Humans reconfigure target and distractor processing to address distinct task demands

Harrison Ritz & Amitai Shenhav

Cognitive, Linguistic & Psychological Science, Brown University, Providence, RI
Carney Institute for Brain Science, Brown University, Providence, RI

Correspondence: Harrison Ritz (harrison.ritz@gmail.com)

Abstract

When faced with distraction, we can focus more on goal-relevant information (targets) or focus less goal-conflicting information (distractors). How people decide to distribute cognitive control across targets and distractors remains unclear. To help address this question, we developed a parametric attentional control task with a graded manipulation to both target discriminability and distractor interference. We find that participants exert independent control over target and distractor processing. We measured control adjustments through the influence of incentives and previous conflict on target and distractor sensitivity, finding that these have dissociable influences on control. Whereas incentives preferentially led to target enhancement, conflict on the previous trial preferentially led to distractor suppression. These distinct drivers of control altered sensitivity to targets and distractors early in the trial, and were promptly followed by reactive reconfiguration towards task-appropriate feature sensitivity. Finally, we provide a process-level account of these findings by show that these control adjustments are well-captured by an evidence accumulation model with attractor dynamics over feature weights. These results help establish a process-level account of control configuration that provides new insights into how multivariate attentional signals are optimized to achieve task goals.

Introduction

Whether we are having a conversation in a crowded coffee shop or writing a paper at our desk while surrounded by browser tabs, most tasks require us to engage in two distinct forms of control. One form of control supports the processing of task-*relevant* information, for instance by paying careful attention to what our conversation partner is sharing with us. The other form of control suppresses the processing of task-*irrelevant* information, particularly that which conflicts with our primary goal (e.g., distraction from a nearby conversation). While past research has extensively studied the processes underpinning these two loci of control, it has done so primarily by focusing on each one separately. As a result, relatively little is known about how control over task-relevant information (targets) interacts with control over task-irrelevant information (distractors), and under what situations one would be prioritized over the other. Here, we bridge previous methodological gaps to gain new insight into how control is differentially allocated to achieve these aims, and provide an integrative model that accounts for the dynamics by which these two forms of control are adjusted within and across trials.

Research into how people enhance the target of their attention versus actively suppress distractors has been largely governed by separate research areas, using different approaches. Studies of perceptual decision-making have characterized the process by which people try to determine the correct response (e.g., which of two categories this stimulus belongs to) based on noisy information about a target stimulus, and how this varies with the difficulty of discriminating that stimulus (e.g., how perceptually similar two stimuli are (Britten, Shadlen, Newsome, & Movshon, 1992; Gold & Shadlen, 2007)). This contrasts with studies of inhibitory control, in which the correct response to a target is typically unambiguous (e.g., respond left when seeing a high-contrast leftward-facing arrow), but a second dimension of the stimulus display (one that is typically processed more automatically; e.g., flanking arrows pointing rightward) triggers a conflicting response (Botvinick & Cohen, 2014; Posner & Snyder, 1975).

Despite the substantial progress that has been made in understanding these two processes in parallel, critical questions remain that can only be addressed by studying them in tandem. Most notably, it is unclear how people decide how to distribute their control between targets and distractors. When the demands or incentives for performing a task change, do people re-direct control towards target enhancement, distractor suppression, or both? For instance, previous work has shown that people are less susceptible to the influence of distractors after overcoming a previously conflicting distractor (the so-called *conflict adaptation* or *congruency sequence* effect; (Egner, 2007; Gratton, Coles, & Donchin, 1992)). Prevailing models have accounted for these findings by assuming that participants increase attention to the target dimension following a high-conflict trial (Botvinick, Braver, Barch, Carter, & Cohen, 2001; Egner, 2007), but limitations of the relevant experiments (in particular, that their focus on variability in distractor congruency) make it difficult to rule out that adaptation is occurring at the level of distractor suppression instead (Lindsay & Jacoby, 1994). It is more generally an open question whether effects of recent task difficulty (e.g., low discriminability or high-conflict) result in control-specific or control-general adaptations and, similarly, whether the motivation to improve performance in such settings leads to preferential engagement of one or both forms of control.

More recent models of controlled decision-making have emphasized the role that within-trial attentional dynamics play in response conflict tasks, offering new insight into the implementation of cognitive control (Servant, Montagnini, & Burle, 2014; Weichart, Turner, & Sederberg, 2020; White, Ratcliff, & Starns, 2011; Yu, Dayan, & Cohen, 2009). These models have largely been used for the Eriksen flanker task, modelling how an attentional spotlight centered on the target item narrows over time. This formulation necessarily yokes target enhancement and distractor suppression due to the spatial spread of attention. As a result, little is known about whether target and distractor processing dynamics can fall under independent control when these are not explicitly yoked, as in the case of feature-based attention. Less still is known about how adjustments driven by factors like conflict adaptation and incentives alter the *dynamics* of target and distractor processing (Adkins & Lee, 2021).

To address these questions, we developed a novel task (parametric attentional control task; PACT) that orthogonally varies target and distractor information, measuring how processing of these two dimensions varies both within and across trials. Our task merges elements of paradigms that have been separately popularized within the two research areas above. To capture variability in target processing, we based our task on the random dot kinematogram paradigm (Danielmeier, Eichele, Forstmann, Tittgemeyer, & Ullsperger, 2011; Kang et al., 2021; Mante, Sussillo, Shenoy, & Newsome, 2013; Shenhav, Straccia, Musslick, Cohen, & Botvinick, 2018). This task parametrically varies the motion discriminability (e.g., percentage of dots moving left) and color discriminability (e.g., percentage of green dots) across an array of dots. Participants were instructed to respond to the color dimension, while ignoring the motion dimension. Critically, whereas color response mappings were arbitrary (e.g., left hand for green), motion responses were exactly aligned with the direction of motion (e.g., left hand for leftward moving stimuli), resulting in potent "Simon-like" response interference from this prepotent distractor. Previous work has shown that this paradigm elicits response conflict and trial-to-trial adjusts for binary conflict (Danielmeier et al., 2011), and our current approach extends these previous findings by parametrizing both target coherence and distractor congruence to gain new insight into feature-selective control.

We find that participants independently and dynamically control target and distractor processing within a trial. Participants preferentially enhance target sensitivity under incentives, and preferentially suppress distractor sensitivity after high conflict trials. Moreover, they implement these control configurations by changing the initial conditions of feature processing dynamics. Finally, we find that these control strategies can be captured by extending classic neural network models of cognitive control to incorporate attractor dynamics that independently regulate the attentional priority of different task features. Together, these results extend our understanding of both decision-making and cognitive control by bridging the methodological and theoretical divides between these fields, providing new assays and process models for the top-down control of information processing.

Results

Participants performed the PACT, a perceptual discrimination task that required them to classify the dominant color in an array of moving dots (Figure 1a). Participants made bimanual responses, for example responding with their left hand when the dominant color was purple or

blue or responding with their right hand when the dominant color was green or beige. Across trials, we varied the extent to which those dots were coherently moving in the same or opposite direction as the correct response (distractor interference; Experiments 1-3) and how easily the participant could determine the dominant color (target discriminability; Experiments 2-3; Figure 1b). Participants performed the main *Attend-Color* PACT in blocks of 100 trials. To enhance the potency of motion as a distracting dimension (Shiffrin & Schneider, 1977) and allow for additional measures of automaticity and feature-specificity, participants alternated between these blocks-of-interest and shorter blocks (20-50 trials) in which participants instead responded to the direction of dot motion (*Attend-Motion* PACT; Figure 1c).

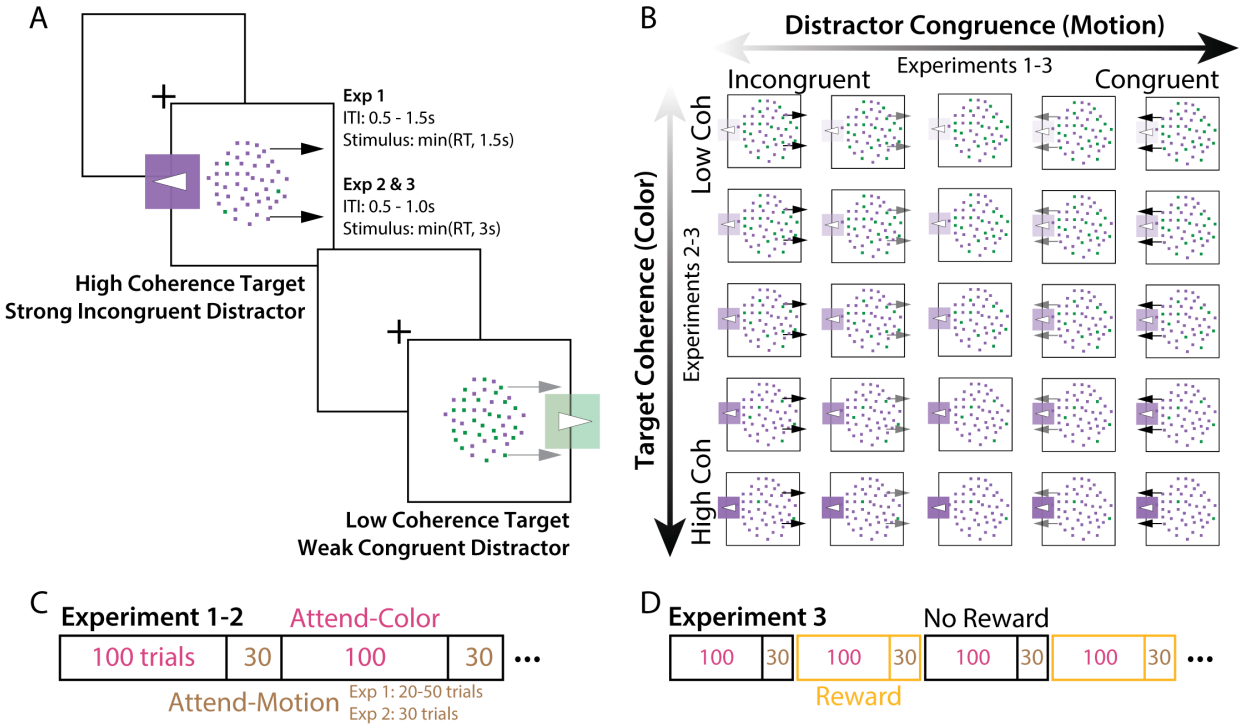


Figure 1. Parametric Attentional Control Task (PACT).
A) On each trial, participants responded to the dominant color in a bivalent random dot kinematogram. This stimulus had a random color (target) coherence, depending on the proportion of dots that were in the majority. This stimulus also had a random motion (distractor) congruence, depending on motion coherence in the same or opposite direction as the color response. B) Across trials, we parametrically and independently varied the coherence of the dominant color (y-axis) and the congruence of the motion direction (x-axis). C) In experiments 1 and 2, participants alternated between longer blocks of Attend-Color trials (target dimension was color, as in A) and shorter blocks of Attend-Motion trials (target dimension was motion). Participants took a self-timed break between blocks. D) In experiment 3, participants alternated between pairs of Reward blocks and No Reward blocks. On Reward blocks, participants could earn a monetary bonus if they were fast and accurate, whereas we just encourage good performance on No Reward blocks. Participants were informed of the reward condition during their break between blocks.

Task performance varies parametrically with distractor interference and target discriminability

In Experiment 1 (N = 58), participants performed the PACT with uniformly colored dots (e.g., all blue or all green), but with the dots moving in a direction either congruent or incongruent with that target response. We varied the strength of this distractor dimension between being fully congruent with the correct color response (100% leftward coherence for a left color response) to being fully incongruent (100% rightward coherence for a left color response; Figure 1b). For trials mid-way between these two extremes (cf. 'neutral' trials), the dots did not move consistently in one direction or another (0% motion coherence).

Consistent with past research on cognitive control, we found that participants were slowest and least accurate when distractors were fully incongruent (median RT = 585ms, mean accuracy = 89%) and fastest and most accurate were fully congruent (median RT = 553ms, mean accuracy = 97%; cf. (Danielmeier et al., 2011)). Performance on neutral trials (0% motion coherence) fell between these two extremes (median RT = 576ms, mean accuracy = 94%). Extending this work, hierarchical regression analyses (see Methods) revealed that performance varied in a graded fashion across this continuum of interference. Both accuracy (Cohen's *d* on regression estimate = -1.47) and reaction time (*d* = 1.25) worsened with parametrically increasing levels of interference (*p*s < 0.001, Figure 1c, Table 1).

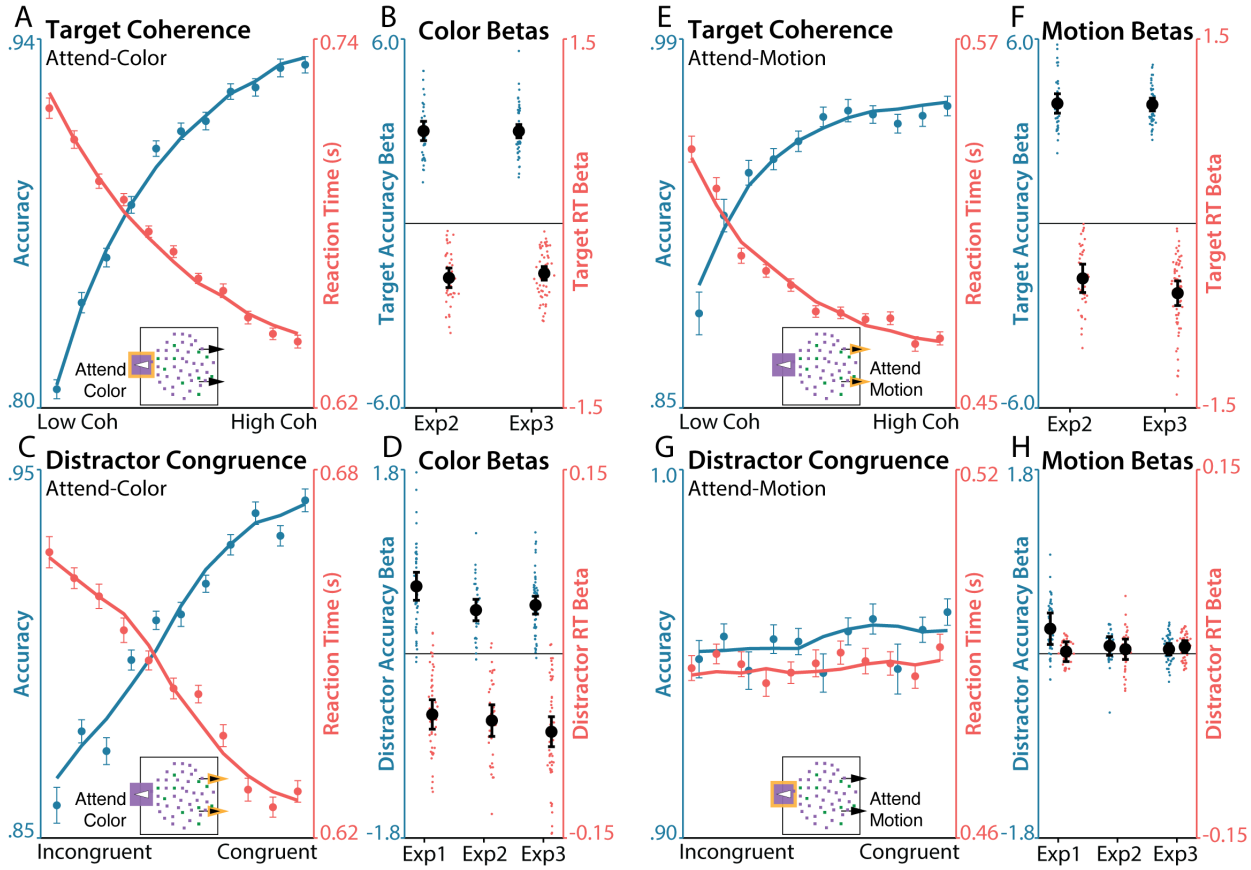


Figure 2. Target and distractor sensitivity.

A) Participants were more accurate (blue, left axis) and responded faster (red, right axis) when the target color had higher coherence. Lines depict aggregated regression predictions. In all graphs, behavior and regression predictions are averaged over participants and experiments. B) Regression estimates for the effect of target coherence on performance within each experiment, plotted for accuracy (blue, left axis) and RT (red, right axis). C) Participants were more accurate and responded faster when the distracting motion had higher congruence (coherence signed relative to target response). In all graphs, behavior and regression predictions are averaged over participants and experiments. D) Regression estimates for the effect of distractor congruence on performance within each experiment, plotted for accuracy and RT. E-F) Similar to A-B, performance (E) and regression estimates (F) for the effects of target coherence during Attend-Motion blocks, in which motion was the target dimension. G-H) As in A-B, performance (G) and regression estimates (H) for the effects of distractor congruence during Attend-Motion blocks, in which color was the distractor dimension. Error bars on behavior reflect within-participant SEM, error bars on regression coefficients reflect 95% CI.

In Experiment 2 (N=40) and Experiment 3 (N=60), participants performed the same task as in Experiment 1, but we additionally varied the discriminability of the target (color) dimension. Across trials, the proportion of the majority color (*color coherence*) varied parametrically to make color discrimination easier (higher coherence) or more difficult (lower coherence). As in Experiment 1, the level of motion interference also varied across trials, independently of targets.

Consistent with past research on perceptual decision-making (Britten et al., 1992; Mante et al., 2013), we found that discrimination performance improved with higher levels of target discriminability. Participants in both studies were faster (Exp 2: $d = -1.90$, Exp 3: $d = -1.99$) and more accurate (Exp 2: $d = 3.27$, Exp 3: $d = 3.73$) with parametrically increasing levels of color coherence (aggregate ps < 0.001; Figure 2a, Table 1). At the same time, we continued to find that participants were slower and less accurate when the goal-irrelevant movement of those dots was increasingly incongruent with the correct color response (see Figure 2b, Table 1).

Performance on our task varied parametrically with both color coherence and motion coherence, but these two coherence manipulations were designed to exert their influence on performance in different ways. Whereas variability in color coherence was intended to influence the stimulus uncertainty directly relevant to goal-directed decision-making (i.e., determining which response is the correct one), motion coherence was intended to exert a more automatic influence on response selection by facilitating responses consistent with the direction of motion (and suppressing the alternate responses. We confirmed this assumption regarding the relative automaticity of motion versus color processing by having participants perform interleaved blocks in which they responded based on motion and ignored color ('Attend-Motion'). We found that participants were more sensitive to the now-relevant motion coherence, but were no longer sensitive to the now-irrelevant color congruence (Supplementary Table 1-2). This asymmetry suggests that participants' decisions were not solely driven by the bottom-up salience of these features, as participants were more sensitive to color when it was relevant and less sensitive to motion when it was irrelevant, reflecting differential engagement of top-down control across the two tasks (Cohen, Servan-Schreiber, & McClelland, 1992).

Table 1. Target and distractor sensitivity

DV	**Predictors**	**Exp 1 (df=45) Effect size (d)**	**Exp 2 (df=25) Effect size (d)**	**Exp 3 (df=45) Effect size (d)**	**Aggregate p-value**
Choice	Target coherence		3.27	3.73	**1.01×10^{-44}**

	Distractor congruence	1.47144	1.42	1.50	**4.89×10^{-32}**
	Target × Distractor		-0.184	-0.344	**0.0226**
RT	Target coherence		-1.90	-1.99	**1.59×10^{-28}**
	Distractor congruence	-1.24983	-1.49	-1.43	**1.26×10^{-29}**
	Target × Distractor		0.230	0.0525	0.437

Effect sizes are calculated from MAP group-level regression estimates. P-values are aggregated across experiments, with statistically significant p-values (two-tailed, $\alpha = 0.05$) shown in bold.

Target discrimination and distractor interference occur in parallel

We found that participants' task performance varied parametrically with both the target discriminability and distractor congruence, both for choice and reaction time. We next sought to further understand the relationships between these changes in performance, within and across features.

First, we tested whether a given feature exerted a similar influence on both accuracy and RT. We found that this was indeed the case, as there was a significant correlation between the effect distractors had on accuracy and RT (rs < -0.87, ps < 0.001). The influences of target discriminability on accuracy and RT were also significantly correlated (rs < -0.54, ps < 0.001; Supplementary Table 3). Thus, participants who became faster with higher levels of a given feature's strength also became more accurate, suggesting that accuracy and RT shared a common underlying process (e.g., evidence accumulation rate, which we return to below).

Second, we tested whether the influences of target discriminability and distractor congruence on performance were independent (e.g., distractors and targets are processed in parallel; (Lindsay & Jacoby, 1994; Servant et al., 2014)) or instead modulatory (e.g., distractor congruence influences target sensitivity). If the two forms of feature processing modulated one another, we would predict that target and distractor coherence would interact in predicting performance. We did not find such an interaction in RTs (ds = 0.05 to 0.23, p = 0.33; Table 1), though we did find a small but significant interaction between target and distractor coherence on accuracy (ds = -0.18 to -0.34, p = 0.023). For both studies, removing target-distractor interactions as predictors in our accuracy regressions improved model fit (Protected exceedance probability on AIC: Exp 2 PXP = 1; Exp 3 PXP = 1). If distractors had an antagonistic influence on target processing, we would also predict that target and distractor sensitivity would be negatively correlated across subjects. Contrary to this prediction, these effects were either not significantly correlated or positively correlated, both for RT (Exp 2: $r(25) = .14$, $p = .48$; Exp 3: $r(45) = .44$, $p = .0019$) and accuracy (Exp 2: $r(25) = -.15$, $p = .45$; Exp 3: $r(45) = .12$, $p = .43$), suggesting that individual differences in target and distractor processing were not antagonistic.

Prior interference preferentially suppresses distractor sensitivity

Within a given trial, we found that performance varies parametrically and independently with the coherence of target and distractor features. We next sought to understand how participants

adapted their information processing *across* trials, to provide insight into the control processes that guide performance in this task. We measured how participants' feature sensitivity changed after difficult (e.g., more incongruent) trials, an index of cognitive control known as conflict adaptation (Egner, 2007; Gratton et al., 1992). The classic effect is that participants show weaker congruence effects after incongruent trials than after congruent trials, with the traditional interpretation being that this reflects upregulated target sensitivity (Botvinick et al., 2001; Egner, 2007). Our task allowed us to build on this work to test whether this adaptation effect varies parametrically with distractor congruence. Critically, we can now test whether adaptation occurs through an influence of prior distractor interference on subsequent enhancement of target processing, suppression of distractor processing, or both. Finally, we can further test whether adaptation occurs due to the discriminability of the *target* on the previous trial.

Across all three of our studies, we found that participants' sensitivity to the distractor dimension was robustly and parametrically influenced by the distractor congruence on the previous trial, as reflected both in their choice (ds = 1.44 to 1.74, $p < .001$; Figure 3a) and RT (ds = 0.83 to 1.79, $p < .001$; Figure 3b; Table 2). When the previous trial had congruent distractors, participants had strong sensitivity to the distractor congruence (Figure 3a-b, navy). When the previous trial had incongruent distractors, participants were much less sensitive to distractors (Figure 3a-b, red). These patterns are consistent with those typically observed in studies of conflict adaptation (Danielmeier et al., 2011; Egner, 2007), and further demonstrate gradations within these classic effects.

When varying both target and distractor features (Experiments 2-3), we found an additional influence of prior distractor congruence on target processing, whereby more incongruent previous trials enhanced the influence of target discriminability on the current trial (Figure 3d-e). However, the influence of prior distraction on target processing was substantially smaller than its effect on distractor processing (see Figure 5), and was only found for accuracy ($p < .001$) and not RT ($p = .57$), Finally, we found that performance adapted to the strength of the previous *target*, with less-discriminable targets yielding *lower* sensitivity to target strength (i.e., poorer performance) on the following trial, potentially due to disengagement (Supplementary Figure 2, Table 2). However, like the distractor-target effect, this target-target effect was much smaller than the distractor-distractor effects and only observable in accuracy ($p < .001$) and not RT ($p = .19$).

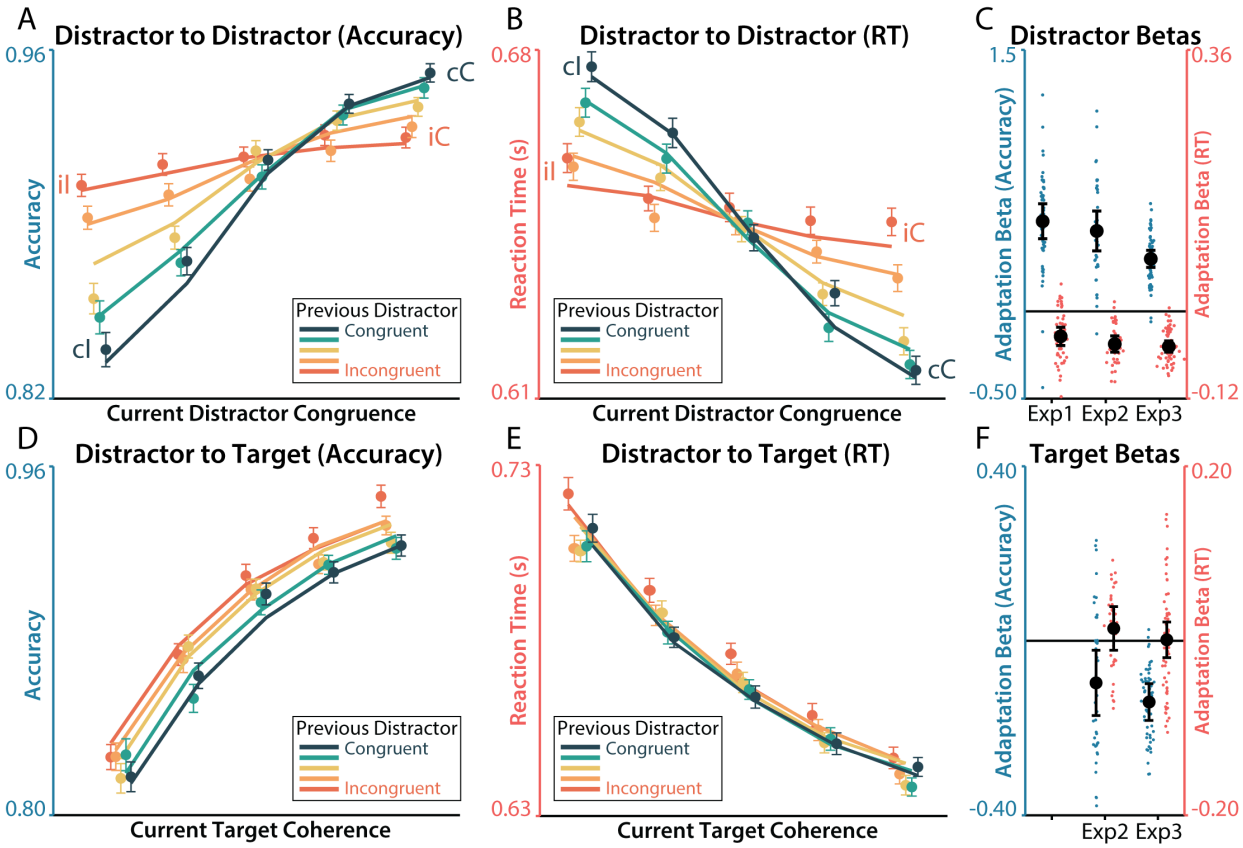


Figure 3. Distractor-dependent adaptation.
A-B) The relationship between distractor congruence and accuracy (A) and RT (B) was weaker when the previous trial was more incongruent (redder colors). Lines depict aggregated regression predictions. C) Regression estimates for the current distractor congruence by previous distractor congruence interaction, within each experiment. D-E) The relationship between target coherence and performance was stronger after more incongruent trials in accuracy (D) but not RT (E). F) Regression estimates for the current target coherence by previous distractor congruence interaction, within each experiment. Error bars on behavior reflect within-participant SEM, error bars on regression coefficients reflect 95% CI.

Table 2. Effects of previous conflict on feature sensitivity

DV	Predictors	Exp 1 (df=41) Effect size (d)	Exp 2 (df=15) Effect size (d)	Exp 3 (df=35) Effect size (d)	Aggregate p-value
Choice	Distractor × Prev Distract	1.59	1.45	1.74	**6.15×10^{-31}**
	Distractor × Prev Target		-0.670	0.103	0.964
	Target × Prev Distract		-0.473	-0.990	**2.83×10^{-8}**
	Target × Prev Target		0.418	0.644	**1.25×10^{-5}**
Lapse Rate	Prev Distract	-0.522	-0.498	-1.04	**1.75×10^{-10}**
	Prev Target		-0.110	-0.494	**0.00934**
RT	Distractor × Prev Distract	-0.836	-1.44	-1.79	**8.99×10^{-24}**

	Distractor × Prev Target		0.174	0.0618	0.520
	Target × Prev Distract		0.210	0.0155	0.726
	Target × Prev Target		0.147	0.154	0.285
	Prev Distract	0.287	0.202	-0.267	0.623
	Prev Target		0.109	-0.275	0.0884

Effect sizes are calculated from MAP group-level regression estimates. P-values are aggregated across experiments, with statistically significant p-values (two-tailed, $\alpha = 0.05$) shown in bold.

A common concern when measuring conflict adaptation effects is the extent to which these reflect control adjustment (as typically assumed) or low-level priming that can occur due to stimulus-stimulus or stimulus-response associations (Braem et al., 2019; Hommel, Proctor, & Vu, 2004; Mayr, Awh, & Laurey, 2003; Schmidt & De Houwer, 2011). For example, in some tasks, if two adjacent trials are both congruent or both incongruent, they are also more likely to share stimulus-response mappings, biasing analyses of sequential adaptation. Our experiment was designed to largely avoid this potential confound by eliminating stimulus-stimulus repetitions (with two colors assigned to each response hand that never repeat), and by using stochastic motion stimuli (versus, e.g., static arrows) that also have very infrequent exact repetitions (probability of coherence repetition is ~9.1%).

However, to further rule out that our key adaptation findings resulted from priming effects, we tested whether adaptation effects were present in our more automatic Attend-Motion blocks. Whereas a priming account would predict similar (within-feature) adaptation effects across both Attend-Color and Attend-Motion blocks, a cognitive control account would predict weaker adaptation effects for Attend-Motion than Attend-Color blocks. We found that adaptation effects during Attend-Motion blocks were overall weak and inconsistently signed (e.g., previous interference led to either increased or decreased sensitivity to distractors across studies; Supplementary Table 4-5). Comparing the adaptation effects across the two types of blocks directly, we found significantly stronger adaptation effects during Attend-Color than Attend-Motion blocks, both for distractors (distractor-distractor adaptation; Choice: $p < .001$; RT: $p < .001$) and for motion (target-target adaptation; Choice: $p < .001$; RT: $p = .34$; Supplementary Table 5). Together, these results suggest that the adaptation effects we observed during Attend-Color trials likely reflected changes in control states rather than stimulus-driven priming.

In addition to affecting sensitivity of choices and RTs to individual features (adaptation effects described above), we found that previous target and distractor information also exerted a small but reliable influence on the likelihood that the participant would respond randomly on the next trial (*lapse rates*). Specifically, higher levels of distractor incongruence and lower levels of target discriminability increased subsequent lapse rates (ps < .001; Table 2), though these changes were subtle (e.g., post-congruent lapse rates ranged from 0.023% to 0.13% across studies; post-incongruent lapse rates ranged from 0.13% to 0.41% across studies). We did not otherwise find consistent main effects of previous targets and distractors on choice behavior (i.e., in the direction of a particular response) or on RT.

Performance incentives preferentially enhance target sensitivity

We found that performance on our task adapted to prior distractor-related interference, and that this influence was observed primarily in subsequent processing of the distractor rather than the target. This may reflect a fundamental bias in the control system towards adjusting distractor processing in our task, but it may also reflect a process that is specialized for conflict adaptation. To disentangle these possibilities, we examined how target and distractor processing are influenced by heightened levels of motivation. In Experiment 3 we incorporated an incentive manipulation, with blocks of trials for which participants could either earn a monetary reward for fast and accurate performance, and blocks where performance was not rewarded (Figure 1d).

We found that participants' accuracy was more sensitive to target discriminability in rewarded blocks than non-rewarded blocks ($d = 0.61$, $p < .001$; Figure 4a, Table 3), but they did not significantly differ in their sensitivity to distractors ($d = -0.016$, $p = 0.91$). This effect of incentives on target sensitivity was specific to accuracy and not RTs ($d = -0.10$, $p = 0.47$), though participants were overall faster in rewarded blocks ($d = -0.41$, $p = 0.0045$). In terms of distractors, we found that in rewarded blocks participants were less sensitive to distractors in RT ($d = 0.35$, $p = 0.012$), albeit with a small effect size, and no effect on accuracy ($d = -0.10$, $p = 0.47$).

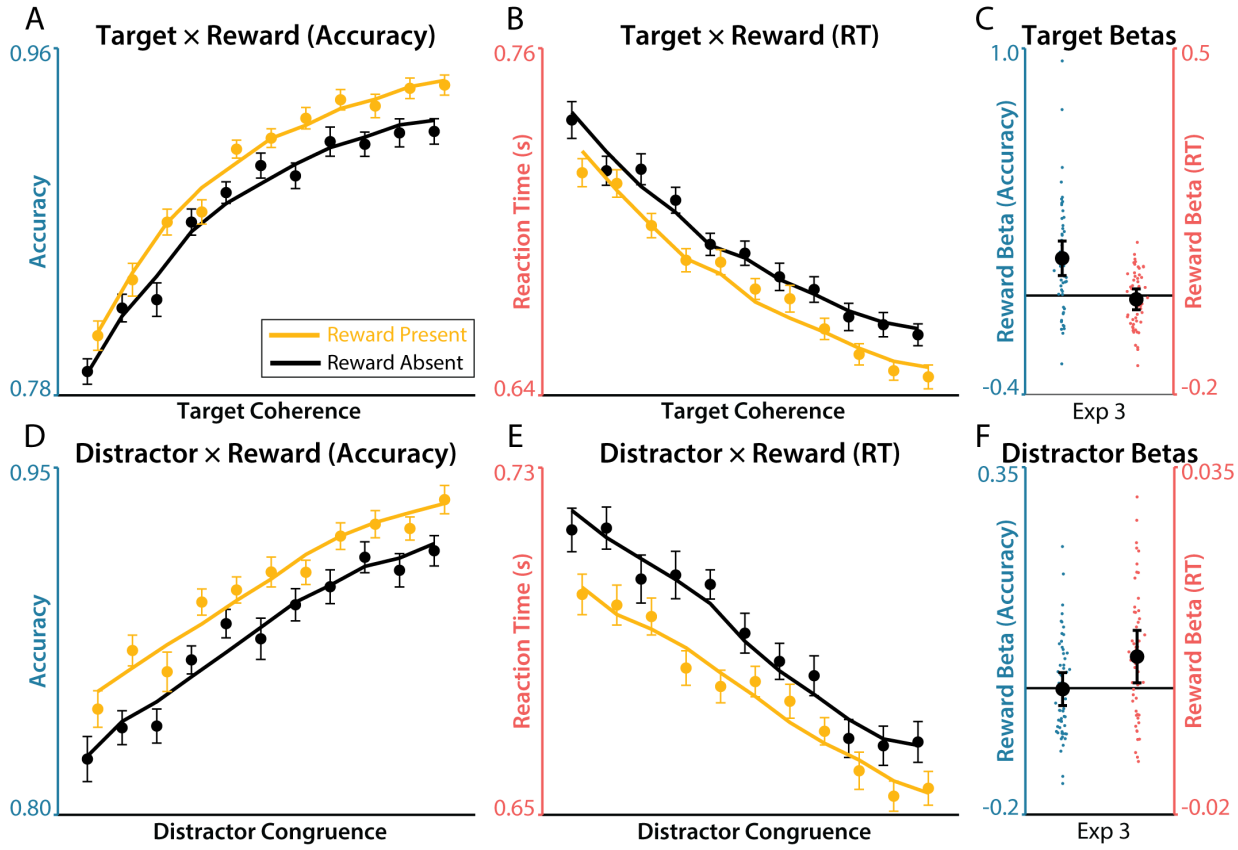


Figure 4. Influence of incentives on target and distractor sensitivity.
A-B) The relationship between target coherence and performance was stronger during incentivized blocks (gold) in the domain of accuracy (A), but not RT (B). Lines depict aggregated regression predictions. C) Regression estimates for the target coherence by incentive interaction. D-E) The relationship between distractor congruence and

performance was weaker on incentivized blocks (gold) in the for RT (E), but not Accuracy (D). F) Regression estimates for the distractor congruence by incentive interaction. Error bars on behavior reflect within-participant SEM, error bars on regression coefficients reflect 95% CI.

We further found that the target-enhancing effects of incentives also were not specific to the color dimension. When motion was the target dimension (attend-motion blocks), incentives preferentially increased sensitivity to motion coherence ($d = 0.70$, $p < .001$). Interestingly, incentives had an even larger influence on target sensitivity in attend-motion relative to attend-color blocks ($t(58.98) = 2.14$, $p = 0.036$; Supplementary Table 6-7).

Table 3. Effects of incentives on feature sensitivity

DV	**Predictors**	**Exp 3 (df=41) Effect size (d)**	**p-value**
Choice	Target × Reward	0.612	**8.56×10^{-5}**
	Distractor × Reward	-0.0156	0.911
Lapse Rate	Reward	0.244	0.0924
RT	Target × Reward	-0.103	0.467
	Distractor × Reward	0.349	**0.0195**
	Reward	-0.411	**0.00447**

Effect sizes are calculated from MAP group-level regression estimates. Statistically significant p-values (two-tailed, $\alpha = 0.05$) are shown in bold.

Incentives and previous interference have dissociable influences on target and distractor processing

Our within-trial results demonstrated that participants are sensitive to target and distractor information, with little interaction between these dimensions. Consistent with this putative independence, we found that previous interference primarily influenced distractor sensitivity (suppressing distractor sensitivity after trials with incongruent distractors), and that rewards primarily influenced target sensitivity (enhancing target sensitivity when incentivized). These findings strongly suggest a dissociation between target and distractor processing.

To confirm these findings, we formally tested the double dissociation between how incentives and previous interference influenced target and distractor choice sensitivity (Figure 5). We found that prior interference had a larger absolute effect on distractor processing than it did on target processing in both accuracy ($t(31.4)=9.54$, $p= 8.36 \times 10^{-11}$) and RT ($t(33.7)= 4.64$, $p= 5.14 \times 10^{-5}$). We found that rewards conversely had a larger influence on targets than distractors in Accuracy ($t(44.5)=5.08$, $p=7.22 \times 10^{-6}$), though not in RT ($t(37.7)=0.25$, $p=0.80$). The cross-over interaction was also significant in both Accuracy ($t(39.6)= 10.2$, $p= 1.36 \times 10^{-12}$) and RT ($t(48.3)= 3.11$, $p=0.0031$), supporting dissociable control over different dimensions of feature processing.

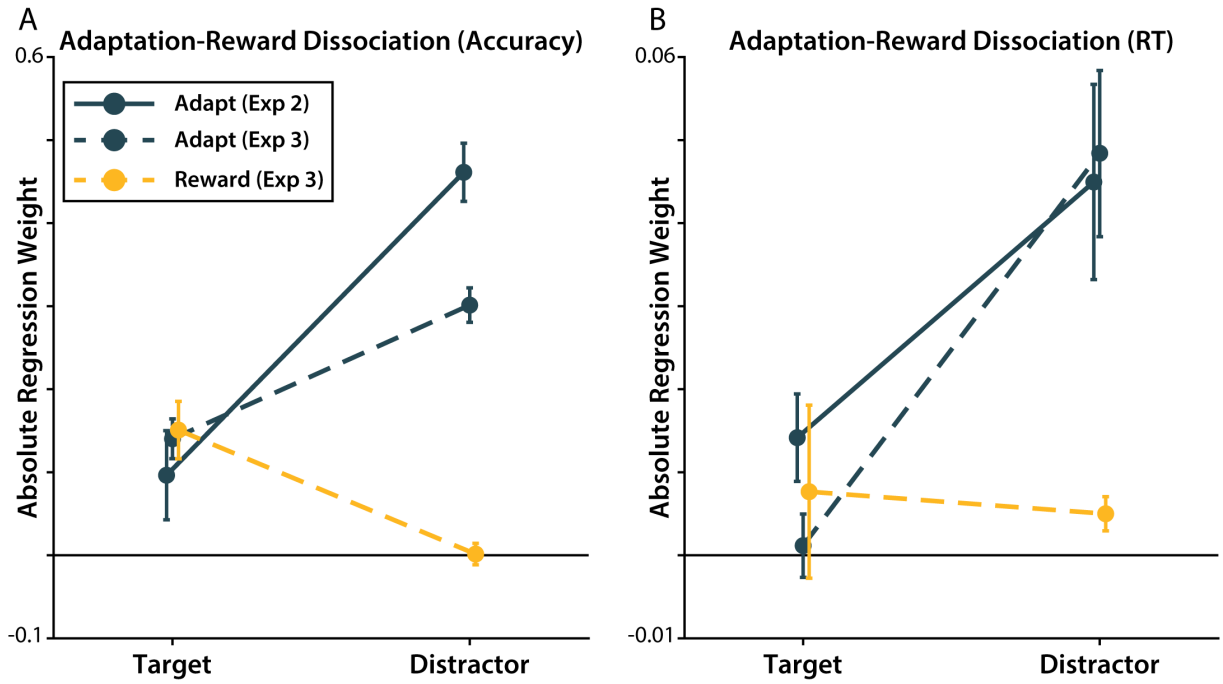


Figure 5. Dissociations between adaptation and incentive effects.
Adaptation effects were significantly larger on distractor sensitivity than target sensitivity in accuracy (A) and RT (B). In contrast, reward effects were significantly larger on target sensitivity than distractor sensitivity in accuracy (A) and similarly large in RT (B). Errors bars show MAP parameter SEM.

Differential within-trial dynamics of target and distractor processing

Our initial results show that participants independently control their sensitivity to target and distractor information. However, previous research has revealed that participants' task processing also dynamically changes within a trial (Servant et al., 2014; Weichart et al., 2020; White et al., 2011), including in response to incentives (Adkins & Lee, 2021). Whereas much of the previous research has focused on dynamics in spatial attention during flanker tasks (e.g., a shrinking spotlight of attention; (Weichart et al., 2020; White et al., 2011)), less is known about the dynamics of attention between features of conjunctive stimuli like those in our task, where target and distractor processing may be more independent (Adkins & Lee, 2021; Servant et al., 2014).

To test how sensitivity to target and distractor features changed within each trial, we measured whether the influence of coherence on participants' choices depended on reaction time (i.e., the *choice* ~ *coherence* × *RT* interaction). These analyses work under the logic that faster RTs reflect earlier epochs of information processing, consistent with the rationale for sequential-sample modelling (Bogacz, Brown, Moehlis, Holmes, & Cohen, 2006) and 'delta function' analyses on how congruence effects differ across RT quantiles (De Jong, Liang, & Lauber, 1994; Ridderinkhof, 2002). In the current study, we extended this methodology to study parametric changes in both target and distractor sensitivity over time.

At the earliest RTs, participants were the least sensitive to targets (Figure 6a) and the most sensitive to distractors (Figure 6d). At later RTs, participants became more sensitive to targets (ds = 0.69 to 0.97, $p < .001$), and less sensitive to distractors (ds = -0.71 to -1.5, $p < .001$; Table 4). This is consistent with an attentional control process that enhances sensitivity to goal-relevant features and suppresses attention towards goal-irrelevant features. Notably, these results suggest that this attentional process occurs 'online' within the course of a trial.

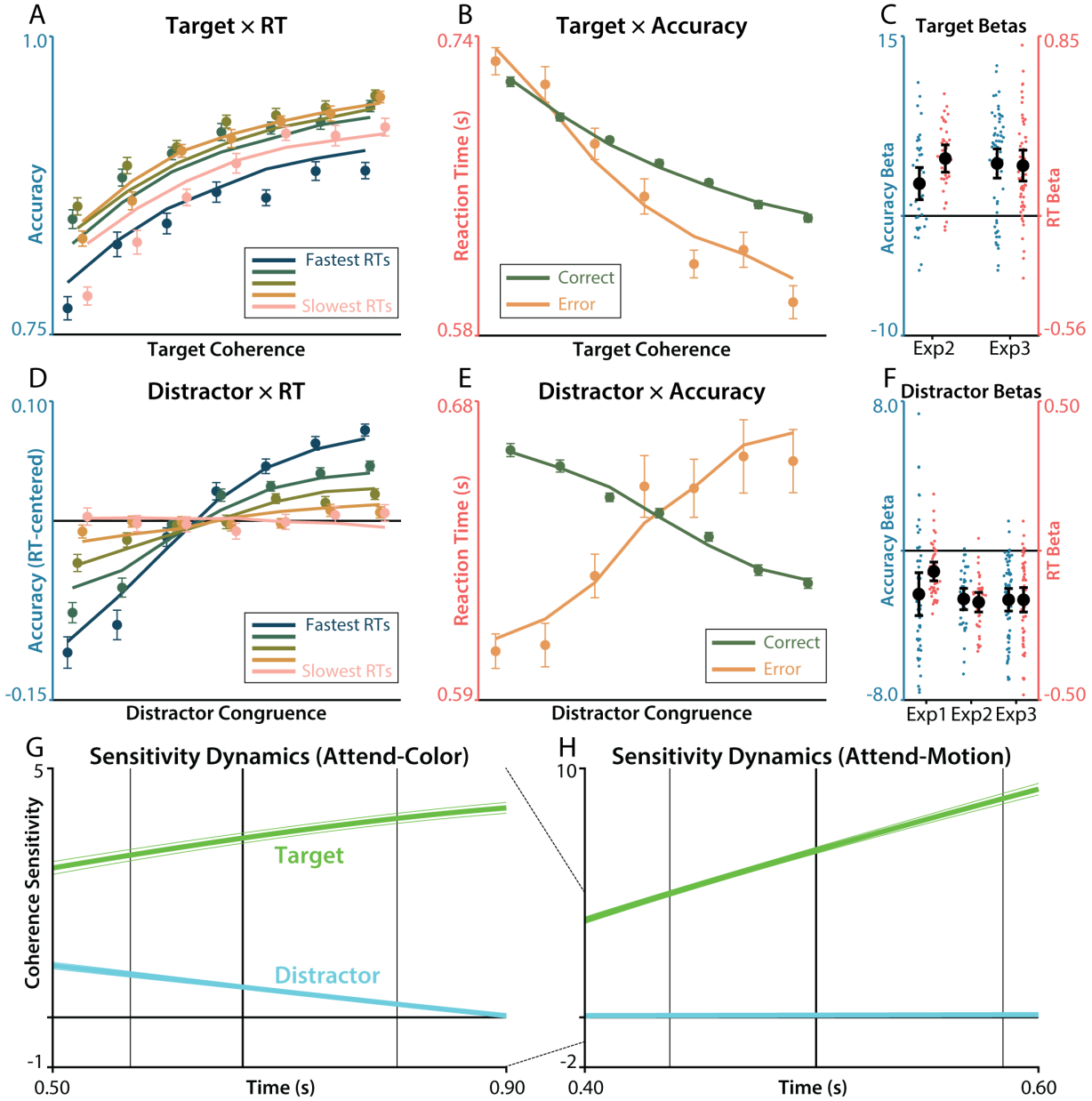


Figure 6. Target and distractor sensitivity dynamics.
A) The relationship between target coherence and accuracy increased at later RTs (pinker color). B) Participants responded faster on error trials that correct trial when target coherence was higher. C) Regression estimates for the interaction between target coherence and RT (blue) and accuracy (red), within each experiment. D) The relationship between distractor congruence and accuracy decreased at later RTs (pinker). Note that these data are mean-centered within each RT bin to remove the target effects from this visualization. E) Participants responded faster on error trials than correct trials when distractors were incongruent. F) Regression estimates for the interaction between

distractor congruence and RT (blue) and accuracy (red), within each experiment. G) Target (green) and distractor (cyan) sensitivity plotted as a function of reaction time, as estimated by our regression model in Attend-Color blocks. Vertical lines indicate quartiles of the RT distribution. H) Same as G, but generated from regression models fit to the Attend-Motion blocks. Note the different scaling of the x-axis and y-axis (see dashed line between plots). Error bars on behavior reflect within-participant SEM, error bars on sensitivity estimates reflect between-participant SEM of the predictions, error bars on regression coefficients reflect 95% CI.

We also fit a complementary analysis for RT (i.e., the *RT ~ coherence × accuracy* interaction). We found that participants had steeper target coherence slopes on error trials (*ds* = 0.89 to 1.5, $p < .001$; Figure 6b), driven by faster errors when the targets were high coherence, consistent with participants responding before their maximal target sensitivity. Likewise, we found that the relationship between RT and distractor congruence inverted on error trials (*ds* = -0.68 to -1.8, $p < .001$; Figure 6e), with participants making faster errors on more incongruent trials, consistent with an early sensitivity to distractors that is suppressed over time.

Table 4. Dynamics of feature sensitivity across response times

DV	Predictors	Exp 1 (df=38) Effect size (d)	Exp 2 (df=21) Effect size (d)	Exp 3 (df=41) Effect size (d)	Aggregate p-value
Choice	Target × RT		0.686	0.975	**4.61×10^{-11}**
	Distractor × RT	-0.709	-1.54	-1.19	**8.40×10^{-20}**
Lapse Rate	RT	0.247	0.928	0.534	**9.25×10^{-13}**
RT	Target × Accuracy		1.46	0.889	**1.99×10^{-14}**
	Distractor × Accuracy	-0.683	-1.82	-1.09	**1.39×10^{-20}**

Effect sizes are calculated from MAP group-level regression estimates. P-values are aggregated across experiments, with statistically significant p-values (two-tailed, $\alpha = 0.05$) shown in bold.

These findings suggest online dynamics in the allocation of top-down attention to facilitate target processing and suppress distractor processing, but it is possible that they instead reflect dynamics inherent to the bottom-up processing of color and motion information. To rule out this alternative hypothesis, we tested whether similar sensitivity dynamics were present during Attend-Motion blocks, when color information serves as a much less potent distractor. During these blocks, we found that participants enhanced target (motion) sensitivity faster than they did during Attend-Color blocks ($p < .001$; Figure 6h; Supplementary Table 8-9). In contrast, participants had slower distractor sensitivity dynamics during Attend-Motion blocks ($p < .001$). Together these results demonstrate that these sensitivity dynamics depend on the task that participants are performing, rather than being exclusively due to stimulus-driven factors.

Interference- and reward-related adaptations influence early trial dynamics

We found that, within a trial, participants dynamically adjusted attention depending on the task at hand, with increasing sensitivity to task-relevant information and decreasing sensitivity to task-

irrelevant information over the course of a trial. This raises the question whether the two forms of adaptation we observed, related to previous conflict and incentives, influenced different components of the within-trial attentional dynamics.

To address this question, we first examined how the dynamics of target and distractor sensitivity were altered by the congruence of the distractor on the previous trial (i.e., *Choice* ~ *PreviousDistractor* × *RT* × *Coherence*). We found that after incongruent trials, participants started the next trial more sensitive to targets and less sensitive to distractors (Figure 7a). Although this means that after congruent trials participants had worse initial conditions (starting less sensitive to targets and more sensitive to distractors), they appeared to compensate for this early disadvantage with faster increases in target enhancement (ds = 0.65 to 1.0, $p < .001$) and distractor suppression (ds = -0.68 to -1.1, $p < .001$; Table 5). Both post-congruent and post-incongruent trials thus reached similar asymptotic levels of feature sensitivity.

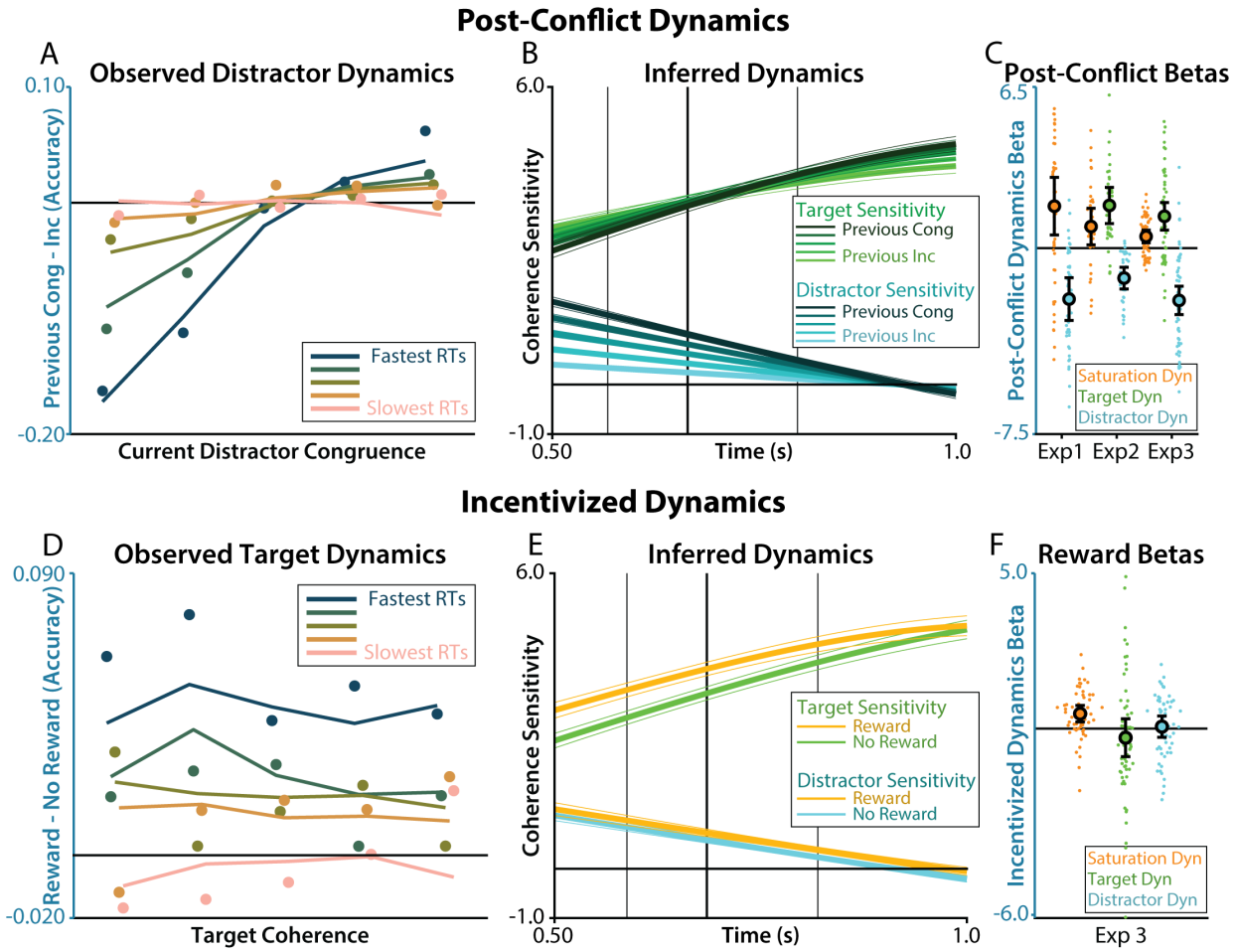


Figure 7. Influence of conflict and incentives on sensitivity dynamics.
A) The relationship between previous distractor congruence and current distractor congruence was strongest for early RTs (pinker color). The y-axis depicts the difference in accuracy between the extreme tertiles of previous congruence, for visualization purposes. B) Target (green) and distractor (cyan) sensitivity plotted as a function of previous congruence (color shade) and reaction time (x-axis), as estimated by our regression model. Vertical lines indicate quartiles of the RT distribution. C) Regression estimates for the interactions between reaction time and previous congruence on lapse rate (orange); or reaction time, previous congruence, and feature coherence on accuracy (target is green, distractor is cyan). D) The relationship between incentives and target coherence was

strongest for early RTs (pinker color). The y-axis depicts the difference in accuracy between blocks where there were rewards vs blocks without rewards. E) Target (green) and distractor (cyan) sensitivity plotted as a function of incentives (gold) and reaction time (x-axis), as estimated by our regression model. Vertical lines indicate quartiles of the RT distribution. F) Regression estimates for the interactions between reaction time and incentives on lapse rate (orange); or reaction time, incentives, and feature coherence on accuracy (target is green, distractor is cyan). Error bars on behavior reflect within-participant SEM, error bars on sensitivity estimates reflect between-participant SEM on the predictions, error bars on regression coefficients reflect 95% CI.

Figure 5. Effects of previous conflict on feature sensitivity dynamics

DV	**Predictors**	**Exp 1 (df=26) Effect size (d)**	**Exp 2 (df=9) Effect size (d)**	**Exp 3 (df=29) Effect size (d)**	**Aggregate p-value**
Choice	Prev Dist × Dist × RT	-0.853	-1.07	-1.01	**1.62×10^{-14}**
	Prev Dist × Targ × RT		1.01	0.646	**1.71×10^{-7}**
Lapse Rate	Prev Dist × RT	0.456	0.463	0.531	**3.52×10^{-6}**
RT	Prev Dist × Dist × Acc	-0.563	-1.15	-1.16	**2.84×10^{-13}**
	Prev Dist × Targ × Acc		-0.00500	-0.524	0.135
	Prev Dist × Acc	0.417	0.175	-0.162	0.881

Effect sizes are calculated from MAP group-level regression estimates. P-values are aggregated across experiments, with statistically significant p-values (two-tailed, $\alpha = 0.05$) shown in bold.

We performed the equivalent analysis for incentive-related adaptation (i.e., *Choice ~ Reward × RT × Coherence*). We found that during incentivized blocks, participants' initial target sensitivity was higher than during non-incentivized blocks, and remained so across much of the trial (see Figure 7d). However, target sensitivity eventually reached an asymptote, such that towards the end of the trial both incentivized and non-incentivized trials had similar levels of target sensitivity (see slowest quantile in Figure 7d). This convergence was accounted for by larger increases in lapse rates later in incentivized trials ($d = 0.52$, $p < .001$; Table 6). The dynamics of distractor sensitivity, by contrast, did not differ between incentivized and non-incentivized trials (d=0.055, p=0.71).

Table 6. Effects of incentives on feature sensitivity dynamics

DV	**Predictors**	**Exp 3 (df=29) Effect size (d)**	**p-value**
Choice	Rew x Target × RT	-0.139	0.331
	Rew x Distractor × RT	0.0546	0.712
Lapse Rate	Rew × RT	0.524	**0.000937**
RT	Rew × Target × Acc	0.437	**0.00432**
	Rew × Distractor × Acc	-0.170	0.260

	Rew × Acc	-0.540	**0.000646**

Effect sizes are calculated from MAP group-level regression estimates. Statistically significant p-values (two-tailed, α = 0.05) are shown in bold.

An accumulator model of attentional control over target and distractor processing

Our results demonstrate that participants independently control the initialization and online adjustment of attention towards target and distractor features. To parsimoniously account for this set of findings, we developed an accumulator model that integrated elements of previous models used to separately account for performance in tasks involving perceptual discrimination (Gold & Shadlen, 2007; Ratcliff & McKoon, 2008) and overriding prepotent distractors (Cohen, Dunbar, & McClelland, 1990; Weichart et al., 2020; White et al., 2011). We used a variant of a feedforward inhibition model, in which inputs provide excitatory inputs to associated response units and inhibitory inputs to alternative response units (Shadlen & Newsome, 2001). Our decision model takes as inputs the color and motion coherence in support of different responses, nonlinearly transforms these inputs, and then integrates evidence for each response in separate rectified accumulators with balanced feedforward excitation and inhibition (Figure 8). The signal-to-noise ratio of the intermediate layer's outputs are determined by control units that determine the gain of a given feature (Cohen et al., 1990; Musslick, Bizyaeva, Agaron, Leonard, & Cohen, 2019). We hand-tuned the parameters of this model to determine whether it could capture our core experimental findings across choice and reaction time.

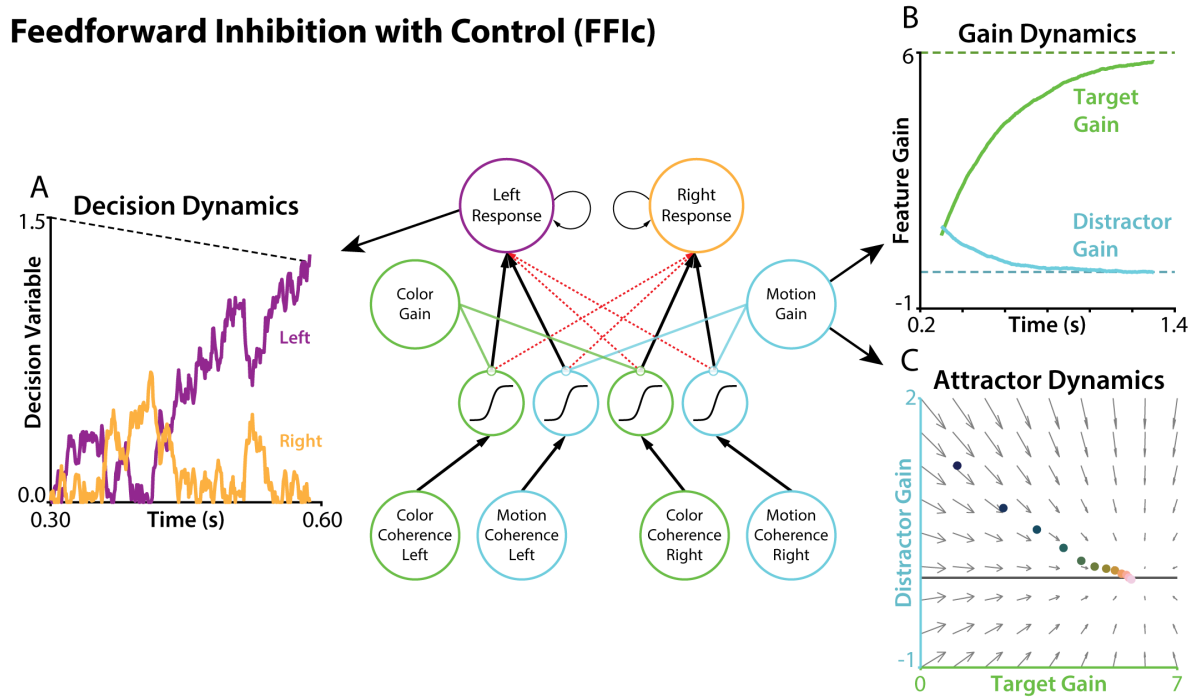


Figure 8. Feedforward inhibition with control.
Color evidence (green) and motion evidence (blue) are transformed and accumulated to make a choice. Balanced excitatory connections (black solid lines) and inhibitory connections (red dashed lines) cause accumulation of the

difference in evidence for each response. A) Evidence for the left response (purple) and right response (orange) are accumulated over time without leak. When one of the accumulators crosses a (linearly collapsing) decision threshold, the model chooses that response. B) Within each trial, the signal-to-noise of each feature pathways is controlled by a feature gain. Over time within a trial, the feature gains for targets (green) and distractors (cyan) exponentially decay to a fixed level (high gain for targets, zero gain for distractors). C) An equivalent visualization of the dynamics in B. Attractor dynamics drive target and distractor gains to their fixed level, shown at different timepoints within the trial (pinker colors are later in the trial). The horizontal line depicts zero distractor gain.

Our accumulator model was able to reproduce our key within-trial findings. During our main Attend-Color trials, it generated responses that were faster and more accurate with increasing color coherence (Figure 9a) and slower and less accurate with increasing motion incongruence (Figure 9b). We simulated Attend-Motion trials by increasing the target gain and decreasing the distractor gain, to capture potential differences in both automaticity and control. Now, our model generated responses that were even faster and more accurate with increasing target coherence (now motion; Figure 9c) but that were insensitive to distractor congruence (now color; Figure 9d), replicating the main behavioral results in Attend-Motion blocks. Notably, distractor effects were not reproduced in an accumulator competition model parameterized to be more 'race-like' (Supplementary Figure 3; (Teodorescu & Usher, 2013)). This occurred because larger inputs (whether congruent or incongruent) drove faster reaction times.

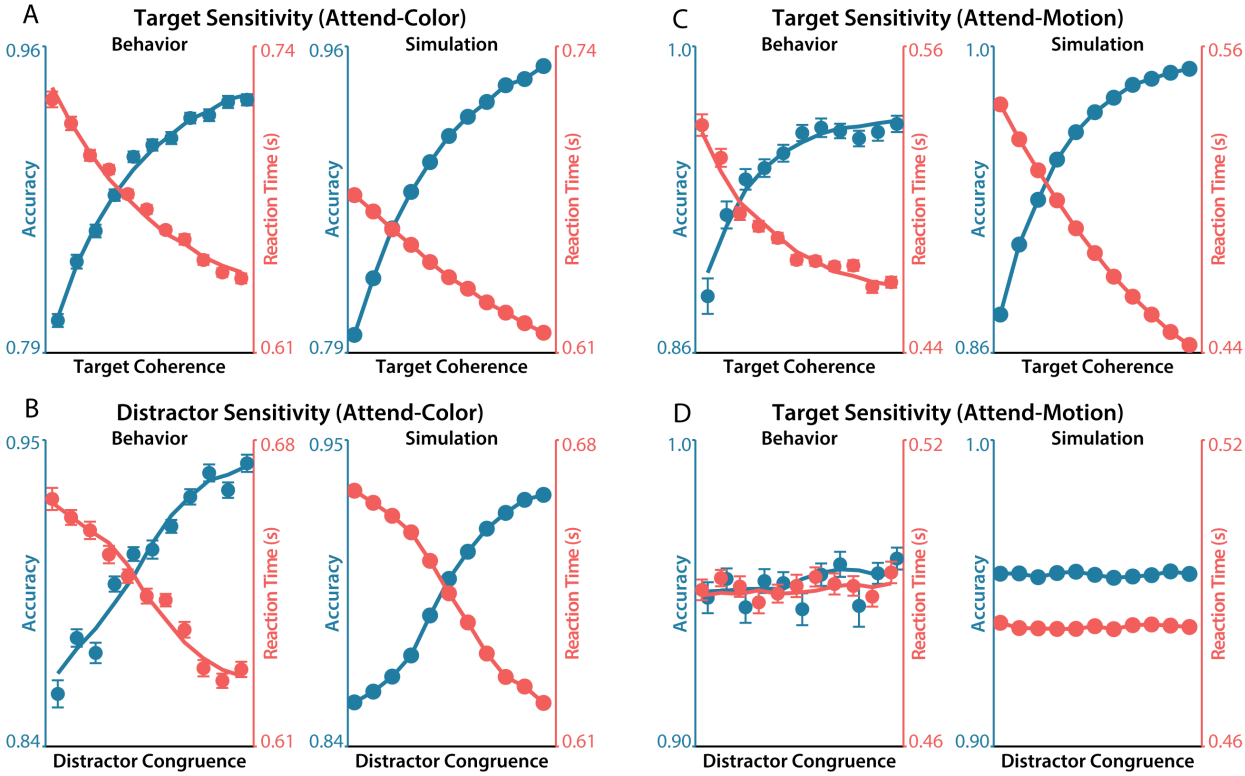


Figure 9. Simulation of target and distractor sensitivity (compare to Figure 2).
A-B) Sensitivity to target coherence (A) and distractor congruence (B) in behavior (left) and in the FFIc simulation (right) for Attend-Color blocks. C-D) Same as A-B, but for Attend-Motion blocks.

We next used this model to test potential mechanisms underlying participants' within- and between-trial control adaptations. First, we tested whether participants' apparent within-trial dynamics in feature sensitivity plausibly resulted from actual within-trial changes in control

gains governing feature sensitivity, or whether such dynamics could result from static control gains. We implemented time-varying feature gains as attractors with an initial gain (e.g., reflecting bottom-up salience or prior learning) that exponentially decays to a fixed point (e.g., determined by the task goals and control; cf. (Musslick et al., 2019; Steyvers, Hawkins, Karayanidis, & Brown, 2019)). We found that incorporating these time-varying gains into our accumulator model allowed it to reproduce participants' behavioral dynamics. In accuracy, our model replicated the shift in target sensitivity over time, with the collapsing bound reducing performance on the slowest trials (Figure 10a). Our model similarly captured participants' decreased target sensitivity at later RTs (Figure 10b). Finally, our model recreated the analogous effects in RT, with faster errors for high coherence targets and incongruent distractors (Figure 10c-d). Critically, we were unable to replicate these qualitative patterns of behavior with a model in which control gains that were frozen throughout the trial (Supplementary Figure 4).

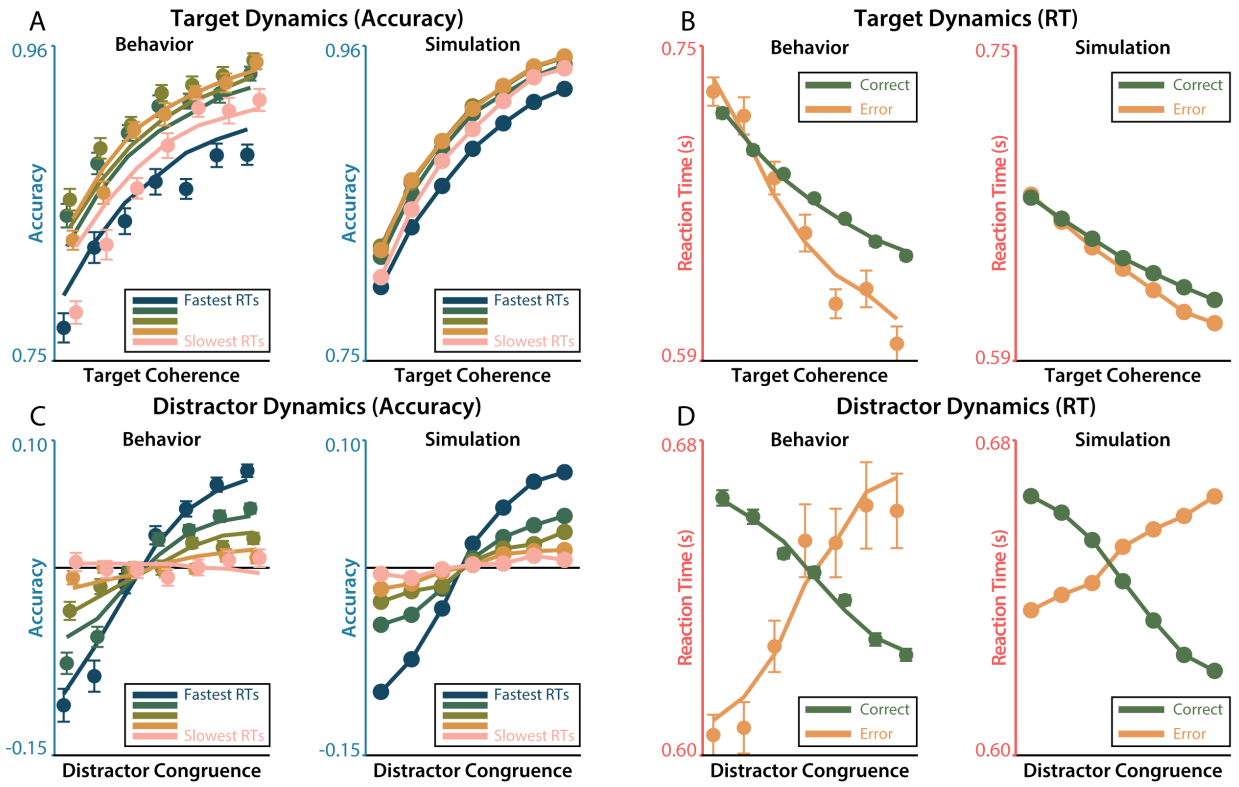


Figure 10. Simulation of target and distractor sensitivity dynamics (compare to Figure 6).
A-B) RT-dependent (A) and accuracy-dependent (B) sensitivity to target coherence in behavior (left) and in the FFIc simulation (right). C-D) Same as A-B, but for distractor congruence.

The parallel feature pathways in this model are designed to capture the independent influences of a target and distractor information (Lindsay & Jacoby, 1994). However, the time-varying feature gains providing an account for the weak interactions we observed between target and distractor sensitivity in accuracy. Despite there being no competition in feature processing in our model, we found these weak target-distractor interactions emerge in *simulated* accuracies but not simulated RTs. This interaction appeared to result from the different time courses of target and distractor sensitivity. As in participants' behavior, the model's errors due to incongruent distractors tend to occur early (Figure 10c-d), end target processing at a lower (early) level of

sensitivity. This interplay between feature sensitivity *dynamics* (but not overall feature sensitivity per se) offers a plausible explanation for the subtle and seeming inconsistent interactions in participants' behavior.

Having provided an account of how each of our stimulus features is processed over the course of the trial depending on the task goal, we next tested a potential model-based account of the two forms of control adaptation we observed across trials. Our participants demonstrated enhanced target sensitivity on rewarded blocks, and suppressed distractor sensitivity after increasingly incongruent trials. In both cases, adaptation appeared to enhance sensitivity to stimulus features at the fastest reaction times. To account for these findings, we modified the initial conditions of our model's gain dynamics (Figure 11a). We simulated post-interference adaptation by initializing the distractor gain closer to its asymptote, and we simulated reward incentivization by initializing the target gain closer to its asymptote. We found that these simulations qualitatively reproduced participants' behavior, with stronger adaptation and reward effects earlier in the trial than later. The exponential decay in our attractor network parsimoniously accounts for the fact that dynamics tended to be faster when they were initialized further from the fixed point (i.e., after trials with higher levels of congruency). Thus, our model was able to capture the range of findings in this experiment, from target-distractor sensitivity, within-trial dynamics, and how the dynamics of target and distractor processing may be influenced by control.

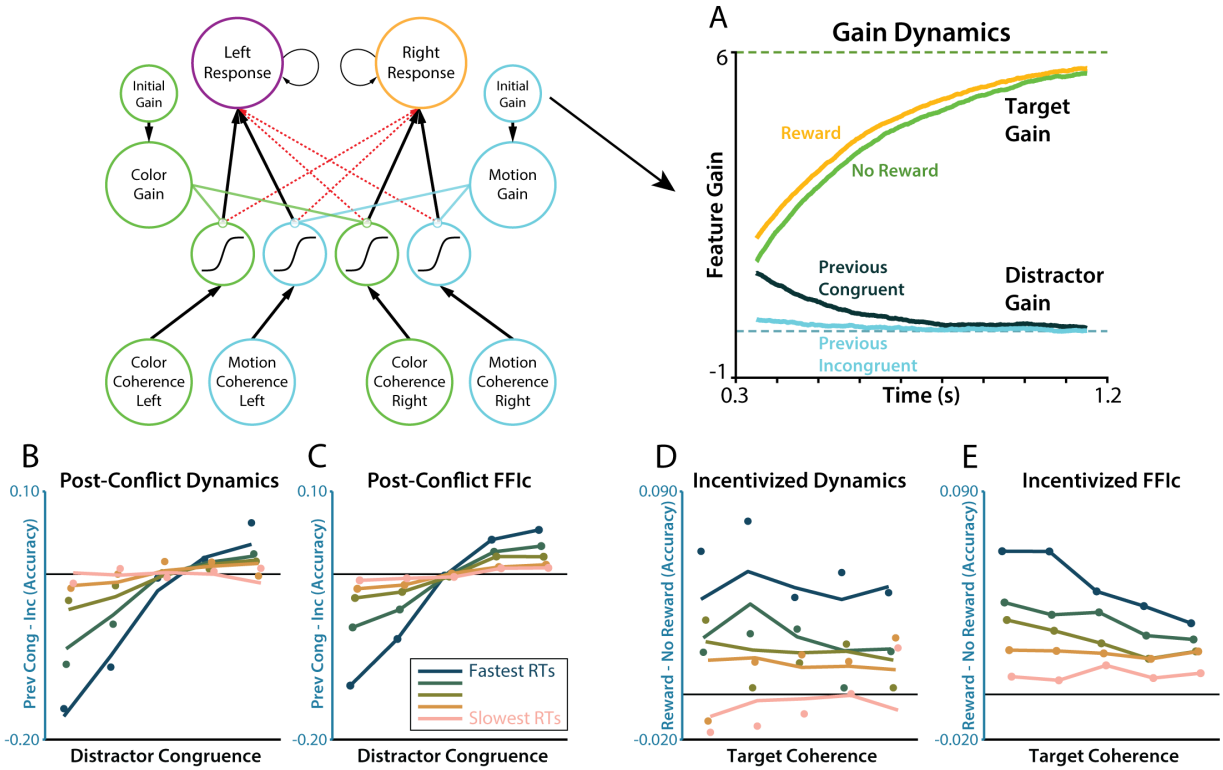


Figure 11. Simulation of post-conflict and incentive effects (compare to Figure 7).
A) The influences of previous congruence (shade) and incentive effects (gold) were implemented through changes to the initial conditions of the feature gain dynamics, with previous congruence influencing initial distractor gain and incentives influence initial target gain. B) The influence of previous congruence on distractor sensitivity dynamics in behavior (left) and in the FFIc simulation (right). C) The influence of incentives on target sensitivity dynamics in behavior (left) and in the FFIc simulation (right).

Discussion

When faced with distraction, we can sustain good performance by engaging with relevant information or ignoring disruptive information. Our experiment revealed that these strategies are under independent cognitive control, and are driven by distinct attentional dynamics. Using a bivalent random dot motion task with parametric target and distractor coherence (PACT), we found that target and distractor information have independent influences on participants' performance. Furthermore, we found that participants' sensitivity to targets and distractors was preferentially modulated by incentives and previous interference, respectively. These adaptations altered the initial conditions of feature-selective gains, which was followed by dynamic enhancement to target gains and suppression of distractor gains. These behavioral phenomena could be parsimoniously explained by a hybrid sequential sampling model with goal-dependent attractor dynamics over feature weights.

Together, these results support a cognitive control architecture that is parametric, multivariate, and dynamic. Previous research has found that cognitive effort is enhanced in response to incentives (Hübner & Schlösser, 2010; Leng, Yee, Ritz, & Shenhav, 2020; Padmala & Pessoa, 2011; Parro, Dixon, & Christoff, 2018) and to prior conflict (Egner, 2007; Gratton et al., 1992). The current experiments extend these previous findings to show that these adaptations are both graded in their intensity, and selective in the allocation. These findings are consistent with a multivariate perspective on cognitive control, in which people optimize a configuration of control signal according to their costs and benefits (Musslick, Shenhav, Botvinick, & Cohen, 2015; Shenhav, Botvinick, & Cohen, 2013). The target and distractor configurations observed here add to a body of work teasing apart the conditions under which people coordinate across multiple control signals (Danielmeier et al., 2011; Leng et al., 2020; Simen et al., 2009).

A core question arising from these results is why there are preferential relationships between prior interference with distractors, and incentives with targets. One possibility is that this is due to credit assignment. A system that could properly assign credit to features based on their contribution to conflict and rewards should allocate control towards distractors and targets. Distractors are a salient source of response conflict, and participants could adjust sensitivity to reduce this conflict. When participants were performing the more automatic Attend-Motion blocks, during which response conflict was absent, this adaptation was also absent. In contrast, reward contingencies were explicitly tied to target discrimination performance. During Attend-Motion blocks, there was a stronger association between target coherence and performance (e.g., due to response compatibility, and that only targets contributed to accuracy), potentially explaining why these blocks had larger incentive effects. This account is consistent with Bayesian models of cognitive control, such as those that predict feature congruence (Jiang, Heller, & Egner, 2014; Yu et al., 2009) or the value of control policies (Bustamante, Lieder, Musslick, Shenhav, & Cohen, 2021; Lieder, Shenhav, Musslick, & Griffiths, 2018).

Our results also provide insight into the dynamic implementation of attentional control. Previous work has shown that within-trial attentional dynamics play an important role in both decision making (Callaway, Rangel, & Griffiths, 2021; Krajbich, Armel, & Rangel, 2010; Li & Ma, 2021; Westbrook et al., 2020) and cognitive control (Adkins & Lee, 2021; Hardwick, Forrence, Krakauer, & Haith, 2019; Servant et al., 2014; Weichart et al., 2020; White et al., 2011). These

foundational experiments have largely focused on spatial attention, with far less known about the dynamics of feature-based attention, where processing of targets and distractors is less mutually constrained. Furthermore, relatively few experiments have studied how attentional dynamics are modified in response to control drivers like incentives or task demands (Adkins & Lee, 2021; Yu et al., 2009).

Our experiment shows that the dynamics of target and distractor sensitivity are independent, and that prior interference and incentives appear to operate through changes to the initial conditions of these feature gains, rather than their speed or asymptotic level. If participants are learning the relevance of different features, it's possible that these initial conditions reflect the prior probability that attention towards targets or distractors will support task goals (Lieder et al., 2018; Yu et al., 2009). Similar to how response priors are reflected in the initial decision state (Bogacz et al., 2006; Simen et al., 2009), priors on feature priority may be reflected in the initial attentional state. In the case of previous interference, this could reflect learning whether distractors enhance performance (e.g., after trials on which congruent distractors led to better performance), or a local estimate of the probability a trial will be congruent (Yu et al., 2009). For incentives this may reflect the expected target-reward contingency. Future research should investigate this account by measuring attentional dynamics as participants learn task contingencies (Shenhav et al., 2018).

Our observations of conflict-dependent and incentive-dependent dynamics also lend stronger support to our interpretation of top-down control. After congruent trials, participants started the next trial with more similar target and distractor gains, that were then more quickly separated within the trial (Figure 7b). If these dynamics were an artifact of the decision process (e.g., due to accumulator attractors (Wong & Wang, 2006)), then we would expect that when target and distractor gains are initially more similar (i.e., post-congruent trials), there would be slower dynamics. Instead, we found that there was the fastest dynamics for post-congruent trials, supporting a role for feedback control that reconfigures attentional gain to align with task goals. Additionally, during incentivized blocks, we saw that participants had modified attentional dynamics for targets, but not distractors. This finding supports the independence of these attentional dynamics, demonstrating that participants can alter dynamics in one feature but not the other. These findings suggests that participants can coordinate attention across features, modifying their initial conditions and reconfiguring attentional gain within a trial.

Such a dynamical process model may help link behavior in response conflict tasks to cognitive dynamics in other domains. In the domain of task-switching, recent cognitive models have developed similar dynamical model of how people reconfigure task sets. Classic work has shown that switch costs exponentially decay with preparation time (Monsell & Mizon, 2006; Rogers & Monsell, 1995), similar to the dynamics in the current task. Computational models have formalized these task sets dynamics during the switch preparation period (Musslick et al., 2019; Ueltzhöffer, Armbruster-Genç, & Fiebach, 2015) across trials (Steyvers et al., 2019). If the within-trial dynamics we observe can be thought of a "micro- adjustment to the task set" (Ridderinkhof, 2002), then this modelling approach may help unify switching dynamics across executive domains. For example, the adjustments to initial conditions in response to task demands and incentives in the current experiment may provide new insight into these other domains. Interestingly, control over initial conditions also play a central role in the neural

dynamics of motor preparation (Churchland, Cunningham, Kaufman, Ryu, & Shenoy, 2010; Remington, Egger, Narain, Wang, & Jazayeri, 2018; Remington, Narain, Hosseini, & Jazayeri, 2018), highlighting the similarities across motor and cognitive domains (Ritz, Frömer, & Shenhav, 2020) and offering potential neural predictions for our task.

This distinction between initialization and dynamics is reminiscent of pre-trial proactive control strategies and peri-trial reactive control strategies, respectively (Braver, 2012; Chiew & Braver, 2017). Our results provide process-level measures of how control strategies can influence separable components of information processing dynamics (e.g., the initial conditions, dynamics, and asymptotic level of attentional gain). Future research should link our measure of attentional dynamics to standard indices of proactive and reactive control (e.g., list-wise and item-specific proportion congruence effects, (Gonthier, Braver, & Bugg, 2016)).

We found that our core results could be well-captured by a sequential sampling model with attentional dynamics. This model combines a standard decision-making model (FFI) with an attractor network that regulates the gain on different stimulus features. A primary touchstone for our model was the classic feedforward inhibition model of Stroop performance (Cohen et al., 1990), as well as recent generations of dynamic drift diffusion models (Evans & Servant, 2020; Ulrich, Schröter, Leuthold, & Birngruber, 2015; Weichart et al., 2020; White et al., 2011). Our parametric manipulations of target discriminability and distractor congruence offer precise measurements of participants' decision and control process, confirming and extending these previous accounts. A key observation from our approach is that drift-diffusion-like dynamics are necessary to account for participants' slowing when incongruent distractors have higher coherence, unlike the prediction of race-like models (Teodorescu & Usher, 2013; Weichart et al., 2020). Our model also provides process-level insight into how attractor networks can account for the dynamics of attentional control, parsimoniously explaining both within-trial gain dynamics and, critically, how adaptation alters the initial conditions of these feature gains. Future research should directly fit this model to participants' behavior, potentially leveraging recent advances in the efficient estimation of parameters for complex decision models (Fengler, Govindarajan, Chen, & Frank, 2021), allowing for finer grained comparisons between different accounts of dynamic attentional control.

Our analyses of attentional dynamics depend on participants' own response times and choices, raising concerns about selection biases. While evidence accumulation modelling typically depends on choice-conditioned reaction times, inferring time-varying influence of targets and distractors present a particularly difficult challenge. To address these concerns, we used simulations to show that the dynamic profiles we observed are not expected when evidence accumulation models have static gains on targets and distractors (Supplementary Figure 4). In contrast, these patterns can be accounted for by the inclusion of dynamic feature gains (Figure 9-11). These results are consistent with previous work showing that DDM estimates of attentional dynamics in conflict tasks are theoretically unbiased (White et al., 2011; White, Servant, & Logan, 2018). Even if these measurements are not biased, using sparse behavioral measures is an inefficient method for measuring latent dynamics, and may combine multiple processes (e.g., accumulation and threshold adjustments). More precise inferences about these latent dynamics should leverage converging measurement about attentional state, such as using interrogation designs (Adkins & Lee, 2021; Hardwick et al., 2019), motor tracking (Erb, Moher, Sobel, &

Song, 2016; Menceloglu, Suzuki, & Song, 2021; Scherbaum, Dshemuchadse, Fischer, & Goschke, 2010), or temporally-resolved neuroimaging (Fischer, Nigbur, Klein, Danielmeier, & Ullsperger, 2018; Scherbaum, Fischer, Dshemuchadse, & Goschke, 2011; Weichart et al., 2020; Yeung, Botvinick, & Cohen, 2004). Integrating across multiple measures of decision and attentional dynamics is needed for the strongest conclusions about the architecture of the decision and control systems.

This experiment provides new insight into how we flexibly adapt to the changing demands of our environment. We find that this flexible control aligns information processing with task goals, and can be captured by a well-defined process model. The developments from this experiment can help extend models of cognitive control towards richer accounts of how multivariate control configurations, such as across targets and distractors, are optimized during goal-directed behavior.

Methods

Participants

All participants provided informed consent in compliance with Brown University's Institutional Review Board, participating for either course credit or pay. We excluded participants from our analyses if they had <70% accuracy during attend-color blocks or completed less than half of the experiment. Fifty-seven individuals participated in Experiment 1 (mean(SD) age: 20.6(2.21); 36 female; 1 excluded), 42 individuals participated in Experiment 2 (age: 19.1(0.971); 31 female; 2 excluded), and 62 individuals participated in Experiment 3 (age: 19.8(1.38); 47 female; 2 excluded), resulting in 156 included participants across the three experiments.

Parametric Attentional Control Task (PACT)

We developed the parametric attentional control task (PACT), extending tasks used to study decision-making (Mante et al., 2013; Shenhav et al., 2018) and cognitive control (Danielmeier et al., 2011). On each trial, participants viewed an array of moving dots (i.e., random dot kinematogram), presented in one of four colors (see Figure 1). Participants were taught to match two colors to a left keypress and two colors to a right keypress (with colors counterbalanced across participants). The majority color did not repeat on adjacent trials to avoid feature priming (Braem et al., 2019).

The direction of the dot motion (leftward or rightward) was task-irrelevant and could be consistent with the color response (distractor congruent trials) or it could be inconsistent with this response (distractor incongruent trials). Uniquely in this experiment, we parametrically varied the degree of distractor congruence on each trial by varying the motion coherence (percentage of dots moving in the same direction). Distractor congruence was linearly spaced between 95% congruence and 95% incongruence, drawn randomly across trials.

In Experiment 1, all of the dots were the same color (100% color coherence), creating a parametric extension of the Simon conflict tasks (Danielmeier et al., 2011). In Experiments 2 and

3, the dots contained a mixture of two colors associated with different responses. Color coherence was linearly spaced between 65% to 95%, drawn randomly across trials.

To maintain the salience of the motion dimension throughout the session (Shiffrin & Schneider, 1977), participants alternated between blocks of the task above ('attend-color' trials, putatively more control-demanding) and blocks where participants were instructed to instead indicate the direction of the dot motion ('attend-motion' trials, putatively less control-demanding). Mirroring the attend-color blocks, in Experiment 1 we held the motion coherence constant (maximal) during attend-motion blocks, while varying the color coherence across trials. In Experiments 2 and 3, we varied the coherence of both dimensions during attend-motion blocks.

Session

Participants first performed 100 motion-only training trials (0% coherent color) and 100 color-only training trials (0% coherent motion; order counterbalanced across participants) to learn the stimulus-response mappings. During training, participants received accuracy feedback on every trial. During the main experiment, participants performed two types of interleaved blocks, without trial-wise feedback. Participants alternated between longer attend-color blocks (100 trials) and shorter attend-motion blocks (Experiment 1: 20-50 trials; Experiment 2-3: 30 trials; order counterbalanced across participants). In Experiments 1 and 2, at the end of each block participants were told their average accuracy and median RT, and encouraged to respond quickly and accurately. Participants were not given this information in Experiment 3 to avoid interactions with the incentive manipulation (see below). Participants took self-timed breaks between blocks.

Stimuli

Participants were seated approximately 60cm from a computer screen, making their responses on a customizable gaming keyboard in a dark testing booth. The random dot motion array was presented in the center of the screen (~15 visual degrees in diameter; ~66.8 dots per visual degree squared; 19" LCD display at 60Hz). The dots colors were approximately (uncalibrated) isoluminant and perceptually equidistant (RGB: [187, 165, 222], [150, 180, 198], [192, 169, 168], [157, 184, 130]; (Teufel & Wehrhahn, 2000)) and moved at ~15 visual degrees per second. Each trial started with a random inter-trial interval (Experiment 1: 0.5 – 1.5s; Experiment 2-3: 0.5 – 1.0s). There was an alerting cue .3s before the trial onset, indicated by the fixation cross turning from grey to white, to minimize non-decision time. The stimuli were then presented until either a response was made, or a deadline was reached (Experiment 1: 3s; Experiment 2-3: 5s).

Task Variants

Experiment 1: These data incorporate several similar versions of this task developed during piloting. The main differences across versions were the number of distractor congruence levels (mean(range) = 13.5(11-15)), the number of trials per attend-motion block (mean(range) = 26(20-50)), and the total number of trials (mean(range) = 469(300-700) attend-color trials). We did not find significant differences in performance across versions, and so our analyses collapsed across these versions. Experiment 1 also included a learning condition in a separate set of blocks, which was outside the scope of the current paper and not included in the analyses we report.

Experiments 2 & 3: These data come from a single task variant (though see Experiment 3's incentive manipulation below). In this variant, we presented participants with 11 levels of target coherence and 11 levels of distractor congruence, linearly spaced within their coherence range and randomly sampled across trials. Participants performed 12 blocks of 100 attend-color trials interleaved with 12 blocks of 30 attend-motion trials.

Incentivized Variant (Experiment 3)

In Experiment 3, we informed participants before the main session that they would be able to earn a monetary reward for good performance. On 'Reward' blocks, we would randomly select trials, and if they were both fast and accurate on those trials, they would receive a bonus payment. On 'No Reward' blocks, participants would not be eligible to earn a reward. We indicated the incentive condition at the beginning of each block with a label and text coloring (gold text for 'Reward', white text for 'No Reward'). Participants performed Attend-Color and Attend-Motion blocks in one incentive condition before alternating to the other incentive condition (order counterbalanced across participants). At the end of the experiment, participants received a bonus calculated from their performance (mean(SD) bonus: $2.5($0.57)USD).

Regression Analyses

Analysis code is available at github.com/xxxxx (TBD). We used a hierarchical nonlinear regression of choice and reaction time as a tractable and minimally theory laden measure of performance (Supplementary Figure 1). In particular, we implemented hierarchical expectation maximization (EM) in MATLAB R2020a (using emfit; available at github.com/mpc-ucl/emfit) to provide a maximum a posteriori (MAP) estimates for the mean and covariance of parameters linking task features to performance. This fitting algorithm alternates between finding the MAP estimates of participants' parameters given the current group-level expectations (M-step; with 5 parameter re-initialization per step), and updating this group-level expectation based on participants' estimated parameters (E-step), repeated until convergence. We fit separate regression to each experiment for independent replications of our findings.

Our regression approach simultaneously estimated parameters for choice and RT, under an additive likelihood function:

$$loglike(total) \ = \ loglike(Choice) + loglike(RT)$$

Our choice sub-function used a lapse-logistic link function:

$$Choice \sim \frac{1 - lapse}{1 + exp(-\beta_{Choice} X_{Choice})} + (lapse \times 0.5)$$

$$lapse \ = \frac{1}{1 + exp(-\beta_{Lapse} X_{Lapse})}$$

Where $\beta_{Choice}$and β_{Lapse} are parameter vectors, and X_{Choice} and X_{Lapse} are design matrices. Our

RT sub-function used a shifted lognormal link function:

$$log(RT - ndt) \sim \beta_{RT} X_{RT}$$

Where again $\beta_{RT} X_{RT}$ is a linear model, and ndt is the estimated non-decision time. Rare RTs less than ndt were assigned a small likelihood.

Finally, the prior probability of the parameters was evaluated under a multivariate normal distribution defined by the group-level parameter mean and covariance, improving the robustness of our estimates through regularization. Critically, we estimated this group-level covariance across both choice and RT parameters, which better regularized our estimates and produced a joint model of performance at the group level.

All regression design matrices included an intercept (choice bias or average RT), an autoregressive component (previous trial's choice or RT), and the transformed target and distractor coherence (scaled between -1 and 1). We transformed feature coherences using a saturating nonlinearity,

$$coh^{*}{}_{feature} = \frac{tanh(\alpha_{feature} \times coh_{feature})}{tanh(\alpha_{feature})}$$

with α_{target} and $\alpha_{distractor}$ fit as free parameters. This approach distinguishes the coherence nonlinearity (α) from how strongly coherence influences performance (β_{Choice} and β_{RT}), with our analyses focused on the latter. To constrain these α parameters, we estimated one parameter for both choice and RT, capturing similar non-linearities across both performance measures.

In our more complex models (e.g., incentives), our primary focus was on how additional task features moderated the influence of feature coherence on performance. Lower order effects of moderating factors (e.g., previous distractor congruence) were included in the lapse rate for choice analysis, and as a main effect in RT analyses. The full parameter sets for all analyses are available in Supplementary Data.

We excluded trials in our regression if they were 1) the first trial of the block, 2) shorter than 200ms or longer than 2s, 3) occurred after an error or after a trial was too fast/slow and 4) if the current trial was an error in reaction time analyses. These exclusion criteria were chosen to be inclusive, while avoiding trials where there were likely to be a mixture of different cognitive processes (e.g., post-error adjustments).

We performed statistical inference on the parameters using an estimate of the group-level error variance from the emfit package, necessary to avoid violations of independence across participants from our hierarchical modelling. Contrast tests across models used Welsh's (unequal variance) t-tests, with contrasts weighting studies by the square root of the sample size. We aggregated p-values across studies using Lipták's method (Lipták, 1958; Zaykin, 2011), weighting studies by the square root of their sample size. Correlations between parameters were calculated by converting the group-level MAP covariance to a correlation matrix.

We generated posterior predictives (trend lines on figures) by generating model predictions for all trials, and then aggregating these predictions in the same way as participants' behavior. This approach allows us to distinguish whether our model systematically deviates from behavior from whether deviations are driven by variability in parameters across participants.

We generated sensitivity dynamics plots (e.g., Figure 6) by computing the regression-estimated coherence effect conditioned on RT. For a range of simulated RTs, the estimated motion sensitivity timeseries is:

$$\beta_{motion}^{RT} = (\beta_{motion} + \beta_{motion:RT} SimRT) \odot (1 - lapse^{RT})$$

$$lapse^{RT} = \frac{1}{1 + exp(-(\beta_{Lapse} + \beta_{RT} SimRT))}$$

Where βs are regression weights estimated in our analysis, $SimRT$ is a vector of simulated RTs (e.g., .5:.01:1), and $\odot$ indicates element-wise multiplication. For control-dependent dynamics (i.e., incentivized dynamics; see Figure 7), we included 2-way and 3-way interactions between feature sensitivity, RT, and control drivers. We generated these sensitivity dynamics for each participant, and then plotted the mean and between-participant standard error.

Feedforward Inhibition with Control Model

Simulation code is available at github.com/xxxxx (TBD). To provide a bridge between our regression analyses and processes models of decision-making, we adopted a generative modeling approach and tested whether participants behavior could be reproduced by a sequential sampling model (Figure 8).

This model takes as inputs the color and motion coherence in support of different responses (e.g., $coh_{colorLeft}$), nonlinearly transforms these inputs (e.g., $coh^{*}{}_{colorLeft}$; see regression analyses above), and then integrates evidence for each response in separate rectified accumulators (x_{left} and x_{right}).

For example, evidence for the left response would be calculated as:

$$dx_{left} = -\lambda x dt + (\beta_{color} coh^{*}{}_{colorLeft} dt + \beta_{motion} coh^{*}{}_{motionLeft} dt + N(0, \sigma_x)_{left} \sqrt{dt}) - \omega(\beta_{color} coh^{*}{}_{colorRight} dt + \beta_{motion} coh^{*}{}_{motionRight} dt + N(0, \sigma_x)_{right} \sqrt{dt})$$

$$if\ x_{left} < 0;\ x_{left} = 0$$

The model makes a choice when one of the accumulators reaches a linearly collapsing decision bound rectified above 0.01. We used a balanced feedforward inhibition model without leak ($\lambda = 0$ and $\omega = 1$), approximating a (rectified) drift diffusion process (Bogacz et al., 2006). Note that parameterizations of a leaky competing accumulator could also approximate the DDM (Bogacz

et al., 2006; Bogacz, Usher, Zhang, & McClelland, 2007), and so are plausible alternatives to our implementation. We preferred the FFI model because it provides a simple interpolation between DDM and race-like decision processes.

To capture dynamics in participants' feature sensitivity, we modified our accumulation model to incorporate an attractor network for the feature weights, a model we call the feedforward inhibition with control model (FFIc model). This control process has a gain setpoint (e.g., aims for zero gain on distractors), and noisily drives the gain from its initial condition (β^0) towards this attractor according to a gain K. For example, the motion gain would be governed by:

$$d\beta_{motion} = -\gamma\beta dt + K_{motion}(fixedpoint_{motion} - \beta_{motion})dt + N(0, \sigma_{gain})\sqrt{dt}$$

With the leak term γ fixed to 0 as in the decision process.

We simulated 10,000 trials for each combination of target discriminability and distractor congruence (11 x 11 x 10,000), and then aggregated simulated behavior in the same way we aggregated participants' behavior.

Supplementary Tables

Supplementary Table 1. Target and distractor sensitivity (Attend-Motion)					
DV	**Predictors**	**Exp 1 (df=45) Effect size (d)**	**Exp 2 (df=25) Effect size (d)**	**Exp 3 (df=45) Effect size (d)**	**Aggregate p-value**
Choice	Target coherence		4.39362	5.43513	**5.694e-53**
	Distractor congruence	0.73214	0.35222	0.26921	**0.0009158**
	Target × Distractor		0.05215	0.145	0.6210
RT	Target coherence		-1.46537	-1.62518	**2.711e-21**
	Distractor congruence	0.12478	0.17862	0.40742	0.1205
	Target × Distractor		0.07099	0.03714	0.8639

Effect sizes are calculated from MAP group-level regression estimates. P-values are aggregated across experiments, with statistically significant p-values (two-tailed, $\alpha = 0.05$) shown in bold.

Supplementary Table 2. Target and distractor sensitivity (Attend-Color - Attend-Motion)					
DV	**Predictors**	**Exp 1 Effect size (d)**	**Exp 2 Effect size (d)**	**Exp 3 Effect size (d)**	**Aggregate p-value**
Choice	Target coherence		-0.58753 (df = 49.999)	-0.61499 (df = 89.779)	**1.003e-10**
	Distractor congruence	0.4317 (df = 88.85)	0.74335 (df = 49.064)	0.97127 (df = 75.762)	**5.786e-19**
RT	Target coherence		0.0095723 (df = 44.421)	0.33943 (df = 69.76)	0.1815
	Distractor congruence	-0.80385 (df = 78.829)	-1.1948 (df = 42.812)	-1.3803 (df = 59.109)	**2.6939e-31**

Effect sizes are calculated from Welsh's contrasts across regression models. P-values are aggregated across experiments, with statistically significant p-values (two-tailed, $\alpha = 0.05$) shown in bold.

Supplementary Table 3. Correlations between RT and accuracy betas			
Experiment	**Correlands**	**MAP r-stat**	**p-value**
Exp 1 (df = 54)	Distractor Betas	-0.8907	**1.9505e-20**
Exp 2 (df = 38)	Target Betas	-0.7123	**1.2673e-07**
	Distractor Betas	-0.8753	**7.4826e-14**
Exp 3 (df = 58)	Target Betas	-0.5402	**4.1956e-06**
	Distractor Betas	-0.9066	**1.0534e-23**

Parameter correlations are calculated from the MAP group-level parameter covariance. Statistically significant p-values (two-tailed, α = 0.05) are shown in bold.

Supplementary Table 4. Effects of previous conflict on feature sensitivity (Attend-Motion)					
DV	**Predictors**	**Exp 1 (df=41) Effect size (d)**	**Exp 2 (df=19) Effect size (d)**	**Exp 3 (df=39) Effect size (d)**	**Aggregate p-value**
Choice	Distractor × Prev Distract	0.01129	-0.17575	0.29325	0.57785
	target × Prev Target		-0.00795	-0.42449	0.26774
RT	Distractor × Prev Distract	0.50266	-0.42662	0.18143	0.29106
	target × Prev Target		0.1312	-0.03334	0.77691

Effect sizes are calculated from MAP group-level regression estimates. P-values are aggregated across experiments, with statistically significant p-values (two-tailed, α = 0.05) shown in bold.

Supplementary Table 5. Effects of previous conflict on feature sensitivity (Attend-Color - Attend-Motion)					
DV	**Predictors**	**Exp 1 Effect size (d)**	**Exp 2 Effect size (d)**	**Exp 3 Effect size (d)**	**Aggregate p-value**
Choice	Distractor adaptation	0.50494 (df=52.735)	0.98418 (df=33.561)	0.27955 (df=47.674)	**4.3873e-08**
	Motion Adaptation		0.44783 (df=21.857)	0.65143 (df=40.186)	**2.2711e-05**
RT	Distractor adaptation	-0.62876 (df=80.178)	-0.95554 (df=29.31)	-1.2012 (df=70.693)	**3.3996e-23**
	Motion Adaptation		-0.23705 (df=19.072)	-0.09294 (df=39.344)	0.3372

Effect sizes are calculated from Welsh's contrasts across regression models. P-values are aggregated across experiments, with statistically significant p-values (two-tailed, α = 0.05) shown in bold.

Supplementary Table 6. Effects of incentives on feature sensitivity (Attend-Motion)			
DV	**Predictors**	**Exp 3 (df=41) Effect size (d)**	**p-value**
Choice	Target × Reward	0.70266	**6.01995e-05**
	Distractor × Reward	0.2888	0.1098
Lapse Rate	Reward	-0.06961	0.6699
RT	Target × Reward	-0.1257	0.4153
	Distractor × Reward	-0.19237	0.2654
	Reward	-0.09546	0.515886

Effect sizes are calculated from MAP group-level regression estimates. Statistically significant p-values (two-tailed, α = 0.05) are shown in bold.

Supplementary Table 6. Effects of incentives on feature sensitivity (Attend-Color - Attend-Motion)			
DV	**Predictors**	**Exp 3 Effect size (d)**	**p-value**
Choice	Target × Reward	-0.27899 (df = 58.976)	**0.036287**
	Distractor × Reward	-0.23007 (df = 48.259)	0.11651
RT	Target × Reward	0.056553 (df = 58.3)	0.66747
	Distractor × Reward	0.27099 (df = 76.967)	**0.01992**

Effect sizes are calculated from Welsh's contrasts across regression models. Statistically significant p-values (two-tailed, α = 0.05) are shown in bold.

Supplementary Table 8. Dynamics of feature sensitivity across response times (Attend-Motion)					
DV	**Predictors**	**Exp 1 (df=38) Effect size (d)**	**Exp 2 (df=21) Effect size (d)**	**Exp 3 (df=41) Effect size (d)**	**Aggregate p-value**
Choice	Target × RT		2.0344	1.3136	**3.3048e-20**
	Distractor × RT	-0.25658	0.1941	-0.48983	**0.022611**
Lapse Rate	RT	0.50911	0.76305	-0.23407	0.94091
RT	Target × Accuracy		0.77152	0.67898	**1.5881e-07**
	Distractor × Accuracy	0.07579	-0.21081	0.01985	0.93086

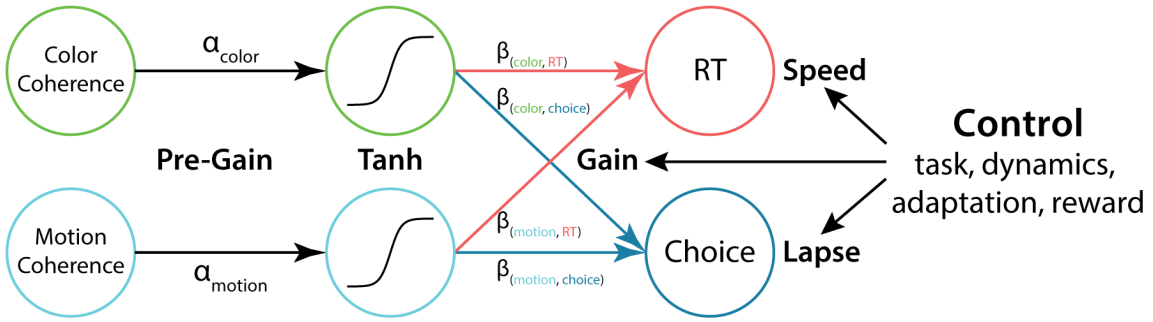
Effect sizes are calculated from MAP group-level regression estimates. P-values are aggregated across experiments, with statistically significant p-values (two-tailed, α = 0.05) shown in bold.

Supplementary Table 9. Dynamics of feature sensitivity across response times (Attend-Color - Attend-Motion)					
DV	**Predictors**	**Exp 1 Effect size (d)**	**Exp 2 Effect size (d)**	**Exp 3 Effect size (d)**	**Aggregate p-value**
Choice	Target × RT		-2.0773 (df=23.03)	-1.0879 (df=46.102)	**3.353e-17**
	Distractor × RT	-0.10198 (df=56.335)	-0.58539 (df=22.396)	-0.53939 (df=78.472)	**4.567e-05**
RT	Target × Accuracy		-0.42922 (df=23.046)	-0.42209 (df=46.164)	**0.001480**

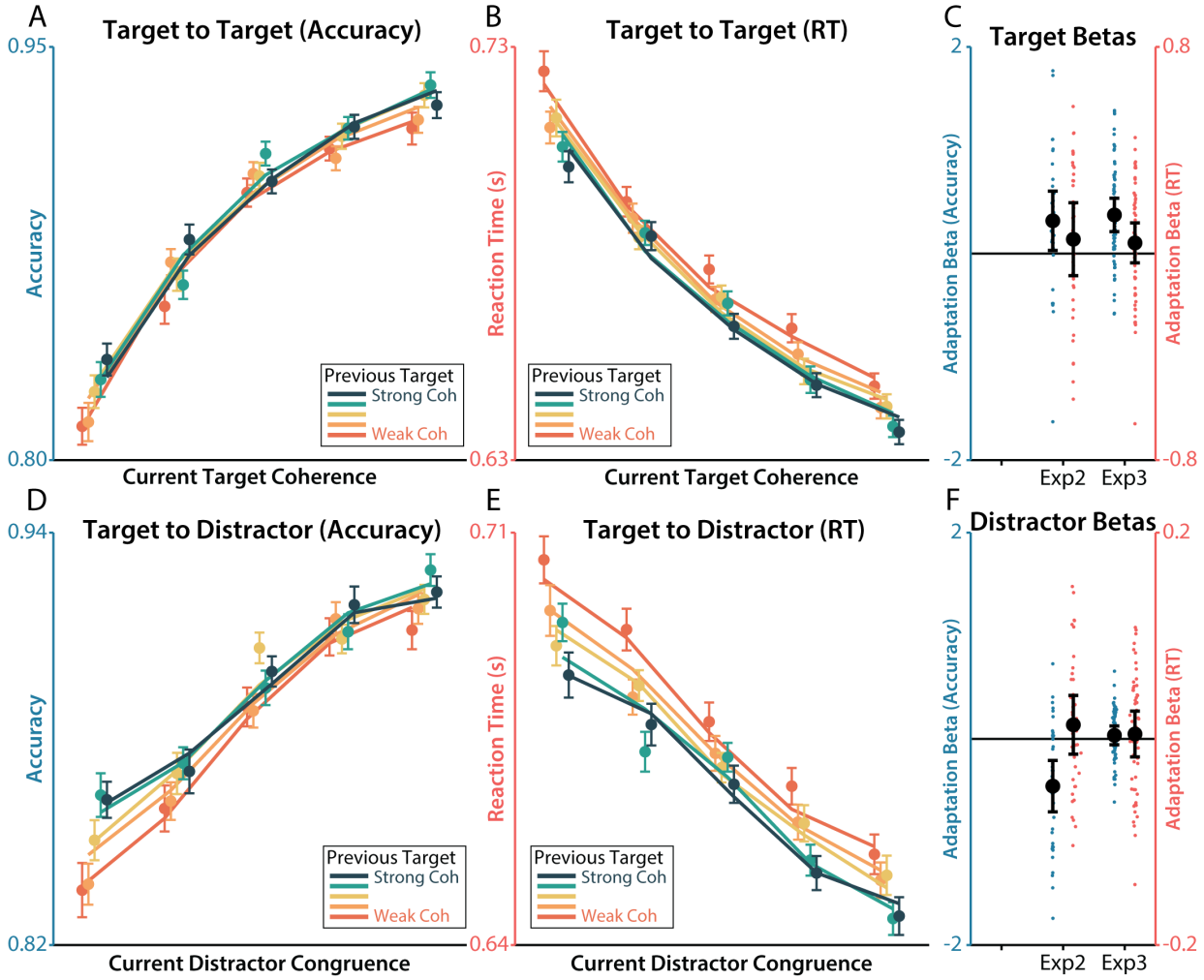
	Distractor × Accuracy	-0.34433 (df=66.41)	-0.64815 (df=29.565)	-0.72317 (df=78.624)	**4.559e-11**

Effect sizes are calculated from Welsh's contrasts across regression models. P-values are aggregated across experiments, with statistically significant p-values (two-tailed, $\alpha = 0.05$) shown in bold.

Supplementary Figures



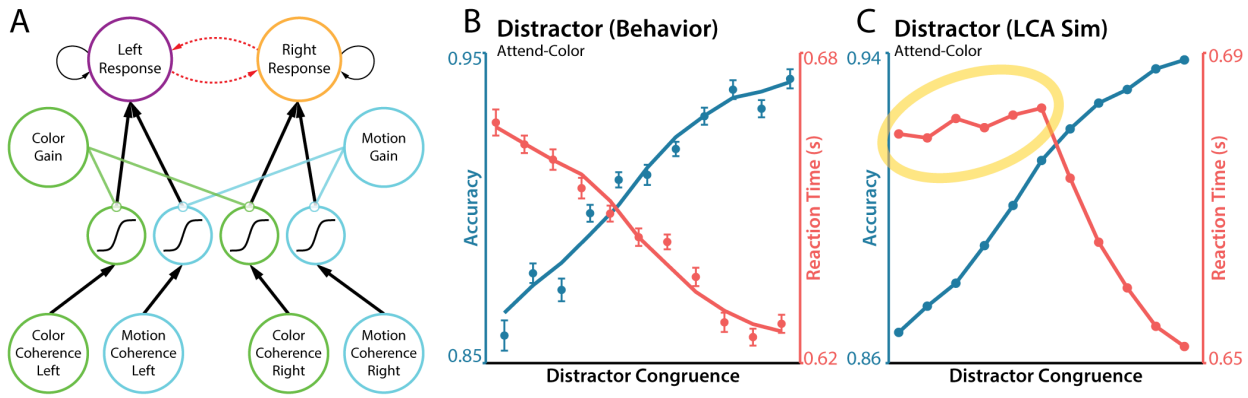
Supplementary Figure 1. Regression schematic. To estimate feature sensitivity, trial-specific color (green) and motion (cyan) coherence levels were passed through a hyperbolic tangent non-linearity (tanh), with the α parameter determining the strength of the non-linearity (see Methods). We estimated the linear mapping from this transformed coherence onto to RT (red) and Choice (blue) as participants feature sensitivity. Our critical analyses tested whether potential indices of control (e.g., task instructions or incentives) interacted with these estimates of feature sensitivity.



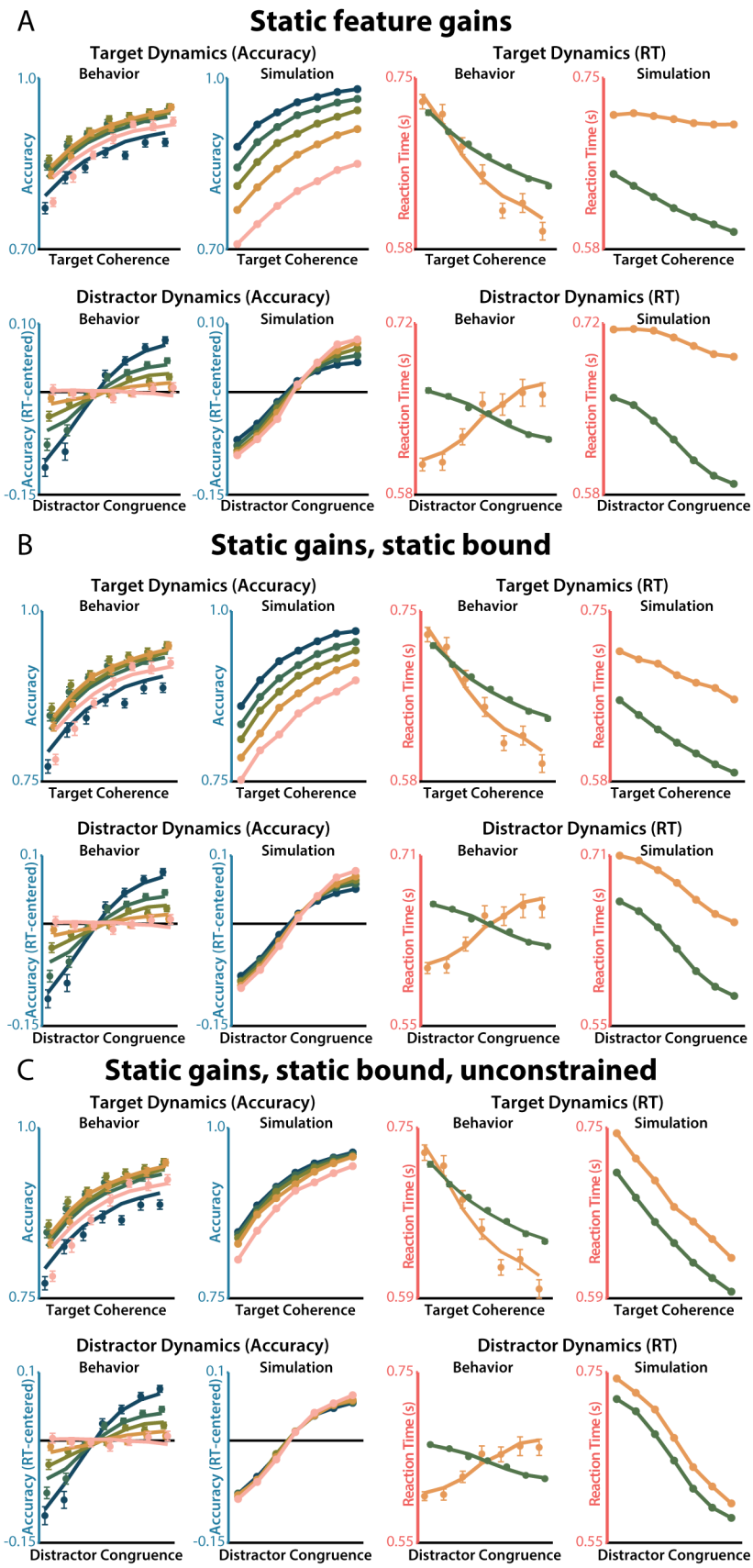
Supplementary Figure 2. Target-dependent adaptation.
A-B) The relationship between target coherence and accuracy (A) was weaker when the previous trial had weaker target coherence (redder colors). There was not a significant effect for RT (B) Lines depict aggregated regression

predictions. C) Regression estimates for the current target coherence by previous target coherence interaction, within each experiment. D-E) There was not a significant relationship between distractor congruence and previous target coherence in accuracy (D) or RT (E). F) Regression estimates for the current distractor congruence by previous target coherence interaction, within each experiment. Error bars on behavior reflect within-participant SEM, error bars on regression coefficients reflect 95% CI.

Leaky Competing Accumulator (LCA)



Supplementary Figure 3. Leak competing accumulator simulation. A) We simulated behavior from a leaky competing accumulator (Usher & McClelland, 2001). In this model, the response accumulators directly compete. In our parameter regime, leak and competition parameters produce race-like accumulation dynamics (Bogacz et al., 2006; Weichart et al., 2020). We found that this parameter regime was unable to capture the effect of distractor congruence on reaction time, as stronger inputs (congruent or incongruent) produce faster RTs in a race-like regime (Teodorescu & Usher, 2013). Other parameter regime, producing DDM-like dynamics, would replicate our main simulation results (Bogacz et al., 2006).


A
Static feature gains
Target Dynamics (Accuracy)
Behavior
Simulation
Target Dynamics (RT)
Behavior
Simulation
Accuracy
Reaction Time (s)
Target Coherence
Distractor Dynamics (Accuracy)
Behavior
Simulation
Distractor Dynamics (RT)
Behavior
Simulation
Accuracy (RT-centered)
Distractor Congruence
B
Static gains, static bound
Target Dynamics (Accuracy)
Target Dynamics (RT)
Distractor Dynamics (Accuracy)
Distractor Dynamics (RT)
C
Static gains, static bound, unconstrained
Target Dynamics (Accuracy)
Target Dynamics (RT)
Distractor Dynamics (Accuracy)
Distractor Dynamics (RT)

Supplementary Figure 4. Static feature gain simulations. We simulated the FFI model under different formulations that lack feature sensitivity dynamics, showing that gain dynamics are necessary to capture the RT- and Accuracy-dependent feature sensitivity we observed in participants' behavior. Feature-specific processes are necessary to capture the opposite-going dynamics on target sensitivity and distractor sensitivity. A) Static model without feature dynamics. B) Static model without feature dynamics or collapse response threshold. C) Static model without feature dynamics, collapsing response threshold, or positive-rectified accumulators.

References

Adkins, T. J., & Lee, T. (2021). *Reward reduces habitual errors by enhancing the preparation of goal-directed actions*. doi:10.31234/osf.io/hv9mz

Bogacz, R., Brown, E., Moehlis, J., Holmes, P., & Cohen, J. D. (2006). The physics of optimal decision making: a formal analysis of models of performance in two-alternative forced-choice tasks. *Psychological Review*, *113*(4), 700–765.

Bogacz, R., Usher, M., Zhang, J., & McClelland, J. L. (2007). Extending a biologically inspired model of choice: multi-alternatives, nonlinearity and value-based multidimensional choice. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *362*(1485), 1655–1670.

Botvinick, M. M., Braver, T. S., Barch, D. M., Carter, C. S., & Cohen, J. D. (2001). Conflict monitoring and cognitive control. *Psychological Review*, *108*(3), 624.

Botvinick, M. M., & Cohen, J. D. (2014). The computational and neural basis of cognitive control: charted territory and new frontiers. *Cognitive Science*, *38*(6), 1249–1285.

Braem, S., Bugg, J. M., Schmidt, J. R., Crump, M. J. C., Weissman, D. H., Notebaert, W., & Egner, T. (2019). Measuring Adaptive Control in Conflict Tasks. *Trends in Cognitive Sciences*, *0*(0). doi:10.1016/j.tics.2019.07.002

Braver, T. S. (2012). The variable nature of cognitive control: a dual mechanisms framework. *Trends in Cognitive Sciences*, *16*(2), 106–113.

Britten, K. H., Shadlen, M. N., Newsome, W. T., & Movshon, J. A. (1992). The analysis of visual motion: a comparison of neuronal and psychophysical performance. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *12*(12), 4745–4765.

Bustamante, L., Lieder, F., Musslick, S., Shenhav, A., & Cohen, J. (2021). Learning to Overexert Cognitive Control in a Stroop Task. *Cognitive, Affective & Behavioral Neuroscience*. doi:10.3758/s13415-020-00845-x

Callaway, F., Rangel, A., & Griffiths, T. L. (2021). Fixation patterns in simple choice reflect optimal information sampling. *PLoS Computational Biology*, *17*(3), e1008863.

Chiew, K. S., & Braver, T. S. (2017). Context processing and cognitive control: From gating models to dual mechanisms. In T. Egner (Ed.), *The Wiley handbook of cognitive control , (pp* (Vol. 642, pp. 143–166). psycnet.apa.org.

Churchland, M. M., Cunningham, J. P., Kaufman, M. T., Ryu, S. I., & Shenoy, K. V. (2010). Cortical preparatory activity: representation of movement or first cog in a dynamical machine? *Neuron*, *68*(3), 387–400.

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

Cohen, J. D., Servan-Schreiber, D., & McClelland, J. L. (1992). A parallel distributed processing approach to automaticity. *The American Journal of Psychology*, *105*(2), 239–269.

Danielmeier, C., Eichele, T., Forstmann, B. U., Tittgemeyer, M., & Ullsperger, M. (2011).

Posterior medial frontal cortex activity predicts post-error adaptations in task-related visual and motor areas. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *31*(5), 1780–1789.

De Jong, R., Liang, C. C., & Lauber, E. (1994). Conditional and unconditional automaticity: a dual-process model of effects of spatial stimulus-response correspondence. *Journal of Experimental Psychology. Human Perception and Performance*, *20*(4), 731–750.

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

Erb, C. D., Moher, J., Sobel, D. M., & Song, J.-H. (2016). Reach tracking reveals dissociable processes underlying cognitive control. *Cognition*, *152*, 114–126.

Evans, N. J., & Servant, M. (2020). *A model-based approach to disentangling facilitation and interference effects in conflict tasks*. doi:10.31234/osf.io/tu8ym

Fengler, A., Govindarajan, L. N., Chen, T., & Frank, M. J. (2021). Likelihood approximation networks (LANs) for fast inference of simulation models in cognitive neuroscience. *ELife*, *10*. doi:10.7554/eLife.65074

Fischer, A. G., Nigbur, R., Klein, T. A., Danielmeier, C., & Ullsperger, M. (2018). Cortical beta power reflects decision dynamics and uncovers multiple facets of post-error adaptation. *Nature Communications*, *9*(1), 5038.

Gold, J. I., & Shadlen, M. N. (2007). The neural basis of decision making. *Annual Review of Neuroscience*, *30*, 535–574.

Gonthier, C., Braver, T. S., & Bugg, J. M. (2016). Dissociating proactive and reactive control in the Stroop task. *Memory & Cognition*, *44*(5), 778–788.

Gratton, G., Coles, M. G. H., & Donchin, E. (1992). Optimizing the use of information: strategic control of activation of responses. *Journal of Experimental Psychology. General*, *121*(4), 480.

Hardwick, R. M., Forrence, A. D., Krakauer, J. W., & Haith, A. M. (2019). Time-dependent competition between goal-directed and habitual response preparation. *Nature Human Behaviour*, *3*(12), 1252–1262.

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.

Hübner, R., & Schlösser, J. (2010). Monetary reward increases attentional effort in the flanker task. *Psychonomic Bulletin & Review*, *17*(6), 821–826.

Jiang, J., Heller, K., & Egner, T. (2014). Bayesian modeling of flexible cognitive control. *Neuroscience and Biobehavioral Reviews*, *46 Pt 1*, 30–43.

Kang, Y. H., Löffler, A., Jeurissen, D., Zylberberg, A., Wolpert, D. M., & Shadlen, M. N. (2021). Multiple decisions about one object involve parallel sensory acquisition but time-multiplexed evidence incorporation. *ELife*, *10*. doi:10.7554/eLife.63721

Krajbich, I., Armel, C., & Rangel, A. (2010). Visual fixations and the computation and comparison of value in simple choice. *Nature Neuroscience*, *13*(10), 1292–1298.

Leng, X., Yee, D., Ritz, H., & Shenhav, A. (2020). Dissociable influences of reward and

punishment on adaptive cognitive control (p. 2020.09.11.294157). doi:10.1101/2020.09.11.294157

Li, Z.-W., & Ma, W. J. (2021). An uncertainty-based model of the effects of fixation on choice. *PLoS Computational Biology*, *17*(8), e1009190.

Lieder, F., Shenhav, A., Musslick, S., & Griffiths, T. L. (2018). Rational metareasoning and the plasticity of cognitive control. *PLoS Computational Biology*, *14*(4), e1006043.

Lindsay, D. S., & Jacoby, L. L. (1994). Stroop process dissociations: the relationship between facilitation and interference. *Journal of Experimental Psychology. Human Perception and Performance*, *20*(2), 219–234.

Lipták, T. (1958). On the combination of independent tests. *Magyar Tud Akad Mat Kutato Int Kozl*, *3*, 171–197.

Mante, V., Sussillo, D., Shenoy, K. V., & Newsome, W. T. (2013). Context-dependent computation by recurrent dynamics in prefrontal cortex. *Nature*, *503*(7474), 78–84.

Mayr, U., Awh, E., & Laurey, P. (2003). Conflict adaptation effects in the absence of executive control. *Nature Neuroscience*, *6*(5), 450–452.

Menceloglu, M., Suzuki, S., & Song, J.-H. (2021). Revealing the effects of temporal orienting of attention on response conflict using continuous movements. *Attention, Perception & Psychophysics*. doi:10.3758/s13414-020-02235-4

Monsell, S., & Mizon, G. A. (2006). Can the task-cuing paradigm measure an endogenous task-set reconfiguration process? *Journal of Experimental Psychology. Human Perception and Performance*, *32*(3), 493–516.

Musslick, S., Bizyaeva, A., Agaron, S., Leonard, N., & Cohen, J. D. (2019). Stability-flexibility dilemma in cognitive control: a dynamical system perspective. *Proceedings of the 41st Annual Meeting of the Cognitive Science Society*.

Musslick, S., Shenhav, A., Botvinick, M., & Cohen, J. (2015, June 7). A Computational Model of Control Allocation based on the Expected Value of Control. *2nd Multidisciplinary Conference on Reinforcement Learning and Decision Making*. Presented at the Multidisciplinary Conference on Reinforcement Learning and Decision Making.

Padmala, S., & Pessoa, L. (2011). Reward reduces conflict by enhancing attentional control and biasing visual cortical processing. *Journal of Cognitive Neuroscience*, *23*(11), 3419–3432.

Parro, C., Dixon, M. L., & Christoff, K. (2018). The neural basis of motivational influences on cognitive control. *Human Brain Mapping*, *39*(12), 5097–5111.

Posner, M., & Snyder, C. (1975). *Attention and cognitive control*.

Ratcliff, R., & McKoon, G. (2008). The diffusion decision model: theory and data for two-choice decision tasks. *Neural Computation*, *20*(4), 873–922.

Remington, E. D., Egger, S. W., Narain, D., Wang, J., & Jazayeri, M. (2018). A Dynamical Systems Perspective on Flexible Motor Timing. *Trends in Cognitive Sciences*, *22*(10), 938–952.

Remington, E. D., Narain, D., Hosseini, E. A., & Jazayeri, M. (2018). Flexible Sensorimotor

Computations through Rapid Reconfiguration of Cortical Dynamics. *Neuron*, *98*(5), 1005-1019.e5.

Ridderinkhof, K. R. (2002). Micro- and macro-adjustments of task set: activation and suppression in conflict tasks. *Psychological Research*, *66*(4), 312–323.

Ritz, H., Frömer, R., & Shenhav, A. (2020). Bridging Motor and Cognitive Control: It's About Time! *Trends in Cognitive Sciences*. Retrieved from https://www.sciencedirect.com/science/article/pii/S1364661319302773?casa_token=lQOhjm4WW9AAAAAA:2xsNe0BxkGw_Tf8J-bpHROiRk44GQKTfBAJSIUJKZd4qf4jHVCoX8xerQS1JluNRJMFNzK9m4Ko

Rogers, R. D., & Monsell, S. (1995). Costs of a predictible switch between simple cognitive tasks. *Journal of Experimental Psychology. General*, *124*(2), 207.

Scherbaum, S., Dshemuchadse, M., Fischer, R., & Goschke, T. (2010). How decisions evolve: the temporal dynamics of action selection. *Cognition*, *115*(3), 407–416.

Scherbaum, S., Fischer, R., Dshemuchadse, M., & Goschke, T. (2011). The dynamics of cognitive control: evidence for within-trial conflict adaptation from frequency-tagged EEG. *Psychophysiology*, *48*(5), 591–600.

Schmidt, J. R., & De Houwer, J. (2011). Now you see it, now you don't: controlling for contingencies and stimulus repetitions eliminates the Gratton effect. *Acta Psychologica*, *138*(1), 176–186.

Servant, M., Montagnini, A., & Burle, B. (2014). Conflict tasks and the diffusion framework: Insight in model constraints based on psychological laws. *Cognitive Psychology*, *72*, 162–195.

Shadlen, M. N., & Newsome, W. T. (2001). Neural basis of a perceptual decision in the parietal cortex (area LIP) of the rhesus monkey. *Journal of Neurophysiology*, *86*(4), 1916–1936.

Shenhav, A., Botvinick, M. M., & Cohen, J. D. (2013). The expected value of control: an integrative theory of anterior cingulate cortex function. *Neuron*, *79*(2), 217–240.

Shenhav, A., Straccia, M. A., Musslick, S., Cohen, J. D., & Botvinick, M. M. (2018). Dissociable neural mechanisms track evidence accumulation for selection of attention versus action. *Nature Communications*, *9*(1), 2485.

Shiffrin, R. M., & Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending and a general theory. *Psychological Review*, *84*(2), 127.

Simen, P., Contreras, D., Buck, C., Hu, P., Holmes, P., & Cohen, J. D. (2009). Reward rate optimization in two-alternative decision making: empirical tests of theoretical predictions. *Journal of Experimental Psychology. Human Perception and Performance*, *35*(6), 1865–1897.

Steyvers, M., Hawkins, G. E., Karayanidis, F., & Brown, S. D. (2019). A large-scale analysis of task switching practice effects across the lifespan. *Proceedings of the National Academy of Sciences of the United States of America*. doi:10.1073/pnas.1906788116

Teodorescu, A. R., & Usher, M. (2013). Disentangling decision models: from independence to

competition. *Psychological Review*, *120*(1), 1–38.

Teufel, H. J., & Wehrhahn, C. (2000). Evidence for the contribution of S cones to the detection of flicker brightness and red-green. *Journal of the Optical Society of America. A, Optics, Image Science, and Vision*, *17*(6), 994–1006.

Ueltzhöffer, K., Armbruster-Genç, D. J. N., & Fiebach, C. J. (2015). Stochastic Dynamics Underlying Cognitive Stability and Flexibility. *PLoS Computational Biology*, *11*(6), e1004331.

Ulrich, R., Schröter, H., Leuthold, H., & Birngruber, T. (2015). Automatic and controlled stimulus processing in conflict tasks: Superimposed diffusion processes and delta functions. *Cognitive Psychology*, *78*, 148–174.

Usher, M., & McClelland, J. L. (2001). The time course of perceptual choice: the leaky, competing accumulator model. *Psychological Review*, *108*(3), 550–592.

Weichart, E. R., Turner, B. M., & Sederberg, P. B. (2020). A model of dynamic, within-trial conflict resolution for decision making. *Psychological Review*. doi:10.1037/rev0000191

Westbrook, A., van den Bosch, R., Määttä, J. I., Hofmans, L., Papadopetraki, D., Cools, R., & Frank, M. J. (2020). Dopamine promotes cognitive effort by biasing the benefits versus costs of cognitive work. *Science*, *367*(6484), 1362–1366.

White, C. N., Ratcliff, R., & Starns, J. J. (2011). Diffusion models of the flanker task: discrete versus gradual attentional selection. *Cognitive Psychology*, *63*(4), 210–238.

White, C. N., Servant, M., & Logan, G. D. (2018). Testing the validity of conflict drift-diffusion models for use in estimating cognitive processes: A parameter-recovery study. *Psychonomic Bulletin & Review*, *25*(1), 286–301.

Wong, K.-F., & Wang, X.-J. (2006). A recurrent network mechanism of time integration in perceptual decisions. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *26*(4), 1314–1328.

Yeung, N., Botvinick, M. M., & Cohen, J. D. (2004). The Neural Basis of Error Detection: Conflict Monitoring and the Error-Related Negativity. *Psychological Review*, *111*(4), 931–959.

Yu, A. J., Dayan, P., & Cohen, J. D. (2009). Dynamics of attentional selection under conflict: toward a rational Bayesian account. *Journal of Experimental Psychology. Human Perception and Performance*, *35*(3), 700–717.

Zaykin, D. V. (2011). Optimally weighted Z-test is a powerful method for combining probabilities in meta-analysis. *Journal of Evolutionary Biology*, *24*(8), 1836–1841.