Control Adaptation through Selective Suppression of Multidimensional Distractors

Davide Gheza (gheza@wustl.edu)

Thea R. Zalabak (thea@wustl.edu)

Wouter Kool (wkool@wustl.edu)

Department of Psychological and Brain Sciences
Washington University in St. Louis
St. Louis, MO, USA

Abstract

Humans manage multiple conflicting sources of information. However, models of cognitive control assume one source of interference and do not explain how we handle multiple distractors. In our multi-dimensional task-set interference paradigm, individuals manage distraction from three independent dimensions. Experiment 1 suggests that people use prior conflict from each dimension to selectively modulate their gain. A neural network, measuring multivariate conflict as energy within each dimension's pathway, captures this effect. Representational similarity analyses of human EEG (Experiment 2) confirmed the selective suppression of distractor representations. These results reveal the striking human ability to simultaneously adjust attention to multiple sources of information. Model predictions converge with recent work suggesting that neural conflict signals emerge from the integration of diverse task variables in medial prefrontal cortex.

Keywords: cognitive control; distractor suppression; neural network model; EEG; RSA

Introduction

The flexible adjustment of attention as a function of past interference is a hallmark feature of cognitive control. Computational models successfully capture such dynamics when people face one source of distraction (Alexander, 2022; Botvinick et al., 2001; Verguts & Notebaert, 2008), but it remains unclear how they orchestrate attention over multiple information streams. Here, we approach this question with a novel multi-dimensional task-set interference (MULTI) paradigm, in which participants need to manage three simultaneous distractors. We first show that people adjust to multidimensional conflict by modulating attention in a distractor-specific way. Next, we report a neural network model that explains this effect through distractor-specific conflict adaptation, with multivariate conflict measured as Hopfield energy in each perceptual pathway. Finally, an EEG study using representational similarity analyses (RSA) shows evidence consistent with dimension-specific attention modulation when distractors carry predictable levels of conflict.

Experiment 1

In the MULTI task, participants have to attend to one of four task dimensions, indicated by a letter cue, with the remaining dimensions acting as distractors (Fig. 1A). For each cued dimension, participants have to find a target feature that is present in one of two stimuli. Target features are randomly chosen at the start of the experiment (Figure 1A; in the example, blue color, oval shape, dashed edge, downward motion). The cued task dimension switches over trials in pseudo-randomized sequences, rendering each dimension periodically relevant. From this design, it follows that task-set interference parametrically varies over a discrete congruency scale from 0 to 3. By analyzing how prior congruency in one dimension affects susceptibility to a current non-cued dimension, we can test whether conflict adaptation generalizes across dimensions, or only adjusts within-dimension attention.

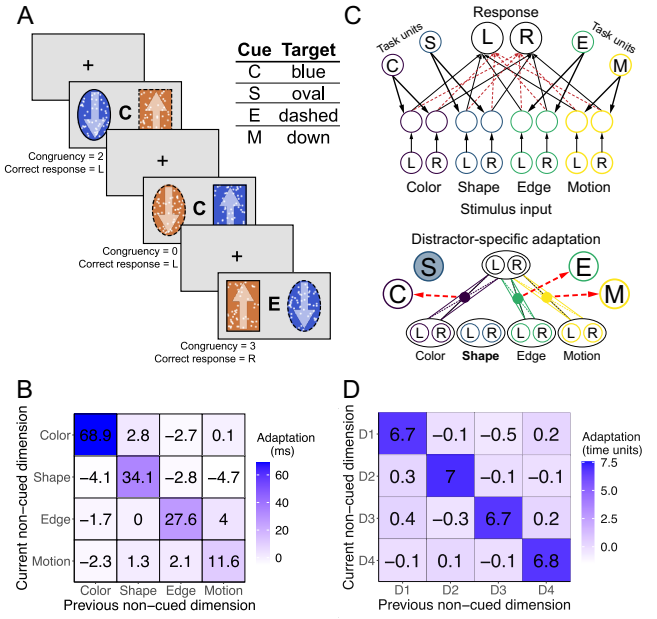

Figure 1. (A) Trial sequence of the MULTI task. (B) Dimension-specific adaptation. (C) Neural network architecture (top), and distractor-specific adaptation mechanism (bottom). (D) Simulation results.

Behavioral and modeling results

Data from Experiment 1 *(N*=104) showed that distractor dimensions produce interference simultaneously, and that humans *parametrically* configure attentional control based on prior congruency levels. Specifically, people became less susceptible to distractors after low-congruency trials (i.e., high conflict; Botvinick et al., 2001), a phenomenon known as the congruency sequence effect (CSE; (Gratton et al., 1992). Critically, we observed dimension-specific CSEs (Figure 1B). When comparing each within-dimension adaptation effect to an average across-dimension adaptation effect, we observed strong evidence for larger within-dimension compared to across-dimension adaptation effects (BFs ≥ 217). We also found some evidence against across-dimension adaptation (range BFs = 2 - 9). This result, replicated in several experiments (Gheza & Kool, 2023), suggests that conflict from a given dimension only affects its own processing.

A neural network model of our task (Fig. 1C), based on prior models of classic interference paradigms

(Cohen et al., 1990; Ritz & Shenhav, 2022), captures this effect. The model processes stimulus information through four feed-forward pathways, biased by individual task nodes. In classic models, conflict is monitored at the response level, where all stimulus features are compressed into two values (e.g., Botvinick et al. 2001). This is incompatible with our results, because conflict adaptation tracks the source of conflict. Instead, we found that dimension-specific CSEs were only captured when including as many conflict monitoring units as distractors (Fig. 1D).

This model measures conflict as Hopfield energy, the coactivation of dimension-specific intermediate units and response units, weighted by their connections:

$$conflict = -\sum_{i,j:i<j} w_{ij} a_i a_j$$

The pathway energy then determines how much control is applied to the corresponding task unit, so that conflict from a given dimension results in its suppression:

$$control_{d,i+1} = \lambda \cdot control_{d,i} + (1-\lambda)(\alpha \cdot conflict_{d,i} + \beta)$$

This modeling work confirms that conflict adaptation requires independent, parallel, conflict-control loops. Thus, we predicted that neural representations of distractors, and not cued dimensions, would shift as function of previous interference (Ebitz et al., 2020).

Experiment 2

In a follow-up EEG version of the MULTI task, we introduced a task-specific proportion congruency (PC) manipulation (Bugg & Crump, 2012), with stable relationships between cued task and experienced conflict. At each point on time, one task was mostly incongruent (MI; 25% probability that any distractor is congruent), one mostly congruent (MC; 75%), and the other two unbiased (Neu; 50%). To access the content of neural representations, we decoded information about dimensions' representations in a time-resolved manner from the EEG signal (*N*=26) using trial-level RSA (Kikumoto & Mayr, 2020). First, we obtained information about the similarity of multivariate neural signals by decoding each combination of the four cues and the locations of their instructed features (left or right), for each time point within each trial. Second, we performed RSAs using the resulting classification profiles as a measure of class similarity. These profiles were simultaneously regressed onto three model vectors, corresponding to representations of the dimension's *rule* (i.e., cue), *side* of the instructed feature (i.e., spatial attention), and their *conjunction* (i.e., feature identity). By using model vectors corresponding to either the cued or each non-cued dimension, we could disentangle dimension-specific representations and measures attention to both target and distractors as it unfolds over time.

Results

Reaction time and accuracy data revealed that humans adapted to PC effects (Figure 2A). For mostly incongruent tasks, the congruency effect was weakened (shallower slopes), reflecting less susceptibility to non-cued dimensions. When tasks had mostly congruent distractors, congruency effects were larger (steeper slopes). RSA confirmed our predictions. The control adaptation elicited by the PC manipulation did not significantly affect representations of the cued dimension (Figure 2B). However, the identity of non-cued features emerged in the response window (2C, SR *conjunction*), compatibly with slips of attention towards distractors. Crucially, this effect was absent for MI blocks, suggesting that narrower attentional control prevented such slips. Additional support for distractor-based adaptation came from correlations between the effect sizes of non-cued rule suppression and of congruency effects (2D), suggesting that if participants were more likely to represent a wrong rule, the more they were susceptible to interference.

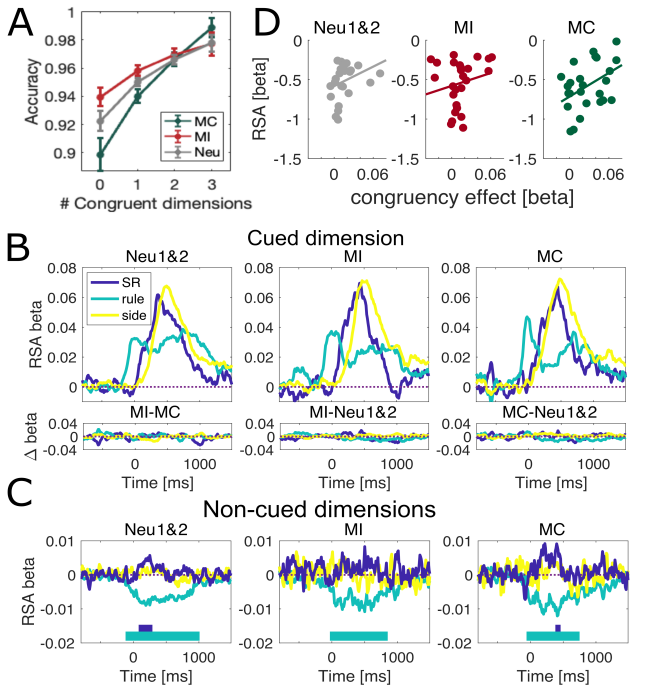

Figure 2. (A) Effect of PC on behavior. (B) Representation of the cued dimension. (C) Averaged representations of the three non-cued dimensions. Bars indicate significant clusters. (D) Brain-behavior correlations.

Conclusion

Our work shows humans adapt control in a distractor-specific way. How can this be implemented neurally? Mixed-selective neuronal populations (Fu et al., 2022; Fusi et al., 2016) may encode multivariate conflict representations and overcome conflict "locally" by orthogonalizing interfering dimensions.

Acknowledgments

WK was supported by ONR/DoD N00014-23-1-2792.

References

Alexander, W. H. (2022). Cognitive control strategies derive from dimension reliability. PsyArXiv. https://doi.org/10.31234/osf.io/a9bmh

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

Bugg, J. M., & Crump, M. (2012). In Support of a Distinction between Voluntary and Stimulus-Driven Control: A Review of the Literature on Proportion Congruent Effects. Frontiers in Psychology, 3. https://www.frontiersin.org/articles/10.3389/fpsyg.2012.00367

Cohen, J. D., McClelland, J. L., & Dunbar, K. (1990). On the Control of Automatic Processes: A Parallel Distributed Processing Account of the Stroop Effect. Psychological Review, 97((3)), 332–361.

Ebitz, R. B., Smith, E., Horga, G., Schevon, C., Yates, M., McKhann, G., Botvinick, M., Sheth, S., & Hayden, B. (2020). Human dorsal anterior cingulate neurons signal conflict by amplifying task-relevant information. https://doi.org/10.1101/2020.03.14.991745

Fu, Z., Beam, D., Chung, J. M., Reed, C. M., Mamelak, A. N., Adolphs, R., & Rutishauser, U. (2022). The geometry of domain-general performance monitoring in the human medial frontal cortex. Science, 376(6593), eabm9922. https://doi.org/10.1126/science.abm9922

Fusi, S., Miller, E. K., & Rigotti, M. (2016). Why neurons mix: High dimensionality for higher cognition. Current Opinion in Neurobiology, 37, 66–74. https://doi.org/10.1016/j.conb.2016.01.010

Gheza, D., & Kool, W. (2023). Distractor-specific control adaptation in multidimensional environments (p. 2023.09.04.556248). bioRxiv. https://doi.org/10.1101/2023.09.04.556248

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, 480–506. https://doi.org/10.1037/0096-3445.121.4.480

Kikumoto, A., & Mayr, U. (2020). Conjunctive representations that integrate stimuli, responses, and rules are critical for action selection. Proceedings of the National Academy of Sciences of the United States of America, 117(19), 10603–10608. https://doi.org/10.1073/pnas.1922166117

Ritz, H., & Shenhav, A. (2022). Humans reconfigure target and distractor processing to address distinct task demands. bioRxiv, 2021.09.08.459546. https://doi.org/10.1101/2021.09.08.459546

Verguts, T., & Notebaert, W. (2008). Hebbian learning of cognitive control: Dealing with specific and nonspecific adaptation. Psychological Review, 115(2), 518–525. https://doi.org/10.1037/0033-295X.115.2.518