training efficacy across individuals (cf. Boot et al., 2013).

In the present study we investigated the role of expectancy in stimulus-stop learning via the inclusion of an additional dependent variable that was sensitive enough (Newell & Shanks, 2014) to detect stimulus-stop learning following task completion (Experiments 1-3) or within the task (Experiment 4).

Experiment 1

In Experiment 1, we combined features of a go/no-go task and a stop-signal task. In standard go/no-go tasks only one stimulus is presented on each trial, determining whether participants have to respond or not. In standard stop-signal tasks participants respond to each stimulus, unless an extra stop signal is presented after a variable delay. In Experiment 1, we used a go/stop task based on those used in studies examining the effects of no-go training effects on food and alcohol consumption (see above). Similar to picture-word Stroop tasks (see e.g. MacLeod, 1991), go and stop signals were superimposed over forty neutral images. The delay between the presentation of the images and the signals was zero ms. A subset of the images was consistently associated with stop signals, another subset was consistently associated with go signals, and the remaining images were control images (not particularly associated with go or stop). After twelve training blocks, the image mappings were reversed, and participants had to respond to the stop-associated images. Participants were not informed about the image mappings, but they were told at the beginning of each block what the go and stop signals were. To discourage the formation of stimulus-signal or stimulus-category associations, we varied the representation of the go and stop signals at the beginning of each block. We predicted that this change manipulation would encourage the formation of image-stop associations (cf. Verbruggen & Logan, 2008b) instead of image-signal associations (i.e. S-R rather than S-S learning). We indexed learning during the task via two measures. The first index was the probability of responding on the stop trials, *p*(respond|stop), which was predicted to be lower for stop-associated images than for the control images. The second index was RT on go trials, which was predicted to be longer for the stop-associated images than for the control images. To examine the role of expectancy in stop learning, participants were asked to rate the extent to which they expected to withhold their response for each of the images presented in the task at the end of the experiment.

Method

Subjects. Thirty-one students from the University of Exeter participated for monetary compensation (£5) or partial course credit (M = 19.43 years, SD = 1.70 years, 17 females, 27 right-handed). Two participants were excluded because they incorrectly executed a response on ≥ 30% of the stop-signal trials (there was no delay between the presentation of the image and the stop signal; consequently, *p*(respond|stop) was expected to be low). The target sample and exclusion criteria were determined before data collection. The data with these participants included are available as Supplementary Material.

Apparatus and stimuli. The experiment was run on an Apple iMac using Psychtoolbox (Brainard, 1997). The stimuli were presented on a 20-in monitor (with a 1680 × 1050 resolution). The experimental paradigm consisted of a go/stop task in which the go/stop rule changed at the beginning of each block. The go and stop signals (a full list of the signals used appears in the Appendix A) were superimposed over forty task-irrelevant neutral images (size: 250 × 250 pixels), which were presented in the centre of the screen on a white background. Each image was presented twice per block. In each block, we used two go signals (e.g. the vowels 'a' or 'e') and two stop signals (e.g. the consonants 't' or 'n'). Participants responded on go trials by pressing the spacebar on a keyboard with their right index finger; they were instructed to withhold their response on stop trials. The signals and the go/stop mapping were shown on the screen at the beginning of each block for a minimum of 5 seconds, and participants had to press a key to start the first trial. The order of the task rules was randomized across the blocks and the response-rule category was counterbalanced across participants (e.g. 'go = vowels, stop = consonants' vs. 'go = consonants, stop = vowels').

Procedure. Unbeknown to the participants, there were two phases in the experimental paradigm that determined the image-go/stop mappings; the first 12 blocks of 80 trials comprised the 'training phase' and the final two blocks of 80 trials comprised the 'test phase'. Participants were verbally instructed to read the task rule screen carefully before starting each block. There was a 15 second break between each block.

There were three image types (Table 1). First, stop-associated images were paired with a stop signal on 75% of presentations in the training phase; in the test phase, they were always paired with a go signal. Second, go-associated images were always paired (100%) with a go signal in the training phase, but they could occur on stop trials in the test phase (eight old go-associated images were paired with a stop signal on 75% of presentations; eight old go-associated images were never paired with a stop signal). Third, control images were paired with a stop signal on 25% of presentations in the training and test phases. The control images were mostly paired with a go signal during training to ensure that the overall probability of a stop trial [p(stop) = 0.25] was the same in the training and the test phases (stopping performance is sensitive to minor variations in signal probability, e.g. see Bissett & Logan, 2011).

All trials began with the concurrent presentation of the image and a go/stop signal (Figure 2), instructing participants to execute (go) or withhold (stop) the spacebar response. After 750 ms (regardless RT), the images and go/stop signal were replaced by a feedback message ('correct', 'incorrect', or 'too slow' in case they did not respond before the end of the trial) which remained on

the screen for 500 ms. The feedback message was presented to encourage fast and accurate responding. Following the feedback message, there was a blank screen for 250 ms, after which the next trial started.

Following completion of the experimental task, each image was again presented on the screen. The order of the images was randomized anew for each participant. Participants were asked to rate 'how much do you expect to withhold your response when this image is presented?' on a scale between 1 ('I definitely do not think this image indicates that I have to withhold my response') and 9 ('I definitely think this image indicates that I have to withhold my response'). As a manipulation check, we also asked participants to rate how much they expected to respond (i.e. go) to each of the images (the order of the respond/withhold ratings was counterbalanced across participants). These go ratings were consistent with the stop expectancy ratings so are not reported further.

Analyses. All data processing and analyses were completed using R (R Development Core Team, 2013). The training and test phase trials were analyzed separately using Analyses of Variance (ANOVA) with image type and block as within-subjects factors. Performance was assessed in terms of average RT for correct go responses, the probability of a missed go response [*p*(miss)] and the probability of responding on a stop trial [*p*(respond|stop)]. RTs < 1 ms were removed prior to analysis. We did not analyze *p*(miss) further as values were very low (Table 2). Table 3 provides an overview of the ANOVAs. For pairwise comparisons, Hedge's g_{av} is the reported effect size measure (Lakens, 2013). All data files and R scripts used for the analyses are deposited in Dropbox (https://www.dropbox.com/sh/3k1346wgvagm9ii/AACzQtru20_GX1wUnIZSycCia?dl=0). [*Note: If this paper is accepted, the data files and R scripts will be deposited on the Open Research Exeter data repository*]

Results

Training phase. The main effect of image type on go RTs was reliable ($p < 0.001$); planned comparisons revealed that responding to the stop-associated images (on the relevant 25% of trials) was slower (414 ms) than to the go-associated images (403 ms), $t(28) = -4.93$, $p < 0.001$, $g_{av} = 0.440$, and to the control images (406 ms), $t(28) = -3.26$, $p = 0.002$, $g_{av} = 0.327$. There was a marginally reliable difference between the go and the control images, $t(28) = -1.99$, $p = 0.055$, $g_{av} = 0.109$ (Figure 3; Table 3). In line with our predictions, the *p*(respond|stop) was lower for the stop-associated images (0.131) than for the control images (0.151), $p = 0.019$ (Figure 3). Thus, performance on go and stop trials suggests that participants acquired the image-stop associations. The effect of block and the interaction between block and image type did not reach significance, suggesting that the effect of

image type was present in most blocks (Table 3). This is consistent with our previous work, which indicates that the effect of stop learning emerges after a single trial presentation, and that it then quickly asymptotes (Verbruggen & Logan, 2008a; Verbruggen & Logan, 2008b). The absence of an overall practice effect is most likely due to the introduction of a novel go/stop rule at the beginning of each block; consistent with this idea, a post-hoc test confirmed that participants responded faster in the second half of a block than in the first half**,** $t(28) = 3.99$, $p < 0.001$, $g_{av} = 0.324$.

Test phase. In the test phase, the stop-associated images were always paired with a go signal, the control images were paired with a stop signal on 25% of the trials (i.e. the control images remained the same in the training and test phases), and the go-associated images were mostly paired with a stop signal (Table 1). Based on the automatic inhibition hypothesis, we predicted that responding on go trials would be slower for the stop-associated images than for the go-associated images and for the control images. Furthermore, *p*(respond|stop) should be higher for the go-associated images than for the control images. However, image type did not influence RT nor *p*(respond|stop) in the test phase (*p's* ≥ 0.557; Table 4). It is possible that the absence of the test phase effect is due to differences in the overall RT (as RTs were faster in the test phase than in the training phase). To investigate this possibility, we plotted RT percentiles for the training and test phases. This revealed that the overall test phase RT cannot account for the absence of the predicted image-stop learning effects (see Supplementary Material).

Expectancy ratings. Due to technical reasons, one participant in Experiment 1 did not complete the expectancy ratings task. The results of the test phase raise some doubts about whether participants learned long-term image-stop associations. However, the analysis of the expectancy ratings obtained following task completion revealed a main effect of image type, $F(2, 54) = 10.06$, $p < 0.001$, $gen.\ \eta 2 = 0.075$. Consistent with the stimulus-stop contingencies during training, participants expected to withhold their response more when the stop-associated images were presented (4.83) than when the go-associated images (3.91) and the control images (4.26) were presented; $t(27) = -3.46$, $p = 0.001$, $g_{av} = 0.653$, and $t(27) = -2.74$, $p = 0.010$, $g_{av} = 0.403$, respectively. The difference between the control and the go-associated images was also reliable, $t(27) = -2.89$, $p = 0.007$, $g_{av} = 0.271$. Thus, participants could distinguish between the images on the basis of their association with the stop and go goals. The 'stop minus control image' expectancy difference correlated with the corresponding RT difference in the test phase, $r(26) = 0.437$, $p = 0.019$: participants who expected to withhold their response more during the presentation of the old stop-associated images slowed more when they had to respond to these images in the test phase. This suggests that expectancies generated on the basis of the acquired image-stop mappings may contribute to the manifestation of an 'automatic' inhibition

effect in the test phase. However, there was no reliable correlation between the 'stop minus control' expectancy difference and the corresponding RT in the training phase, $r(26) = 0.010$, $p = 0.961$. There was also no reliable correlation between the RT and expectancy differences for the stop- and the go-associated images in the training phase, $r(26) = -0.040$, $p = 0.841$, or the test phase, $r(26) = 0.272$, $p = 0.161$ (Note that uncorrected p's are reported).

Discussion

In Experiment 1, we investigated two questions highlighted in our recent review article (Verbruggen, Best, et al., 2014): (1) can participants learn direct associations between stimuli and stopping; and (2) what is the role of expectancy in stimulus-stop learning? The results provide some answers to both questions. Task performance during the training phase showed that participants could acquire direct stimulus-stop associations when the rules (and consequently, signals) constantly changed throughout the task. This indicates that the learning effects were not mediated via signal representations (as each image was only presented twice per block and there were two stop signals and two go signals per block). Furthermore, the expectancy data obtained following task completion showed that participants generated expectancies that were consistent with the stimulus-stop contingencies acquired during training.

However, the results of Experiment 1 raised a new question: why did stimulus-stop associations not influence performance in the test phase? We found an associative effect on behavior that appeared early in training but then disappeared again in the later training blocks and in the test phase (Figure 3; for similar results in another action control paradigm, see Gaschler & Nattkemper, 2012), even though the expectancy data measured at the end of the experiment indicated that the associations were not forgotten. We attribute this to an interaction between attention and learning. The role of attention in stimulus-stop learning has not yet been considered (and, indeed, is something we did not discuss in our recent review; Verbruggen, Best et al., 2014). In previous studies demonstrating stimulus-stop learning (e.g. Verbruggen & Logan, 2008b), the go/stop items were task-relevant as they determined the required response; consequently, optimal task performance in these studies depended on participants attending to the stop items (as opposed to the signals). In the present study, we adapted a paradigm frequently used in applied research (e.g. Houben & Jansen, 2011) whereby go/stop signals were superimposed on a series of images. This was advantageous as it allowed us to vary the representation of the go/stop signals throughout the task whilst independently manipulating the image-stop contingencies. However, a consequence of this procedure is that optimal task performance does not depend on attending to the stop-associated images. Initially, the task-irrelevant images may

have captured attention because they were novel, allowing the effects of learning to emerge. But habituation to the images and reduced salience may have reduced attentional capture, and consequently, weakened or even eliminated the effects of stop-learning on behavior in later blocks.

The hypothesized role of attention in the acquisition of stimulus-stop associations is consistent with the associative-learning literature. For example, a review by Kruschke (2003) indicates that attention is crucial in explaining associative learning phenomena. Following the principles first enunciated by Mackintosh (1975), he argued that attending to informative cues whilst ignoring irrelevant cues will accelerate learning. Furthermore, the amount of attention that is paid to the cues will determine the influence of acquired associations on behavior. In a similar vein, Instance Theory assumes that attention determines what is learned and what is retrieved (Logan & Etherton, 1994; Logan, 1988). But attention can also be influenced by learning. For example, the learned predictability of the outcome relative to other concurrently presented cues may influence the extent to which cues are considered informative or salient, and consequently, the extent to which participants attend to them (see Mackintosh, 1975). Consistent with this suggestion, Livesey & McLaren (2007) demonstrated that stimuli that were better predictors of an outcome became relatively more salient than stimuli that were worse predictors of the outcome over practice (see also Le Pelley & McLaren, 2003)[2]. In other words, previous research indicates that attention and associative learning go hand in hand.

In Experiment 1, the stop-associated images could be considered relatively worse predictors of the stop goal when presented with a stop signal. After all, the stop-associated images were associated with the stop goal (i.e. the outcome in this case) on 75% of the trials, whereas any given stop signal (e.g. the consonants 't' or 'n') was associated with the stop goal on 100% of presentations. Similarly, control images could occur on both go and stop trials. Therefore, attentional accounts of associative learning predict that the images would decrease in salience with exposure; consequently, their contribution to performance would also diminish with increased image exposure (see Le Pelley, Suret & Beesley, 2009). The suggestion that the relative salience of the images diminished during training is also consistent with conflict monitoring accounts (e.g. Botvinick, Braver, Barch, Carter & Cohen, 2001). These accounts predict decreased attention to the images due to response conflict triggered by the inconsistency in the predictability of these images. For instance, Egner & Hirsch (2005) have demonstrated that when response conflict is detected, task-relevant information is amplified. Hence, conflict detection accounts predict that participants should increase their attention to the go/stop signals relative to the task-irrelevant images. Thus, in this regard, the main difference between the associative learning and conflict monitoring accounts is the detailed mechanism by which

the cognitive system adjusts attentional settings. The conflict account requires conflict to drive this change in attention whereas the associability account does not. All the latter requires is that one stimulus (in this case the stop signal itself) has a greater associative strength to the outcome (stopping) than the other stimulus present (the image).

In sum, the findings of Experiment 1 show that participants can acquire direct associations between specific stimuli and the stop goal. However, despite reliable learning effects in the training phase and in expectancy ratings obtained following task completion, we found no evidence of learning in the test phase when the stimulus-stop mappings were reversed. We hypothesize that attention plays a role in determining the influence of stimulus-stop learning on behavior. This idea could put important constraints on current theories of the automaticity of control processes. Therefore, we conducted three more experiments to replicate and extend the findings of Experiment 1, and to explore the role of attention in the influence of stimulus-stop associations on behavior.

Experiment 2

In Experiment 1, we hypothesized that habituation and the predictability of the signal-stop contingency relative to the image-stop contingency decreased the amount of attention that was paid to the stop-associated images over practice. To investigate the predictability hypothesis, in Experiment 2, we manipulated the contingency between the images and stopping, to ensure that the stop-associated images were paired with a stop signal and were predictive of the stop goal on 100% of presentations during training (cf. 75% of presentations in Experiment 1). This should prevent conflict driving down attention, but it would not abolish any associability effects as the stop signal would still tend to be the stimulus with the strongest connection to stopping. All that an associability theory requires for the images to lose attention is that they are worse predictors of the outcome relative to the stop signal(s). This will occur when the stop signal(s) always predicts the outcome whereas the images only predict the stop goal on the trials on which they occur. As a result, image associability will be driven down in a block, and will not have time to recover when the stop signal changes at the beginning of each block.

Method

Subjects. Thirty students from the University of Exeter participated for monetary compensation (£5) or partial course credit (M = 19.97 years, SD = 2.81, 23 females, 27 right-handed). No participants were excluded.

Apparatus, stimuli, procedure, and analyses. The apparatus, stimuli and procedure were identical to those of Experiment 1, except for the following changes: the stop-associated images (10 images) were paired with a stop signal on 100% of trials during the training phase, and were never paired with a stop signal in the test phase; the go-associated images (30 images) were never paired with a stop signal in the training phase, but some of these images were paired with a stop signal in the test phase (20 old go-associated images were never paired with a stop signal; 10 old go-associated images were paired with a stop signal on 100% of the trials). The analyses were identical to those of Experiment 1, except that the contingencies meant that, for obvious reasons, we could not examine the effect of image type on go RTs or *p*(respond|stop) in the training phase of this experiment (see Table 1).

Results

Training phase. In the training phase, the RT for the go-associated images reliably decreased as a function of block ($p = 0.038$). This suggests that participants acquired the stimulus-go associations during the training. The *p*(respond|stop) for the stop-associated images did not reliably decrease as a function of practice (Figure 4, Table 3), which could be due to a floor effect.

Test phase. Contrary to the predictions of the automatic inhibition hypothesis, go RT was not influenced by image type in the test phase when the image-stop mappings were reversed (Table 4). As in Experiment 1, the absence of an effect in the test phase cannot be accounted for by the overall speeding of RTs (for RT distributions, see Supplementary Material).

Expectancy ratings. Despite the absence of an effect of image-stop learning in the test phase, expectancy ratings obtained following task completion revealed a main effect of image type: participants expected to withhold their response more for the stop-associated images (5.99) than for the go-associated images (3.86), $t(29) = -5.17$, $p < 0.001$, $g_{av} = 1.436$. This suggests that participants had learned the image-stop contingencies during training, even though these contingencies did not significantly influence performance in the test phase. The 'stop minus go' image expectancy difference did not significantly correlate with the RT difference in the test phase, $r(28) = 0.262$, $p = 0.162$. Note that the 'stop minus go' expectancy difference was larger in Experiment 2 than in Experiment 1 (in which stop items could occur on 25% of go trials in the training phase), $t(49) = -2.47$, $p = 0.017$, Cohen's $d = 0.644$. In other words, this between-experiment comparison indicates that the image-stop contingency (100% in Experiment 2 relative to 75% in Experiment 1) influenced expectancy ratings but it did not influence performance during the test phase.

Discussion

In Experiment 2, we investigated whether the relative predictability of the stop-associated images influenced the extent to which the acquired stimulus-stop associations influenced task performance when these mappings were reversed. Therefore, the stop-associated images were paired with stopping on 100% of presentations during training (cf. 75% of presentations in Experiment 1).

Consistent with Experiment 1, the decrease in go RT for the go-associated images shows that participants acquired the stimulus associations during training (i.e. they associated the go-associated images with responding), and the expectancy ratings obtained following task completion show that participants expected to stop their responses more for the stop-associated images than for the go-associated images. Furthermore, these expectancy ratings were sensitive to the increased predictability of the stop-associated images as the expectancy difference between stop-associated and go-associated images was larger in Experiment 2 than in Experiment 1. However, as in Experiment 1, RTs were comparable for the old stop-associated images and the old go-associated images in the test phase, which indicates that the acquired associations did not influence performance in the test phase when the image-stop mappings reversed. On the face of it, these results do not support the conflict account of attentional modulation (e.g. Botvinick, Braver, Barch, Carter & Cohen, 2001). However, it is possible that participants quickly learned to ignore the images in the test phase when the mapping had reversed. Consistent with this idea, participants were slower to respond to the stop-associated images (382 ms) than to the go-associated images (376 ms) in the first half of block 13, but this was in the opposite direction in the second half of block 13 (stop-associated images: 374 ms, go-associated images: 380 ms; this reversal could be due to an increased error signal in the first half of the test phase). This suggests that participants may have quickly re-learnt the new mappings in the test phase. Note that we did not conduct any inferential statistics on this difference due to low numbers of trials ($\leq$ 20 trials per cell). An alternative possibility is that participants habituated to the images and stopped paying attention to them because the images were less novel. We tested the habituation hypothesis in Experiment 3.

Experiment 3

The aim of Experiment 2 was to investigate whether the relative predictiveness of the stop-associated images influenced the extent to which the stimulus-stop mappings acquired during training influenced task performance in the test phase. However, even though participants acquired the stimulus-stop mappings, these mappings did not modulate performance in the test phase. It is possible that did not prevent participants 'tuning-out' attention to these images over practice because they became less

novel. Therefore, in Experiment 3 we investigated whether stimulus exposure influenced the extent to which participants attended to the stop-associated images. To this end, we halved the number of stimulus presentations in the training phase, such that there were 12 presentations prior to the test phase (cf. 24 presentations in Experiments 1 & 2).

Method

Subjects. Thirty-two students from the University of Exeter participated for monetary compensation (£5) or partial course credit (M = 19.19 years, SD = 1.49, 26 females, 29 right-handed). One participant was excluded because they incorrectly executed a response on ≥ 30% of stop trials. The data with this participant included are available as Supplementary Material.

Apparatus, stimuli, procedure, and analyses. The apparatus, stimuli and procedure were identical to those of Experiments 1 and 2, except for the following changes: each image was presented once per block (i.e. 14 presentations in total). To ensure that the overall p(stop) was the same as in Experiments 1 and 2, the reduced number of image presentations meant that the stimulus-stop contingencies for the go and the control images in the test phase had to be altered (for the specific contingencies, see Table 1). As in Experiment 1, the stop-associated images were paired with a stop signal on 75% of presentations during the training phase to provide an index of image-stop learning during training. For comparison with Experiments 1 and 2, in the analyses the blocks were collapsed to ensure that the number of observations per cell was comparable.

Results

Training phase. In the training phase, the main effect of image type on go RTs was marginally significant (p = 0.058); planned comparisons revealed marginally significant differences between the stop-associated images (428 ms) and the go-associated images (422 ms), $t(30) = -1.99$, $p = 0.055$, $g_{av} = 0.234$, and between the stop-associated images and the control images (422 ms), $t(30) = -1.92$, $p = 0.064$, $g_{av} = 0.242$. There was no reliable difference between the control and the go-associated images, $t(30) = 0.28$, $p = 0.777$, $g_{av} = 0.017$**.** However, Figure 5 shows that RTs were longer for the stop-associated images than for the control and the go-associated images in blocks 1-3, but this difference disappeared from block 4 onwards. This conclusion was supported by a reliable interaction between image type and block (p = 0.005). The overall main effect of block was reliable, suggesting that participants improved as a function of task practice ($p < 0.001$). There were no reliable differences in p(respond|stop).

Test phase. As in Experiments 1 and 2, there was no main effect of image type on go RT in the test phase ($p = 0.479$). However, the difference in p(respond|stop) between the go-associated images (0.183) and the control images (0.125) was marginally significant, $p = 0.062$, suggesting that the image-go associations did influence test phase performance to some extent (Table 4).

Expectancy ratings. Consistent with the previous experiments, image type influenced expectancy ratings, $F(2, 60) = 11.44$, $p < 0.001$, $gen.\ \eta 2 = 0.136$. Expectancy ratings were greater for the stop-associated images (5.54) than for the go-associated images (4.57), $t(30) = -3.50$, $p = 0.001$, $g_{av} = 0.850$, and the control images (4.76), $t(30) = -3.44$, $p = 0.001$, $g_{av} = 0.687$. There was no reliable difference between the control images and the go-associated images, $t(30) = -1.84$, $p = 0.075$, $g_{av} = 0.199$. However, the expectancy differences did not correlate with the corresponding RT differences (r's ≤ 0.136, p's ≥ 0.464).

Discussion

In Experiment 3, we investigated whether the amount of exposure to the stop-associated images influenced the extent to which the stimulus-stop mappings acquired during training affected task performance in the test phase when the stimulus-stop mappings were reversed.

Consistent with Experiment 1-2, our results indicate that participants acquired the stimulus-stop mappings during training; participants were slower to respond to the stop-associated images than to the go images and the control images. However, this effect appeared and then disappeared again throughout practice; this conclusion was supported by a significant interaction between block and image type. This is consistent with the (numerically) diminished learning effect observed at the end of the training phase in Experiment 1. Furthermore, participants were not slower to respond to the stop-associated images than to the go-associated images and to the control images in the test phase (although we observed a marginally significant difference between go and control images). This suggests that the amount of habituation to the images cannot entirely account for the absence of the test phase effect. This leaves an associability mechanism controlling attention to the stimuli as the most plausible explanation for the results of our experiments so far.

As in Experiments 1-2, we find clear evidence that participants acquired the stimulus-stop contingencies in the expectancy ratings obtained following task completion; participants expected to stop their response more for the stop-associated images than for the go-associated images and the control images. This suggests that participants did not forget the stimulus-stop contingencies, despite the disappearance of the learning effect on task performance towards the end of the training phase and during the test phase.

Experiment 4

In the final experiment, we presented the image before the go and stop signals, and asked participants to rate whether they expected to stop or not. Furthermore, we presented the go and stop signals around the image, at one of four possible locations (one of four corners of the image; for a similar procedure see Houben & Jansen, 2011). These manipulations served two purposes. First, the results of Experiments 1-3 suggested that participants stopped paying attention to the task-irrelevant images. We tried to increase attention to the images by making them perfect predictors of the outcome (Experiment 2) or by decreasing image habituation (Experiment 3). These manipulations were only moderately effective: some behavioral indices indicate that our manipulation influenced learning, but the effect of learning on test performance still disappeared over training. By presenting the images before the go and stop signals, and asking participants to rate their stop expectancy, participants were less likely to ignore the images in Experiment 4 (however, subjects were not explicitly instructed to attend to the images so as to keep the image-stop mappings implicit as in Experiments 1-3). Furthermore, the images initially did not have the stop signal present as a competitor driving their associability down. If our attentional account is correct, we should observe the effects of stop training in the later blocks of the training phase and in the test phase. Second, in Experiments 1-3, we found that participants generated expectancies based on the image-stop associations acquired during training. In Experiment 1, expectancy correlated with some aspects of performance in the test phase, but we could not replicate this finding in Experiments 2-3. It is possible that obtaining the expectancy ratings following task performance meant that these expectancies were contaminated by the re-learning of the new (inconsistent) mappings in the test phase. Therefore, in Experiment 4, we further investigated the role of expectancy in stimulus-stop learning by obtaining expectancy ratings during task performance (for a similar procedure, see e.g. McAndrew, Jones, McLaren & McLaren, 2012; Perruchet, Cleeremans & Destrebecqz, 2006).

Method

Subjects. Thirty-two students from the University of Exeter participated for partial course credit (M = 18.47 years, SD = 0.62 years, 27 females, 31 right-handed). Four participants were excluded because they incorrectly executed a response on ≥ 30% of stop trials. The data with these participants included are available as Supplementary Material.

Apparatus, stimuli, procedure & analyses. The apparatus, stimuli and procedure were identical to those of Experiment 3, except for the following changes: All trials began with the

presentation of the image in the centre of the screen. The word ‘RATING’ was presented above and below the image to instruct participants to rate ‘how much do you expect to withhold your response?’. Participants inputted their ratings on a scale between 1 ('I definitely do not think that I will have to withhold my response') and 9 ('I definitely think that I will have to withhold my response') using the number keys of the keyboard with their right index finger (latency rating response: M = 969 ms; sd = 681 ms). After participants made their expectancy rating, a go/stop signal appeared at one of four locations on the screen (top-left, bottom-left, top-right, or bottom-right corner of the image). The delay between the expectancy response and the presentation of the go/stop signals varied randomly between 500 and 1250 ms. Participants responded on go trials by pressing the spacebar on a keyboard with their left index finger. To allow for the presentation of the signals at each location on the screen, task rules used in Experiments 1-3 that were based on signal location (e.g. ‘X on the left/right of the image’) or signal shape (e.g. ‘shape bigger/smaller than a fifty pence piece’) were excluded and, of the remaining rules, seven rules were selected on the basis of response latencies in Experiments 1-3 using a non-parametric box and whisker method (Tukey, 1977). A full list of the signals used appears in Appendix A. The expectancy ratings data in the training and test phase trials were analyzed separately using ANOVAs with image type and block as within-subjects factors.

Results

Training phase. In the training phase, there was a reliable interaction between image type and block on go RTs ($p = 0.03$), reflecting slower responding for the stop-associated images than for the go-associated images and the control images in the second half of the training phase (see Figure 6). The p(respond|stop) was also lower for the stop-associated images (0.152) than for the control images (0.185) ($p = 0.011$). The interaction between image type and block in the p(respond|stop) was not reliable.

The analysis of the online expectancy ratings also revealed a reliable image type by block interaction ($p = 0.005$), reflecting higher stopping expectancies for the stop-associated images in the second half of the training phase (blocks 4-6; see Figure 6C). There was also a reliable main effect of block on the expectancy ratings ($p = 0.012$): overall mean expectancy ratings decreased with task practice, which is consistent with the overall p(stop) of 0.25 (note, the increase in expectancy ratings across block for the stop-associated images was not reliable, $p = 0.261$). Combined, these findings indicate that participants were generating appropriate expectancies during the acquisition of the stimulus-stop mappings. Importantly, the overall ‘stop minus go’ expectancy ratings difference reliably correlated with the corresponding RT difference in the training phase, $r(26) = 0.575$, $p =$

0.001; the overall 'stop minus control' expectancy ratings difference also correlated with the corresponding RT difference, $r(26) = 0.498$, $p = 0.006$.

Test phase. Unlike in Experiments 1-3, we found a main effect of image type on go RTs in the test phase ($p = 0.004$). Planned comparisons revealed that responding to the old stop-associated images was slower (443 ms) than to the go-associated images (422 ms), $t(27) = -2.84$, $p = 0.008$, $g_{av} = 0.517$, and to the control images (424 ms), $t(27) = -2.87$, $p = 0.007$, $g_{av} = 0.542$. There was no reliable difference between the go and control images, $t(27) = -0.31$, $p = 0.756$, $g_{av} = 0.040$, (Figure 6; Table 3). Image type did not reliably influence p(respond|stop) in the test phase (however, the means were in the predicted direction, see Figure 6; Table 4).

There was also a reliable main effect of image type on test phase expectancies ($p = 0.002$); planned comparisons revealed that participants expected to stop more for the old stop-associated images (4.80) than for the go-associated images (3.86), $t(27) = -2.65$, $p = 0.013$, $g_{av} = 0.807$, and the control images (4.01), $t(27) = -2.83$, $p = 0.008$, $g_{av} = 0.719$. There was no reliable difference between the go-associated and the control images, $t(27) = -1.37$, $p = 0.181$, $g_{av} = 0.143$. As in the training phase, we found that the 'stop minus go' expectancy ratings difference reliably correlated with the corresponding RT difference, $r(26) = 0.624$, $p < 0.001$; the 'stop minus control' expectancy ratings difference also correlated with the corresponding RT difference, $r(26) = 0.653$, $p < 0.001$. Hence, participants who had a stronger expectancy to stop their response when the stop-associated images were presented displayed greater response slowing for these images than for the go-associated images and for the control images upon signal presentation.

To further investigate to what extent the expectancy to stop determined response slowing for the stop-associated images, we conducted a median-split analysis on the expectancy ratings of the test phase (we could not perform a similar analysis in the training phase because there were not enough trials in each block). We calculated the median for each image type and participant separately. Ratings greater than the median were classified as a 'stop' expectancy whereas ratings less than or equal to the median were classified as a 'go' expectancy. Four participants were excluded from these analyses as they always entered the same expectancy rating for one or more of the image types (consequently, we could not perform a median split). We analyzed the data with a 2 (expectancy: stop vs. go) by 3 (image type) ANOVA. Consistent with previous work on proactive control (see e.g. Verbruggen & Logan, 2009), responding was slower for trials on which participants expected a stop signal (445 ms) compared with trials on which participants expected a go signal (420 ms), $F(1, 23) = 13.96$, $p = .001$, $gen.\ \eta 2 = .088$. As discussed above, image type also had a reliable main effect on performance. Importantly, the effects of stimulus-stop learning and expectancy were additive; i.e. the two-way

interaction between expectancy and image-type was not reliable, $F(2, 46) = .08$, $p = .915$, gen. $\eta 2 < .001$ (for descriptive statistics, see Table 5). Thus, the slowing for the stop-associated images is unlikely to reflect an entirely strategic, expectancy-driven effect.

Discussion

Consistent with the results of Experiments 1-3, we find evidence that participants acquired the stimulus-stop associations. In the training phase, responding became slower for the stop-associated images than for the go-associated images and the control images with task practice, and the *p*(respond|stop) was lower for the stop-associated images than for the control images. In addition, the expectancy ratings showed that participants generated expectancies that were consistent with the trained stimulus-stop contingencies in the second half of the training phase. These expectancies correlated with task performance in the training phase: participants who expected to withhold their response more to the stop-associated images responded more slowly to these images than to the go-associated images and to the control images during training. Unlike in Experiments 1-3, we find that learning also influenced performance in the test phase: participants were slower to respond to the stop-associated images than to the go-associated images and the control images during the test phase.

Our results suggest that presenting the images before the go/stop signals and asking participants to rate their expectancy on each trial increased the extent to which participants attended to these images. In order to ensure that attention to the task-irrelevant images was maximized, we combined these manipulations in the same procedure. As a consequence, we cannot determine the relative contributions of these manipulations to the observed slowing for the stop-associated images in the test phase. One could speculate that the observed slowing reflects an entirely strategic, expectancy-driven effect, rather than the implicit retrieval of the acquired stimulus-stop associations (as predicted by the automatic inhibition account). We argue that this explanation is unlikely for several reasons. First, our median split analysis on expectancy ratings in the test phase shows that the slowing for the stop-associated images occurred even when stop signal expectancy was relatively low. This result suggests that expectancy ratings cannot account for the whole data pattern[3]. Second, previous studies have demonstrated stop-learning effects using procedures in which the stop-associated stimuli are presented prior to stop-signal onset but, unlike the present experiment, without expectancy ratings on each trial. For example, in a recent study we presented the stop-associated stimuli as 'warning cues' for a variable duration prior to the presentation of the stop signal, and observed stop learning effects during the training and test phases (Bowditch, Verbruggen & McLaren, 2015). Similarly, Veling and colleagues have conducted two experiments using go/no-go designs in

which food images were presented 100 ms (Veling, van Koningsbruggen, Aarts & Stroebe, 2014) or 500 ms (Veling, Aarts & Stoebe, 2013) prior to the onset of the go/no-go signal. They found that when the food images were consistently presented on no-go trials, subsequent choice of the food items was reduced (Veling, Aarts & Stroebe, 2013) and weight loss was facilitated (Veling et al., 2014). Finally, research in the wider action control literature is consistent with the pattern of findings in the present study. For example, Frings and Moeller (2012) found that associations between old distractor stimuli and the previously required target response only interfered with responding when the distractors were presented prior to the target stimuli. Combined, these studies suggest that presenting the task-irrelevant image before the go or no-go signal increases attention to the images, and consequently, the probability that the image-stop association are retrieved. However, future research is required to determine the relative contributions of increased attention and expectancies (see e.g. Best, Stevens, et al., 2015; Footnote 3).

To conclude, the presence of a learning effect in the test phase is consistent with our hypothesis that attention to the images determines whether acquired stimulus-stop associations influence behavior in the test phase. Now that the images are task-relevant and associability is no longer driven down for the images by virtue of their competition for attention with the stop signal, we see a strong effect on test phase go RTs. Furthermore, the test-phase expectancy ratings show that participants continued to generate expectancies consistent with the image-stop mappings acquired during training, despite the reversal of these mappings. As in the training phase, these expectancies reliably correlated with task performance: participants who expected to withhold their response more for the stop-associated images responded more slowly to these images than to the go-associated images and to the control images in the test phase. However, the median split also suggested a contribution of implicit (non-expectancy related) processes.

General Discussion

In the present study, we investigated three outstanding issues relating to the mechanisms of stimulus-stop learning. The first two issues were highlighted in our recent review on stimulus-stop learning (Verbruggen, Best, et al., 2014): (1) are associations between stimuli and stopping direct, and (2) what is the role of expectancy in stimulus-stop learning? Based on the results of Experiment 1, Experiments 2-4 also investigated a third issue: (3) does attention to the stop items affect the extent to which stimulus-stop learning influences behavior? Based on our findings, we can answer each of these questions.

Are associations between stimuli and stopping direct?

Across four experiments where the specific stop signals and rules were always changing, we provide strong evidence for the idea that participants can learn direct stimulus-stop associations (Verbruggen & Logan, 2008b). During training, we found that responding was slower (Experiments 1, 3, and 4; in Experiment 2, we could not compare stop and go-associated images in the training phase) and the *p*(respond|stop) was lower (Experiment 1 & Experiment 4) for images that were consistently associated with stopping than for images associated with going and for control images that were not particularly associated with stopping or going.

In recent experiments, we have observed that learning can influence the *p*(respond|stop) but not response latencies on go trials (see e.g. Experiment 2 in Verbruggen, Best, et al., 2014). Based on previous findings in the conditioning literature (for a review, see Hall, 2002), we hypothesized that participants in these experiments learned an association between an item and a representation of a no-go or stop signal. Hence, when the item was repeated, it primed the signal so that it was detected sooner on stop-signal trials, resulting in improved response inhibition and, consequently, a lower *p*(respond|stop). The signal priming idea explains why it can be that learning influences the probability of stopping on signal trials without influencing response latencies on go trials. In the present study, both RTs and *p*(respond|stop) were influenced even though the go/stop signals and task rules constantly changed (and there were two go signals and two stop signals in each block). This indicates that learning was not (solely) mediated via image-signal associations. The most parsimonious account is that the effects in the present study reflect the direct association of the stop-associated images with a stop goal, rather than the association of the stop-associated images with the representation of a single stop signal. Therefore, the present study provides the strongest evidence to date for the original automatic inhibition hypothesis of stimulus-stop (goal) learning. In situations where the task rules do not constantly change, it is likely that individuals will acquire both stimulus-goal and stimulus-signal associations (indeed, research in the conditioning literature suggests that the acquisition of multiple associations is the norm; Hall, 2002). It is possible that experimental factors, such as the perceptual properties of the stop signal, will influence which association dominates behavior.

It is important to note that the learning effects demonstrated in the present study are assumed to reflect the acquisition of stimulus-stop associations rather than the absence of stimulus-go learning on stop trials. Whilst the 'absence of go learning' explanation may initially seem parsimonious, it cannot account for several findings previously reported in the stop-learning literature. First, we have previously demonstrated that responding to old stop items is slowed compared with novel items that

were not presented during training (hence, these items were not associated with going or stopping; Verbruggen & Logan, 2008b, Exp 1). Second, neuroimaging work has shown that the presentation of old stop items activates the neural inhibitory control network (Lenartowicz et al., 2011; but see also below). Third, brain stimulation studies have shown that even when the probability of go and no-go signals is equal (i.e. 50/50), motor-evoked potentials are below baseline 200-300 ms following no-go stimulus presentation (indicating that responding is suppressed; Leocani, Cohen & Wassermann, 2000). In other words, successful performance on a no-go trial requires the activation of a no-go or stop response, and not just the absence of a go response. Fourth, short-term after-effects of stopping further support the idea that participants can learn stimulus-stop associations that can have a (global) inhibitory effect on responding (Giesen & Rothermund, 2014). Finally, in the present experiments, response latencies decrease for go and control images but we observe an initial increase in response latencies for stop-associated images over practice (Experiment 1). In Experiments 3 and 4, this conclusion is further supported by a reliable interaction between image type and block. Finally, the comparison of expectancy ratings in Experiments 1 and 2 revealed that expectancy ratings were altered when the image-stop consistencies had changed (even though the image-go contingencies did not change). Therefore, previous results and the findings reported in the present study are consistent with the idea that participants can learn go associations on go trials and stop associations on stop trials (which interfere with responding).

What is the role of expectancy in stimulus-stop learning?

In the present study, we show that participants generated expectancies that were consistent with the stimulus-stop mappings acquired during training: participants expected to withhold their responses more when stop-associated images were presented than when go and control images were presented. Furthermore, these expectancy ratings were sensitive to the specific contingencies in play: participants expected to withhold their responses more for the stop-associated images that were reinforced on 100% of presentations (Experiment 2) than for the stop-associated images that were reinforced on 75% of presentations (Experiment 1). Finally, we found that these expectancies correlated with task performance both during the acquisition of the stimulus-stop mappings in the training phase (Experiment 4) and following the reversal of these mappings in the test phase (Experiment 1 & Experiment 4).

The role of expectancies in stimulus-stop learning has not been previously investigated. Therefore, the present study provides the first evidence that stimulus-stop learning is partly mediated via explicit knowledge of the stimulus-stop contingencies in play (although the median split analysis and the absence of significant correlations in some of the experiments indicate that implicit processes

must play a role as well). This could indicate that the response slowing observed for the stop-associated images is caused by top-down control processes. First, the slowing could be partly due to proactive control. According to this proactive control account of stimulus-stop learning, stop items could become predictive cues (e.g. if image X then *p*(stop) is high) that indicate that participants should adjust their response strategies accordingly. If this were the case, this would suggest that earlier findings that have demonstrated response slowing and neural activation of the inhibitory control network by old stop items (Lenartowicz et al., 2011) could be due to proactive control (i.e. another algorithmic process), rather than the direct activation of the stop response via memory-retrieval (i.e. picture X = stop). Therefore, whilst the *retrieval* of the stimulus-stop association may still be automatic, the subsequent *slowing* observed following the reversal of the stimulus-stop mapping would be due to a top-down control process (rather than a bottom-up process as is currently assumed). Second, stop items could effectively become a new stop signal (the direct stopping account). In other words, the only difference between the stop items and an external stop signal is that the association with stopping is acquired via learning in the case of the stop items, whereas it is acquired via instructions in the case of the stop signal. Thus, in both cases, response inhibition is a deliberate act of control. But the advantage of the former form of control is that the go and stop processes in stop-signal tasks could be initiated simultaneously and, therefore, start the race at the same time (Logan & Cowan, 1984); consequently, response inhibition is more likely to succeed.

It is important to note, however, that the proactive control route and the direct stopping route are both compatible with the idea that associative learning plays a key role in response inhibition paradigms; indeed, both accounts still assume that stimulus-specific learning influences stop performance. Learning offers participants another route to control their behavior. The key difference between these two top-down accounts and the 'automatic' inhibition account is the nature of the process that occurs following the retrieval of the stimulus-stop association; either this association directly activates the stop goal via an S-R based link (in the automatic stopping account) or this association indirectly activates the stop goal via a top-down (algorithmic and deliberate) control process. Future research is required to distinguish between these accounts (see e.g. Best, Stevens, et al., 2015).

Does attention to the stop items affect the extent to which stimulus-stop learning influences behavior?

In Experiments 1-3, the acquired stimulus-stop associations did not influence performance in the test phase, despite effects of learning on task performance in the training phase and on expectancies

following task completion (suggesting that participants had not forgotten the stimulus-stop associations).

A potential explanation for this finding is that the images used in the present study were task-irrelevant so participants may have begun to ignore the images as they became less novel and as they learned that they were less predictive. In Experiments 1-3, task performance did not require participants to attend to the stop-associated stimuli (unlike in our previous work; see e.g. Verbruggen & Logan, 2008b), so participants may have started ignoring all the images over time. In line with this possibility, the effect of image type reliably interacted with block (in Experiment 3) and visual inspection of the data shows that the influence of image-stop learning on performance began to disappear at the end of the training phase (Experiment 1). Since there were no differences between the image types in the final block of the training phase, this may explain why we did not find any effect of image-stop learning in the test phase[4]. Several associative learning accounts suggest that the reduced predictiveness of the images relative to the go/stop signals (in Experiments 1 & 3) may have decreased the extent to which they were considered informative or salient and, consequently, the extent to which participants attended to them and the extent to which they can influence performance (Mackintosh, 1975). Effects that point to this conclusion have been previously observed in animals (see e.g. Sutherland & Mackintosh, 1971 for a review of this literature) and, importantly, also in humans (Le Pelley & McLaren, 2003; Livesey & McLaren, 2007; Suret & McLaren, 2005). For example, Le Pelley and McLaren (2003) showed that foods that were worse predictors of an outcome than other foods present on a trial in an allergy discrimination task became less salient, resulting in slower learning of a new association to these stimuli in a later training phase (cf. learned irrelevance; Mackintosh, 1975). Note that the majority of our results of Experiments 1-3 are also consistent with conflict monitoring accounts (e.g. Botvinick, Braver, Barch, Carter & Cohen, 2001), which predict that participants will ignore task-irrelevant information that produces response conflict or choice errors. However, unlike the associative learning accounts, these conflict monitoring accounts do not easily explain the absence of a learning effect in the test phase found in Experiment 2 when conflict should have been minimized by the use of 100% contingencies.

It is possible that the use of neutral images increased the extent to which participants began to 'tune out' their attention. Motivationally-salient images capture attention even if they are task-irrelevant (e.g. Anderson, Laurent & Yantis, 2011). Consequently, if task-irrelevant, but motivationally-salient images are used as the stop-associated stimuli, the attentional capture to the images would be increased, and the 'tuning out' of attention could be slowed. Thus, the salience of

task-irrelevant stop-associated stimuli could be a key consideration for applied studies examining the effects of no-go training effects on food and alcohol consumption.

Importantly, we found a clear effect of stimulus-stop learning on test phase performance when attention to the images was increased in Experiment 4 (as a result of presenting them before the go/ stop signals and the requirement to make an online expectancy rating on each trial). This finding is consistent with the Instance Theory (Logan, 1988; Logan & Etherton, 1994) and other theories of associative learning. For example, Instance Theory suggests that processing episodes will only be stored and retrieved from memory when participants attend to each stimulus presentation (Logan, 1988; Logan & Etherton, 1994). Thus, by encouraging subjects to attend to the image in Experiment 4, the image-stop associations were more likely to be retrieved, and performance was influenced in the test phase. Therefore, the present study strongly indicates that the influence of image-stop learning on behavior is likely to be determined by the interplay of both attentional control and associative learning systems (see also Logan, 1988; Verbruggen, McLaren & Chambers, 2014).

Wider implications

In addition to contributing to our theoretical understanding of stimulus-stop learning, the present study has implications for more applied research. First, our results indicate that attentional settings influence learning in response inhibition tasks. Even when salient images are used as stimuli (e.g. as in the food studies mentioned above), participants may still adjust their attentional settings, and ignore the images to a certain degree. Currently, the task-relevance of the stop-associated images used in current stop-training studies varies. Whilst the task-relevance of the images may not influence engagement in impulsive behaviors (e.g. impulsive eating can be prompted by implicit processing of food cues in the environment), our results suggest that designs in which participants must attend to the images should produce 'stronger' stimulus-stop associations that will have a more pronounced influence on stop-learning. Second, the present study indicates that it is possible to learn a direct association between a stimulus and a stop 'goal' or the act of withholding a response when multiple signals are used. When only one signal is used, there is the possibility that participants will learn stimulus-signal associations (as our recent results suggest; see above). Thus, if the aim is to obtain 'inhibition' training effects that transfer to real-world settings where stop signals are no longer present, multiple signals may be preferable.

In order to maximize the inhibitory control training effects, it is important to consider other features of the stop learning task as well. In the present study, we devised a novel task that combined features of the go/no-go task and the stop-signal task. In Experiments 1-3, the delay between the

presentation of the images and the go/stop signal was zero ms; in Experiment 4, the go and stop signal also occurred at the same moment (i.e. there was no delay between the go and the stop signal). But to avoid that subjects would simply wait on all trials, we used a low overall proportion of stop trials (. 25), imposed a relatively strict response deadline (750 ms) and provided feedback if the participant did not respond in time. We believe that this hybrid design is optimal to investigate stop learning it allows us to manipulate the go/stop signal representation whilst maximizing the number of correct stop trials. After all, our previous work indicates that stimulus-stop associations are less likely to be learned when inhibition is unsuccessful (Verbruggen & Logan, 2008a, 2008b; Verbruggen et al., 2008). The idea that the stop outcome is important is further supported by studies in the applied domain. Stop learning effects on task performance and on food- and alcohol consumption have been observed after both go/no-go and stop-signal training (see above). However, a recent meta-analysis indicates that go/no-go training has stronger effects on appetitive behavior than stop training (Jones et al., 2015). This could be due to generally higher success rates in the go/no-go task (Jones et al., 2015; Verbruggen & Logan, 2008b). Neuroimaging research also shows that, despite some overlap, there may be several differences in the neural substrates of the go/no-go and stop-signal tasks (for a discussion; see e.g. Eagle, Bari & Robbins, 2008; Swick, Ashlet & Turken, 2011). Thus, it is possible that the differences between the training protocols could be due to other factors as well. Future research is required to investigate the specific action control processes influenced by stop learning in these tasks.

Our results indicate that expectancies also play a role in stop-learning paradigms. It is possible that differences in the expectancy to stop are present in applied studies, especially as the go/stop rule is typically simpler (and remains the same throughout the task), the image-stop mappings are more explicit, and the stimulus set smaller than in the present study. In applied studies, expectancies and demand characteristics may play an important role (Boot et al, 2013). However, it is currently unclear the extent to which the expectancy effects observed in the present study relate to the demand characteristics identified by Boot and colleagues (2013) and, indeed, the behavioral findings of applied stop-training studies. For example, there are some procedural differences between the present study and stop-training studies (e.g. in Experiment 4, we obtained an expectancy rating on every trial). Similarly, whilst our results show a relationship between expectancy and go RT, it is unclear the extent to which expectancy equates to other dependent variables used in the stop-training studies, such as food intake or stimulus devaluation (see e.g. Wessel, O'Doherty, Berkebile, Linderman, Aron, 2014). For example, a recent stop-training study from our lab suggests that whilst a substantial proportion of participants became aware of the stimulus-stop contingencies during training (in a

funneled debrief, 83% of participants in Experiment 1 and 74% of participants in Experiment 2 reported knowledge that specific images were associated with stopping), the majority of participants did not expect these image-stop associations to influence their subsequent food intake (Lawrence et al., 2015). Nevertheless, we believe that future applied research should include a dependent variable that is sensitive enough (see Newell & Shanks, 2014), such as expectancy ratings (e.g. Stothart, Simons, Boot & Kramer, 2014), to examine the extent to which the behavioral effects observed both during and following inhibitory control training relate to the expectancy to stop.

Conclusion

In sum, the present findings indicate that participants can learn direct associations between stimuli and a stop goal when the go/stop rule changes at the beginning of each block. Exposure to the image-stop associations influenced task performance during training, and expectancies following task completion. However these results also suggest that attention to stimulus attributes is key for retrieval of processing episodes; if participants do not attend to the stop stimulus then the previously acquired stimulus-stop associations will not influence behavior. Our results are consistent with the Instance Theory and other attentional accounts of associative learning.

References

Anderson, B.A. & Folk, C. (2013). Conditional Automaticity in Response Selection Contingent Involuntary Response Inhibition With Varied Stimulus-Response Mapping. *Psychological Science*, *25*(2), 547–554.

Anderson, B.A., Laurent, P.A., & Yantis, S. (2011). Value-driven attentional capture. *Proceedings of the National Academy of Sciences USA, 108,* 10367-10371.

Anselme, P., Robinson, M. J. F. & Berridge, K. C. (2013). Reward uncertainty enhances incentive salience attribution as sign-tracking. *Behavioural Brain Research*, *238*, 53–61.

Bechara, A., Noel, X. & Crone, E. (2006). Loss of willpower: Abnormal neural mechanisms of impulse control and decision making in addiction. In R. W. Wiers & A. W. Stacy (Eds.), *Handbook of implicit cognition and addiction* (pp. 225–232). Thousand Oaks, CA: Sage Publications.

Bissett, P. G. & Logan, G. D. (2011). Balancing cognitive demands: control adjustments in the stop-signal paradigm. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*(2), 392–404.

Boot, W. R., Simons, D. J., Stothart, C.R. & Stutts, C. (2013). The Pervasive Problem With Placebos in Psychology: Why Active Control Groups Are Not Sufficient to Rule Out Placebo Effects. *Perspectives on Psychological Science*, *8*(4), 445–454.

Botvinick, M. M., Braver, T. S., Barch, D. M., Carter, C. S. & Cohen, J. D. (2001). Evaluating the demand for control: Anterior cingulate cortex and conflict monitoring. *Psychological Review*, *108*(3), 624–652.

Boucher, L., Palmeri, T. J., Logan, G. D. & Schall, J. D. (2007). Inhibitory control in mind and brain: an interactive race model of countermanding saccades. *Psychological Review*, *114*(2), 376–97.

Bowditch, W. A., Verbruggen, F., & McLaren, I. P. L. (2015). Associatively-Mediated Stopping: Training Stimulus-Specific Inhibitory Control. Manuscript Submitted for Publication.

Bowley, C., Faricy, C., Johnstone, S. J. & Smith, J. L. (2013). The effects of inhibitory control training on alcohol consumption, implicit alcohol-related cognitions and brain electrical activity. *International Journal of Psychophysiology*, *89*, 342–348.

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, *10*(4), 433–436.

Crews, F. T. & Boettiger, C. A. (2009). Impulsivity, frontal lobes and risk for addiction. *Pharmacology, Biochemistry, and Behaviour*, *93*(3), 237–247.

De Wit, H. (2009). Impulsivity as a determinant and consequence of drug use: a review of underlying processes. *Addiction Biology*, *14*, 22–31.

Diamond, A. (2013). Executive functions. *Annual Review of Psychology*, *64*, 135–68.

Donders, F. C. (1969). On the speed of mental processes. In W. G. Koster (Ed.), *Attention and performance II* (pp. 412–431). Amsterdam: North-Holland.

Eagle, D.M., Bari, A., & Robbins, T.W. (2008). The neuropharmacology of action inhibition: cross species translation of the stop-signal and go/no-go tasks. *Psychopharmacology (Berl.), 199(3)*, 439-456.

Egner, T. & Hirsch, J. (2005). Cognitive control mechanisms resolve conflict through cortical amplification of task-relevant information. *Nature Neuroscience*, *8*(12), 1784–90.

Enticott, P. G., Bradshaw, J. L., Bellgrove, M. a, Upton, D. J. & Ogloff, J. R. P. (2009). Stop task after-effects: the extent of slowing during the preparation and execution of movement. *Experimental Psychology*, *56*(4), 247–251.

Fernie, G., Peeters, M., Gullo, M. J., Christiansen, P., Cole, J. C., Sumnall, H. & Field, M. (2013). Multiple behavioural impulsivity tasks predict prospective alcohol involvement in adolescents. *Addiction*, *108*(11), 1916–1923.

Frings, C. & Moeller, B. (2012). The horserace between distractors and targets: Retrieval-based probe responding depends on distractor-target asynchrony. *Journal of Cognitive Psychology, 25*(5), 582-590.

Garavan, H. & Stout, J. C. (2005). Neurocognitive insights into substance abuse. *Trends in Cognitive Sciences*, *9*(4), 195–201.

Gaschler, R. & Nattkemper, D. (2012). Instructed task demands and utilization of action effect anticipation. *Frontiers in Psychology*, *3*, 1-14.

Giesen, C. & Rothermund, K. (2014). You better stop! Binding "stop" tags to irrelevant stimulus features. *Quarterly Journal of Experimental Psychology (2006)*, *67*(4), 809–832.

Hall, G. (2002). Associative structures in Pavlovian and instrumental conditioning. In C. R. Gallistel (Ed.), *Stevens' handbook of experimental psychology* (3rd ed., pp. 1–45). New York: John Wiley & Sons.

Henson, R. N., Eckstein, D., Waszak, F., Frings, C. & Horner, A. J. (2014). Stimulus-response bindings in priming. *Trends in Cognitive Sciences*, *18*(7), 376–384.

Hogarth, L., Dickinson, A., Austin, A., Brown, C. & Duka, T. (2008). Attention and expectation in human predictive learning: the role of uncertainty. *Quarterly Journal of Experimental Psychology (2006)*, *61*(11), 1658–1668.

Hommel, B. (1998). Event files: Evidence for automatic integration of stimulus-response episodes. *Visual Cognition*, *5*(1/2), 183–216.
Hommel, B. (2004). Event files: feature binding in and across perception and action. *Trends in Cognitive Sciences*, *8*(11), 494–500.
Hommel, B., Proctor, R. W. & Vu, K.-P. L. (2004). A feature-integration account of sequential effects in the Simon task. *Psychological Research*, *68*(1), 1–17.
Houben, K. (2011). Overcoming the urge to splurge: influencing eating behaviour by manipulating inhibitory control. *Journal of Behaviour Therapy and Experimental Psychiatry*, *42*(3), 384–388.
Houben, K. & Jansen, A. (2011). Training inhibitory control. A recipe for resisting sweet temptations. *Appetite*, *56*(2), 345–349.
Houben, K., Nederkoorn, C., Wiers, R. W. & Jansen, A. (2011). Resisting temptation: decreasing alcohol-related affect and drinking behaviour by training response inhibition. *Drug and Alcohol Dependence*, *116*(1-3), 132–136.
Jahfari, S., Verbruggen, F., Frank, M. J., Waldorp, L. J., Colzato, L., Ridderinkhof, K. R. & Forstmann, B. U. (2012). How preparation changes the need for top-down control of the basal ganglia when inhibiting premature actions. *The Journal of Neuroscience*, *32*(32), 10870–10878.
Jasinska, A. J. (2013). Automatic inhibition and habitual control: alternative views in neuroscience research on response inhibition and inhibitory control. *Frontiers in Behavioural Neuroscience*, *7*, 1–4.
Jones, A. & Field, M. (2013). The effects of cue-specific inhibition training on alcohol consumption in heavy social drinkers. *Experimental and Clinical Psychopharmacology*, *21*, 8–16.
Jones, A., Di Lemma, L. C. G., Robinson, E., Christiansen, P., Nolan, S., Tudur-Smith, C., & Field, M. (2015). Inhibitory control training for appetitive behaviour change: A meta-analysis . Manuscript Submitted for Publication.
Koch, I. & Allport, A. (2006). Cue-based preparation and stimulus-based priming of tasks in task switching. *Memory & Cognition*, *34*(2), 433–444.
Kruschke, J. K. (2003). Attention in learning. *Current Directions in Psychological Science*, *12*(5), 171–175.
Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: A practical primer for t-tests and ANOVAs. *Frontiers in Psychology*, 4:863.
Lawrence, N. S., Verbruggen, F., Morrison, S., Adams, R. C. & Chambers, C. D. (2015). Stopping to food can reduce intake. Effects of stimulus-specificity and individual differences in dietary restraint. *Appetite*, *85C*, 91–103.
Le Pelley, M. E. & McLaren, I. P. L. (2003). Learned associability and associative change in human causal learning. *The Quarterly Journal of Experimental Psychology. B, Comparative and Physiological Psychology*, *56*(1), 68–79. doi:10.1080/02724990244000179
Le Pelley, M. E., Suret, M. B. & Beesley, T. (2009). Learned predictiveness effects in humans: a function of learning, performance, or both? *Journal of Experimental Psychology: Animal Behaviour Processes*, *35*(3), 312–327.
Lenartowicz, A., Verbruggen, F., Logan, G. D. & Poldrack, R. A. (2011). Inhibition-related activation in the right inferior frontal gyrus in the absence of inhibitory cues. *Journal of Cognitive Neuroscience*, *23*(11), 3388–3399.
Leocani, L., Cohen, L. & Wassermann, E. (2000). Human corticospinal excitability evaluated with transcranial magnetic stimulation during different reaction time paradigms. *Brain*, *123*, 1161–1173.
Livesey, E. J. & McLaren, I. P. L. (2007). Elemental associability changes in human discrimination learning. *Journal of Experimental Psychology: Animal Behaviour Processes*, *33*(2), 148–159.
Logan, G. (1988). Toward an Instance Theory of Automatization. *Psychological Review*, *95*(4), 492–527.
Logan, G. (1990). Repetition Priming and Automaticity: Common underlying mechanisms? *Cognitive Psychology*, *35*, 1–35.
Logan, G. D. & Cowan, W. B. (1984). On the ability to inhibit thought and action: A theory of an act of control. *Psychological Review*, *91*(3), 295–327.
Logan, G. D. & Etherton, J. L. (1994). What Is Learned During Automatization ? The Role of Attention in Constructing an Instance. *Journal of Experimental Psychology: Learning Memory, and Cognition*, *20*(5), 1022–1050.
Logan, G. D., Van Zandt, T., Verbruggen, F. & Wagenmakers, E.-J. (2014). On the ability to inhibit thought and action: general and special theories of an act of control. *Psychological Review*, *121*(1), 66–95.
Logan, G. D., Yamaguchi, M., Schall, J. D., & Palmeri, T. J. (2015). Inhibitory control in mind and brain 2.0: Blocked-Input models of saccadic countermanding. *Psychological Review*, 122, 115-147. doi:10.1037/a0038893
Mackintosh, N.J. (1975). A Theory of Attention: Variations in the Associability of Stimuli with Reinforcement. *Psychological Review*, *82*(4), 276–298.

MacLeod, C. (1991). Half a Century of Research on the Stroop Effect: An Integrative Review. *Psychological Bulletin*, *109*(2), 163–203.
Manuel, A. L., Bernasconi, F. & Spierer, L. (2013). Plastic modifications within inhibitory control networks induced by practicing a stop-signal task: an electrical neuroimaging study. *Cortex*, *49*(4), 1141–1147.
Manuel, A. L., Grivel, J., Bernasconi, F., Murray, M. M. & Spierer, L. (2010). Brain dynamics underlying training-induced improvement in suppressing inappropriate action. *The Journal of Neuroscience*, *30*(41), 13670–13678.
McAndrew, A., Jones, F. W., McLaren, R. P. & McLaren, I. P. L. (2012). Dissociating expectancy of shock and changes in skin conductance: an investigation of the Perruchet effect using an electrodermal paradigm. *Journal of Experimental Psychology: Animal Behaviour Processes*, *38*(2), 203–208.
McLaren, I. P. L., Forrest, C. L. D., McLaren, R. P., Jones, F. W., Aitken, M. R. F. & Mackintosh, N. J. (2014). Associations and propositions: the case for a dual-process account of learning in humans. *Neurobiology of Learning and Memory*, *108*, 185–195.
McLaren, I. P. L., Green, R. & Mackintosh, N. J. (1994). Animal learning and the implicit/explicit distinction. In N. C. Ellis (Ed.), *Implicit and explicit learning of languages* (pp. 313–332). London: Academic Press.
McLaren, I. P. L., Wills, A. J. & Graham, S. (2010). Attention and perceptual learning. In C. Mitchell & M. Le Pelley (Eds.), *Attention and associative learning: From brain to behaviour* (pp. 131–158). Oxford: Oxford: Oxford University Press.
Mitchell, C., De Houwer, J. & Lovibond, P. (2009). The propositional nature of human associative learning. *Behavioural and Brain Sciences, 32*, 183–246.
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., Murray Thomson, W., & Caspi, A. (2011). A gradient of childhood self-control predicts health, wealth, and public safety. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(7), 2693–2698.
Nederkoorn, C., Jansen, E., Mulkens, S. & Jansen, A. (2007). Impulsivity predicts treatment outcome in obese children. *Behaviour Research and Therapy*, *45*(5), 1071–1075.
Neill, W. & Valdes, L. (1992). Persistence of Negative Priming: Steady State or Decay? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18*(3), 565–576.
Neill, W., Valdes, L., Terry, K. & Gorfein, D. (1992). Persistence of negative priming: II. Evidence for episodic trace retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*(5), 993–1000.
Newell, B. R. & Shanks, D. R. (2014). Unconscious influences on decision making: a critical review. *The Behavioural and Brain Sciences*, *37*(1), 1–19.
Nigg, J. (2001). Is ADHD a Disinhibitory Disorder? *Psychological Bulletin*, *127*(5), 571–598.
Nigg, J. T., Wong, M. M., Martel, M. M., Jester, J. M., Puttler, L. I., Glass, J. M., . . . Zucker, R. A. (2006). Poor response inhibition as a predictor of problem drinking and illicit drug use in adolescents at risk for alcoholism and other substance use disorders. *Journal of the American Academy of Child and Adolescent Psychiatry*, 45, 468-475.
Noël, X., Brevers, D. & Bechara, A. (2013). A neurocognitive approach to understanding the neurobiology of addiction. *Current Opinion in Neurobiology*, *23*, 1–7.
Pearce, J. & Hall, G. (1980). A model for Pavlovian learning: variations in the effectiveness of conditioned but not of unconditioned stimuli. *Psychological Review*, *87*(6), 532–552.
Pearce, J. & Mackintosh, N. (2010). Two theories of attention: a review and a possible integration. In C. Mitchell & M. E. Le Pelley (Eds.), *Attention and Associative Learning: From Brain to Behaviour* (pp. 11–39). Oxford: Oxford University Press.
Perruchet, P., Cleeremans, A. & Destrebecqz, A. (2006). Dissociating the effects of automatic activation and explicit expectancy on reaction times in a simple associative learning task. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *32*(5), 955–65.
R Core Team (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.
Ramautar, J. R., Kok, a & Ridderinkhof, K. R. (2004). Effects of stop-signal probability in the stop-signal paradigm: the N2/P3 complex further validated. *Brain and Cognition*, *56*(2), 234–252.
Rieger, M. & Gauggel, S. (1999). Inhibitory after-effects in the stop signal paradigm. *British Journal of Psychology*, *90*, 509–518.
Salinas, E. & Stanford, T. R. (2013). The countermanding task revisited: fast stimulus detection is a key determinant of psychophysical performance. *The Journal of Neuroscience*, *33*(13), 5668–5685.
Shiffrin, R. & Schneider, W. (1977). Controlled and Automatic Human information Processing: II. Perceptual Learning, Automatic Attending, and a General Theory. *Psychological Review*, *84*(2), 127–190.
Spierer, L., Chavan, C. F. & Manuel, A. L. (2013). Training-induced behavioural and brain plasticity in inhibitory control. *Frontiers in Human Neuroscience*, *7*, 1–9.

Stothart, C. R., Simons, D. J., Boot, W. R. & Kramer, A. F. (2014). Is the effect of aerobic exercise on cognition a placebo effect? *PloS One*, *9*(10), e109557.
Suret, M. & McLaren, I. (2005). Elemental representation and associability: An integrated model. In A. J. Wills (Ed.), *New directions in human associative learning* (pp. 155–188). New Jersey: Lawrence Erlbaum Associates, Inc.
Sutherland, N. & Mackintosh, N. (1971). *Mechanisms of animal discrimination learning*. New York: Academic Press.
Swick, D., Ashley, V., & Turken, U. (2011). Are the neural correlates of stopping and not going identical? Quantitative meta-analysis of two response inhibition tasks. *Neuroimage, 56,* 1655-1665.
Tukey, J. (1977). *Exploratory data analysis*. Reading, MA: Addison-Wesley.
Veling, H., Aarts, H. & Papies, E. K. (2011). Using stop signals to inhibit chronic dieters' responses toward palatable foods. *Behaviour Research and Therapy*, *49*(11), 771–780.
Veling, H., Aarts, H. & Stroebe, W. (2013). Using stop signals to reduce impulsive choices for palatable unhealthy foods. *British Journal of Health Psychology*, *18*(2), 354–368.
Veling, H., van Koningsbruggen, G. M., Aarts, H. & Stroebe, W. (2014). Targeting impulsive processes of eating behaviour via the internet. Effects on body weight. *Appetite*, *78*, 102–109.
Verbruggen, F., Best, M., Bowditch, W., Stevens, T. & McLaren, I. P. L. (2014). The inhibitory control reflex. *Neuropsychologia*, *65*, 263–278.
Verbruggen, F. & Logan, G. D. (2008a). Long-term aftereffects of response inhibition: memory retrieval, task goals, and cognitive control. *Journal of Experimental Psychology: Human Perception and Performance*, *34*(5), 1229–1235.
Verbruggen, F. & Logan, G. D. (2008b). Automatic and controlled response inhibition: associative learning in the go/no-go and stop-signal paradigms. *Journal of Experimental Psychology: General*, *137*(4), 649–672.
Verbruggen, F., & Logan, G. D. (2008c). Response inhibition in the stop-signal paradigm. *Trends in Cognitive Sciences*, 12, 418-424.
Verbruggen, F. & Logan, G. D. (2009). Proactive adjustments of response strategies in the stop-signal paradigm. *Journal of Experimental Psychology: Human Perception and Performance*, *35*(3), 835–854.
Verbruggen, F., Logan, G. D., Liefooghe, B. & Vandierendonck, A. (2008). Short-term aftereffects of response inhibition: repetition priming or between-trial control adjustments? *Journal of Experimental Psychology: Human Perception and Performance*, *34*(2), 413–426.
Verbruggen, F., McLaren, I. P. L. & Chambers, C. D. (2014). Banishing the control homunculi in studies of action control and behaviour change. *Perspectives on Psychological Science.*, *9*(5), 497–524.
Verbruggen, F., Stevens, T. & Chambers, C. (2014). Proactive and Reactive Stopping When Distracted: An Attentional Account. *Journal of Experimental Psychology: Human Perception and Performance*, *40*(4), 1295–1300.
Waszak, F., Hommel, B. & Allport, A. (2003). Task-switching and long-term priming: Role of episodic stimulus–task bindings in task-shift costs. *Cognitive Psychology*, *46*(4), 361–413.
Waszak, F., Hommel, B. & Allport, A. (2004). Semantic generalization of stimulus-task bindings. *Psychonomic Bulletin & Review*, *11*(6), 1027–1033.
Waszak, F., Hommel, B. & Allport, A. (2005). Interaction of task readiness and automatic retrieval in task switching: Negative priming and competitor priming. *Memory & Cognition*, *33*(4), 595–610.
Wessel, J. R., O'Doherty, J. P., Berkebile, M. M., Linderman, D., & Aron, A. R. (2014). Stimulus devaluation induced by stopping action. *Journal of Experimental Psychology: General*, 143, 2316-2329.
Zandbelt, B. B., Bloemendaal, M., Neggers, S. F. W., Kahn, R. S. & Vink, M. (2013). Expectations and violations: delineating the neural network of proactive inhibitory control. *Human Brain Mapping*, *34*(9), 2015–2024.

Footnotes

[1] In this context, a task goal refers to an abstract representation of going or stopping; in other words, it does not specify which specific response or motor program has to be executed. Consistent with the goal account, Giesen & Rothermund (2014) recently demonstrated that stop associations may have global effects on responding.

[2] There is some evidence that suggests that inconsistent reinforcement can increase attention to, and motivation salience of, conditioned stimuli. For example, the Pearce-Hall (1980) model suggests that associability is maintained for stimuli that are followed by unpredictable outcomes. However, despite animal data in support of this effect (e.g. Anselme, Robinson & Berridge, 2013), there is relatively little data showing this effect in humans (see Hogarth, Dickinson, Austin, Brown & Duka, 2008). The weight of evidence using humans participants is in favour of the Mackintosh (1975) model outlined above (but for a combination of both algorithms in one model, see Pearce & Mackintosh, 2010).

[3] This conclusion is further supported by a recent study in which we directly manipulated knowledge of the stimulus-stop contingencies whilst measuring attentional focus during task performance (Best, Stevens, McLaren & Verbruggen, 2015). The data pattern in the explicit condition was similar to the pattern observed in a proactive control study using the same paradigm (Verbruggen, Stevens, et al., 2014). Importantly, the data pattern looked qualitatively different in the implicit condition, suggesting that the response slowing for stop-associated items was not (entirely) due to proactive control adjustments resulting from the expectancy to stop.

[4] We can rule out the possibility that the absence of a test phase effect in Experiments 1-3 is due to the use of images or the frequent rule switching. In Appendix B, we present the results of an experiment in which we used a word version of the go/stop task with a single rule. In this experiment, reversing the word-go/stop mapping in the test phase did not influence performance either.

Tables

Table 1: Proportion of stop-signal trials as a function of experiment, image type and phase. The overall p(stop-signal) both across experiments and within the experimental phases was 0.25.

	# images	% stop-signal trials	
		Training phase	Test phase
Experiment 1			
stop-associated	8	75%	0%
go-associated	16	0%	8 images: 75%; 8 images: 0%
control	16	25%	25%
Experiment 2			
stop-associated	10	100%	0%
go-associated	30	0%	20 images: 0%; 10 images: 100%
Experiment 3 & Experiment 4			
stop-associated	8	75%	0%
go-associated	16	0%	4 images: 0%; 12 images: 50%
control	16	25%	8 images: 0%; 8 images: 50%

Table 2: Probability of a missed go response [*p*(miss)] as a function of experiment, image type and experimental phase. *P*(miss) is the ratio of the number of omitted responses to the total number of no-stop-signal trials: *p*(miss) = missed/ (correct + missed). M = mean; sd = standard deviation.

	Training phase		*Test phase*	
	M	*sd*	*M*	*sd*
Experiment 1				
stop-associated	0.020	0.071	0.013	0.033
go-associated	0.015	0.024	0.014	0.024
control	0.016	0.028	0.017	0.036
Experiment 2				
stop-associated	-	-	0.024	0.043
go-associated	0.016	0.025	0.013	0.031
Experiment 3				
stop-associated	0.028	0.083	0.038	0.060
go-associated	0.023	0.034	0.037	0.062
control	0.018	0.031	0.036	0.052
Experiment 4				
stop-associated	0.028	0.088	0.018	0.033
go-associated	0.020	0.032	0.007	0.018
control	0.021	0.032	0.025	0.036

Table 3: Overview of repeated Analyses of Variance performed to compare go and stop training phase performance. Image type (Experiments 1, 3, & 4: stop-associated, go-associated, control) and block (Experiments 1 & 2: 1-12; Experiments 3 & 4: 1-6) are the within-subjects factors. We did not analyze *p*(miss) because values were low.

	df 1	df 2	SS1	SS2	*F*	*p*	*gen. η2*
Experiment 1							
Go Reaction Time							
image type	2	56	21980	41878	14.70	< 0.001	0.009
block	11	308	43427	1438376	0.85	0.575	0.017
image type:block	22	616	17475	431981	1.13	0.331	0.007
p(respond|stop)							
image type	1	28	0.071	0.323	6.17	0.019	0.005
block	11	308	0.547	7.616	2.01	0.043	0.040
image type:block	11	308	0.154	3.384	1.28	0.238	0.011
Experiment 2							
Go Reaction Time							
block	11	319	27502	405062	1.97	0.039	0.037
p(respond|stop)							
block	11	319	0.199	3.846	1.50	0.136	0.037
Experiment 3							
Go Reaction Time							
image type	2	60	5589	47836	3.51	0.058	0.006
block	5	150	64257	437275	4.41	< 0.001	0.061
image type:block	10	300	16048	162531	2.96	0.005	0.016
p(respond|stop)							
image type	1	30	0.010	0.200	1.48	0.232	0.002
block	5	150	0.053	2.688	0.60	0.703	0.009
image type:block	5	150	0.055	1.780	0.93	0.461	0.009
Experiment 4							
Go Reaction Time							
image type	1	27	6447	70581	2.47	0.128	0.012
block	5	135	51108	157194	8.78	0.000	0.085
image type:block	5	135	12088	128773	2.53	0.037	0.021
p(respond|stop)							
image type	1	27	0.087	0.317	7.45	0.011	0.015
block	5	135	0.109	1.857	1.58	0.173	0.019
image type:block	5	135	0.180	2.363	2.05	0.077	0.031

Table 4: Overview of repeated Analyses of Variance performed to compare go and stop test phase performance. Image type (Experiments 1, 3 & 4: stop-associated, go-associated, control, Experiment 2: stop-associated, go-associated) and block (Experiments 1 & 2: 13-14; Experiments 3 & 4: 7) are the within-subjects factors. We did not analyse *p*(miss) because values were low.

	df 1	df 2	SS1	SS2	*F*	*p*	*gen. η2*
Experiment 1							
Go Reaction Time							
image type	2	56	471	22425	0.59	0.557	0.002
block	1	28	2318	53598	1.21	0.281	0.010
image type:block	2	56	771	13040	1.66	0.200	0.003
p(respond|stop)							
image type	1	28	0.001	0.268	0.16	0.695	< 0.001
block	1	28	0.058	0.380	4.24	0.048	0.028
image type:block	1	28	0.007	0.316	0.64	0.429	0.004
Experiment 2							
Go Reaction Time							
image type	1	29	160	10621	0.44	0.513	< 0.001
block	1	29	390	47352	0.24	0.629	0.002
image type:block	1	29	60	7087	0.25	0.624	< 0.001
p(respond|stop)							
block	1	29	0.020	0.337	1.73	0.198	0.019
Experiment 3							
Go Reaction Time							
image type	2	60	711	28880	0.74	0.479	0.005
p(respond|stop)							
image type	1	30	0.51	0.416	3.73	0.062	0.043
Experiment 4							
Go Reaction Time							
image type	2	54	7329	28228	7.01	0.004	0.064
p(respond|stop)							
image type	1	27	0.050	0.473	2.83	0.104	0.034

Table 5: Go reaction times (in ms) in the test phase as a function of expectancy (go, stop) and image type (stop-associated, go-associated, control) in Experiment 4. M = mean; sd = standard deviation.

	Stop expectancy		Go expectancy	
	M	*sd*	*M*	*sd*
stop-associated	453	56	429	51
go-associated	437	43	411	36
control	440	40	411	33

Figure captions

Figure 1: Overview of the architecture of the associative stop system (for a more detailed overview, see Verbruggen, Best et al., 2014). There are two associative routes to activating the stop-goal; a direct association between the stimulus or cue and the go/stop goal, or indirect association between the stimulus or cue and the go/stop goal that is mediated via a representation of the go/stop signal. Excitatory and inhibitory connections are represented on the diagram with arrows.

Figure 2: Example go/stop trial sequence. The task rule changed at the beginning of each block (e.g. Block *n*: vowel = stop; consonant = go, Block *n* + 1: > 5 = stop; < 5 = go). In Experiments 1-3, the go/stop signals were superimposed on top of the image (as shown). In Experiment 4, the signals were presented in one of the four corners of the image (top-left, bottom-left, top-right, bottom-right).

Figure 3: go RTs (in ms; upper panel) and *p*(respond|stop) data (lower panel) for the three image types (stop; go; control) as a function of the block (blocks 1-12 = training phase; blocks 13- 14= test phase) in Experiment 1. Error bars are 95% confidence intervals.

Figure 4: go RTs (in ms; upper panel) and p(respond|stop) data (lower panel) for the two image types (stop; go) as a function of the block (blocks 1-12 = training phase; blocks 13- 14 = test phase) in Experiment 2. Error bars are 95% confidence intervals.

Figure 5: go RTs (in ms; upper panel) and p(respond|stop) data (lower panel) for the three image types (stop; go; control) as a function of the block (blocks 1-6 = training phase; block 7 = test phase) in Experiment 3. Error bars are 95% confidence intervals.

Figure 6: go RTs (in ms; upper panel), p(respond|stop) data (middle panel) and expectancy ratings (lower panel) for the three image types (stop; go; control) as a function of the block (blocks 1-6 = training phase; block 7 = test phase) in Experiment 4. Error bars are 95% confidence intervals.

Figure 1

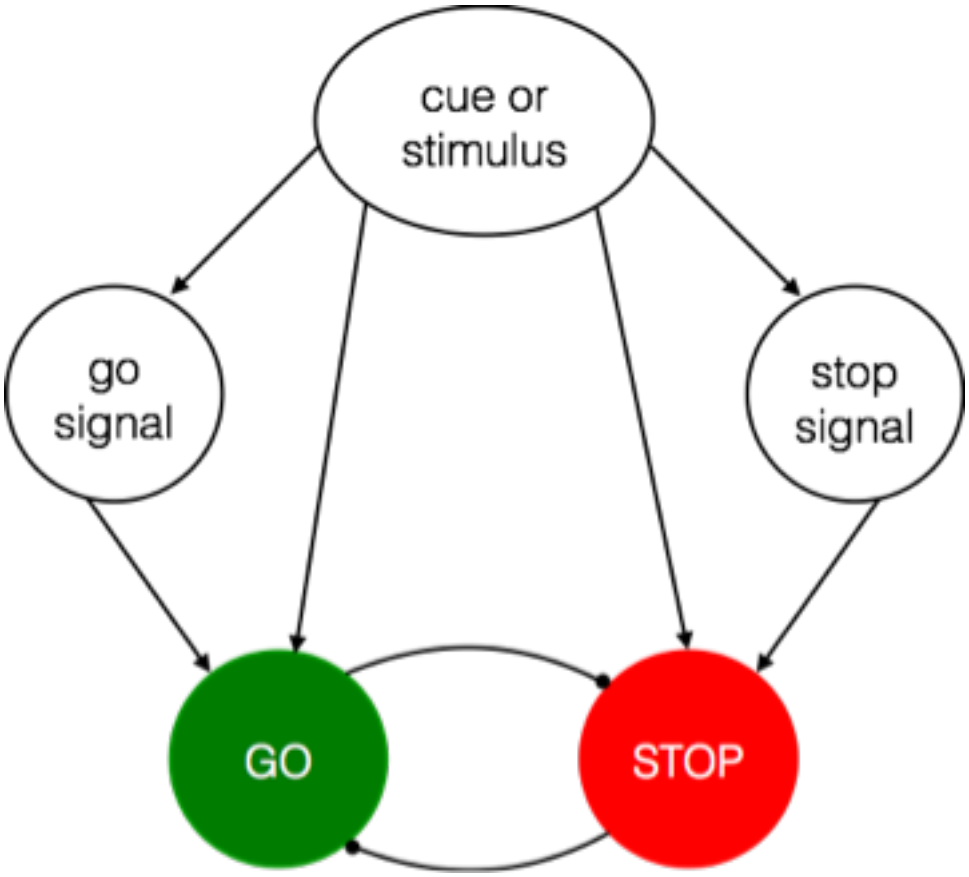


Figure 2

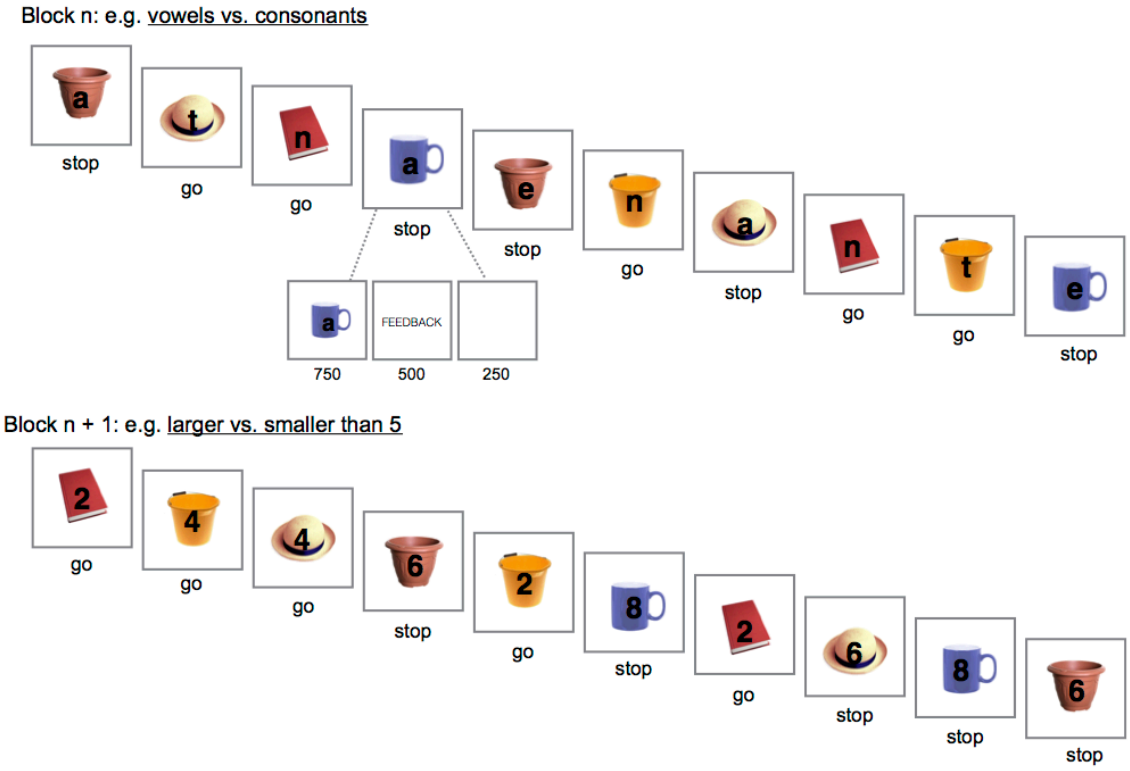

Block n: e.g. vowels vs. consonants
stop
go
go
stop
stop
go
stop
go
go
stop
a
FEEDBACK
750
500
250
Block n + 1: e.g. larger vs. smaller than 5
go
go
go
stop
go
stop
go
stop
stop
stop

Figure 3

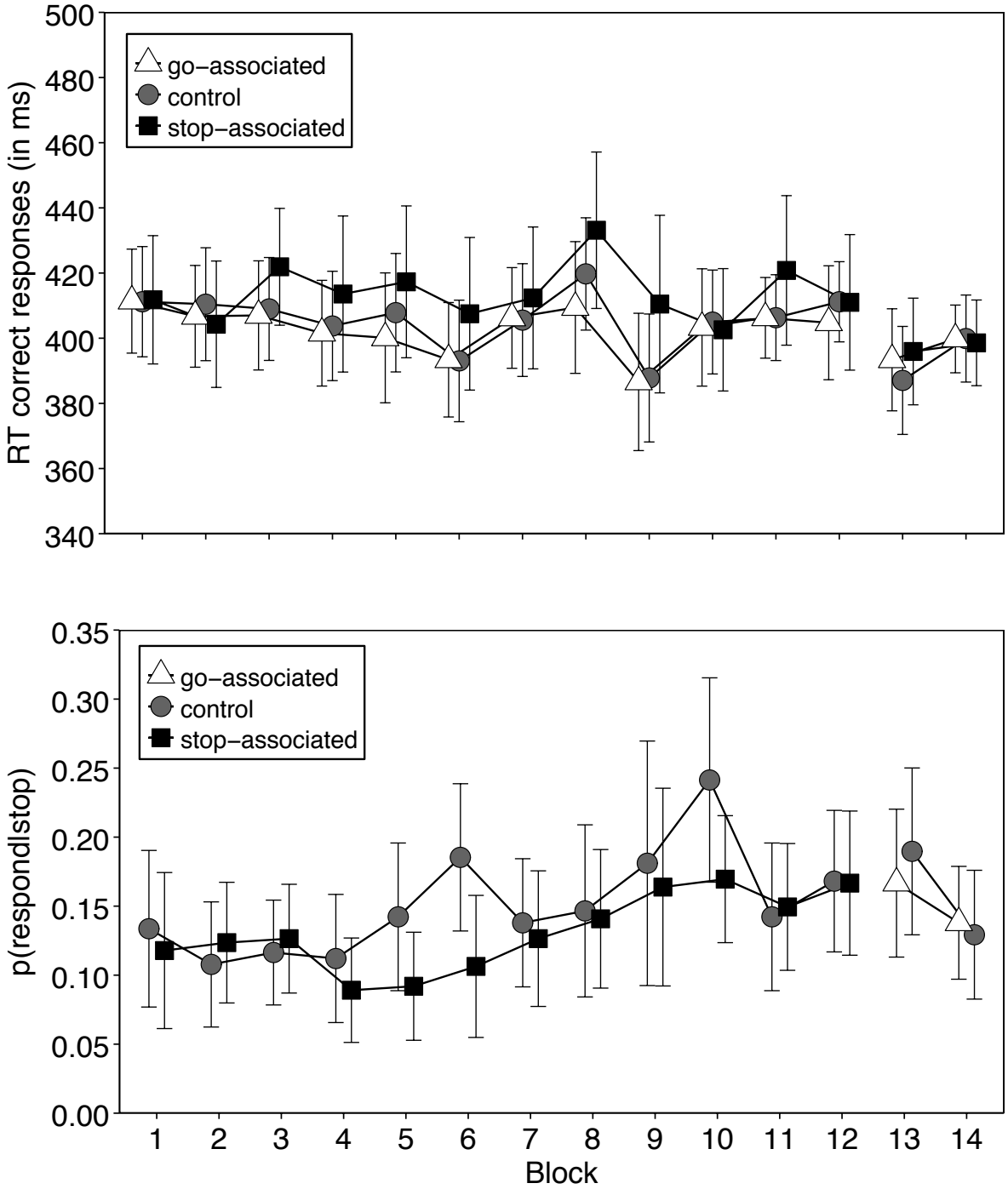

RT correct responses (in ms)
500
480
460
440
420
400
380
360
340
go−associated
control
stop−associated
p(respondIstop)
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
1
2
3
4
5
6
7
8
9
10
11
12
13
14
Block

Figure 4

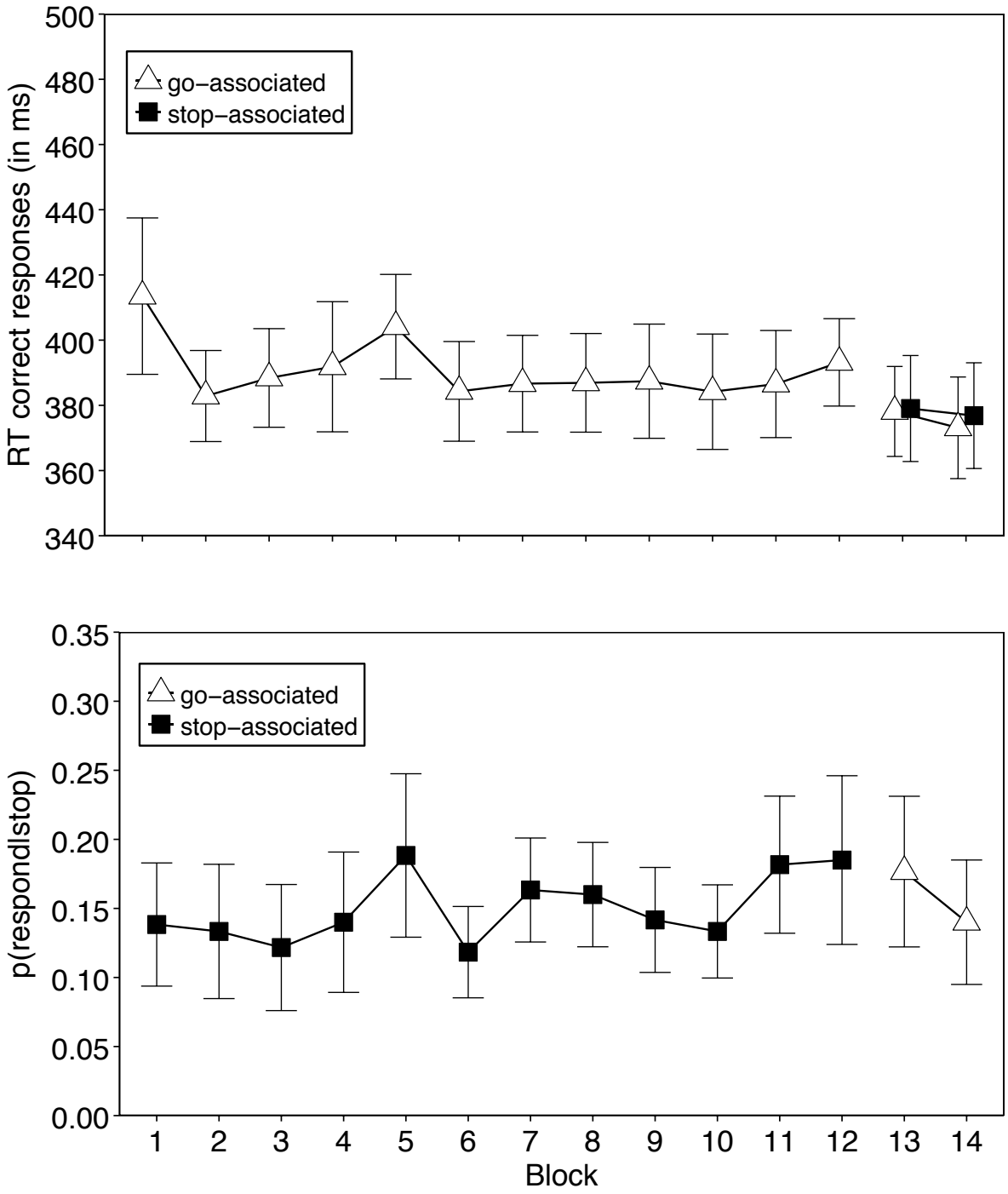

500
480
460
440
420
400
380
360
340
RT correct responses (in ms)
go–associated
stop–associated
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
p(respondIstop)
go–associated
stop–associated
1
2
3
4
5
6
7
8
9
10
11
12
13
14
Block

Figure 5

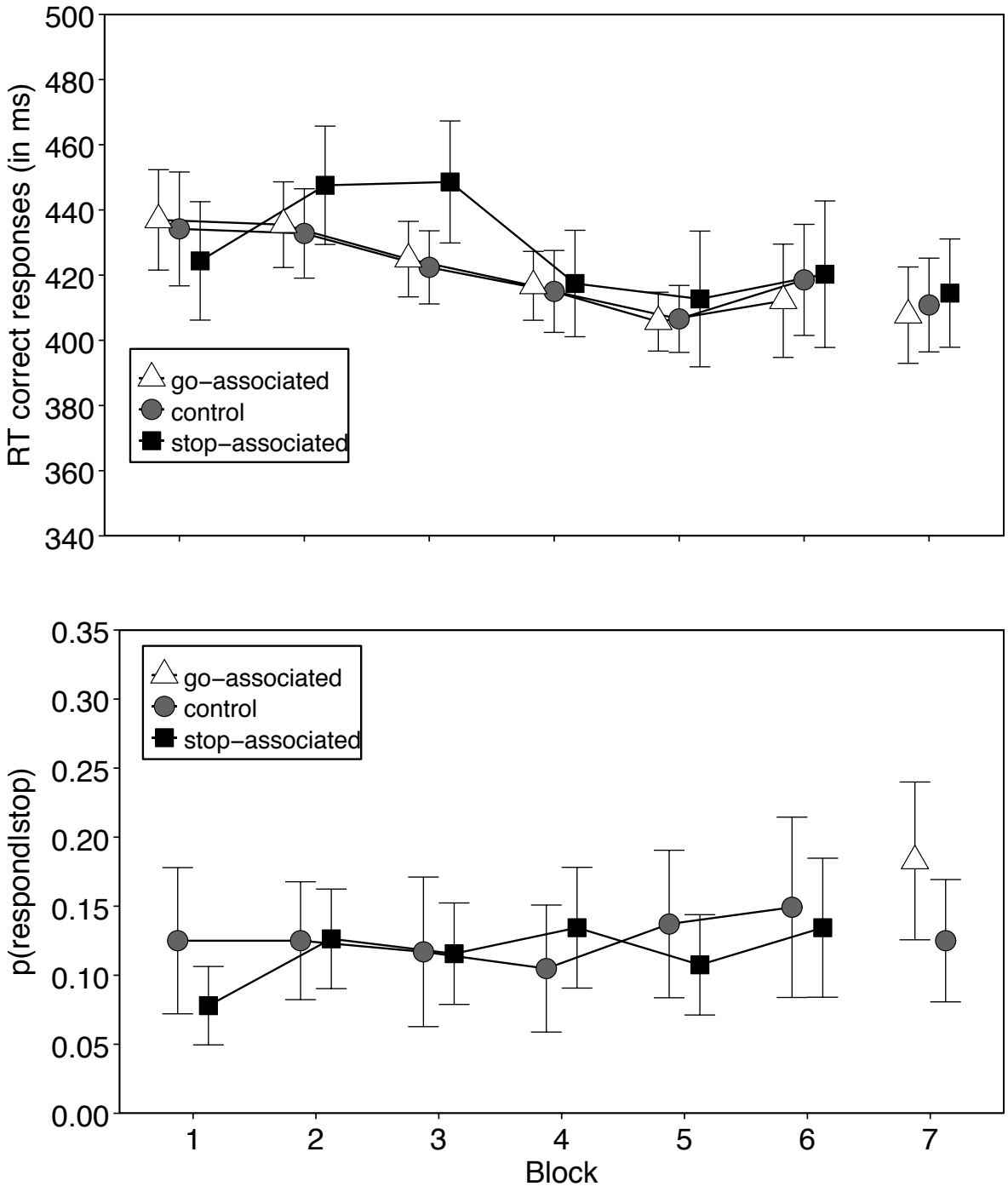
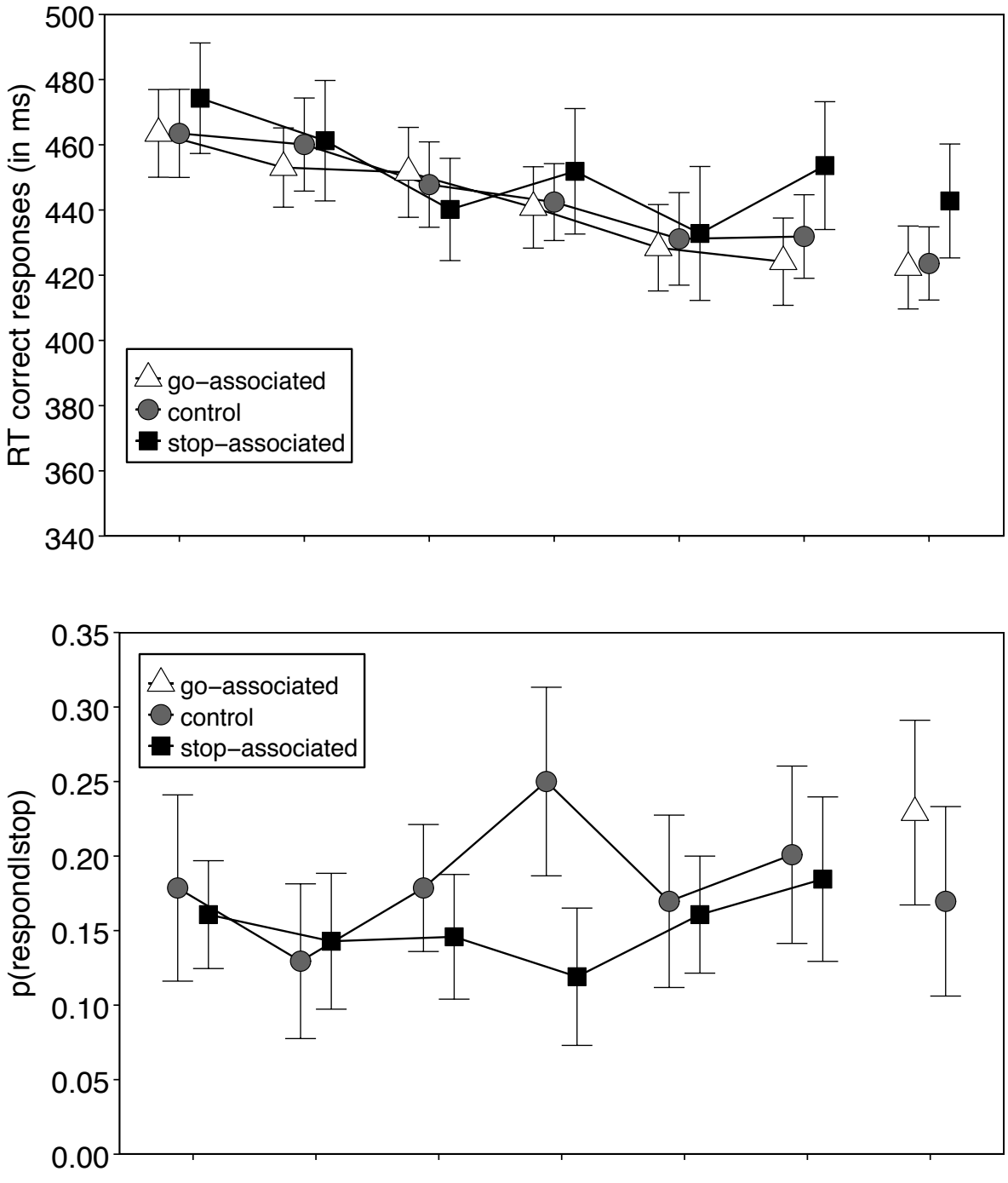
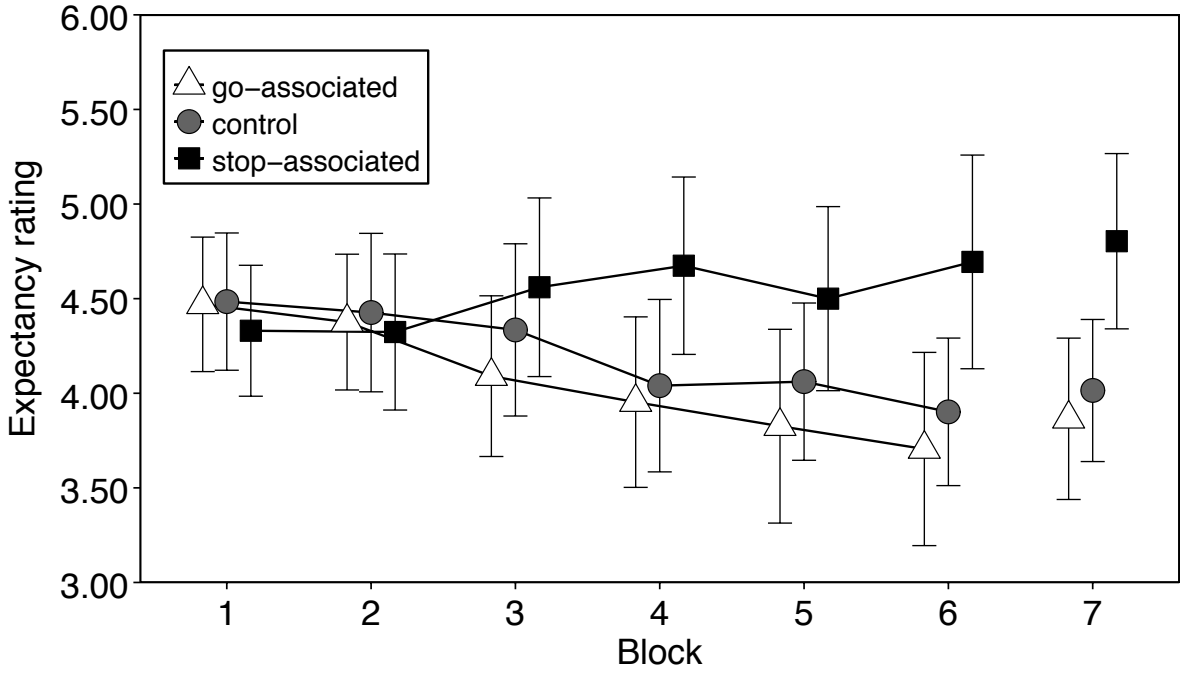

RT correct responses (in ms)
500
480
460
440
420
400
380
360
340
go–associated
control
stop–associated
p(respondIstop)
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
1
2
3
4
5
6
7
Block

Figure 6


RT correct responses (in ms)
500
480
460
440
420
400
380
360
340
go–associated
control
stop–associated
p(respondIstop)
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
go–associated
control
stop–associated


6.00
5.50
5.00
4.50
4.00
3.50
3.00
Expectancy rating
go−associated
control
stop−associated
1
2
3
4
5
6
7
Block

Appendix A:

A full list of the rules and the stimuli used in Experiments 1-4

In Experiments 1-3 we used fourteen go/stop rules. In Experiment 4, we used rules 1-7 only. The signals used in rules 1-11 were presented in Arial (font size = 50). The sizes of the signals used in rules 12-14 are provided in pixels below (screen resolution: 1680 × 1050).

1. Vowels ('a' or 'e') vs. consonants ('t' or 'n').
2. Symbols that are the same ('@@' or '&&') vs. symbols that are different ('@&' or '&@').
3. Uppercase letters ('H' or 'R') vs. lowercase letters ('h' or 'r').
4. Long symbol strings ('£%£%' or '%£%£') vs. short symbol strings ('£%' or '%£').
5. Curved letters ('S' or 'C') vs. angled letters ('K' or 'W').
6. Digits smaller than 5 ('2' or '4') vs. digits bigger than 5 ('6' or '8').
7. Curly brackets ('{' or '}') vs. square brackets ('[' or ']').
8. Words that refer to animals ('horse' or 'sheep') vs. words that refer to fruit ('lemon' or 'apple').
9. Symmetric letter strings ('UYYU' or 'YUUY') vs. asymmetric letter strings ('YYUU' or 'UUYY').
10. Crosses on the left of image vs. crosses on the right of the image relative to the center (crosses appeared at the top and bottom of the image).
11. Asterisks on the top of the image vs. asterisks on the bottom of the image relative to the center (asterisks appeared on the left and right of the image).
12. Horizontal lines (lines appeared across the top or bottom of the image relative to the center) [width: 240 pixels] vs. vertical lines (lines appeared along the left or right of the image relative to the center) [height: 240 pixels].
13. Shapes bigger than a fifty pence piece (square or circle) [100 × 100 pixels] vs. shapes smaller than a fifty pence piece (square or circle) [40 × 40 pixels].
14. Lines thicker than a matchstick vs. lines thinner than a matchstick (lines appeared horizontally [width = 240 pixels] or vertically [height = 240 pixels] about the center of the image).

Appendix B:
A word version of the go/stop task with a single rule

We also ran a word version of the go/stop task. On each trial, a colored word was presented. Half of the participants were instructed to respond when the ink color was green (in other words, green was the go signal), but to refrain from responding when the ink color was red (in other words, red was the stop signal). The color-go/stop mapping was reversed for the other participants (i.e. red=go; green=stop). In the training phase, half of the words always occurred on go trials; the other words always occurred on stop trials. In the test phase, the word-go/stop mapping was reversed.

Method

Subjects. Twenty subjects from Vanderbilt University participated for monetary compensation ($12). All subjects reported normal or corrected-to-normal vision and were native speakers of English.

Apparatus and stimuli. The experiment was run on a PC running Tscope (Stevens, Lammertyn, Verbruggen & Vandierendonck, 2006) and the stimuli were presented on a 21-in monitor. A list of 32 words was drawn from a list of 640 words used by Arrington and Logan (2004): *bamboo, barn, bead, blackboard, boar, buffalo, cabin candle, cattle, cup, elephant, elm, eucalyptus, ferret, grasshopper, hammer, hamster, holly, leash, lemon, mushroom, paddle, rose, scissors, shamrock, shoelace, ski, slug, suitcase, toothbrush, tripod, trombone*. For every subject, two random subsets of 16 words were selected. The first subset was presented on go trials in the training phase and on stop trials in the test phase; the second subset was presented on stop trials in the training phase and on go trials in the test phase. All stimuli were presented in a white lower case Courier font on a black background and ranged from 12 to 52 mm in width (approximately 1.1° to 5.0°) and 4 to 7 mm (approximately 0.4° to 0.7°) in height.

Procedure. Subjects were seated individually in private testing rooms after providing informed consent. The experimenter left the room after giving instructions and watching the first few practice trials.

Unbeknown to the participants, there were two phases in the experiment. The training phase consisted of 16 blocks of 32 trials. In each training block, each word was presented once. The training phase was followed by a test phase in which the word-go/stop

mapping was reversed (e.g., for Participant 1, 'bamboo' always occurred on stop trials in the training phase, but it occurred on go trials in the test phase). The test phase consisted of 6 blocks of 32 trials.

In both phases of both conditions, all trials started with the presentation of the colored word in the center of the screen. Half of the participants were instructed to press the space bar of a QWERTY keyboard with the index finger of the dominant hand as quickly as possible when the ink color was green, but to refrain from responding when the ink color was red. The color-go/stop mapping was reversed for the other participants. The word remained on the screen for 750 ms, regardless of go RT in order to equate study time for go and stop stimuli. A response could be given only while the stimulus was on the screen. The intertrial-interval was 750 ms. At the end of each block, the mean RT on go trials, the number of missed responses on go trials, and the number of incorrect responses on stop trials were displayed and subjects had to pause for 10 seconds, after which they could continue by pressing the space bar.

Results and Discussion

Table B1 shows that go performance in the test phase was comparable to performance in the last six training blocks. This was confirmed by a repeated measures ANOVA that examined the effect of block number (excluding blocks 1-10), $F(11, 209) = 1.24$, $p = .26$, *gen. η2 = .016*. Furthermore, we compared go RT in the last training block with go RT in the first test block using a Bayesian t-test (Rouder, Speckman, Sun, Morey & Iverson, 2009), and found substantial support for the null hypothesis, B = .23. Thus, go performance was not influenced by the reversal of the word-go/stop mapping. Again, we attribute the absence of an effect of associative learning on performance to the interplay between attention and automaticity. In this experiment, we did not measure expectancy.

P(respondlsignal) was very low in both the training phase and the test phase (Table B1), so we did not analyze it further.

Table B1: Mean go RT (and standard deviation) and probability of responding on stop trials (p(R|S) for each block. The test blocks are in underlined.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
Go RT																						
M	341	324	327	329	322	323	331	322	324	325	324	319	325	322	316	318	317	323	316	325	321	311
sd	31	25	25	35	33	33	36	32	33	33	29	34	35	39	36	33	34	37	35	31	36	31
p(R|S)																						
M	.03	.01	.02	.01	.01	.02	.03	.01	.02	.01	.01	.03	.03	.01	.01	.02	.02	.03	.03	.02	.01	.02
sd	.05	.02	.03	.03	.02	.03	.05	.02	.05	.03	.03	.04	.05	.03	.03	.03	.04	.04	.04	.03	.03	.03

Supplementary Material

Outlier Data

Subjects were excluded who incorrectly executed a response on ≥ 30% of the stop-signal trials. In Experiment 1, two subjects were excluded; in Experiment 3, one subject was excluded; in Experiment 4, four subjects were excluded. No subjects were excluded from Experiment 2. Exclusion of these subjects did not substantially alter the overall pattern of data. The descriptive statistics with these subjects are presented in Tables S1 and S2.

Table S1: Probability of a response on a stop-signal trial [*p*(respond|stop] and RT on go trials as a function of experiment, image type and experimental phase. M = mean; sd = standard deviation.

	p(respond|stop)				*Go RTs*			
	Training phase		*Test phase*		*Training phase*		*Test phase*	
	M	*sd*	*M*	*sd*	*M*	*sd*	*M*	*sd*
Experiment 1								
stop-associated	0.149	0.152	-	-	410	61	392	43
go-associated	-	-	0.175	0.164	399	48	392	39
control	0.167	0.169	0.188	0.185	402	47	388	45
Experiment 3								
stop-associated	0.124	0.127	-	-	428	54	413	46
go-associated	-	-	0.182	0.153	421	37	406	40
control	0.134	0.156	0.125	0.119	421	39	410	39
Experiment 4								
stop-associated	0.177	0.136	-	-	450	51	435	50
go-associated	-	-	0.263	0.183	438	40	418	36
control	0.214	0.174	0.215	0.209	442	39	420	31

Table S2: Expectancy ratings as a function of experiment and image type. M = mean; sd = standard deviation.

	Training phase		*Test phase*		*End of task*	
	M	*sd*	*M*	*sd*	*M*	*sd*
Experiment 1						
stop-associated	-	-	-	-	4.88	1.47
go-associated	-	-	-	-	3.99	1.27
control	-	-	-	-	4.32	1.25
Experiment 3						
stop-associated	-	-	-	-	5.54	1.31
go-associated	-	-	-	-	4.62	0.94
control	-	-	-	-	4.79	0.93
Experiment 4						
stop-associated	4.52	1.15	4.77	1.15	-	-
go-associated	4.06	1.13	3.91	1.08	-	-
control	4.23	1.03	4.02	0.91	-	-

RT Percentiles

To investigate the possibility that the absence of an effect of image type in the test phase of Experiments 1-3 is due to response latencies (responding was faster in the test phase than in the training phase) we plotted RT percentiles for the training and test phases. These RT percentiles revealed that the overall response latency cannot account for the absence of a learning effect in the test phase.

Furthermore, in Experiments 1-3, visual inspection of the percentile plots suggests that the slowing for the stop-associated images emerges in the slow end of the RT distribution. This conclusion is supported by a reliable two-way interaction between image type (stop; go; control) and percentile in the training phase of Experiment 1, $F(4, 112) = 8.03$, $p = .001$, $gen.\,\eta 2 = .005$. However, in Experiment 4, the slowing for the stop-associated images emerges in the fast end of the RT distribution. This conclusion is also supported by a reliable two-way interaction between image type and percentile in the training phase, $F(4, 108) = 28.05$, $p < .001$, $gen.\,\eta 2 = .034$.

In Experiments 1-3, processing the image could slow overall RT; but for stop-associated items, processing the image would also lead to retrieval of the stop associations, and consequently, automatic inhibition of the response. Alternatively, only on slower trials, the stimulus-stop associations could be retrieved in time and affect performance. In Experiment 4, attention to the images prior to signal presentation meant that there was more time for the acquired stimulus-stop associations to be retrieved and thus influence performance.

Figure S1: go RTs (in ms) in the training phase (blocks 1-12; upper panel) and *the test phase* (blocks 13-14; lower panel) for the three image types (stop-associated; go-associated; control) as a function of percentile in Experiment 1.

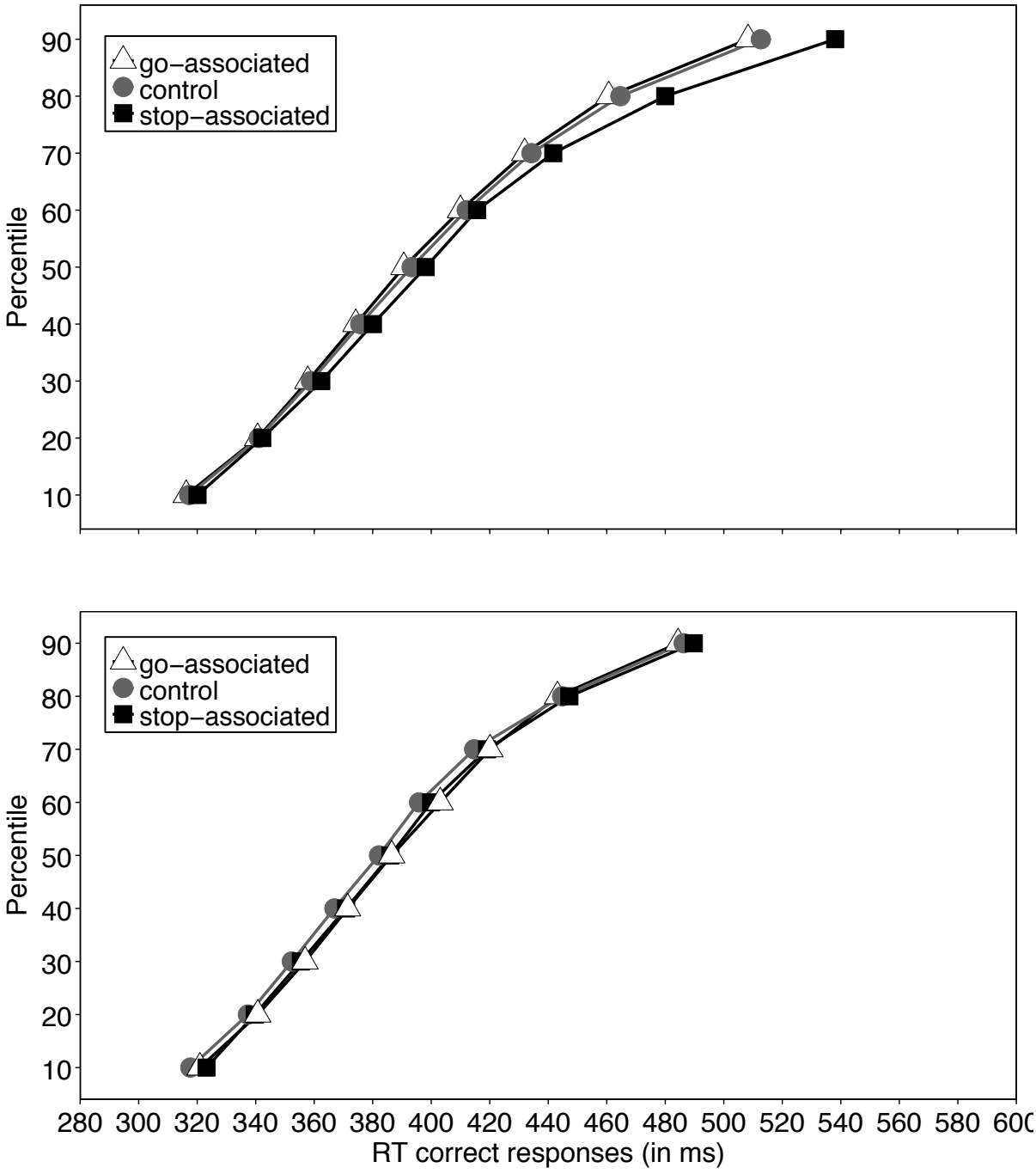


Figure S2: go RTs (in ms) in the test phase (blocks 13-14) for the two image types (stop-associated; go-associated) as a function of percentile in Experiment 2. For obvious reasons, we could not plot RT percentiles for the training phase data.

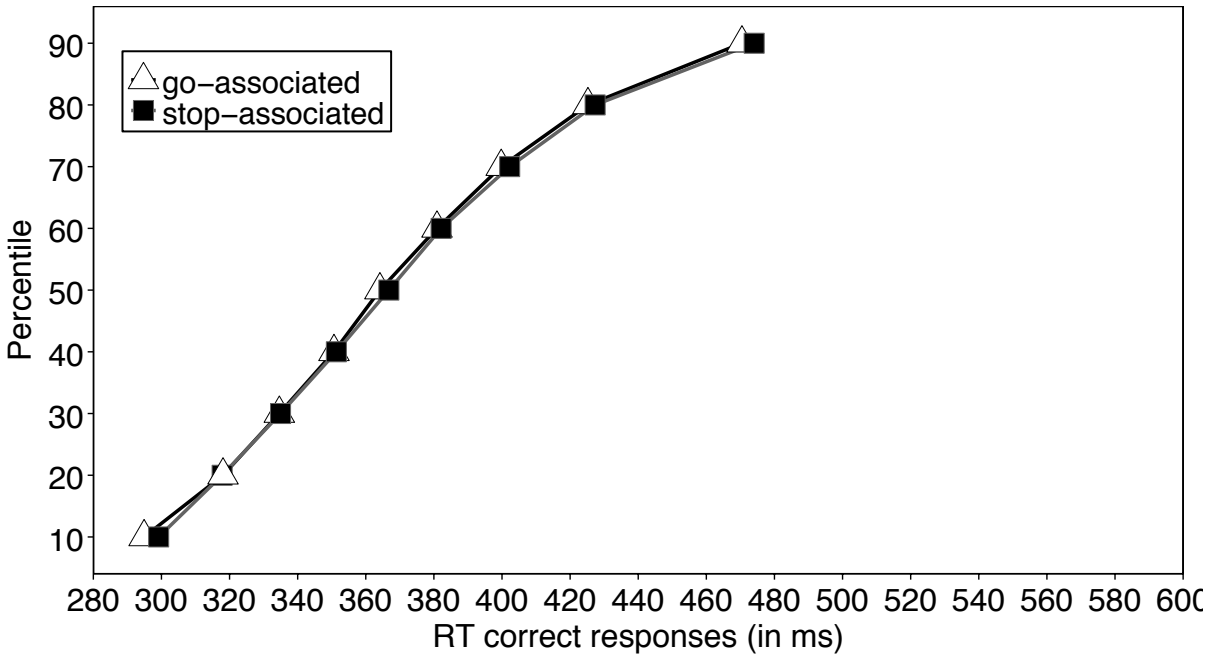


Figure S3: go RTs (in ms) in the training phase (blocks 1-6; upper panel) and *the test phase* (block 7; lower panel) for the three image types (stop-associated; go-associated; control) as a function of percentile in Experiment 3.

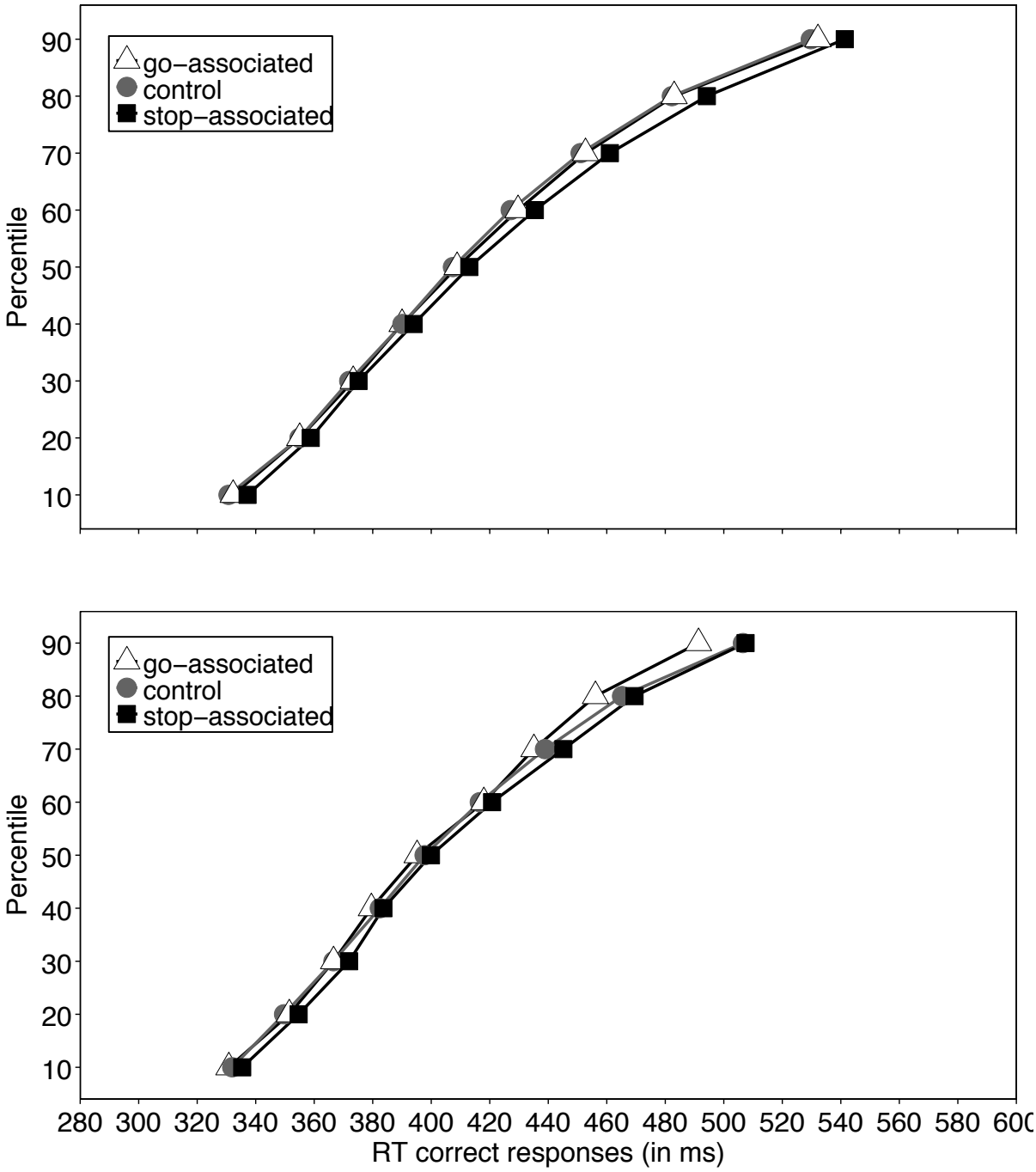
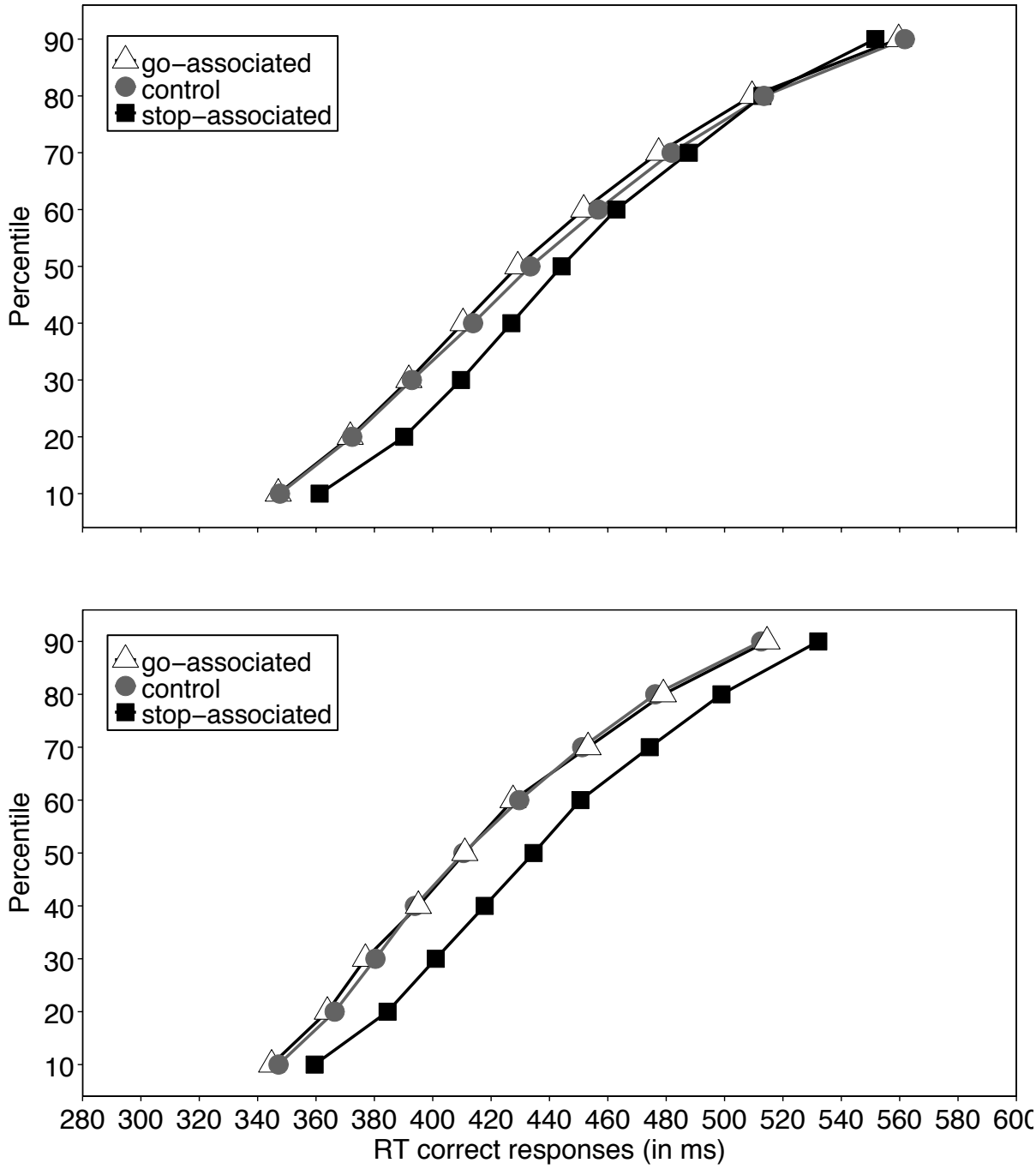


Figure S4: go RTs (in ms) in the training phase (blocks 1-6; upper panel) and *the test phase* (block 7; lower panel) for the three image types (stop-associated; go-associated; control) as a function of percentile in Experiment 4.



Expectancy/RT correlation plots

Figure S5: Expectancy/RT correlations in the training phase (blocks 1-12; upper panels) and *the test phase* (blocks 13-14; lower panels) in Experiment 1. Note, 'stop-associated minus control image' expectancy difference reliably correlated with the corresponding RT difference in the test phase, $r(26) = 0.437$, $p = 0.019$. All other correlations were not reliable (r's ≤ 0.272, p's ≥ 0.161).

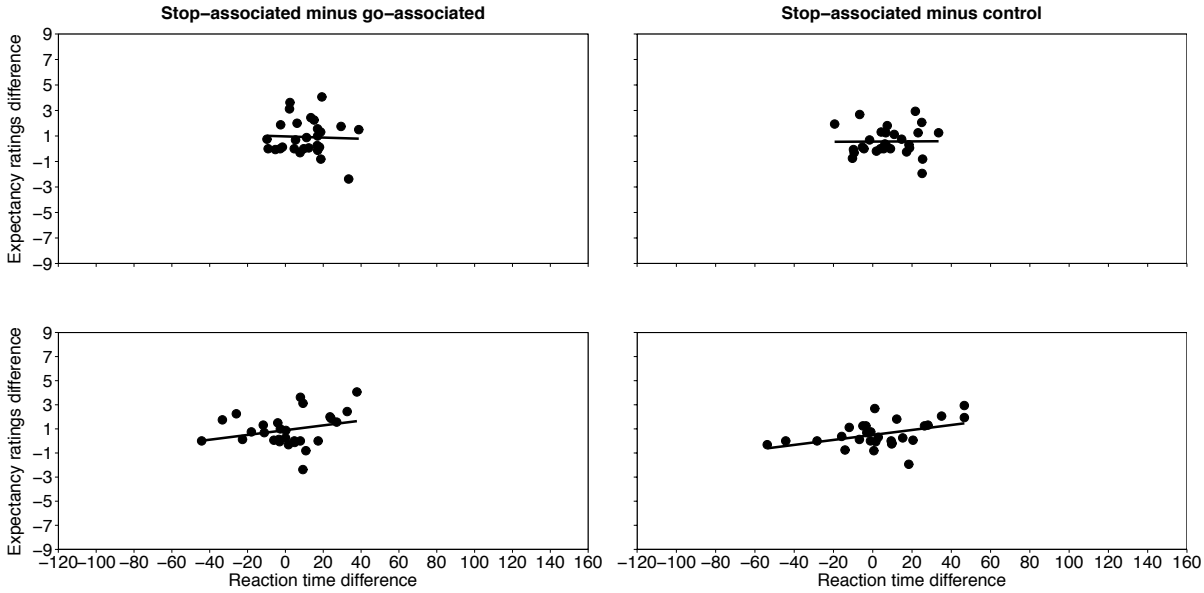


Figure S6: Expectancy/RT correlations in *the test phase* (blocks 13-14; lower panel) in Experiment 2. Due to the stimulus-stop contingencies used, we could not run these correlations on the training phase data. Note, the stop-associated minus go-associated correlation was not reliable ($r(28) = 0.262$, $p = 0.162$).

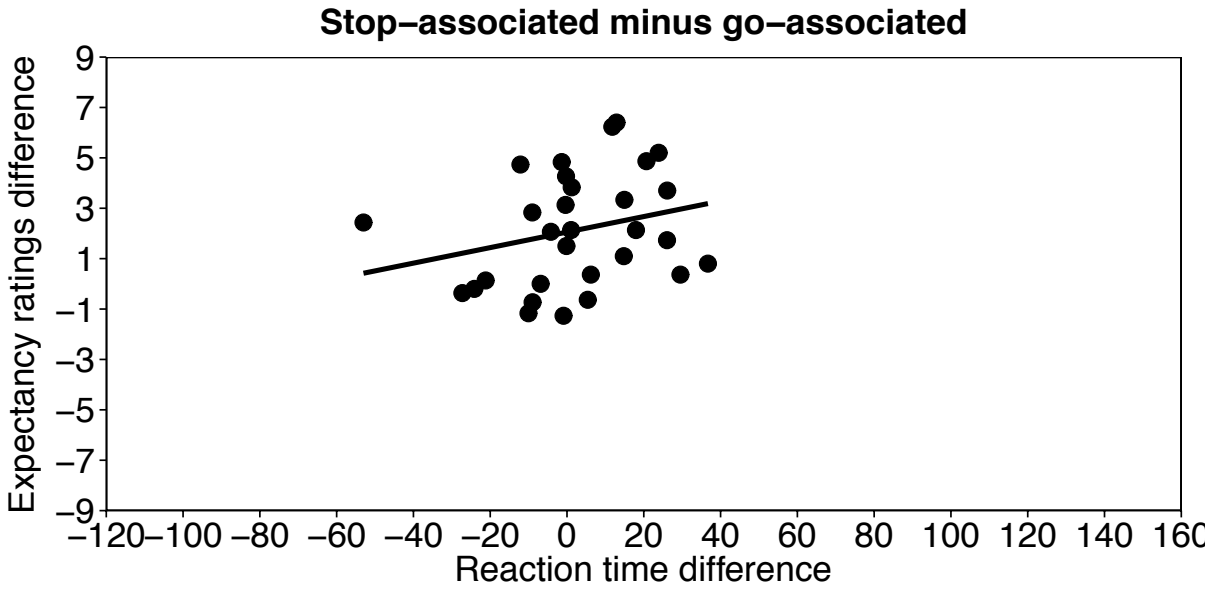


Figure S7: Expectancy/RT correlations in the training phase (blocks 1-6; upper panels) and *the test phase* (block 7; lower panels) in Experiment 3. Note, all correlations were not reliable (r's ≤ 0.136, p's ≥ 0.464).

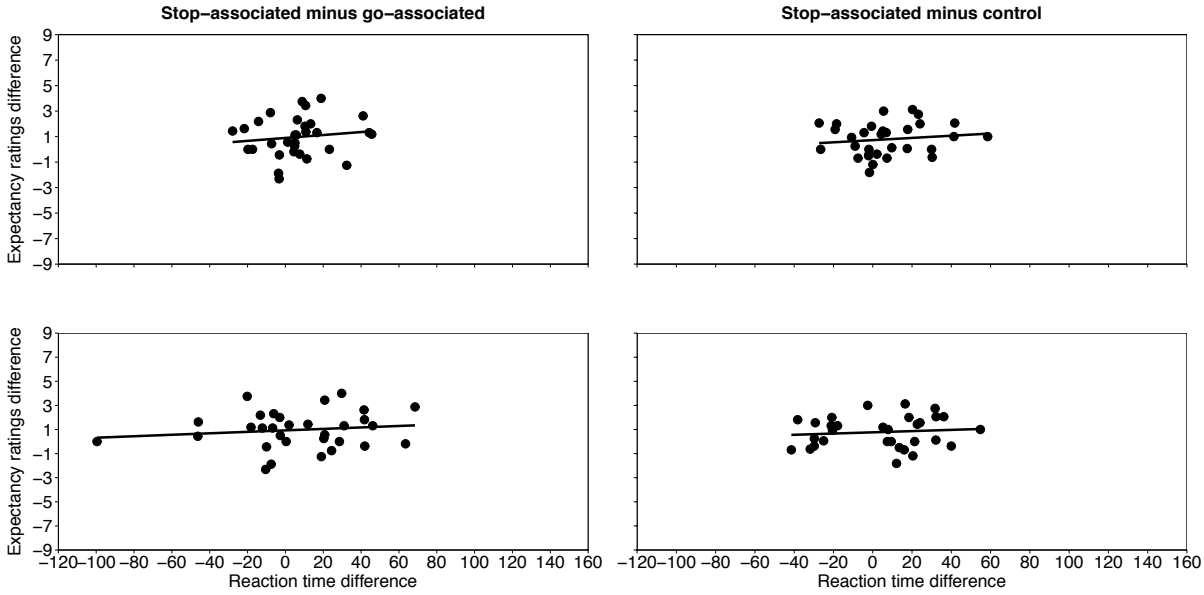


Figure S8: Expectancy/RT correlations in the training phase (blocks 1-6; upper panels) and *the test phase* (block 7; lower panels) in Experiment 4. Note, all correlations were reliable (r's ≥ 0.498, p's ≤ 0.006).