

Psychological Review

Humans Reconfigure Target and Distractor Processing to Address Distinct Task Demands

Harrison Ritz and Amitai Shenhav

Online First Publication, September 4, 2023. <https://dx.doi.org/10.1037/rev0000442>

CITATION

Ritz, H., & Shenhav, A. (2023, September 4). Humans Reconfigure Target and Distractor Processing to Address Distinct Task Demands. *Psychological Review*. Advance online publication. <https://dx.doi.org/10.1037/rev0000442>

Humans Reconfigure Target and Distractor Processing to Address Distinct Task Demands

Harrison Ritz^{1, 2, 3} and Amitai Shenhav^{1, 2}

¹ Department of Cognitive, Linguistic and Psychological Science, Brown University

² Carney Institute for Brain Science, Brown University

³ Princeton Neuroscience Institute, Princeton University

When faced with distraction, we can focus more on goal-relevant information (targets) or focus less on goal-conflicting information (distractors). How people use cognitive control to distribute attention across targets and distractors remains unclear. We address this question by developing a novel Parametric Attentional Control Task that can “tag” participants’ sensitivity to target and distractor information. We use these precise measures of attention to develop a novel process model that can explain how participants control attention toward targets and distractors. Across three experiments, we find that participants met the demands of this task by independently controlling their processing of target and distractor information, exhibiting distinct adaptations to manipulations of incentives and conflict. Whereas incentives preferentially led to target enhancement, conflict in the previous trial preferentially led to distractor suppression. These distinct drivers of control altered sensitivity to targets and distractors early in the trial, promptly followed by reactive reconfiguration toward task-appropriate feature sensitivity. To provide a process-level account of these empirical findings, we develop a novel neural network model of evidence accumulation with attractor dynamics over feature weights that reconfigure target and distractor processing. These results provide a computational account of control reconfiguration that provides new insights into how multivariate attentional signals are optimized to achieve task goals.

Keywords: cognitive control, decision making, attention, evidence accumulation

Supplemental materials: <https://doi.org/10.1037/rev0000442.supp>


Whether we are having a conversation in a crowded coffee shop or writing an article at our desk while surrounded by browser tabs, most tasks require us to engage in two distinct forms of attentional control.¹ One form of control enhances 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., a distraction from a nearby conversation). While past research has extensively studied target and distractor processing, 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). Can people control multiple forms of information processing, and if so, how do they regulate these information streams over time? Here, we bridge previous methodological gaps to gain new insight into the top-down control of target and distractor processing, providing an integrative model of how dynamic control adjustments could occur 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 have examined how this varies with the difficulty of discriminating that stimulus (e.g., how perceptually similar two stimuli are; Britten et al., 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).

¹ Throughout, we refer to “cognitive control” as the process that configures information processing to achieve task goals (Botvinick & Cohen, 2014). Whereas cognitive control refers to all such adjustments, such as changes to stimulus sensitivity or decision threshold, we reserve “attentional control” for just the top-down control over stimulus sensitivity.

Harrison Ritz  <https://orcid.org/0000-0003-3011-2946>

Amitai Shenhav  <https://orcid.org/0000-0002-0222-0774>

This work was supported by the Daniel Cooper Graduate Student Fellowship and the C.V. Starr Fellowship (Harrison Ritz), as well as Grants National Institutes of Health R01MH124849 and NSF CAREER Award 2046111 (Amitai Shenhav). The authors are grateful to Kia Sadahiro, William McNelis, Allison Loynd, and Savannah Doelfel for assistance in data collection and to Michael J. Frank, Matthew N. Nassar, Jonathan Cohen, David Badre, Tobias Egner, Senne Braem, Sebastian Musslick, and both the Shenhav Lab and LNCC for helpful discussions on these topics.

The authors have no conflicts of interest to disclose.

Correspondence concerning this article should be addressed to Harrison Ritz, Princeton Neuroscience Institute, Princeton University, Washington Road, Princeton, NJ 08544, United States. Email: hritz@princeton.edu

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 (Ritz et al., 2022; see also Frömer & Shenhav, 2022). 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 redirect control toward 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; Gratton et al., 1992). Prevailing models have accounted for these findings by assuming that participants increase attention to the target dimension following a high-conflict trial (Botvinick et al., 2001; Egner, 2007), but limitations of the relevant experiments (e.g., most experiments do not manipulate target salience; through see Lindsay & Jacoby, 1994; Servant et al., 2014; Stafford et al., 2011) make it difficult to rule out that adaptation may also occur at the level of distractor suppression (Lindsay & Jacoby, 1994; Tzelgov et al., 1992). 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. Cognitive control is fundamentally an adaptive process, so people's specific control policies should depend on the task structure (Botvinick & Cohen, 2014; Egner, 2008). However, understanding how people coordinate multiple forms of information processing can help inform the architecture of the underlying control process (Ritz et al., 2022).

One way that previous research has studied target and distractor adjustments is to measure changes in brain activity associated with task-relevant stimulus processing. For example, some previous work has suggested that conflict-triggered control preferentially enhances sensitivity in regions associated with target stimuli (e.g., faces in fusiform face gyrus; Egner & Hirsch, 2005). Other studies have found evidence for both target and distractor control by using similar stimulus-tagging methods (Gazzaley et al., 2005; Soutschek et al., 2015) or by exploiting lateralized electroencephalogram responses (Noonan et al., 2016; Wöstmann et al., 2019). The range of results across these neuroimaging experiments may come from the different tasks that have been deployed (Egner, 2008) but may also arise from noisy or complex correspondence between neuroimaging methods and underlying cognitive processes. In the current experiment, we provide new methods for indexing target and distractor sensitivity from behavior alone, enabling us to provide new insight into the cognitive architecture of feature-selective control.

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 et al., 2014; Weichart et al., 2020; White et al., 2011; Yu et al., 2009). These models have largely focused on the Eriksen flanker task, modeling 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 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 et al., 2011; Kang et al., 2021; Kayser et al., 2010; Mante et al., 2013; Shenhav et al., 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"² response interference from this prepotent distractor. A salient incongruent distractor provokes an erroneous response, providing a qualitatively different form of difficulty from how low-coherence targets make it harder to choose the correct answer (Norman & Bobrow, 1975).

Previous work has demonstrated response conflict and trial-to-trial adjustments in a color–motion kinematogram with full target coherence and binary distractor congruence (Danielmeier et al., 2011). We extended this task by parametrizing both target coherence and distractor congruence. In doing so, we are able to obtain more precise measures of feature sensitivity by accounting for global performance factors (e.g., lapse rate; Wichmann & Hill, 2001). Importantly, however, we can also isolate how participants simultaneously configure attention toward each of these feature dimensions. Using standard elicitors of cognitive control, namely performance-contingent incentives and response conflict, we examine how people dynamically configure both target and distractor gain to maximize their performance. We then use the precision afforded by these methodological advances to inform an explicit process model of attentional control.

We find that participants independently and dynamically control target and distractor processing over the course of a trial. To meet the demands of this task, participants preferentially enhanced target sensitivity under incentives and preferentially suppressed distractor sensitivity after high-conflict trials. Moreover, they implement these control strategies by changing the initial conditions of a dynamic process that enhances task-relevant feature processing and suppresses task-irrelevant feature processing. Finally, we find that these control strategies can be captured by extending classic neural network models of cognitive control to incorporate an attractor network that dynamically regulates the influence of different task features on choice. Together, these results extend our understanding of both decision making and cognitive control by bridging the methodological and theoretical divides between these fields,

² The Simon task is a classic cognitive control task in which participants must ignore a response-affording stimulus feature (e.g., respond "left" to a leftward spatial location) and instead respond to a less prepotent stimulus feature (e.g., respond "right" to a blue stimulus). The classic pattern of results is that participants perform more poorly when these two features disagree than when they correspond to the same response (see Egner, 2007).

providing new insight into how we control multiple forms of information processing.

Method

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, $M(SD)$ age: 20.6(2.21); 36 female; one excluded, 42 individuals participated in Experiment 2, age: 19.1(0.971); 31 female; two excluded, and 62 individuals participated in Experiment 3, age: 19.8(1.38); 47 female; two excluded, resulting in 156 included participants across the three experiments. Sample sizes were guided by piloting in Experiment 1 and experimental standards in cognitive control research (commonly $n = 20-40$; e.g., Adkins & Lee, 2021; Danielmeier et al., 2011; Jiang et al., 2015; Vogel et al., 2020; White et al., 2011).

Parametric Attentional Control Task

We developed the Parametric Attentional Control Task (PACT), extending tasks used to study decision making (Kang et al., 2021; 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 of color did not repeat on adjacent trials to avoid 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 vs. moving in a random direction). Distractor congruence was linearly spaced between 95% congruence and 95% incongruence, sampled randomly across trials. For variants with 11 levels of congruence, the congruence levels were $-95, -76, -57, -38, -19, 0, 19, 38, 57, 76, 95$, with negative values being incongruent and positive values being congruent. We made the motion highly salient to maximize the conflict induced by this distracting dimension (Wöstmann et al., 2021), with dots moving quickly across a large aperture.

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% and 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. In attend-motion trials, we allowed distractor colors to repeat on consecutive trials, mirroring the stimulus repetitions that occurred in attend-color blocks.

Comparing performance across tasks that are matched for visual and motoric demands also allows us to test whether behavioral effects depend on stimulus or response confounds. For example, participants' behavior may be influenced by eye movement confounds (e.g., bottom-up attentional capture by motion coherence), response repetition biases (e.g., due to responses switching more often than repeating), or stimulus-response priming (e.g., due to how response switching coincides with stimulus transitions). Critically, attend-color and attend-motion tasks differ in their putative control demands, allowing us to isolate stimulus-response confounds from goal-directed control.

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 accurate 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; Experiments 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 reaction time (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 60 cm 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 at 60 Hz). The dots colors were approximately (uncalibrated) isoluminant and perceptually equidistant (RGB values: [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 intertrial interval (Experiment 1: 0.5–1.5 s; Experiments 2–3: 0.5–1.0 s). There was an alerting cue 300 ms before the trial onset, indicated by the fixation cross turning from gray to white, to minimize nondecision time. The stimuli were then presented until either a response was made or a deadline was reached (Experiment 1: 3 s; Experiments 2–3: 5 s).

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, $M(\text{range}) = 13.5(11-15)$, the number of trials per attend-motion block, $M(\text{range}) = 26(20-50)$, and the total number of trials, $M(\text{range}) = 469(300-700)$ 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 article and not included in the analyses we report.

Experiments 2 and 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. Illustrative task instructions are provided in Supplemental Material Note 1.

Incentivized Variant (Experiment 3)

In Experiment 3, we studied task performance under monetary incentives to provide a convergent measure of control adjustments, to test where task processing was limited by motivation rather than hard constraints like stimulus information (Norman & Bobrow, 1975). We informed participants before the main session that they would be able to earn a monetary reward for good performance. On “reward” blocks, we randomly selected trials at the end of the experiment, and participants earned bonus payment for trials on which they were both fast (<75% of their RT distribution) and accurate. On “no reward” blocks, participants would not be eligible to earn a reward but were encouraged to be fast and accurate. 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 were not instructed on the reward algorithm, only that they would earn rewards from being fast and accurate on randomly selected trials. Participants were not informed which trials were selected to avoid biasing postreward trials. 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, $M(SD)$ bonus: \$2.5(\$0.57)USD.

Regression Analyses

We used a hierarchical nonlinear regression of choice and reaction time as a tractable and minimally theory-laden measure of performance (Supplemental Figure 1). We designed these regression models to quantify changes to target and distractor sensitivity while controlling for global factors like behavioral autocorrection and how task factors may change lapse rates. The results of these regression analyses then provided the basis for our explicit process modeling (see below). We confirmed that our regression models are identifiable using Belsley’s collinearity diagnostics (collintest in MATLAB; Supplemental Table 10).

In particular, we implemented hierarchical expectation maximization in MATLAB R2020a (using emfit; available at [https://github](https://github.com/mpc-ucl/emfit)

.com/mpc-ucl/emfit) to provide a maximum a posteriori (MAP) estimates for the mean and covariance of parameters linking task features to participants’ reaction time and accuracy. This fitting algorithm alternates between finding the MAP estimates of participants’ parameters given the current group-level expectations (M-step; with five parameter reinitialization 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. The analysis code is available at <https://github.com/shenhav/vlab/PACT-public>.

Our regression approach simultaneously estimated parameters for choice and RT:

$$\begin{aligned} \log\text{Post} &= \log\text{Like}(\text{Choice}) + \log\text{Like}(\text{RT}) \\ &+ \log\text{Prior}(\text{Choice}, \text{RT}). \end{aligned} \quad (1)$$

Our choice subfunction used a lapse-logistic likelihood function, as previous work has shown that un-modeled lapse rates can mimic changes in psychometric slope (Wichmann & Hill, 2001). Our choice subfunction had the form:

$$\text{Choice} \sim \frac{1 - \text{lapse}}{1 + \exp(-\beta_{\text{Choice}}X_{\text{Choice}})} + (\text{lapse} \times 0.5), \quad (2)$$

$$\text{lapse} = \frac{1}{1 + \exp(-\beta_{\text{Lapse}}X_{\text{Lapse}})}, \quad (3)$$

Where β_{Choice} and β_{Lapse} are parameter vectors, and X_{Choice} and X_{Lapse} are design matrices. Our RT subfunction used a shifted log-normal likelihood function:

$$\log(\text{RT} - \text{ndt}) \sim \beta_{\text{RT}}X_{\text{RT}}, \quad (4)$$

Where again $\beta_{\text{RT}}X_{\text{RT}}$ is a linear model, and ndt is the estimated nonddecision time. Rare RTs less than ndt were assigned a small likelihood. This helped avoid one fast RT from unduly influencing this parameter, while still capturing these informative trials.

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 included autoregressive components to capture well-established behavioral features like choice repetition and RT autocorrelation (Egner, 2007; Laming, 1979; Lau & Glimcher, 2005; Urai et al., 2019). We transformed feature coherences using a saturating nonlinearity,

$$\text{coh}_{\text{feature}}^* = \frac{\tanh(\alpha_{\text{feature}} \times \text{coh}_{\text{feature}})}{\tanh(\alpha_{\text{feature}})}, \quad (5)$$

with α_{target} and $\alpha_{\text{distractor}}$ fit as free parameters. This nonlinearity was inspired by classical work on psychophysical scaling laws (i.e., Fechner–Weber–Stevens scaling, Krueger, 1989; Nieder &

Miller, 2003) and more recent work demonstrating this scaling during cognitive control experiments (Servant et al., 2014; Stafford et al., 2011). 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 nonlinearities 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 tanh-transformed 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 Supplemental Material.

We excluded trials in our regression if they were (a) the first trial of the block, (b) shorter than 200 ms or longer than 2 s, (c) occurred after an error or after a trial was too fast/slow, and (d) in reaction time analyses, if the current trial was an error. 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., posterror 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 modeling. Contrast tests across models used Welch'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 matrix to a correlation matrix.

We generated posterior predictive checks (trend lines on figures) by generating regression model predictions for all trials and then aggregating these predictions in the same way as participants' raw 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. To provide finer-grained insights into our model fit, we generated additional posterior predictive checks that aggregate trends across all participants (Supplemental Figure 6) and that highlight single participants (Supplemental Figure 7). To provide further validation of the robustness of our parameter estimation procedure, we performed parameter recovery (simulated behavior from our best-fitting regression parameters, refit our model to this simulated behavior, and the compared data-generating and recovered parameters; Supplemental Figure 8) and parameter knockout analyses (refit models with key nuisance regressors removed; Supplemental Figure 9). These robustness checks provided convergent evidence that our key parameters had good identifiability.

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 time series is:

$$\beta_{\text{motion}}^{\text{RT}} = (\beta_{\text{motion}} + \beta_{\text{motion:RT}} \text{SimRT}) \odot (1 - \text{lapse}^{\text{RT}}), \quad (6)$$

$$\text{lapse}^{\text{RT}} = \frac{1}{1 + \exp(-(\beta_{\text{Lapse}} + \beta_{\text{RT}} \text{SimRT}))}, \quad (7)$$

Where β s 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 two-way and three-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

To provide a bridge between our regression analyses and process models of decision making, we adopted a generative modeling approach and tested whether participant behavior could be reproduced by a sequential sampling model (Figure 8). This model was inspired by two theoretical traditions. The first was a classic connectionist model of cognitive control (Cohen et al., 1990), which demonstrated how—top-down adjustments to target and distractor sensitivity in evidence accumulation framework can capture a wide range of behavioral phenomena. To capture apparent within-trial adjustments to feature processing (see the Results section), our second inspiration was from dynamical models of task set reconfiguration, both across trial (Gilbert & Shallice, 2002; Musslick et al., 2018; Steyvers et al., 2019) and within trial (Mante et al., 2013; Pagan et al., 2022). In these dynamic models, adjustments in feature gain behave as a dynamical system, starting at some initial condition and exponentially approaching a fixed point.

This model takes as inputs the color and motion coherence in support of different responses (e.g., color coherence supporting a left response: $\text{coh}_{\text{colLeft}}$), nonlinearly transforms these inputs (e.g., $\text{coh}_{\text{colLeft}}^*$; see the Regression Analyses section in Methods), and then integrates evidence for each response in separate rectified accumulators (x_{left} and x_{right}).

For example, the change in evidence for the left response is modeled as:

$$\begin{aligned} dx_{\text{left}} = & -\lambda x_{\text{left}} dt + (\beta_{\text{col}} \text{coh}_{\text{colLeft}}^* dt + \beta_{\text{mot}} \text{coh}_{\text{motLeft}}^* dt + \epsilon_{\text{left}} \sqrt{dt}) \\ & - \omega (\beta_{\text{col}} \text{coh}_{\text{colRight}}^* dt + \beta_{\text{mot}} \text{coh}_{\text{motRight}}^* dt + \epsilon_{\text{right}} \sqrt{dt}), \end{aligned} \quad (8)$$

where $\epsilon_{\text{left}} \sim N(0, \sigma_x)$, $\epsilon_{\text{right}} \sim N(0, \sigma_x)$, and $x_{\text{left}} = \max(0, x_{\text{left}})$. 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 model (DDM; Bogacz et al., 2006). Note that parameterizations of a leaky competing accumulator could also approximate the DDM (Bogacz et al., 2006, 2007) and so are plausible alternatives to our implementation. We preferred the feedforward inhibition (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 (Mante et al., 2013), a model we call the feedforward inhibition with control model (FFIc model). In this model, control acts like a stochastic dynamical system. The system starts at an initial level of feature gain (β^0 ; e.g., due to bottom-up salience or learning). This feature gain exponentially approaches an asymptotic gain level (its "fixed point"; e.g., a set point on zero distractors gain), according to a decay

rate K (e.g., control gain). For example, the change in motion gain is modeled as:

$$d\beta_{\text{mot}} = -\gamma\beta_{\text{mot}}dt + K_{\text{mot}}(\text{fixedpoint}_{\text{mot}} - \beta_{\text{mot}}) + \varepsilon_{\text{gain}}\sqrt{dt},$$

$$\varepsilon_{\text{gain}} \sim N(0, \sigma_{\text{gain}}), \quad (9)$$

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 ($11 \times 11 \times 10,000$) and then aggregated simulated behavior in the same way we aggregated participants' behavior. Simulation code and parameter sets are available at <https://github.com/shenhavlab/PACT-public>.

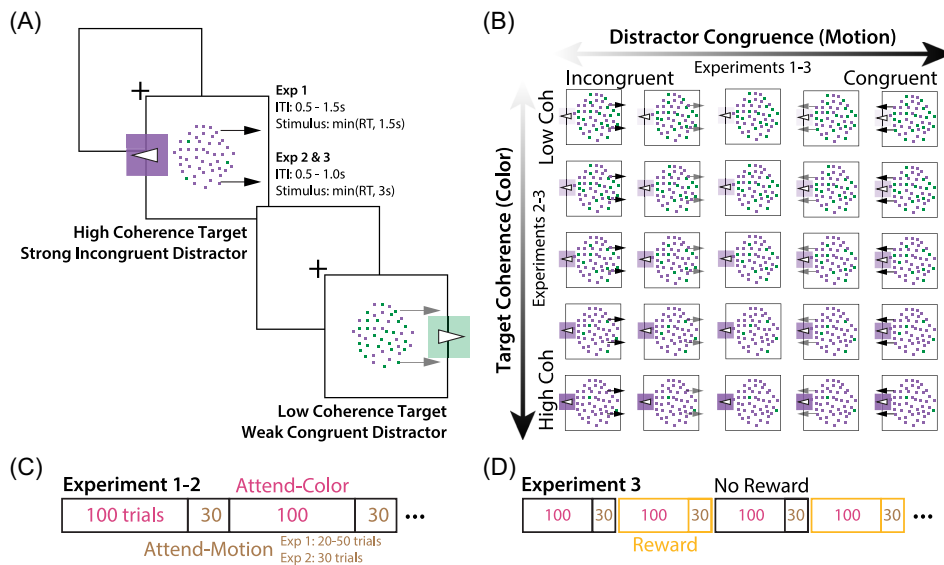
Transparency and Openness

We report how we determined our sample size, all data exclusions, all manipulations, and all measures in the study, and we follow APA Journal Article Reporting Standards (Appelbaum et al., 2018). All data and analysis code are available at <https://github.com/shenhavlab/PACT-public>. This study's design and its analysis were not preregistered.

Results

Participants performed 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. To avoid stimulus repetition priming (Braem et al., 2019; Mayr et al., 2003), two colors were assigned to each response and the majority of color did not repeat across sequential trials. 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)



Note. (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. Exp = experiment; ITI = intertrial interval; RT = reaction time. See the online article for the color version of this figure.

Task Performance Varies Parametrically With Target Discriminability and Distractor Interference

In Experiment 1 ($N = 56$), 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) and being fully incongruent (100% rightward coherence for a left color response; Figure 1D). For trials midway between these two extremes (“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 = 585 ms, mean accuracy = 89%) and were fastest and most accurate when distractors were fully congruent (median RT = 553 ms, mean accuracy = 97%) (cf. Danielmeier et al., 2011). Performance on neutral trials (0% motion coherence) fell between these two extremes (median RT = 576 ms, mean accuracy = 94%). Extending past work, hierarchical regression analyses (see the Method section) revealed that performance varied in a graded fashion across this continuum of interference. Both accuracy (Cohen’s d on regression estimate; $d = -1.47$) and reaction time ($d = 1.25$) worsened with parametrically increasing levels of interference ($ps < .001$, Figure 2C, Table 1).

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 (Experiment 2: $d = -1.90$, Experiment 3: $d = -1.99$) and more accurate (Experiment 2: $d = 3.27$, Experiment 3: $d = 3.73$) with parametrically increasing levels of color coherence (aggregate $ps < .001$; 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 2A, 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. 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 (Figure 2E) but were no longer sensitive to the now-irrelevant color congruence (Figure 2G; Supplemental Tables 1 and 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 et al., 1992).

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.

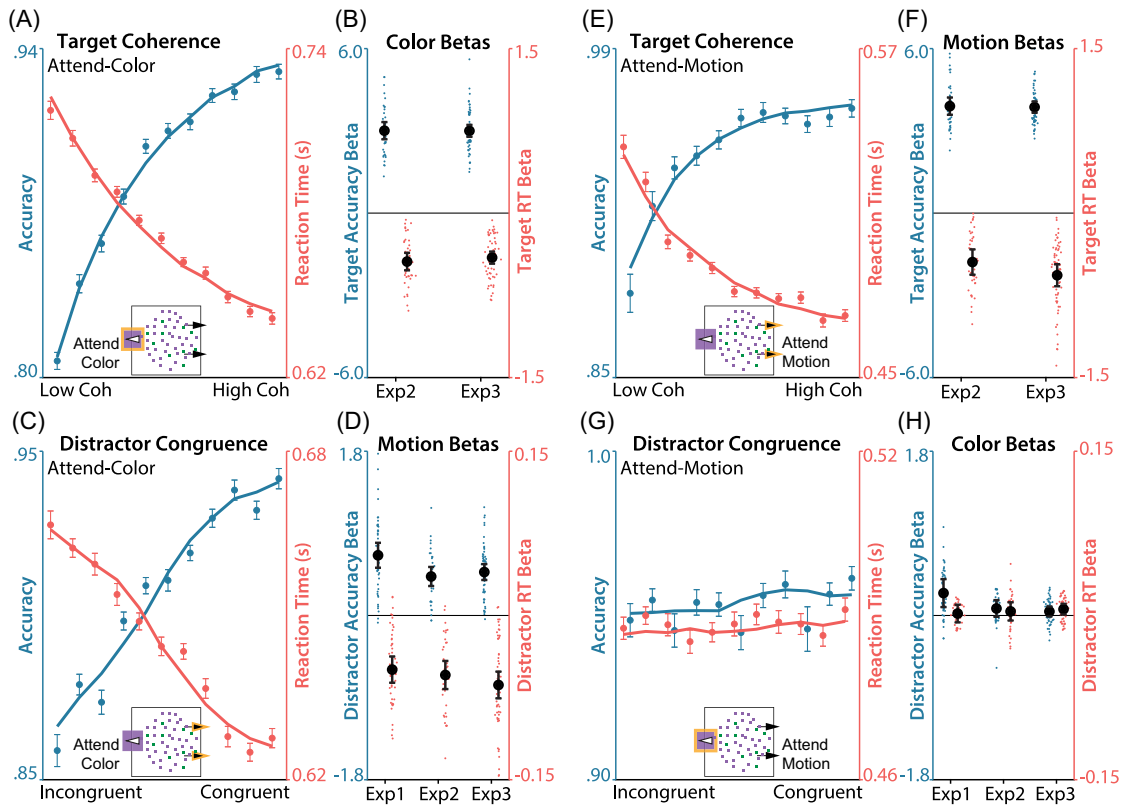
First, we tested whether a given feature exerted a similar influence on both accuracy and RT. We found that this was indeed the case, as there was a significant correlation between the effect distractors had on accuracy and RT ($rs < -0.87$, $ps < .001$). The influences of target discriminability on accuracy and RT were also significantly correlated ($rs < -0.54$, $ps < .001$; Supplemental 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 ($ds = 0.05$ – 0.23 , $p = .33$; Table 1), though we did find a small but significant interaction between target and distractor coherence on accuracy ($ds = -0.18$ to -0.34 , $p = .023$). For both studies, removing target–distractor interactions as predictors in our accuracy regressions improved model fit (protected exceedance probability on Akaike information criteria: Experiment 2 protected exceedance probability = 1; Experiment 3 protected exceedance probability = 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, Experiment 2: $r(25) = .14$, $p = .48$; Experiment 3: $r(45) = .44$, $p = .0019$, and accuracy, Experiment 2: $r(25) = -.15$, $p = .45$; Experiment 3: $r(45) = .12$, $p = .43$, suggesting that individual differences in target and distractor processing were not antagonistic.

Previous Conflict Preferentially Suppresses Distractor Sensitivity

Within a given trial, we found that performance varies parametrically and independently with the coherence of target (color)

Figure 2
Target and Distractor Sensitivity



Note. (A) Participants were more accurate (blue, left axis) and responded faster (red, right axis) when the target color had higher coherence. Circles depict participant behavior, and lines depict aggregated regression predictions. In all graphs, behavior and regression predictions are averaged over participants and experiments. Target sensitivity aggregated across Experiments 2 and 3. (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. Distractor sensitivity aggregated across Experiments 1–3. (D) Regression estimates for the effect of distractor congruence on performance within each experiment, plotted for accuracy and RT. (E and F) Similar to A and B, performance (E) and regression estimates (F) for the effect of target coherence during attend-motion blocks, in which motion was the target dimension. (G and H) Similar to A and B, performance (G) and regression estimates (H) for the effect of distractor congruence during attend-motion blocks, in which color was the distractor dimension. The y-axis range is matched within-feature across tasks, see Supplemental Figure 10 for matched y-axes across all features and tasks. Error bars on behavior reflect within-participant SEM, and error bars on regression coefficients reflect 95% CI. Psychometric functions are jittered on the x-axis for ease of visualization. CI = confidence interval; Exp = experiment; RT = reaction time; SEM = standard error of the mean. See the online article for the color version of this figure.

and distractor (motion) 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 also test whether adaptation occurs through an influence of previous conflict

on subsequent target enhancement, distractor suppression, or both. Finally, we can further test whether adaptation occurs due to the discriminability of the *target* in 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 in the previous trial, as reflected both in their choice ($d_s = 1.44$ – 1.74 , $p < .001$; Figure 3A) and RT ($d_s = 0.83$ – 1.79 , $p < .001$; Figure 3B; Table 2). When the previous trial had congruent distractors, participants had a strong sensitivity to the congruence of distractors on the current trial (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

Table 1
Target and Distractor Sensitivity

DV	Predictors	Exp 1 (<i>df</i> = 45) Effect size (<i>d</i>)	Exp 2 (<i>df</i> = 25) Effect size (<i>d</i>)	Exp 3 (<i>df</i> = 45) Effect size (<i>d</i>)	Aggregate <i>p</i> value
Choice	Target coherence		3.27	3.73	1.01×10^{-44}
	Distractor congruence	1.47	1.42	1.50	4.89×10^{-32}
	Target × Distractor		-0.184	-0.344	.0226
RT	Target coherence		-1.90	-1.99	1.59×10^{-28}
	Distractor congruence	-1.25	-1.49	-1.43	1.26×10^{-29}
	Target × Distractor		0.230	0.0525	.437

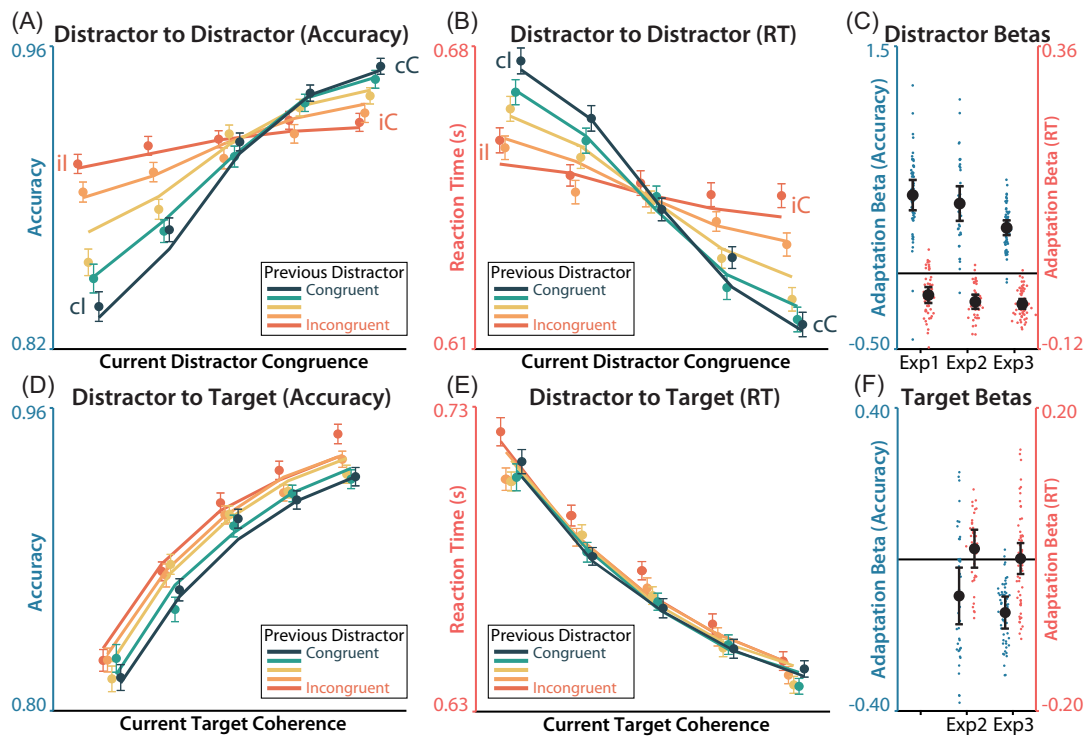
Note. Effect sizes are calculated from MAP group-level regression estimates. The *p* values are aggregated across experiments, with statistically significant *p* values (two-tailed, $\alpha = 0.05$) shown in bold. DV = dependent variable; RT = reaction time; MAP = maximum a posteriori; Exp = experiment.

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 previous 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 previous distraction on target processing was substantially smaller than its effect on distractor processing (see Figure 5) and was only found for accuracy ($p < .001$)

Figure 3
Distractor-Dependent Adaptation



Note. (A and B) The relationship between distractor congruence and accuracy (A) and RT (B) was weaker when the previous trial was more incongruent (redder colors). Circles depict participant behavior, and lines depict aggregated regression predictions. The traditional binary adaptation effects are highlighted with their short forms, e.g., ‘iC’ means the previous trial was incongruent (i) and the current trial is congruent (C). (C) Regression estimates for the current distractor congruence interaction, within each experiment. (D and 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, and error bars on regression coefficients reflect 95% CI. Psychometric functions are jittered on the *x*-axis for ease of visualization. Feature coherence was rank ordered and binned into quantiles with equal numbers of trials at each level of target coherence, distractor congruence, or previous distractor congruence. RT = reaction time; SEM = standard error of the mean; CI = confidence interval; ns = nonsignificant; Exp = experiment. See the online article for the color version of this figure.

Table 2
Effects of Previous Conflict on Feature Sensitivity

DV	Predictors	Exp 1 (<i>df</i> = 41) Effect size (<i>d</i>)	Exp 2 (<i>df</i> = 15) Effect size (<i>d</i>)	Exp 3 (<i>df</i> = 35) Effect size (<i>d</i>)	Aggregate <i>p</i> value
Choice	Distractor × Previous Distractor	1.59	1.45	1.74	6.15×10^{-31}
	Distractor × Previous Target		-0.670	0.103	.964
	Target × Previous Distractor		-0.473	-0.990	2.83×10^{-8}
Lapse rate	Target × Previous Target		0.418	0.644	1.25×10^{-5}
	Previous distractor	-0.522	-0.498	-1.04	1.75×10^{-10}
	Previous target		-0.110	-0.494	.00934
RT	Distractor × Previous Distractor	-0.836	-1.44	-1.79	8.99×10^{-24}
	Distractor × Previous Target		0.174	0.0618	.520
	Target × Previous Distractor		0.210	0.0155	.726
	Target × Previous Target		0.147	0.154	.285
	Previous distractor	0.287	0.202	-0.267	.623
	Previous target		0.109	-0.275	.0884

Note. Effect sizes are calculated from MAP group-level regression estimates. The *p* values are aggregated across experiments, with statistically significant *p* values (two-tailed, $\alpha = 0.05$) shown in bold. DV = dependent variable; RT = reaction time; MAP = maximum a posteriori; Exp = experiment.

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 (Supplemental 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 ($p = .19$).

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 et al., 2004; Mayr et al., 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 (Schmidt, 2019). Our experiment was designed to largely avoid potential priming confounds by eliminating stimulus repetitions (with two colors assigned to each response hand that never repeat) and by using stochastic motion stimuli (vs., e.g., static arrows) that also have very infrequent exact repetitions. For example, the probability that two trials will have the same motion coherence was only 9%.

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 (Moeller & Frings, 2014), 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; Supplemental Tables 4 and 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. Distractor adaptation was weaker during attend-motion than attend-color, despite including color repetitions during attend-motion blocks (choice: $p < .001$; RT: $p < .001$). Critically, we can directly compare trial-to-trial changes in motion sensitivity when motion is

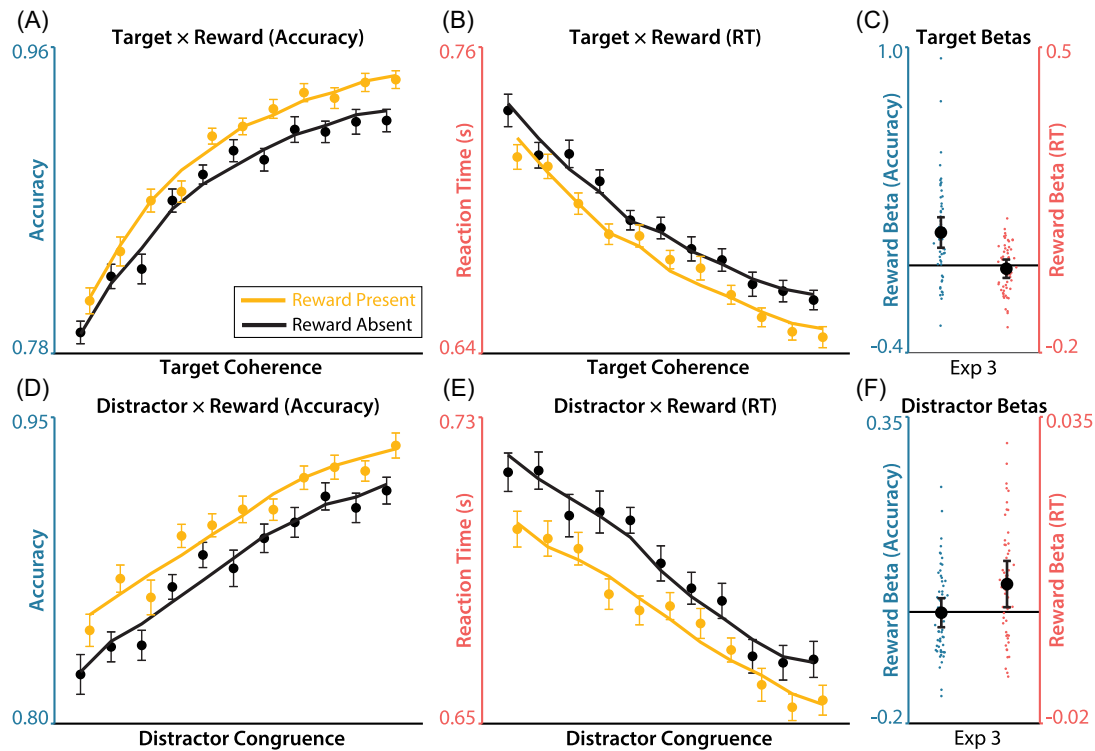
task relevant (attend-motion) and task irrelevant (attend-color), matching the salience of this motion dimension across tasks (Giesen et al., 2012). Target adaptation during attend-motion was not significant (choice: $p = .268$; RT: $p = .777$; Supplemental Table 4) and was weaker than distractor adaptation during attend-color (choice: $p < .001$; RT: $p = .34$; Supplemental 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 influencing the 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 rate*, see the Regression Analysis section). 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., postcongruent lapse rates ranged from 0.023% to 0.13% across studies; postincongruent 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 previous distractor-related interference and that this influence was observed primarily in subsequent processing of the (motion) distractor rather than the (color) target. This may reflect a fundamental bias in the control system toward 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).

Figure 4
Influence of Incentives on Target and Distractor Sensitivity



Note. (A and B) The relationship between target coherence and performance was stronger during incentivized blocks (gold) in the domain of accuracy (A), but not RT (B). Circles depict participant behavior, and lines depict aggregated regression predictions. (C) Regression estimates for the target coherence by incentive interaction. (D and E) The relationship between distractor congruence and performance was weaker during incentivized blocks (gold) in the domain of RT (E), but not accuracy (D). (F) Regression estimates for the distractor congruence by incentive interaction. Error bars on behavior reflect within-participant SEM, and error bars on regression coefficients reflect 95% CI. Psychometric functions are jittered on the *x*-axis for ease of visualization. RT = reaction time; SEM = standard error of the mean; CI = confidence interval; Exp = experiment. See the online article for the color version of this figure.

We found that participants' accuracy was more sensitive to target discriminability in rewarded blocks than in nonrewarded blocks ($d = 0.61, p < .001$; Figure 4A, Table 3). This effect of incentives on target sensitivity was specific to choice and not RTs ($d = -0.10, p = .47$), though participants were overall faster in rewarded blocks ($d = -0.41,$

$p = .0045$). Participants were also marginally more likely to make lapses responses during rewarded blocks ($d = 0.244, p = .092$). In terms of distractors, we found that in rewarded blocks, participants were less sensitive to distractors in RT ($d = 0.35, p = .012$), albeit with a small effect size, and that incentives did not significantly influence distractor sensitivity in choice ($d = -0.016, p = .91$).

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, $t(59.0) = 2.14, p = .036$; Supplemental Tables 6 and 7.

Previous Conflict and Incentives Have Dissociable Influences on Target and Distractor Processing

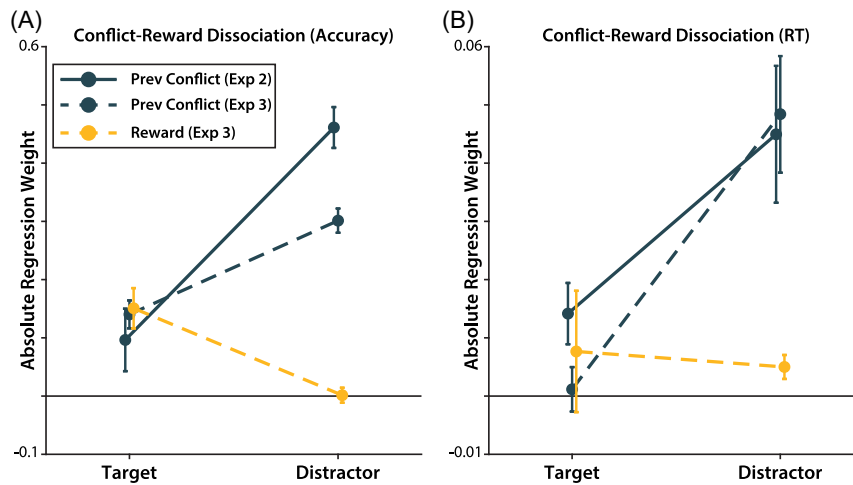
Our within-trial results demonstrated that participants are sensitive to target (color) and distractor (motion) information, with little interaction between these dimensions. Consistent with this putative independence, we found that previous interference

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	.911
Lapse rate	Reward	0.244	.0924
	Target × Reward	-0.103	.467
RT	Distractor × Reward	0.349	.0195
	Reward	-0.411	.00447

Note. Effect sizes are calculated from MAP group-level regression estimates. Statistically significant p values (two-tailed, $\alpha = 0.05$) are shown in bold. DV = dependent variable; RT = reaction time; MAP = maximum a posteriori; Exp = experiment.

Figure 5
Dissociations Between Previous Conflict and Incentive Effects



Note. Postconflict 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 SEM. RT = reaction time; SEM = standard error of the mean; MAP = maximum a posteriori; Exp = experiment. See the online article for the color version of this figure.

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 previous conflict 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, $t(37.7) = 0.25$, $p = .80$. Critically, the difference-of-differences was also significant in both accuracy, $t(39.6) = 10.2$, $p = 1.36 \times 10^{-12}$, and RT, $t(48.3) = 3.11$, $p = .0031$, supporting dissociable control over different dimensions of feature processing.

These findings are consistent with a previous neuroimaging experiment that found incentives enhanced responses in target-related areas (visual word form area for text targets) and mostly incongruent blocks suppressed responses in distractor-related areas (fusiform face area for face distractors; Soutschek et al., 2015). In the following sections, we extend these convergent findings to explore how previous conflict and incentives influence the dynamics of control implementation.

Differential Within-Trial Dynamics of Target and Distractor Processing

Our initial results show that participants independently control their sensitivity to target (color) and distractor (motion) information. However, previous research has revealed that participants' task processing also dynamically changes within a trial (Servant et al.,

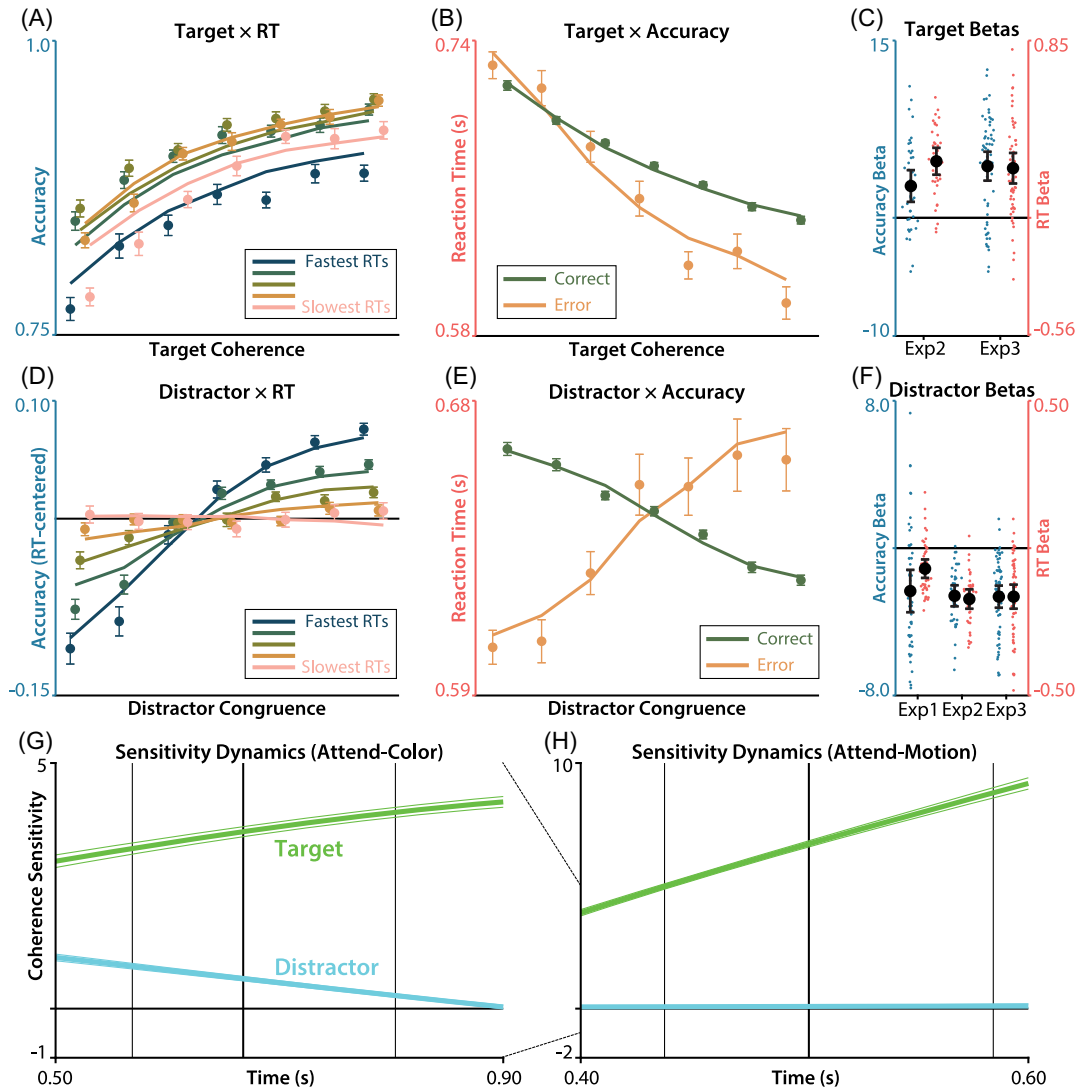
2014; Weichart et al., 2020; White et al., 2011), including in response to incentives (Adkins & Lee, 2021). Whereas much of the previous research has focused on dynamics in spatial attention during flanker tasks (e.g., a shrinking spotlight of attention; Weichart et al., 2020; White et al., 2011), less 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 \sim Coherence \times RT interaction). These analyses work under the logic that faster RTs reflect earlier epochs of information processing, which we confirm through subsequent evidence accumulation simulations (see An Accumulator Model of Attentional Control Over Target and Distractor Processing section; Supplemental Figures 4 and 5). Our approach builds on "delta function" analyses of how congruence effects differ across RT quantiles (De Jong et al., 1994; Ridderinkhof, 2002; van den Wildenberg et al., 2010),³ extending this methodology with a general linear modelling approach that estimates 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 ($ds = 0.69$ – 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

³ Previous work on delta-plot analyses has investigated how RT difference scores (e.g., congruent – incongruent) vary across RT quantiles. This work has been criticized based on the inherent mean–variance relationship in skewed RT distributions (Zhang & Kornblum, 1997). Instead, our analyses investigate how accuracy effects vary as a function of RT instead, avoiding this concern.

Figure 6
Target and Distractor Sensitivity Dynamics



Note. (A) The relationship between target coherence and accuracy increased at later RTs (pink color). (B) Participants responded faster on error trials than on correct trials 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 (pink). Note that these data are mean-centered within each RT bin to remove the target effects in (A) from this visualization of distractor sensitivity. (E) Participants responded faster on error trials than on 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 are 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, and error bars on regression coefficients reflect 95% CI. Psychometric functions are jittered on the x-axis for ease of visualization. Feature coherence and RT were rank ordered and binned into quantiles with equal numbers of trials at each level of target coherence, distractor congruence, or RT bin. RT = reaction time; SEM = standard error of the mean; CI = confidence interval; Exp = experiment. See the online article for the color version of this figure.

process that enhances sensitivity to goal-relevant features and suppresses attention toward goal-irrelevant features. Notably, these results suggest that this attentional process occurs “online” within the course of a trial.

We also fit a complementary analysis for RT (i.e., the $RT \sim Coherence \times Accuracy$ interaction). We found that participants had steeper target coherence slopes on error trials ($ds = 0.89-1.5$, $p < .001$; Figure 6B), driven by faster errors when the targets were high

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 \times RT		0.686	0.975	4.61×10^{-11}
	Distractor \times RT	-0.709	-1.54	-1.19	8.40×10^{-20}
Lapse rate RT	RT	0.247	0.928	0.534	9.25×10^{-13}
	Target \times Accuracy		1.46	0.889	1.99×10^{-14}
	Distractor \times Accuracy	-0.683	-1.82	-1.09	1.39×10^{-20}

Note. Effect sizes are calculated from MAP group-level regression estimates. The p values are aggregated across experiments, with statistically significant p values (two-tailed, $\alpha = 0.05$) shown in bold. DV = dependent variable; RT = reaction time; MAP = maximum a posteriori; Exp = experiment.

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.

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; Supplemental Tables 8 and 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.

Finally, we tested whether participants' within-trial attentional dynamics changed over the course of the experiment, modeling the linear change in parameters across blocks of trials. We found that later in the experiment, participants' overall sensitivity to distractors was higher (in choice), and their sensitivity to targets was lower (in reaction time; see Supplemental Table 11). However, later in the experiment, participants also had faster target and distractor dynamics, such that maladaptive sensitivity was most prominent in the earliest phase of the trial (Supplemental Figure 11). These results speculatively suggest that over time participants shift from maintaining initial sensitivity to reactively reconfiguring attention, potentially due to fatigue or proactive interference from attend-motion blocks.

Previous Conflict and Incentives 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 color information and decreasing sensitivity to task-irrelevant motion 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.

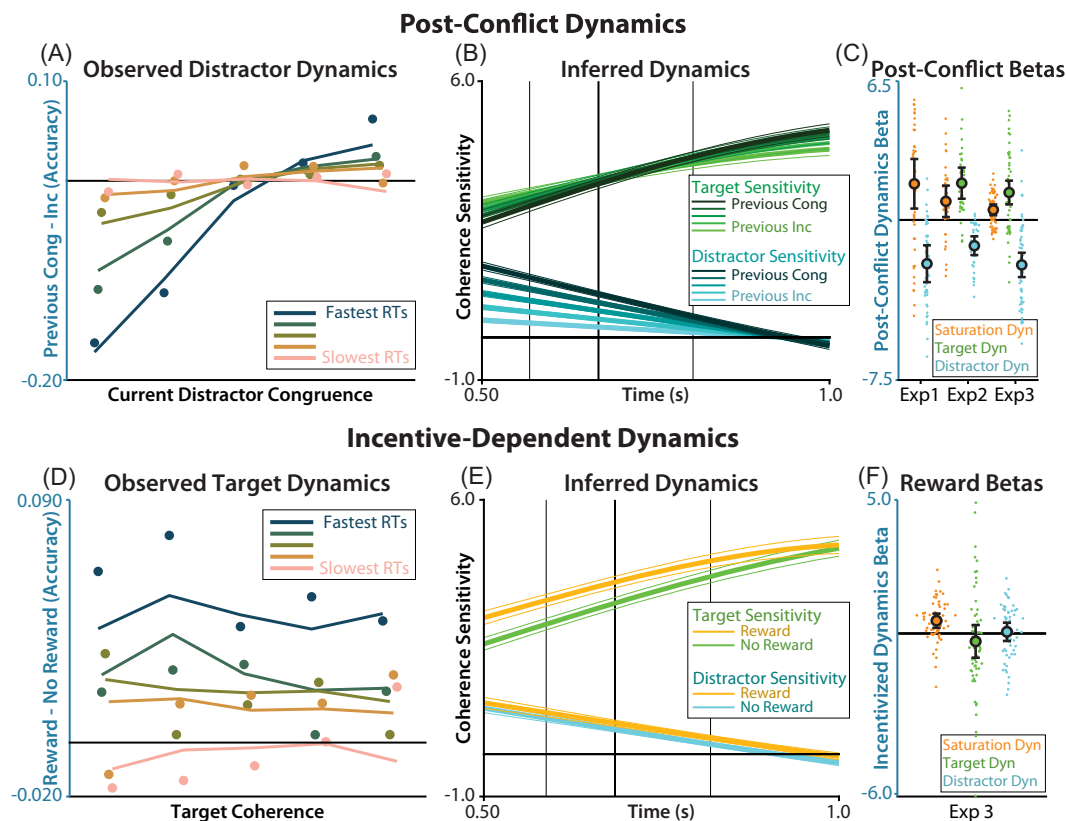
To address this question, we first examined how the dynamics of target and distractor sensitivity were altered by the congruence of the distractor in the previous trial (i.e., *Choice PreviousDistractor \times RT \times Coherence*). We found that after incongruent trials, participants started the next 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 ($ds = 0.65$ – 1.0 , $p < .001$) and distractor suppression ($ds = -0.68$ to -1.1 , $p < .001$; Table 5). Both postcongruent and postincongruent trials thus reached similar asymptotic levels of feature sensitivity. This early influence of previous conflict on congruence sensitivity is consistent with previous experiments on the time course of conflict adaptation (Stins et al., 2008; Wylie et al., 2010), with the current work extending these findings to show concurrent, albeit weaker, target-enhancement dynamics.

We performed the equivalent analysis for incentive-related adaptation (i.e., *Choice Reward \times RT \times Coherence*), and found that during incentivized blocks, participants' initial target sensitivity was higher than during nonincentivized blocks and remained so across much of the trial (see Figure 7D). However, target sensitivity eventually reached an asymptote, such that toward the end of the trial, both incentivized and nonincentivized 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 significantly differ between incentivized and nonincentivized trials ($d = 0.055$, $p = .71$).

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 toward 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 et al., 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

Figure 7
Influence of Conflict and Incentives on Sensitivity Dynamics



Note. (A) The relationship between previous distractor congruence and current distractor congruence was strongest for early RTs (bluer 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 are 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 (“saturation dynamics,” 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 (bluer color). The y-axis depicts the difference in accuracy between blocks where there were rewards versus 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, and error bars on regression coefficients reflect 95% CI. Psychometric functions are jittered on the x-axis for ease of visualization. Feature coherence and RT were rank ordered and binned into quantiles with equal numbers of trials at each level of target coherence, distractor congruence, or RT bin. RT = reaction time; SEM = standard error of the mean; CI = confidence interval; Exp = experiment. See the online article for the color version of this figure.

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 is determined by control units that determine the gain of a given feature (Cohen et al., 1990; Musslick et al., 2019). We hand tuned the parameters of this model to determine whether it could capture our core experimental findings across choice and reaction time.

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”

Table 5
Effects of Previous Conflict on Feature Sensitivity Dynamics

DV	Predictors	Exp 1 (<i>df</i> = 26) Effect size (<i>d</i>)	Exp 2 (<i>df</i> = 9) Effect size (<i>d</i>)	Exp 3 (<i>df</i> = 29) Effect size (<i>d</i>)	Aggregate <i>p</i> value
Choice	Previous Distractor × Distractor × RT	−0.853	−1.07	−1.01	1.62 × 10^{−14}
	Previous Distractor × Target × RT		1.01	0.646	1.71 × 10^{−7}
Lapse rate	Previous Distractor × RT	0.456	0.463	0.531	3.52 × 10^{−6}
RT	Previous Distractor × Distractor × Accuracy	−0.563	−1.15	−1.16	2.84 × 10^{−13}
	Previous Distractor × Target × Accuracy		−0.00500	−0.524	.135
	Previous Distractor × Accuracy	0.417	0.175	−0.162	.881

Note. Effect sizes are calculated from MAP group-level regression estimates. The *p* values are aggregated across experiments, with statistically significant *p* values (two-tailed, $\alpha = 0.05$) shown in bold. DV = dependent variable; RT = reaction time; MAP = maximum a posteriori; Exp = experiment.

(Supplemental Figure 3; Teodorescu & Usher, 2013). This occurred because larger inputs (whether congruent or incongruent) drove faster reaction times.

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 learning) that exponentially approaches a fixed point (e.g., determined by the task goals and control; cf. Musslick et al., 2019; Steyvers et al., 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 and 10D). Critically, we were unable to replicate these qualitative patterns of behavior with FFI models in which control gains were frozen throughout the trial (Supplemental Figure 4). Drift-diffusion models with across-trial variability in gain, noise, or threshold, and drift-diffusion models with within-trial dynamics in noise or threshold, were also unable to capture our key effects without within-trial gain dynamics (Supplemental Figure 5).

At later RTs, participants were more likely to exhibit lapses in performance (i.e., choose randomly; $ds = 0.25\text{--}0.93$, $p < .001$, see Table 4), which were estimated with a separate term in our regression models (see the Regression Analyses section in Methods). This is evident in poorer overall performance in the slowest RT bin, relative to the second and fourth bins (see Figure 10A, left panel, pink line). A similar "hook" is often observed in RT-conditioned accuracy functions, with gradually better performance followed by poorer performance for the slowest RTs (van den Wildenberg et al., 2010; Weichart et al., 2020). Our simulation captured this global reduction in accuracy by including a collapsing boundary (Drugowitsch et al., 2012; Rosenbaum et al., 2022), which leads to late errors irrespective of feature coherence (see Supplemental Figure 5 for contrast to fixed bound). Notably, even though overall accuracy is reduced over time, target sensitivity is stronger at the slowest RT bin relative to the earliest RT bin (compare navy and pink psychometric slopes in Figure 10A), consistent with both feature-selective dynamics (gain control) and global dynamics (collapsing bound). By including a theory-driven mechanism for reductions in overall accuracy, our FFIc model captures performance trends in this slowest RT quantile that were difficult to capture with for more model-agnostic regression analyses.

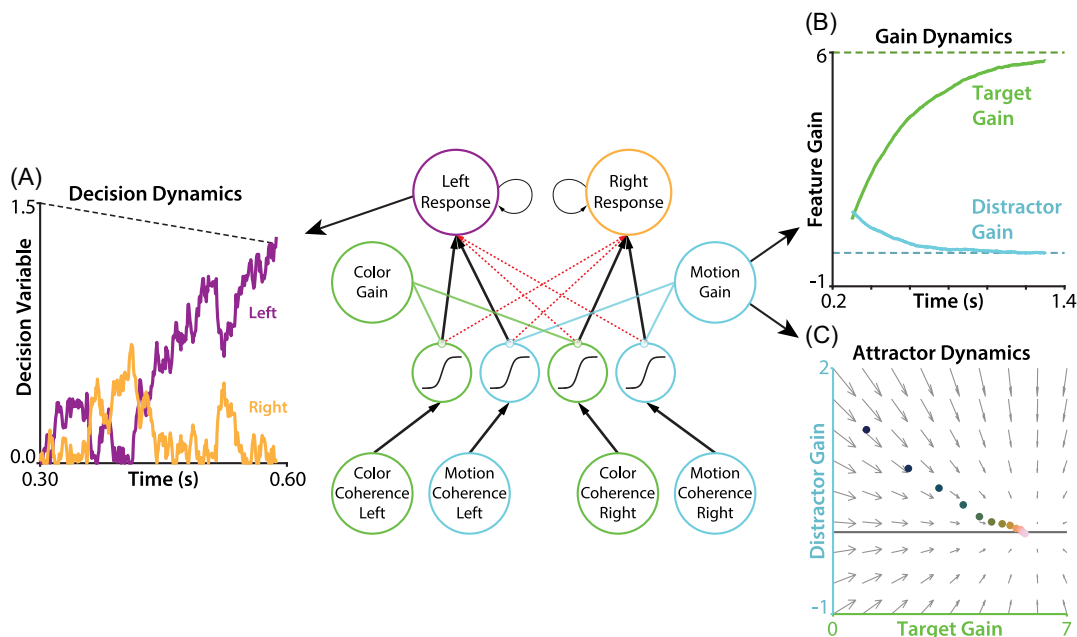
The parallel feature pathways in this model are designed to capture the independent influences of target and distractor information (Lindsay & Jacoby, 1994). However, the time-varying feature gains provide an account for the weak interactions we observed between target and distractor sensitivity in accuracy. Despite there being no

Table 6
Effects of Incentives on Feature Sensitivity Dynamics

DV	Predictors	Exp 3 (<i>df</i> = 29) Effect size (<i>d</i>)	<i>p</i> value
Choice	Reward × Target × RT	−0.139	.331
	Reward × Distractor × RT	0.0546	.712
Lapse rate	Reward × RT	0.524	.000937
RT	Reward × Target × Accuracy	0.437	.00432
	Reward × Distractor × Accuracy	−0.170	.260
	Reward × Accuracy	−0.540	.000646

Note. Effect sizes are calculated from MAP group-level regression estimates. Statistically significant *p* values (two-tailed, $\alpha = 0.05$) are shown in bold. DV = dependent variable; RT = reaction time; MAP = maximum a posteriori; Exp = experiment.

Figure 8
Feedforward Inhibition With Control



Note. 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 the 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 a 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 pathway is controlled by a feature gain. Over time within a trial, the feature gains for targets (green) and distractors (cyan) exponentially approach to a fixed level (high gain for targets, zero gain for distractors). Note the difference in *x*-axis scaling compared to Figure 6G. (C) An equivalent visualization of the dynamics in B. Attractor dynamics drive target and distractor gains to their fixed level, shown at different time points within the trial (pink colors are later in the trial). The horizontal line depicts zero distractor gain. FFIC = feedforward inhibition with control. See the online article for the color version of this figure.

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 and 10D), censoring 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.

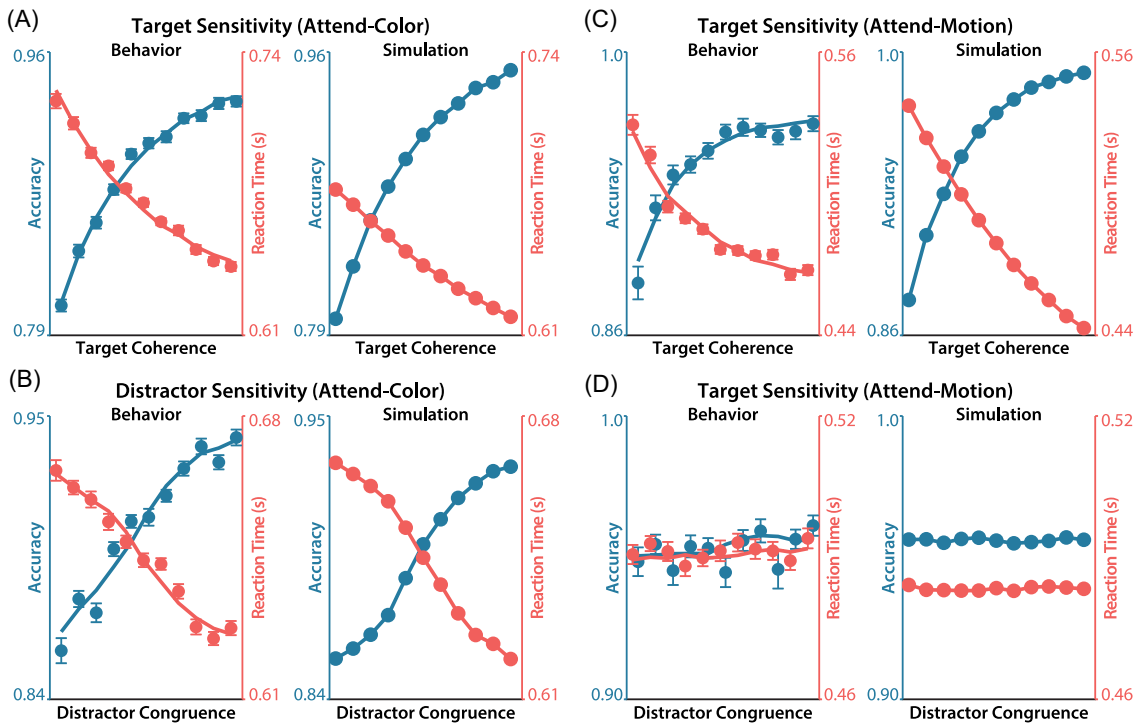
To account for the early effects of conflict and incentives, we modified the initial conditions of our model's gain dynamics (Figure 11A). We simulated postinterference adaptation by initializing the distractor gain closer to its asymptote, and we simulated reward incentivization by initializing the 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 dynamics in our attractor network parsimoniously account for the fact that dynamics tended to be faster when they were initialized further from the fixed point (i.e., postcongruent trials). Thus, our model was able to capture the range of findings in this experiment: target–distractor sensitivity, within-trial dynamics, and how the dynamics of target and distractor processing may be influenced by control.

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

Figure 9
Simulation of Target and Distractor Sensitivity (See Figure 2)



Note. (A and B) Sensitivity to target coherence (A) and distractor congruence (B) in behavior (left) and in the FFIC simulation (right) for attend-color blocks. (C and D) Same as A and B, but for attend-motion blocks. See the online article for the color version of this figure.

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.

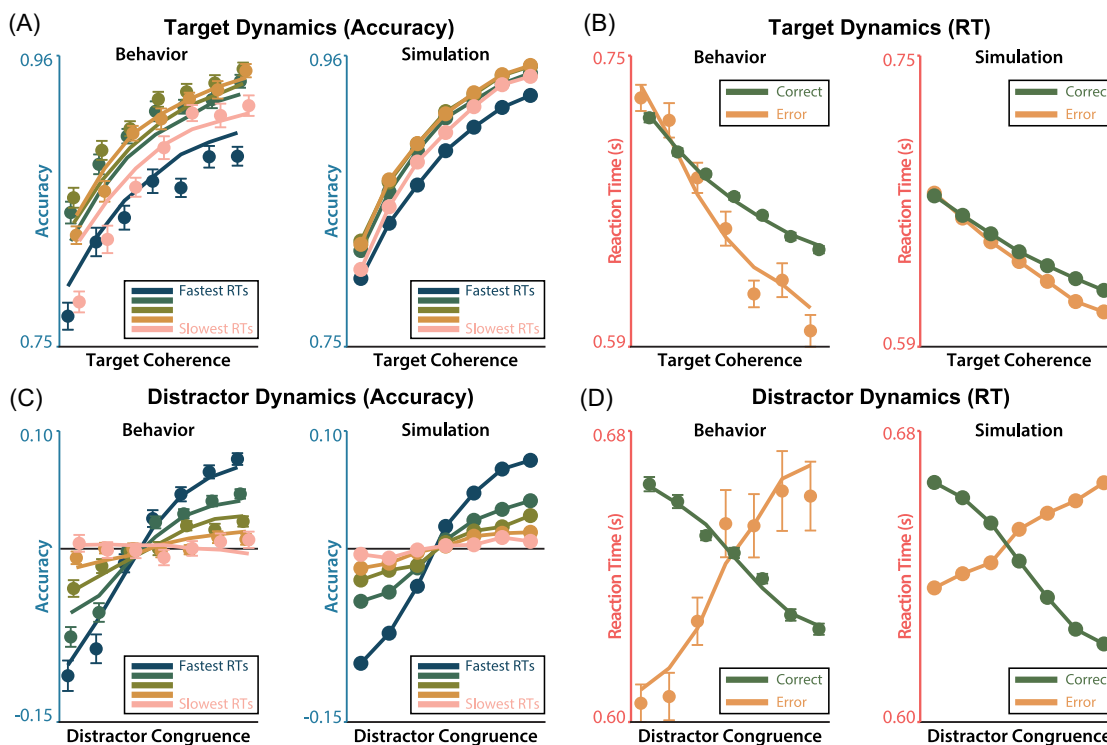
Together, these results support a cognitive control architecture that is parametric, multivariate, and dynamic. Previous research has found that cognitive effort is enhanced in response to incentives (Parro et al., 2018; Yee & Braver, 2018) and previous conflict (Egner, 2007; Gratton et al., 1992). The current experiments extend these previous findings to show that these adaptations are both graded in their intensity and selective in their allocation. These findings are consistent with a multivariate perspective on cognitive control (Ritz et al., 2022), in which people optimize a configuration of control signals according to their costs and benefits (Musslick et al., 2015; Shenhav et al., 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., 2021; Noonan et al., 2016; Simen et al., 2009; Soutschek et al., 2015; Wöstmann et al., 2019).

A core question arising from these results is why there are preferential relationships between previous conflict 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 incentives should allocate

control toward 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 the fact 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 et al., 2014; Yu et al., 2009) or the value of control policies (Bustamante et al., 2021; Lieder et al., 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 et al., 2021; Krajbich et al., 2010; Li & Ma, 2021; Westbrook et al., 2020) and cognitive control (Adkins & Lee, 2021; Hardwick et al., 2019; Servant et al., 2014; Ulrich et al., 2015; 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 the processing of targets and distractors is less mutually constrained. Whereas previous work has modeled within-trial dynamics as simplified impulse functions (Ulrich et al., 2015), our modeling approach extends these accounts with

Figure 10
Simulation of Target and Distractor Sensitivity Dynamics (See Figure 6)



Note. (A and B) RT-dependent (A) and accuracy-dependent (B) sensitivity to target coherence in behavior (left) and in the FFIC simulation (right). (C and D) Same as A and B, but for distractor congruence. FFIC = feedforward inhibition with control; RT = reaction time. See the online article for the color version of this figure.

a more process-oriented focus on how a neural network could be parameterized to produce key patterns of within-trial attentional dynamics. Furthermore, relatively few experiments have studied how attentional dynamics are modified in response to control drivers like incentives or task demands (though see Adkins & Lee, 2021; van den Wildenberg et al., 2010; Yu et al., 2009).

Our experiments show that the dynamics of target and distractor sensitivity are independent and that previous conflict and incentives appear to operate through changes to the initial conditions of these feature gains.⁴ These findings are broadly consistent with influential theories of attentional dynamics, which propose that early task processing is largely driven by feature salience and statistical or reinforcement learning, whereas attentional control has a relatively slower time course (Awh et al., 2012; Theeuwes, 2010, 2018; see also van den Wildenberg et al., 2010).⁵ If participants are learning the relevance of different features, these initial conditions in part may reflect the prior probability that attention toward 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

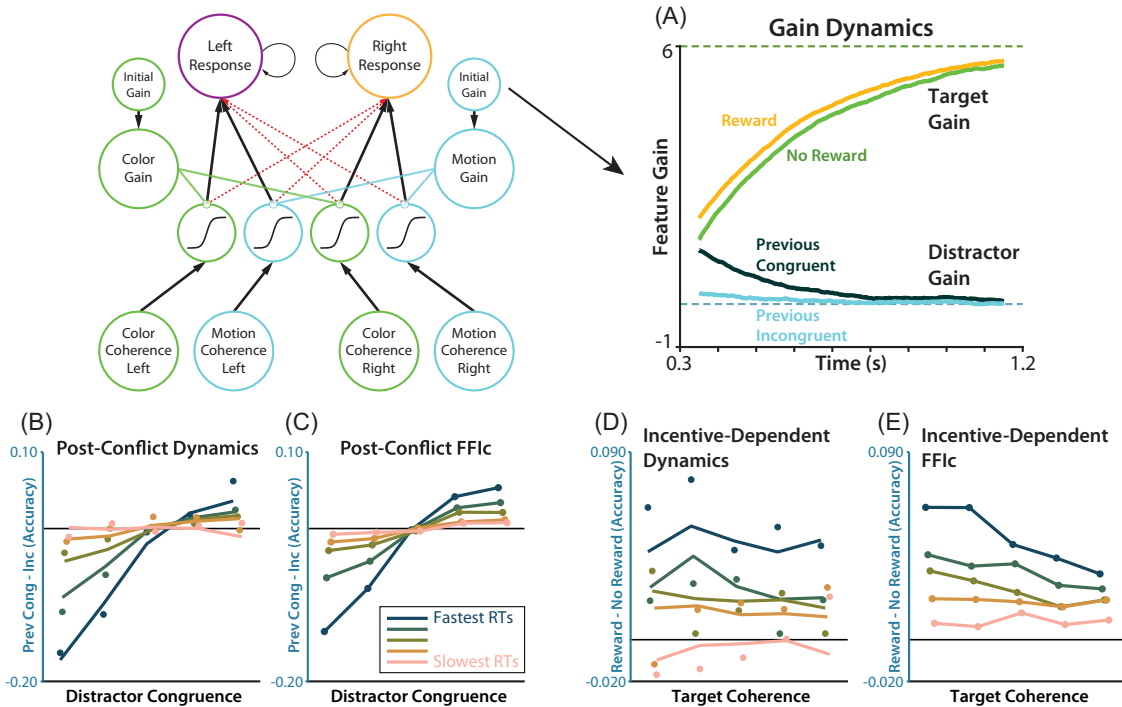
may reflect the expected target-reward contingency. Future research should investigate this account by measuring attentional dynamics as participants learn task contingencies (Shenhav et al., 2018).

Our patterns of conflict- and incentive-dependent dynamics help rule out stimulus-driven dynamics and support independent control over feature processing. After congruent trials, participants started the next trial with more similar target and distractor gains, which 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, there would be slower dynamics. Instead, we found faster dynamics, supporting a role for feedback control that reconfigures attentional gain to align with task goals. Additionally, during incentivized blocks, we saw that participants modified attentional

⁴ We found that just modifying feature gains' initial conditions parsimoniously accounted for incentive and previous conflict effects. Note that we do not explicitly compare this model to more complex models incorporating changes to parameters like decay rate and/or asymptotic gain, which should be more thoroughly investigated in future experiments.

⁵ We assume that "early" and "late" processing do not reflect discrete stages (Hübner et al., 2010) but different time points in a continual process. While this is consistent with previous work showing that gradual attentional adjustments are a better model of flanker task performance (White et al., 2011), future work should experimentally confirm the continuous nature of these attentional dynamics.

Figure 11
Simulation of Postconflict and Incentive Effects (See Figure 7)



Note. (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 influencing initial target gain. (B and C) The influence of the previous congruence on distractor sensitivity dynamics in behavior (B) and in the FFlc simulation (C). (D and E) The influence of incentives on target sensitivity dynamics in behavior (D) and in the FFlc simulation. FFlc = feedforward inhibition with control. See the online article for the color version of this figure.

dynamics for targets, but not distractors. This finding further supports the independence of these attentional dynamics, demonstrating that participants can alter attention toward individual features one at a time. This pattern of incentives enhancing sensitivity to target information, while also producing faster responses and a marginally higher lapse rate, is consistent with previous work on motivated attention. A recent experiment used drift-diffusion modeling to show that participants increase their rate of evidence accumulation and decrease their response threshold when faced with higher rewards, consistent with the reward rate-optimal policy (Leng et al., 2021). The current experiment extends these findings by revealing how specific attentional adjustments improve evidence accumulation, providing a more process-oriented account of motivated cognitive control.

Our dynamical process model may help link behavior in response to conflict tasks to cognitive dynamics in other domains. In the domain of task switching, recent cognitive models have developed similar dynamical accounts 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 set dynamics during the switch preparation period (Gilbert & Shallice, 2002; Jongkees et al., 2023; Musslick et al., 2019; Ueltzhöffer et al., 2015; Yeung & Monsell, 2003) and across trials (Grahek et al., 2022; Jaffe et al., 2023;

Steyvers et al., 2019). If the within-trial dynamics we observe here reflect such “task set micro-adjustments” (Ridderinkhof, 2002), then our results highlight the computational similarities between different forms of cognitive flexibility. Both within trials and across tasks, reconfiguration appears to be well captured by a common class of dynamical systems in which task configurations exponentially approach an appropriate set point. In this experiment, we show that these dynamics are multivariate and adjusted to meet local task demands through changes to initial conditions. Interestingly, control over initial conditions also plays a central role in the neural dynamics of motor preparation (Churchland et al., 2010; Kao et al., 2020; Remington et al., 2018), highlighting the broader similarities across motor and cognitive domains (Ritz et al., 2020, 2022) and generating predictions for the neural implementation of dynamic cognitive control.

The evidence we provide for dissociable control over target and distractor processing is consistent with previous neuroscience experiments that used neural correlates of stimulus processing to argue for independent enhancement and suppression processes (Gazzaley et al., 2005; Noonan et al., 2016; Soutschek et al., 2015; Wöstmann et al., 2019). Our results extend these findings by exploring how different factors can contribute to the dynamic reconfiguration of target and distractor attention, which we formalize in an explicit process model. Notably, our findings diverge from neuroimaging experiments that have suggested that control primarily

acts through enhancements to target processing (Egner & Hirsch, 2005). One potential source of this divergence may be that people's control strategies differ depending on the source of task conflict (Braem et al., 2014; Egner, 2008; Egner, 2007). For example, tasks evoking stimulus–stimulus conflict (e.g., semantic competition in the Stroop task) may require different strategies than tasks evoking stimulus–response conflict (e.g., distractors driving competing responses, as in PACT). Although previous work using Stroop-like tasks has found similar patterns of control adjustments as in the current experiment (Soutschek et al., 2015), this raises the broader question of whether the specific feature–control relationships in this experiment should generalize to other tasks. According to the expected value of control theory and the learned value of control model that builds upon it, control strategies are adapted to specific task contexts (Lieder et al., 2018; Ritz et al., 2022; Shenhav et al., 2013). This framework predicts that there will be strategic or learned control–feature mappings, rather than a rigid relationship between task features and control policies. The current results show that participants can independently control target and distractor processing when these features are independent, and future work should explore whether control strategies appropriately accommodate other tasks.

Interestingly, participants appeared to suppress distractor sensitivity even on congruent trials, evident in the right half of Figure 6D (see also Mante et al., 2013; Pagan et al., 2022), suggesting that they are not reactively adjusting this control policy when the trial conditions deem it unnecessary or even detrimental. On its face, this finding presents a challenge to models that propose control allocation on the basis of response conflict (Botvinick et al., 2001; Yu et al., 2009), though much of the evidence for these theories comes from across-trial adjustments (Botvinick et al., 2001; Egner & Hirsch, 2005; Kerns et al., 2004; Yeung et al., 2004). The current results may thus inform understanding of the timescale over which people plan reactive control adjustments. In some cases, this decision process may take more time than would be helpful for fast within-trial reconfiguration.

Our analyses of attentional dynamics depend on participants' own response times and choices, raising concerns about selection biases (i.e., lack of experimental control over reaction times). While evidence accumulation modeling typically depends on choice-conditioned reaction times, inferring the time-varying influence of targets and distractors presents a particular challenge. To address these concerns, we used simulations to show that the dynamic profiles we observed cannot be accounted for by an evidence accumulation model with static gains on target and distractor processing (Supplemental Figure 4) or models with dynamic changes to nonselective components like decision threshold (Supplemental Figure 5). Introducing dynamic feature gains allowed us to account for those same patterns (Figures 9–11; Supplemental Figure 5). These results are consistent with previous work validating DDM estimates of attentional dynamics in conflict tasks (White et al., 2011, 2018). Even if these measurements are valid, using sparse behavioral measures is an inefficient method for measuring latent dynamics and may combine multiple processes (e.g., accumulation and threshold adjustments). By integrating across multiple convergent measures of decision and attentional dynamics—including interrogation protocols (Adkins & Lee, 2021; Hardwick et al., 2019), motor tracking (Erb et al., 2016; Menciloglu et al., 2021; Scherbaum et al., 2010), and/or temporally resolved neuroimaging (Fischer et al., 2018; Scherbaum et al., 2011; Weichart et al., 2020; Yeung et al., 2004)—future work can help

strengthen and build on our understanding of continuous changes in the configuration of multiple control processes.

The evidence accumulation modeling in the current experiment was able to categorically rule out several alternative architectures, demonstrating the necessity and sufficiency of feature-specific adjustments for capturing the full array of putative attentional dynamics. Our model validation approach supports our interpretation of feature-selective adjustments while committing less strongly to the specific formulation of attentional control (e.g., a specific model parameterization or the functional form of the collapsing bound). An important direction for future research should be to leverage emerging methods for parameter estimation to directly fit our accumulator model to participants' behavior (Fengler et al., 2021; Weichart et al., 2020). This approach will help extend insights from the current experiment, such as enabling participant-specific parameters to reveal individual differences in attentional control.

Together, these experiments provide new insight into how we flexibly adapt to the changing demands of our environment. We find evidence for flexible control that aligns multiple forms of information processing with task goals and can be captured by a computationally explicit process model. The developments from this experiment can help extend models of cognitive control toward richer accounts of how multivariate control configurations, such as across targets and distractors, are optimized during goal-directed behavior.

References

- Adkins, T. J., & Lee, T. (2021). *Reward reduces habitual errors by enhancing the preparation of goal-directed actions*. PsyArXiv. <https://doi.org/10.31234/osf.io/hv9mz>
- Appelbaum, M., Cooper, H., Kline, R. B., Mayo-Wilson, E., Nezu, A. M., & Rao, S. M. (2018). Journal article reporting standards for quantitative research in psychology: The APA Publications and Communications Board task force report. *American Psychologist*, 73(1), 3–25. <https://doi.org/10.1037/amp0000191>
- Awh, E., Belopolsky, A. V., & Theeuwes, J. (2012). Top-down versus bottom-up attentional control: A failed theoretical dichotomy. *Trends in Cognitive Sciences*, 16(8), 437–443. <https://doi.org/10.1016/j.tics.2012.06.010>
- 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. <https://doi.org/10.1037/0033-295X.113.4.700>
- 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 B: Biological Sciences*, 362(1485), 1655–1670. <https://doi.org/10.1098/rstb.2007.2059>
- 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), 624–652. <https://doi.org/10.1037/0033-295X.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. <https://doi.org/10.1111/cogs.12126>
- Braem, S., Abrahamse, E. L., Duthoo, W., & Notebaert, W. (2014). What determines the specificity of conflict adaptation? A review, critical analysis, and proposed synthesis. *Frontiers in Psychology*, 5, Article 1134. <https://doi.org/10.3389/fpsyg.2014.01134>
- 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*, 23(9), 769–783. <https://doi.org/10.1016/j.tics.2019.07.002>

- 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*, *12*(12), 4745–4765. <https://doi.org/10.1523/JNEUROSCI.12-12-04745.1992>
- Bustamante, L., Lieder, F., Musslick, S., Shenhav, A., & Cohen, J. (2021). Learning to overexert cognitive control in a stroop task. *Cognitive, Affective & Behavioral Neuroscience*, *21*(3), 453–471. <https://doi.org/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), Article e1008863. <https://doi.org/10.1371/journal.pcbi.1008863>
- 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. <https://doi.org/10.1016/j.neuron.2010.09.015>
- 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. <https://doi.org/10.1037/0033-295X.97.3.332>
- 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. <https://doi.org/10.2307/1423029>
- 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*, *31*(5), 1780–1789. <https://doi.org/10.1523/JNEUROSCI.4299-10.2011>
- 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. <https://doi.org/10.1037/0096-1523.20.4.731>
- Drugowitsch, J., Moreno-Bote, R., Churchland, A. K., Shadlen, M. N., & Pouget, A. (2012). The cost of accumulating evidence in perceptual decision making. *The Journal of Neuroscience*, *32*(11), 3612–3628. <https://doi.org/10.1523/JNEUROSCI.4010-11.2012>
- Egner, T. (2007). Congruency sequence effects and cognitive control. *Cognitive, Affective & Behavioral Neuroscience*, *7*(4), 380–390. <https://doi.org/10.3758/CABN.7.4.380>
- Egner, T. (2008). Multiple conflict-driven control mechanisms in the human brain. *Trends in Cognitive Sciences*, *12*(10), 374–380. <https://doi.org/10.1016/j.tics.2008.07.001>
- Egner, T., & Hirsch, J. (2005). Cognitive control mechanisms resolve conflict through cortical amplification of task-relevant information. *Nature Neuroscience*, *8*(12), 1784–1790. <https://doi.org/10.1038/nn1594>
- Erb, C. D., Moher, J., Sobel, D. M., & Song, J.-H. (2016). Reach tracking reveals dissociable processes underlying cognitive control. *Cognition*, *152*, 114–126. <https://doi.org/10.1016/j.cognition.2016.03.015>
- Fengler, A., Govindarajan, L. N., Chen, T., & Frank, M. J. (2021). *Likelihood approximation networks (LANs) for fast inference of simulation models in cognitive neuroscience*. Cold spring harbor laboratory. <https://doi.org/10.1101/2020.11.20.392274>
- 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), Article 5038. <https://doi.org/10.1038/s41467-018-07456-8>
- Frömer, R., & Shenhav, A. (2022). Filling the gaps: Cognitive control as a critical lens for understanding mechanisms of value-based decision-making. *Neuroscience & Biobehavioral Reviews*, *134*, Article 104483. <https://doi.org/10.1016/j.neubiorev.2021.12.006>
- Gazzaley, A., Cooney, J. W., Rissman, J., & D'Esposito, M. (2005). Top-Down suppression deficit underlies working memory impairment in normal aging. *Nature Neuroscience*, *8*(10), 1298–1300. <https://doi.org/10.1038/nn1543>
- Giesen, C., Frings, C., & Rothermund, K. (2012). Differences in the strength of distractor inhibition do not affect distractor–response bindings. *Memory & Cognition*, *40*(3), 373–387. <https://doi.org/10.3758/s13421-011-0157-1>
- Gilbert, S. J., & Shallice, T. (2002). Task switching: A PDP model. *Cognitive Psychology*, *44*(3), 297–337. <https://doi.org/10.1006/cogp.2001.0770>
- Gold, J. I., & Shadlen, M. N. (2007). The neural basis of decision making. *Annual Review of Neuroscience*, *30*(1), 535–574. <https://doi.org/10.1146/annurev.neuro.29.051605.113038>
- Grahek, I., Leng, X., Fahey, M. P., Yee, D., & Shenhav, A. (2022). *Empirical and computational evidence for reconfiguration costs during within-task adjustments in cognitive control* [Conference session]. Proceedings of the Annual Meeting of the Cognitive Science Society, Toronto, CA, United States.
- 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), 480–506. <https://doi.org/10.1037/0096-3445.121.4.480>
- 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. <https://doi.org/10.1038/s41562-019-0725-0>
- 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. <https://doi.org/10.1007/s00426-003-0132-y>
- Hübner, R., Steinhauser, M., & Lehle, C. (2010). A dual-stage two-phase model of selective attention. *Psychological Review*, *117*(3), 759–784. <https://doi.org/10.1037/a0019471>
- Jaffe, P. I., Poldrack, R. A., Schafer, R. J., & Bissett, P. G. (2023). Modelling human behaviour in cognitive tasks with latent dynamical systems. *Nature Human Behaviour*, *7*, 986–1000. <https://doi.org/10.1038/s41562-022-01510-8>
- Jiang, J., Beck, J., Heller, K., & Egner, T. (2015). An insula-frontostriatal network mediates flexible cognitive control by adaptively predicting changing control demands. *Nature Communications*, *6*(1), Article 8165. <https://doi.org/10.1038/ncomms9165>
- Jiang, J., Heller, K., & Egner, T. (2014). Bayesian modeling of flexible cognitive control. *Neuroscience and Biobehavioral Reviews*, *46*(1), 30–43. <https://doi.org/10.1016/j.neubiorev.2014.06.001>
- Jongkees, B., Todd, M., Lloyd, K., Dayan, P., & Cohen, J. D. (2023). *When it pays to be quick: Dissociating control over task preparation and speed–accuracy trade-off in task switching*. PsyArXiv. <https://doi.org/10.31234/osf.io/kuhns>
- 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*, Article e63721. <https://doi.org/10.7554/eLife.63721>
- Kao, T.-C., Sadabadi, M. S., & Hennequin, G. (2020). *Optimal anticipatory control as a theory of movement preparation: A thalamo-cortical circuit model*. bioRxiv. <https://doi.org/10.1101/2020.02.02.931246>
- Kayser, A. S., Erickson, D. T., Buchsbaum, B. R., & D'Esposito, M. (2010). Neural representations of relevant and irrelevant features in perceptual decision making. *The Journal of Neuroscience*, *30*(47), 15778–15789. <https://doi.org/10.1523/JNEUROSCI.3163-10.2010>
- Kerns, J. G., Cohen, J. D., MacDonald, A. W., III, Cho, R. Y., Stenger, V. A., & Carter, C. S. (2004). Anterior cingulate conflict monitoring and adjustments in control. *Science*, *303*(5660), 1023–1026. <https://doi.org/10.1126/science.1089910>
- 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. <https://doi.org/10.1038/nn.2635>
- Krueger, L. E. (1989). Reconciling Fechner and Stevens: Toward a unified psychophysical law. *Behavioral and Brain Sciences*, *12*(2), 251–267. <https://doi.org/10.1017/S0140525X0004855X>
- Laming, D. (1979). Autocorrelation of choice–reaction times. *Acta Psychologica*, *43*(5), 381–412. [https://doi.org/10.1016/0001-6918\(79\)90032-5](https://doi.org/10.1016/0001-6918(79)90032-5)

- Lau, B., & Glimcher, P. W. (2005). Dynamic response-by-response models of matching behavior in rhesus monkeys. *Journal of the Experimental Analysis of Behavior*, 84(3), 555–579. <https://doi.org/10.1901/jeab.2005.110-04>
- Leng, X., Yee, D., Ritz, H., & Shenhav, A. (2021). Dissociable influences of reward and punishment on adaptive cognitive control. *PLOS Computational Biology*, 17(12), Article e1009737. <https://doi.org/10.1371/journal.pcbi.1009737>
- Li, Z.-W., & Ma, W. J. (2021). An uncertainty-based model of the effects of fixation on choice. *PLOS Computational Biology*, 17(8), Article e1009190. <https://doi.org/10.1371/journal.pcbi.1009190>
- Lieder, F., Shenhav, A., Musslick, S., & Griffiths, T. L. (2018). Rational metareasoning and the plasticity of cognitive control. *PLOS Computational Biology*, 14(4), Article e1006043. <https://doi.org/10.1371/journal.pcbi.1006043>
- 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. <https://doi.org/10.1037/0096-1523.20.2.219>
- Lipták, T. (1958). On the combination of independent tests. *Magyar Tud Akad Mat Kutató Int Közl*, 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. <https://doi.org/10.1038/nature12742>
- Mayr, U., Awh, E., & Laurey, P. (2003). Conflict adaptation effects in the absence of executive control. *Nature Neuroscience*, 6(5), 450–452. <https://doi.org/10.1038/nn1051>
- 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*, 83(4), 1463–1478. <https://doi.org/10.3758/s13414-020-02235-4>
- Moeller, B., & Frings, C. (2014). Attention meets binding: Only attended distractors are used for the retrieval of event files. *Attention, Perception & Psychophysics*, 76(4), 959–978. <https://doi.org/10.3758/s13414-014-0648-9>
- 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. <https://doi.org/10.1037/0096-1523.32.3.493>
- Musslick, S., Bizyaeva, A., Agaron, S., Leonard, N., & Cohen, J. D. (2019). *Stability-flexibility dilemma in cognitive control: A dynamical system perspective* [Paper presentation]. Proceedings of the 41st Annual Meeting of the Cognitive Science Society.
- Musslick, S., Jang, S. J., Shvartsman, M., Shenhav, A., & Cohen, J. D. (2018). *Constraints associated with cognitive control and the stability-flexibility dilemma* *CogSci*. <https://shenhavlab.org>.
- Musslick, S., Shenhav, A., Botvinick, M., & Cohen, J. (2015). *A computational model of control allocation based on the expected value of control* 2nd multidisciplinary conference on reinforcement learning and decision making [Paper presentation]. The Multidisciplinary Conference on Reinforcement Learning and Decision Making.
- Nieder, A., & Miller, E. K. (2003). Coding of cognitive magnitude: Compressed scaling of numerical information in the primate prefrontal cortex. *Neuron*, 37(1), 149–157. [https://doi.org/10.1016/S0896-6273\(02\)01144-3](https://doi.org/10.1016/S0896-6273(02)01144-3)
- Noonan, M. P., Adamian, N., Pike, A., Printzlau, F., Crittenden, B. M., & Stokes, M. G. (2016). Distinct Mechanisms for distractor suppression and target facilitation. *The Journal of Neuroscience*, 36(6), 1797–1807. <https://doi.org/10.1523/JNEUROSCI.2133-15.2016>
- Norman, D. A., & Bobrow, D. G. (1975). On data-limited and resource-limited processes. *Cognitive Psychology*, 7(1), 44–64. [https://doi.org/10.1016/0010-0285\(75\)90004-3](https://doi.org/10.1016/0010-0285(75)90004-3)
- Pagan, M., Tang, V. D., Aoi, M. C., Pillow, J. W., Mante, V., Sussillo, D., & Brody, C. D. (2022). *A new theoretical framework jointly explains behavioral and neural variability across subjects performing flexible decision-making*. <https://doi.org/10.1101/2022.11.28.518207>
- Parro, C., Dixon, M. L., & Christoff, K. (2018). The neural basis of motivational influences on cognitive control. *Human Brain Mapping*, 39(12), 5097–5111. <https://doi.org/10.1002/hbm.24348>
- Posner, M., & Snyder, C. (1975). Attention and cognitive control. In R. L. Solso (Ed.), *Information processing and cognition* (pp. 55–85). Erlbaum.
- Ratcliff, R., & McKoon, G. (2008). The diffusion decision model: Theory and data for two-choice decision tasks. *Neural Computation*, 20(4), 873–922. <https://doi.org/10.1162/neco.2008.12-06-420>
- 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. <https://doi.org/10.1016/j.tics.2018.07.010>
- Ridderinkhof, K. R. (2002). Micro- and macro-adjustments of task set: Activation and suppression in conflict tasks. *Psychological Research*, 66(4), 312–323. <https://doi.org/10.1007/s00426-002-0104-7>
- Ritz, H., Frömer, R., & Shenhav, A. (2020). Bridging motor and cognitive control: It's about time! *Trends in Cognitive Sciences*, 24(1), 6–8. <https://doi.org/10.1016/j.tics.2019.11.005>
- Ritz, H., Leng, X., & Shenhav, A. (2022). Cognitive control as a multivariate optimization problem. *Journal of Cognitive Neuroscience*, 34(4), 569–591. https://doi.org/10.1162/jocn_a_01822
- Rogers, R. D., & Monsell, S. (1995). Costs of a predictable switch between simple cognitive tasks. *Journal of Experimental Psychology: General*, 124(2), 207–231. <https://doi.org/10.1037/0096-3445.124.2.207>
- Rosenbaum, D., Glickman, M., Fleming, S. M., & Usher, M. (2022). The cognition/metacognition trade-off. *Psychological Science*, 33(4), 613–628. <https://doi.org/10.1177/09567976211043428>
- Scherbaum, S., Dshemuchadse, M., Fischer, R., & Goschke, T. (2010). How decisions evolve: The temporal dynamics of action selection. *Cognition*, 115(3), 407–416. <https://doi.org/10.1016/j.cognition.2010.02.004>
- 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. <https://doi.org/10.1111/j.1469-8986.2010.01137.x>
- Schmidt, J. R. (2019). Evidence against conflict monitoring and adaptation: An updated review. *Psychonomic Bulletin & Review*, 26(3), 753–771. <https://doi.org/10.3758/s13423-018-1520-z>
- 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. <https://doi.org/10.1016/j.actpsy.2011.06.002>
- 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. <https://doi.org/10.1016/j.cogpsych.2014.03.002>
- 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. <https://doi.org/10.1152/jn.2001.86.4.1916>
- 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. <https://doi.org/10.1016/j.neuron.2013.07.007>
- 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), Article 2485. <https://doi.org/10.1038/s41467-018-04841-1>
- 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), 127–190. <https://doi.org/10.1037/0033-295X.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. <https://doi.org/10.1037/a0016926>

- Soutschek, A., Stelzel, C., Paschke, L., Walter, H., & Schubert, T. (2015). Dissociable effects of motivation and expectancy on conflict processing: An fMRI study. *Journal of Cognitive Neuroscience*, *27*(2), 409–423. https://doi.org/10.1162/jocn_a_00712
- Stafford, T., Ingram, L., & Gurney, K. N. (2011). Piéron's Law holds during stroop conflict: Insights into the architecture of decision making. *Cognitive Science*, *35*(8), 1553–1566. <https://doi.org/10.1111/j.1551-6709.2011.01195.x>
- 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*, *116*(36), 17735–17740. <https://doi.org/10.1073/pnas.1906788116>
- Stins, J. F., Polderman, J. C. T., Boomsma, D. I., & de Geus, E. J. C. (2008). Conditional accuracy in response interference tasks: Evidence from the Eriksen flanker task and the spatial conflict task. *Advances in Cognitive Psychology*, *3*(3), 409–417. <https://doi.org/10.2478/v10053-008-0005-4>
- Teodorescu, A. R., & Usher, M. (2013). Disentangling decision models: From independence to competition. *Psychological Review*, *120*(1), 1–38. <https://doi.org/10.1037/a0030776>
- 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. <https://doi.org/10.1364/JOSAA.17.000994>
- Theeuwes, J. (2010). Top-down and bottom-up control of visual selection. *Acta Psychologica*, *135*(2), 77–99. <https://doi.org/10.1016/j.actpsy.2010.02.006>
- Theeuwes, J. (2018). Visual selection: Usually fast and automatic; seldom slow and volitional. *Journal of Cognition*, *1*(1), Article 29. <https://doi.org/10.5334/joc.13>
- Tzelgov, J., Henik, A., & Berger, J. (1992). Controlling Stroop effects by manipulating expectations for color words. *Memory & Cognition*, *20*(6), 727–735. <https://doi.org/10.3758/BF03202722>
- Ueltzhöffer, K., Armbruster-Genç, D. J. N., & Fiebach, C. J. (2015). Stochastic dynamics underlying cognitive stability and flexibility. *PLOS Computational Biology*, *11*(6), Article e1004331. <https://doi.org/10.1371/journal.pcbi.1004331>
- 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. <https://doi.org/10.1016/j.cogpsych.2015.02.005>
- Urai, A. E., de Gee, J. W., Tsetsos, K., & Donner, T. H. (2019). Choice history biases subsequent evidence accumulation. *eLife*, *8*, Article e46331. <https://doi.org/10.7554/eLife.46331>
- Usher, M., & McClelland, J. L. (2001). The time course of perceptual choice: The leaky, competing accumulator model. *Psychological Review*, *108*(3), 550–592. <https://doi.org/10.1037/0033-295X.108.3.550>
- van den Wildenberg, W. P. M., Wylie, S. A., Forstmann, B. U., Burle, B., Hasbroucq, T., & Ridderinkhof, K. R. (2010). To head or to heed? Beyond the surface of selective action inhibition: A review. *Frontiers in Human Neuroscience*, *4*, Article 222. <https://doi.org/10.3389/fnhum.2010.00222>
- Vogel, T. A., Savelson, Z. M., Otto, A. R., & Roy, M. (2020). Forced choices reveal a trade-off between cognitive effort and physical pain. *eLife*, *9*, Article e59410. <https://doi.org/10.7554/eLife.59410>
- Weichart, E. R., Turner, B. M., & Sederberg, P. B. (2020). A model of dynamic, within-trial conflict resolution for decision making. *Psychological Review*, *127*(5), 749–777. <https://doi.org/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. <https://doi.org/10.1126/science.aaz5891>
- 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. <https://doi.org/10.1016/j.cogpsych.2011.08.001>
- 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. <https://doi.org/10.3758/s13423-017-1271-2>
- Wichmann, F. A., & Hill, N. J. (2001). The psychometric function: I. Fitting, sampling, and goodness of fit. *Perception & Psychophysics*, *63*(8), 1293–1313. <https://doi.org/10.3758/BF03194544>
- Wong, K.-F., & Wang, X.-J. (2006). A recurrent network mechanism of time integration in perceptual decisions. *The Journal of Neuroscience*, *26*(4), 1314–1328. <https://doi.org/10.1523/JNEUROSCI.3733-05.2006>
- Wöstmann, M., Alavash, M., & Obleser, J. (2019). Alpha oscillations in the human brain implement distractor suppression independent of target selection. *The Journal of Neuroscience*, *39*(49), 9797–9805. <https://doi.org/10.1523/JNEUROSCI.1954-19.2019>
- Wöstmann, M., Störmer, V. S., Obleser, J., Addleman, D. A., Andersen, S., Gaspelin, N., Geng, J., Luck, S. J., Noonan, M., Slagter, H. A., & Theeuwes, J. (2021). *Ten simple rules to study distractor suppression*. PsyArXiv. <https://doi.org/10.31234/osf.io/vu2k3>
- Wylie, S. A., Ridderinkhof, K. R., Elias, W. J., Frysinger, R. C., Bashore, T. R., Downs, K. E., van Wouwe, N. C., & van den Wildenberg, W. P. M. (2010). Subthalamic nucleus stimulation influences expression and suppression of impulsive behaviour in Parkinson's disease. *Brain: A Journal of Neurology*, *133*(12), 3611–3624. <https://doi.org/10.1093/brain/awq239>
- Yee, D. M., & Braver, T. S. (2018). Interactions of motivation and cognitive control. *Current Opinion in Behavioral Sciences*, *19*, 83–90. <https://doi.org/10.1016/j.cobeha.2017.11.009>
- 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. <https://doi.org/10.1037/0033-295X.111.4.931>
- Yeung, N., & Monsell, S. (2003). Switching between tasks of unequal familiarity: The role of stimulus-attribute and response-set selection. *Journal of Experimental Psychology: Human Perception and Performance*, *29*(2), 455–469. <https://doi.org/10.1037/0096-1523.29.2.455>
- 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. <https://doi.org/10.1037/a0013553>
- 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. <https://doi.org/10.1111/j.1420-9101.2011.02297.x>
- Zhang, J., & Kornblum, S. (1997). Distributional analysis and De Jong, Liang, and Lauber's (1994) dual-process model of the Simon effect. *Journal of Experimental Psychology: Human Perception and Performance*, *23*(5), 1543–1551. <https://doi.org/10.1037/0096-1523.23.5.1543>

Received August 10, 2022

Revision received May 18, 2023

Accepted June 10, 2023 ■