Précis to **Multivariate Cognitive Control**

Harrison Ritz
Brown University, 2022
Readers: Amitai Shenhav, Michael J. Frank, & Jörn Diedrichsen

Overview

How we control our body and how we control our minds have long been studied separately (Broadbent, 1977). These parallel streams of inquiry have proceeded with minimal interaction, despite fundamental computational similarities across domains. In *motor control*, researchers study how we use representations of our bodies (Kawato, 1999; Shadmehr & Mussa-Ivaldi, 1994) to optimize our movement (Todorov & Jordan, 2002; Uno et al., 1989). In *cognitive control*, researchers focus on analogous problems, studying how we represent tasks (Kornblum et al., 1990; Musslick & Cohen, 2021) to optimize task processing (Shenhav et al., 2013; Silvetti et al., 2018). Cross-domain studies have suggested that these domain share features like automaticity (Haith & Krakauer, 2013) or the minimization of costly effort (Chong et al., 2017; Manohar et al., 2015). However, an integrative account of motor and cognitive control has been largely neglected (Ritz et al., 2020).

I argue that the core theoretical link between motor and cognitive domains is their relationship to normative models from engineering control theory. Since the early days of post-war cybernetics, feedback control has been proposed to underly our ability to pursue our goals (Ashby, 1954; Powers, 1973; Rosenblueth et al., 1943). Modern models of 'optimal control' cast feedback as an optimization problem, such as finding the control policy that best achieves a goal using the least amount of work. The optimal (feedback) control approach has been central to motor control research for decades (Chow & Jacobson, 1971; Flash & Hogan, 1985; Nubar & Contini, 1961; Uno et al., 1989), with a particular reliance on a class of algorithms with tractable analytic solutions (Loeb et al., 1990; Todorov & Jordan, 2002). Cognitive control shares many features with motor control that have been explored through a control-theoretic lens, most notably a central role for optimization (Shenhav et al., 2013). While optimal control algorithms have begun to influence models of goal-directed neural dynamics (Athalye et al., 2019; Tang & Bassett, 2018), there has been limited appreciation of the broader implications of the optimal control framework for theories of cognitive control. Here, I explore whether the regulation of body and mind are intrinsically connected through analogous control algorithms.

A hallmark of optimal motor control is our capacity for coordinated movement: dozens of joints and muscles must work together to pick up rice with a pair of chopsticks. This coordination is a central component of optimal control theories, in which multiple bodily effectors are bound together through how they achieve a common objective function (Diedrichsen et al., 2010; Todorov, 2004). Despite the important role of coordination in optimal theories of motor control, little is known about how people coordinate multiple forms of cognition. This thesis provides an initial bridge between motor and cognitive domains by examining how the brain coordinates multiple *cognitive* effectors.

In **Chapter 1**, I explore how people regulate multiple streams of information using a novel behavioral paradigm that can 'tag' sensitivity to different stimulus dimensions. Finding strong evidence for multiple independent control processes, I develop a neural network model to explain how the brain could use feedback control to dynamically regulate different processing channels. In **Chapter 2**, I explore the neural representations that allow people to regulate multiple cognitive processes. Using novel multivoxel fMRI analyses, I find that cognitive control regions independently encode multiple information streams, but only when they are needed for control. In **Chapter 3**, I outline the broader implications of multiple cognitive effectors in an integrative theory and review. I highlight a core challenge to coordinated control arising from multiple realizability, and how solutions to similar problems in motor control may provide a unified account of physical and mental effort costs.

Together, this work expands our understanding of dexterous human cognition, setting the stage for domain-general theories of how we orchestrate our thoughts and action. Combining empirical, analytic, and theoretical innovations with cross-disciplinary influences from motor control, systems neuroscience, and engineering control theory, this thesis aims to advance our understanding of how we coordinate our mind to reach our goals.

Chapter 1: *Humans reconfigure target and distractor processing to address distinct task demands*

Under revision at *Psychological Review* [preprint DOI: 10.1101/2021.09.08.459546]

In this chapter, I explore how people regulate multiple streams of information processing during decision-making. A topic of long-standing interest in cognitive control is how the brain controls task-relevant information (targets) and task-conflicting information (distractors; Egner, 2008; Lindsay & Jacoby, 1994). A major barrier to answering this question has been that existing methods for testing cognitive control can only poorly dissociate target and distractor processing. Here, I develop a novel task that can 'tag' sensitivity to different features, finding that people can independently and dynamically regulate target and distractor processing. Moreover, I find that these control processes can be reproduced in a neural network model of top-down attention. Together, these results suggest that cognitive control may deploy multivariate feedback control over information processing.

A novel task reveals independent sensitivity to target and distractor information

The Parametric Attentional Control Task (PACT) combines empirical traditions from perceptual decision-making (Kayser et al., 2010; Mante et al., 2013) and inhibitory control

(Danielmeier et al., 2011) to independently measure target and distractor processing. Participants (N=156; ~930 trials each) performed a random dot color-motion task, viewing an array of colored dots that moved left or right (Figure 1A). In the critical condition, participants reported with a key press which color was in the majority. Across trials, I parametrically varied the color coherence (% dots of the same color) to manipulate target discriminability. Independently, I also varied motion congruence (% dots moving in the same direction as the color response) to manipulate target-distractor compatibility (Figure 1B).

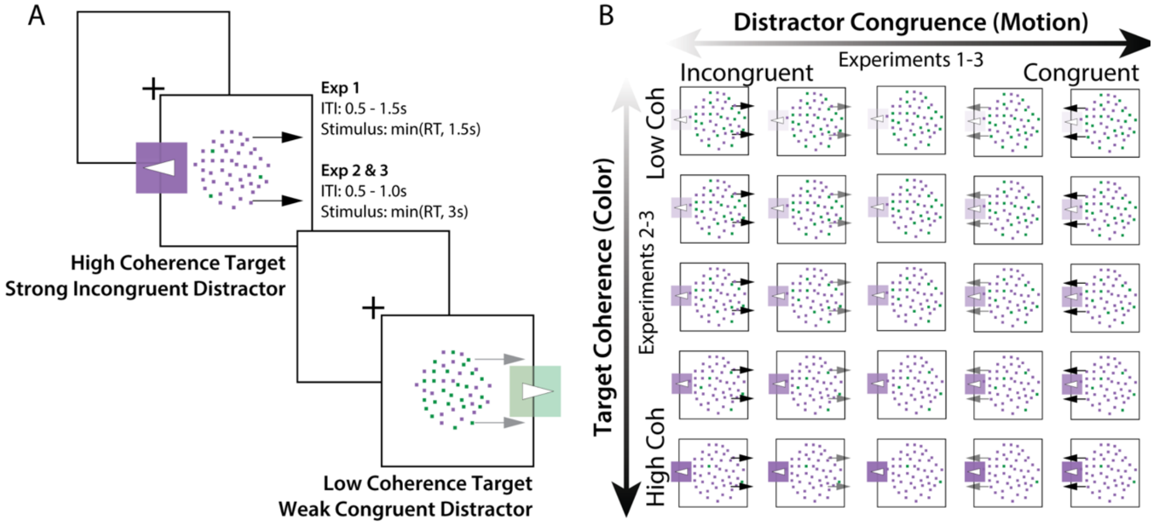


Figure 1. Parametric Attentional Control Task. A) *Participants saw moving colored dots. They responded based on which color was in the majority, ignoring the color.* ***B)*** *Color coherence (% dots same color; y axis) and distractor congruence (% dots moving the same direction as color response; x axis) were parametrically and independently manipulated across trials.*

Participants' performance parametrically improved when they had higher coherence targets (Figure 2A) and higher congruence distractors (Figure 2B). Target and distractor sensitivity (i.e., the slope of the psychometric function) was uncorrelated across participants, and targets and distractors difficulty had little or no interaction in the single-trial prediction of performance. These distinct sensitivity profiles suggest that people independently process these features.

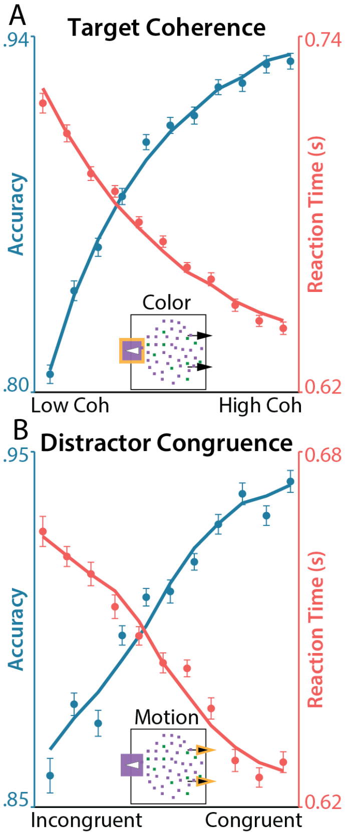


Figure 2. Target and distractor sensitivity. A) *Participants were faster (red) and more accurate (blue) when targets had higher coherence.* ***B)*** *Participants were faster (red) and more accurate (blue) when distractors were more congruent with targets. Dots reflect behavior and lines reflect regression predictions.*

Incentives and conflict dissociate adjustments to target and distractor sensitivity

I explored whether target and distractor processing were under independent control by testing whether standard triggers for cognitive control differentially influenced feature sensitivity. Participants' behavior supported such a dissociation: whereas previous conflict primarily suppressed distractor sensitivity (Figure 3A), performance incentives primarily enhanced target sensitivity (Figure 3B). This dissociation demonstrates that adjustments to target sensitivity are not necessarily linked to adjustments in distractor sensitivity (Figure 3C), supporting the independent control of multiple streams of information processing.

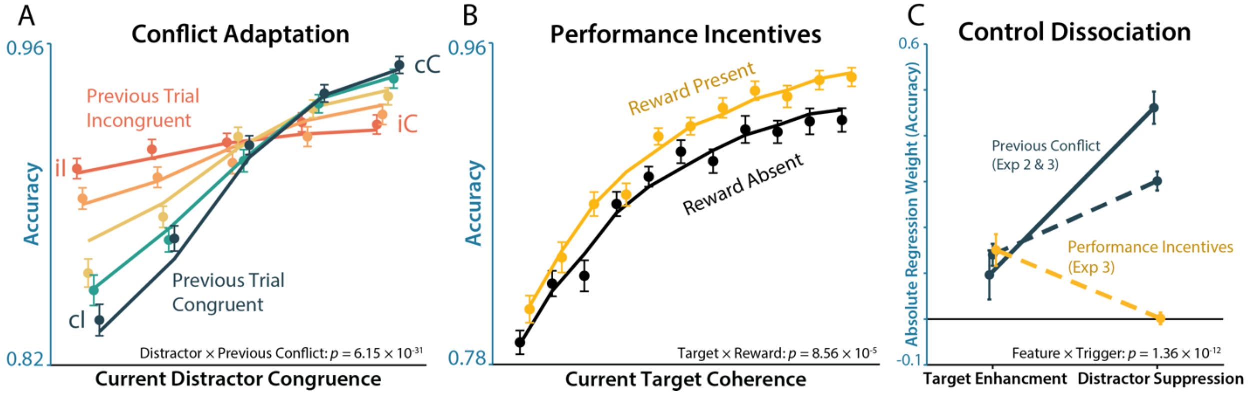


Figure 3. Feature-dependent control. A) *Participants were less sensitive to distractors after high conflict trials.* ***B)*** *Participants were more sensitive to targets on incentivized runs. Dots reflect behavior and lines reflect regression predictions.* ***C)*** *Conflict and incentives differentially influenced target and distractor sensitivity.*

These adjustments may influence steady-state target and distractor sensitivity, but we can gain further insight into the control mechanism by examining the within-trial dynamics of target enhancement and distractor suppression. Examining how control changes as a function of reaction time (De Jong et al., 1994; van den Wildenberg et al., 2010), I found that over time participants increased sensitivity to targets and decreased sensitivity to distractors. Extensive evidence accumulation simulations confirmed that this analysis was not biased by conditioning on reaction times, instead being a specific prediction of within-trial changes to attention (White et al., 2018). When I looked at how conflict and incentives influenced these dynamics, I found that they changed the earliest sensitivity to target and distractor processing, again showing feature specificity (Figure 4). Interestingly, despite changes to this initial feature processing, participants' sensitivity appeared to asymptote at a similar level.

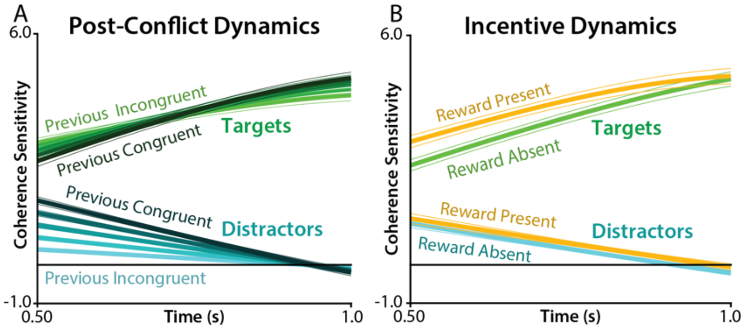


Figure 4. Within-trial dynamics. ***A)*** *Previous conflict changed early sensitivity to distractors, which decreased to the same asymptotic level.* ***B)*** *Incentives increased initial sensitivity to targets, which increased to the same asymptotic level. Lines reflect predicted coherence sensitivity from regression analyses on choice.*

An attractor neural network model captures dissociable within-trial attentional dynamics. To provide a process-level explanation for how people dynamically control feature sensitivity, I developed a novel neural network model that could reproduce participants' core behavioral signatures of cognitive control (Figure 5). Building on a classic neural network model of top-down attention (Cohen et al., 1990), I incorporated an attractor network that dynamically adjusted feature gain over time (Figure 5B; cf. Musslick et al., 2019; Steyvers et al., 2019). I found that a single parameterization of this model could qualitatively reproduce participants' feature sensitivity and within-trial dynamics. Moreover, I found that I could parsimoniously capture conflict- and incentive-dependent attentional adjustments by only changing the initial conditions of this attractor network (Figure 5D). This model provides an explicit computational account for how a neural feedback control system could regulate multiple control targets, dynamically enhancing task-relevant channels and suppressing task-irrelevant channels.

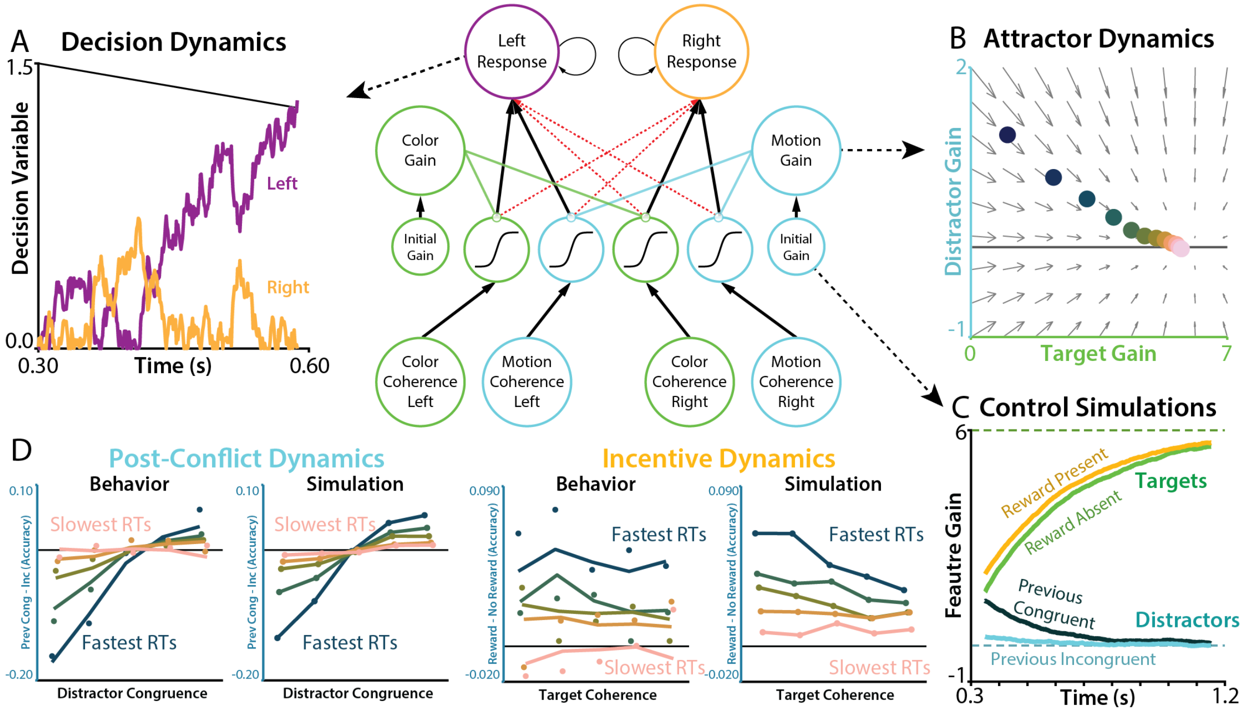


*Figure 5. **Attractor neural network model of dynamic attention. A)** Decision evidence is accumulated over time until it reaches a collapsing bound **B)** The hidden layer gains on targets and distractors were dynamically controlled through an attractor network, defined through exponential decay into a fixed point **C)** I modeled conflict and incentive effects as adjustments to the initial conditions of these gain dynamics **D)** Behavioral signatures of conflict- and incentive-dependent attentional dynamics were qualitatively similar to behavior generated from the neural network model.*

Together, Chapter 1 provides an empirical foundation and theoretical scaffolding for how our cognitive control system may dynamically regulate attention across multiple streams of information processing. These results demonstrate that the brain can control multiple processes when necessary and to the extent the task allows (e.g., when features are separable enough as to allow for independent regulation; cf. Egner, 2008). In a different task, the brain may use a different control strategy, but we show that it has the capacity for multivariate control. Speculatively, the specific relationships between conflict to distractors and incentives to targets may reflect an inferential process that reconfigures processing depend on the feature-outcome relationship (i.e., distractors cause conflict and targets cause rewards). Consistent with this account, normative modeling has proposed that prior beliefs about feature utility should bias the initial sensitivity to those features (Yu et al., 2009). Inspired by these behavioral findings, I next sought to better understand the neural mechanisms predicted by our process model of multivariate control.

Chapter 2: *Orthogonal neural encoding of targets and distractors supports attentional control*

Under revision at *Nature Human Behavior* [preprint DOI: 10.1101/2022.12.01.518771]

In Chapter 1, I found behavioral evidence that people can independently control their processing of multiple information streams, enhancing target sensitivity and suppressing distractor sensitivity in response to different task demands. These findings make strong predictions for how the brain must encode different sources of task information. To independently regulate target and distractor processing, the brain's cognitive control system must have independent representations for feature-dependent monitoring and adjustment.

In the process model developed in Chapter 1, target and distractor gains were independently controlled using separate, non-interacting units in our attractor network. Our core hypothesis was that control operates similarly in the brain, using independent representations of target and distractor information distributed across multiple functional units (i.e., independent 'encoding subspaces'; Mante et al., 2013). I hypothesized this would occur both for monitoring feature-specific difficulty and gating feature-specific processing. I tested these hypotheses using fMRI, which uniquely allows us to noninvasively measure activation patterns at high spatial resolution across the whole brain. Developing a novel multivoxel analysis, I explored how targets and distractors are represented across critical nodes for cognitive control.

Distinct encoding of target and distractor difficulty in cingulate cortex

Human participants (N=29; 900 trials each) performed the same task as Chapter 1 during fMRI, again replicating our core behavior patterns. A large literature links activity in the dorsal anterior cingulate cortex (dACC) to monitoring the need for control (Cavanagh & Frank, 2014; Nee et al., 2007; Shenhav et al., 2013), but whether these signals reflect global factors (e.g., error likelihood) or more specific factors (e.g., assigning credit to different sources of difficulty) remains controversial (Ebitz et al., 2020; Kragel et al., 2018). Looking within dACC, I found support for feature-specific monitoring, with target difficulty (low coherence targets) and distractor difficulty (incongruent distractors) encoded in distinct locations of dACC along the rostrocaudal axis (Figure 6). Target and distractor encoding patterns were uncorrelated, confirming that dACC could track feature-specific task demands.

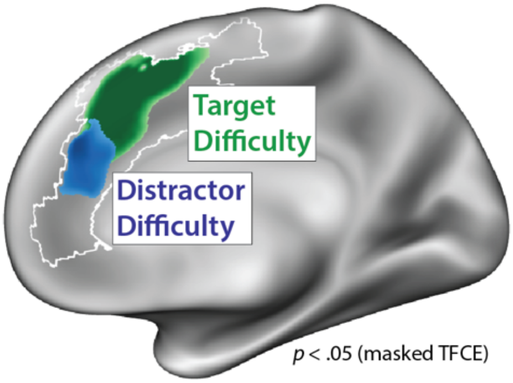


Figure 6. Difficulty encoding within dACC. *Within an a priori mask of dACC (white outline), I found distinct parametric signals reflecting lower target coherence (green) and more incongruent distractors (blue). Note that figures in this précis chapter have been updated to match the current publication and may deviate from thesis figures.*

Orthogonal encoding of target and distractor gain in parietal cortex

I next examined how the neural control system might regulate target and distractor gain, analogous to the attractor network in Chapter 1. I hypothesized that the brain independently encodes the amount of information in each dimension ('absolute coherence'), and that this would occur in critical regions for top-down attention.

To measure whether target and distractor coherence was encoded independently, I developed a novel multivariate fMRI analysis called 'Encoding Geometry Analysis' (EGA), combining the strengths of multivariate encoding analyses and representational similarity analyses (Figure 7A). In brief, I first estimate each voxel's coherence encoding using regression. Next, in cortical parcels where target and distractor coherence are both reliably encoded, I estimate how patterns of regression weights (i.e., encoding subspaces) are aligned using cross-validated correlation. Finally, I use Bayesian tests at the group level to estimate whether it's more likely that target and distractor coherence subspaces are correlated (e.g., spatial salience) or orthogonal (e.g., feature-specific priority).

I found that the neural encoding of target and distractor coherence was jointly reliable throughout visual and parietal cortex. Looking at the alignment between target and distractor encoding, I found that the medial parietal cortex had correlated encoding of targets and distractors. In contrast, the intraparietal sulcus (IPS; a core node for cognitive control), had orthogonal encoding of target and distractor coherence (Figure 7B-C). At the ROI level, this orthogonal relationship between targets and distractors was clear in a low-dimensional embedding of coherence representations (Figure 7C).

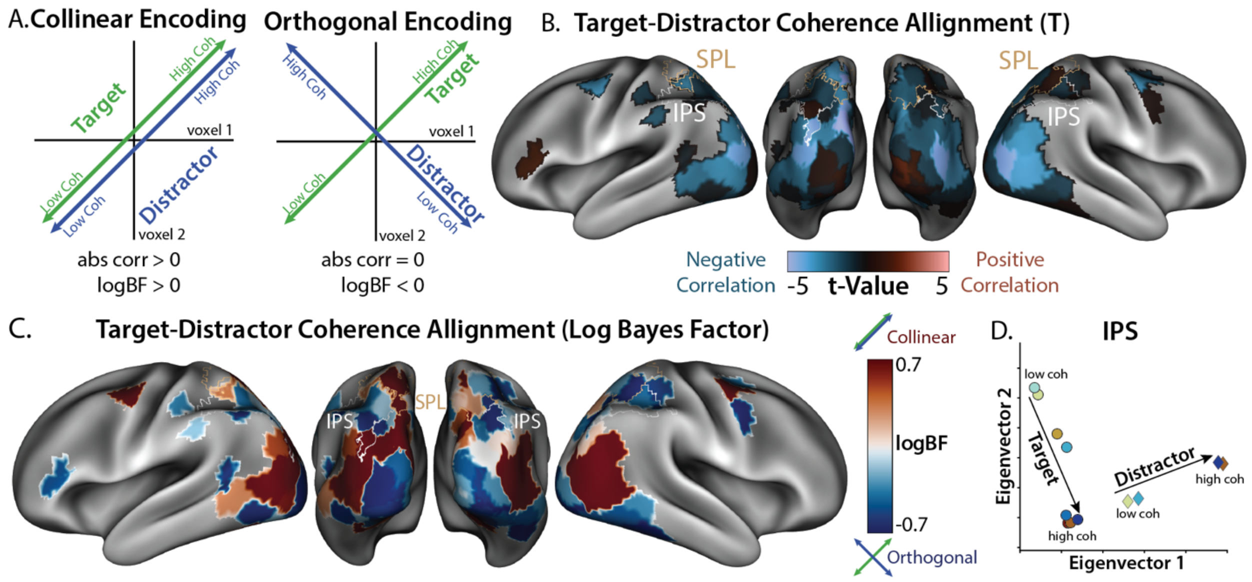


Figure 7. Independent encoding of target and distractor coherence. A) *Encoding Geometry Analysis (EGA) tests the alignment between encoding profiles, such as whether they are correlated (left) or orthogonal (right).* ***B)*** *Visual and parietal cortex encoded both target and distractor coherence, and in intraparietal sulcus (IPS) these encoding profiles had a correlation near zero.* ***C)*** *Bayesian tests of orthogonality supported independent feature encoding in IPS.* ***D)*** *A low-dimensional embedding of condition-specific representations supported independent axes for target and distractor coherence. Colors reflect coherence (light/dark) and response (blue/brown) conditions, with arrows visualizing the axis of coherence encoding.*

Target and distractor gain depend on difficulty and predict performance

Target and distractor coherence was orthogonally encoded in intraparietal cortex, a potential cortical mechanism for independent feature processing in a well-established system for goal-directed attention (Gottlieb, 2014; Posner & Petersen, 1990; Yantis & Serences, 2003). I next sought to validate whether these representations reflected top-down attention, or bottom-up salience. Our experiment had a built-in test for this: on alternating runs, participants performed an easier version of the task, responding to the motion dimension instead of the color dimension. This alternative task has precisely matched stimuli and responses, but places much lower demands on cognitive control (as was also found in Chapter 1).

When participants performed the easier 'Attend-Motion' version of our task, I found similarly strong encoding of response information, with aligned encoding across tasks (Flesch et al., 2022; Mante et al., 2013). In stark contrast, during this easier task I did not find any encoding of target and distractor coherence, dissociating putative measures of decision making and control (Figure 8). This suggests that coherence representations depend on top-down attention (and not bottom-up salience), supporting a mechanism for independently regulating multiple forms of information process.

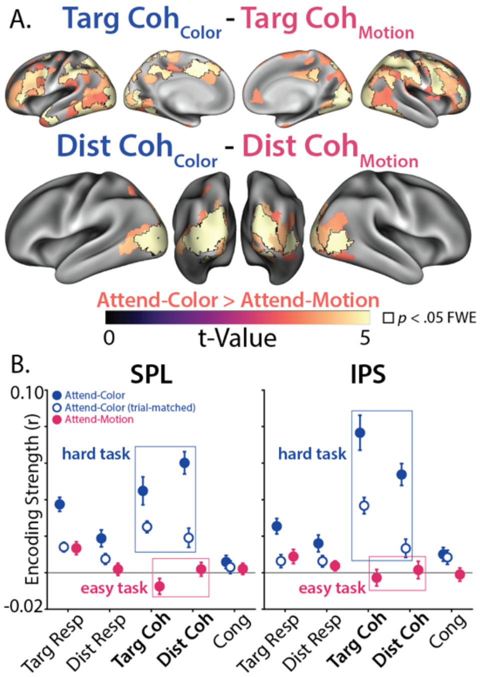


Figure 8. Stronger coherence encoding during control-demanding tasks. A) *Target coherence (top) and distractor coherence (bottom) were encoded more reliably in the difficult task (Attend-Color; Navy) than the easy task (Attend-Motion; Purple), despite matched stimuli and responses.* ***B)*** *Feature encoding strength compared across tasks in key ROIs.*

Finally, I confirmed that coherence representations were related to task performance. I tested whether coherence subspaces were aligned with performance subspaces, inspired by recent methods in systems neuroscience (Stringer et al., 2019). I found that target coherence aligned with better performance (faster RT and better accuracy), and distractor coherence aligned with poorer performance (slower RT and poorer accuracy), consistent with coherence representations reflecting the sensitivity to each feature (Figure 9).

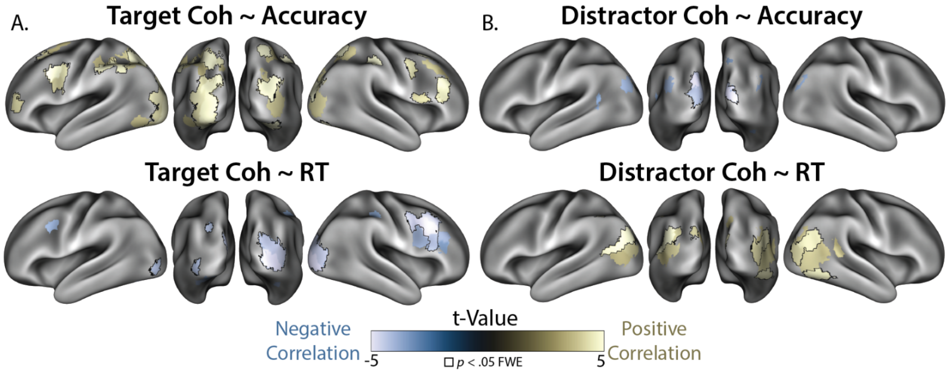


Figure 9. Alignment between coherence and performance encoding. A) *Target coherence encoding aligned with better performance.* ***B)*** *Distractor coherence encoding aligned with poorer performance.*

In sum, I found that the neural encoding of task information during cognitive control had the independent feature representations that were predicted by our computational model. This principle of independent control representations is fundamental, reflected in the classical cybernetics theorem of ‘Requisite Variety’: a good controller must have the complexity of the process it aims to regulate (Ashby, 1961). This fMRI project supports our evidence in Chapter 1 for multivariate control over information processing, while providing neural mechanisms in the form of macro-level circuits (monitoring and regulation across the frontoparietal network) and local representations (orthogonal encoding subspaces). These neural mechanisms were elicited through novel extensions of recent methods in system neuroscience (Ebitz et al., 2020; Mante et al., 2013; Rust & Cohen, 2022), providing a principled test of neuro-cognitive theory.

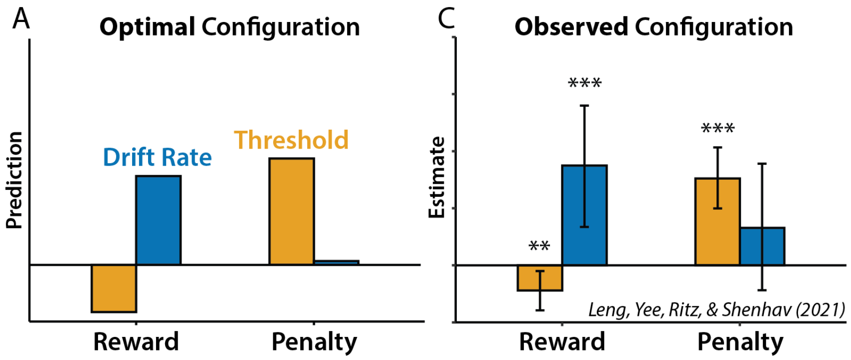
Chapter 3: *Cognitive control as a multivariate optimization problem*

Ritz, Leng, & Shenhav (2022); *Journal of Cognitive Neuroscience*

In Chapters 1 and 2, I provided new evidence for how people independently control multiple streams of information processing over time (Chapter 1), and a neural mechanism for how independent control occurs in the human brain (Chapter 2). In Chapter 3, I examined the theoretical implications of multivariate cognitive control, and how this might inform major outstanding questions in cognitive control relating to the origins of mental effort costs. Through an integrative theoretical review, I outlined how our understanding of motor coordination can inform both multivariate cognitive control and the origins of mental effort.

The many-to-many mapping between cognition and goals makes cognitive control ill-posed

A growing literature demonstrates that people take a multi-faceted approach to cognitive control. For example, Multiple Object Tracking classically demonstrates that people can track multiple simultaneous objects as they move in space (Pylyshyn & Storm, 1988). People also deploy a variety of overlapping cognitive strategies in response to errors (Danielmeier & Ullsperger, 2011), conflict (Egner, 2008), and incentives (Leng et al., 2021). The experiments in this thesis contribute novel process-oriented insights to this literature. Despite this complexity, however, peoples' control strategies are often reasonable. In recent work alongside this thesis, I found that people jointly configure attention and response caution for different incentives in a way that closely mirrors the optimal multivariate policy (Figure 10; Leng et al., 2021).



Leng, Yee, Ritz, & Shenhav (2021)

Figure 10. Optimal multivariate control. A) *Reward-rate optimal adjustments to drift diffusion parameters in response to greater rewards for correct responses or greater penalties for errors.* ***B)*** *Estimated adjustments to drift diffusion parameters in response to greater rewards and penalties.*

I propose that these overlapping many-to-many relationships between goals and cognitive effectors highlights a fundamental challenge to goal-directed cognition. Our brain and mind are often much more complex than many of the tasks we face (especially in the laboratory). More degrees of freedom in the brain than in a task makes planning cognitive control fundamentally *ill-posed*, meaning that there is an unstable or non-unique solution to the optimization problem (Hadamard, 1902). This is present at the cognitive level (e.g., parameter degeneracy in evidence accumulation; Figure 11), but is even more severe at the neural level (e.g., multiple realizability; Krakauer et al., 2017; Putnam, 1967). If cognitive control is ill-posed, how can the brain effectively plan how to think?

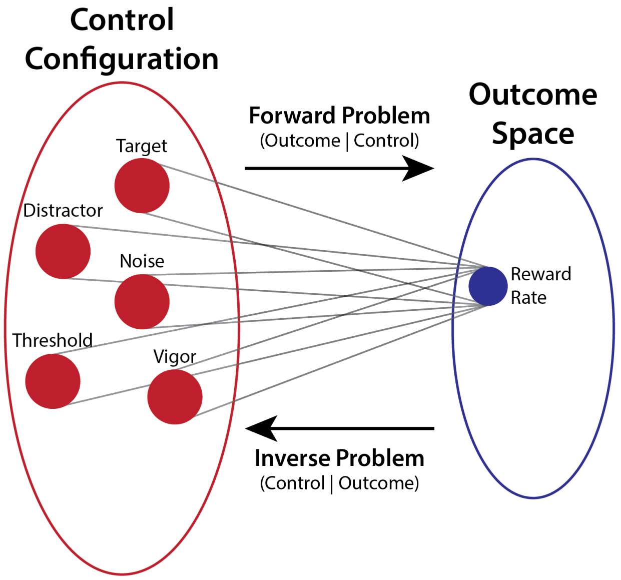


Figure 11. Ill-posed cognitive control. *On the left, a high-dimensional control space, here the parameters of an evidence accumulation model. On the right, a low-dimensional outcome space, here an aggregation of good performance (correct responses per second). Several control parameters are degenerate (have similar influence on performance), leading to an ill-posed inverse problem (the mapping from desired outcomes to control).*

Effort as a solution to ill-posed control

I suggest that a possible solution may come from motor control, a goal-directed behavior with fundamental similarities to cognitive control. The ill-posed nature of motor planning has been a focus of the field for almost a century (Bernstein, 1935), reflecting the fact that there are more degrees of freedom in the skeletomotor system than required by many tasks. An exciting solution proposed in the computational motor control literature is that muscle force is costly (effortful) not only because of metabolic costs. Such costs would also provide regularization to motor planning (Jordan, 1988; Kawato et al., 1990), a classic solution to ill-posed inverse problems (Tikhonov, 1943). In other words, effort costs create a stable and unique control solution for achieving goals by incorporating a least-effort constraint.

Like motor control, cognitive control is ill-posed and subjectively effortful. I propose that mental effort may provide the regularization necessary to orchestrate a complex system like the brain towards a common goal. If we cast regularized control as a form of Bayesian inference (Botvinick & Toussaint, 2012), this aligns mental effort with other domains of ill-posed cognition like perception (Poggio et al., 1985) and knowledge induction (Tenenbaum et al., 2011). Thinking about effort as a control prior (rather than a constraint) may also help explain paradoxical effort-seeking behaviors as changes in expectations (Bustamante et al., 2021; Inzlicht et al., 2018).

This regularization theory of mental effort contributes to a growing list of (non-exclusive) theories of cognitive control costs (Kool & Botvinick, 2018; Musslick & Cohen, 2021; Piray & Daw, 2021; Zénon et al., 2019), but uniquely integrates well-established computational theories of physical effort from the motor domain. This analogy to motor control may fruitfully extend to *algorithmic* similarities between domains, such as the specific optimal feedback control models that have been the gold standard in motor control for the past several decades

(Diedrichsen et al., 2010; Haar & Donchin, 2020; Loeb et al., 1990; Todorov & Jordan, 2002). While here too there are similarities across domains, future work should explicitly test integrative theories of goal-directed control over thoughts and actions (Athalye et al., 2021; Braun et al., 2021).

Conclusions

These thesis chapters help characterize the brain's ability to act as a multiple-input multiple-output control system, pursuing goals with every tool at its disposal. While the complexity of the brain's information processing capacity is overwhelming, one hope for progress relies on a combination of precise neural-behavioral analyses with normative models of tractable optimization (Anderson, 1991; Van Rooij, 2008). This thesis uses empirical methods from psychophysics and fMRI, neural modelling of attractor dynamics and encoding geometry, and normative theories of optimization and control to triangulate fundamental models of cognitive control. This thesis sets the groundwork for explicit domain-general theories of *purposeful* cognition and action (Rosenblueth et al., 1943), which will require even richer measures of brain and behavior during the pursuit of our goals.

References

Anderson, J. R. (1991). Is human cognition adaptive? *The Behavioral and Brain Sciences, 14*(3), 471–485.

Ashby, R. (1954). *Design for a Brain*. Chapman & Hall London.

Ashby, W. R. (1961). *An introduction to cybernetics*. Chapman & Hall Ltd.

Athalye, Vivek R., Carmena, J. M., & Costa, R. M. (2019). Neural reinforcement: re-entering and refining neural dynamics leading to desirable outcomes. *Current Opinion in Neurobiology, 60*, 145–154.

Athalye, Vivek Ravindra, Khanna, P., Gowda, S., Orsborn, A. L., Costa, R. M., & Carmena, J. M. (2021). The brain uses invariant dynamics to generalize outputs across movements. In *bioRxiv* (p. 2021.08.27.457931). https://doi.org/10.1101/2021.08.27.457931

Bernstein, N. A. (1935). *The co-ordination and regulation of movements.*

Botvinick, M., & Toussaint, M. (2012). Planning as inference. *Trends in Cognitive Sciences, 16*(10), 485–488.

Braun, U., Harneit, A., Pergola, G., Menara, T., Schäfer, A., Betzel, R. F., Zang, Z., Schweiger, J. I., Zhang, X., Schwarz, K., Chen, J., Blasi, G., Bertolino, A., Durstewitz, D., Pasqualetti, F., Schwarz, E., Meyer-Lindenberg, A., Bassett, D. S., & Tost, H. (2021). Brain network dynamics during working memory are modulated by dopamine and diminished in schizophrenia. *Nature Communications, 12*(1), 3478.

Broadbent, D. E. (1977). Levels, Hierarchies, and the Locus of Control. In *Quarterly Journal of Experimental Psychology* (Vol. 29, Issue 2, pp. 181–201). https://doi.org/10.1080/14640747708400596

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

Cavanagh, J. F., & Frank, M. J. (2014). Frontal theta as a mechanism for cognitive control. *Trends in Cognitive Sciences, 18*(8), 414–421.

Chong, T. T.-J., Apps, M., Giehl, K., Sillence, A., Grima, L. L., & Husain, M. (2017). Neurocomputational mechanisms underlying subjective valuation of effort costs. *PLoS Biology, 15*(2), e1002598.

Chow, C. K., & Jacobson, D. H. (1971). Studies of human locomotion via optimal programming. *Mathematical Biosciences, 10*(3), 239–306.

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.

Danielmeier, C., Eichele, T., Forstmann, B. U., Tittgemeyer, M., & Ullsperger, M. (2011). Posterior medial frontal cortex activity predicts post-error adaptations in task-related visual and motor areas. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience, 31*(5), 1780–1789.

Danielmeier, C., & Ullsperger, M. (2011). Post-error adjustments. *Frontiers in Psychology, 2*, 233.

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.

Diedrichsen, J., Shadmehr, R., & Ivry, R. B. (2010). The coordination of movement: optimal feedback control and beyond. *Trends in Cognitive Sciences, 14*(1), 31–39.

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

Egner, T. (2008). Multiple conflict-driven control mechanisms in the human brain. *Trends in Cognitive Sciences, 12*(10), 374–380.

Flash, T., & Hogan, N. (1985). The coordination of arm movements: an experimentally confirmed mathematical model. *Journal of Neuroscience, 5*(7), 1688–1703.

Flesch, T., Juechems, K., Dumbalska, T., Saxe, A., & Summerfield, C. (2022). Orthogonal representations for robust context-dependent task performance in brains and neural networks. *Neuron, 110*(7), 1258-1270.e11.

Gottlieb, J. (2014). Neuronal Mechanisms of Attentional Control. In Anna C. (Kia) Nobre and Sabine Kastner (Ed.), *The Oxford Handbook of Attention*. Oxford University Press.

Haar, S., & Donchin, O. (2020). A Revised Computational Neuroanatomy for Motor Control. *Journal of Cognitive Neuroscience, 32*(10), 1823–1836.

Hadamard, J. (1902). Sur les problèmes aux dérivées partielles et leur signification physique. *Princeton University Bulletin*, 49–52.

Haith, A. M., & Krakauer, J. W. (2013). Model-based and model-free mechanisms of human motor learning. *Advances in Experimental Medicine and Biology, 782*, 1–21.

Inzlicht, M., Shenhav, A., & Olivola, C. Y. (2018). The Effort Paradox: Effort Is Both Costly and Valued. *Trends in Cognitive Sciences, 22*(4), 337–349.

Jordan, M. I. (1988). *Supervised learning and systems with excess degrees of freedom* [Technical Report]. University of Massachusetts. https://dl.acm.org/citation.cfm?id=896594

Kawato, M. (1999). Internal models for motor control and trajectory planning. *Current Opinion in Neurobiology, 9*(6), 718–727.

Kawato, M., Maeda, Y., Uno, Y., & Suzuki, R. (1990). Trajectory formation of arm movement by cascade neural network model based on minimum torque-change criterion. *Biological Cybernetics, 62*(4), 275–288.

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: The Official Journal of the Society for Neuroscience, 30*(47), 15778–15789.

Kool, W., & Botvinick, M. (2018). Mental labour. *Nature Human Behaviour, 2*(12), 899–908.

Kornblum, S., Hasbroucq, T., & Osman, A. (1990). Dimensional overlap: cognitive basis for stimulus-response compatibility--a model and taxonomy. *Psychological Review, 97*(2), 253–270.

Kragel, P. A., Kano, M., Van Oudenhove, L., Ly, H. G., Dupont, P., Rubio, A., Delon-Martin, C., Bonaz, B. L., Manuck, S. B., Gianaros, P. J., Ceko, M., Reynolds Losin, E. A., Woo, C.-W., Nichols, T. E., & Wager, T. D. (2018). Generalizable representations of pain, cognitive control, and negative emotion in medial frontal cortex. *Nature Neuroscience, 21*(2), 283–289.

Krakauer, J. W., Ghazanfar, A. A., Gomez-Marin, A., MacIver, M. A., & Poeppel, D. (2017). Neuroscience Needs Behavior: Correcting a Reductionist Bias. *Neuron, 93*(3), 480–490.

Leng, X., Yee, D., Ritz, H., & Shenhav, A. (2021). Dissociable influences of reward and punishment on adaptive cognitive control. *PLoS Computational Biology, 17*(12), e1009737.

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.

Loeb, G. E., Levine, W. S., & He, J. (1990). Understanding sensorimotor feedback through optimal control. *Cold Spring Harbor Symposia on Quantitative Biology, 55*, 791–803.

Manohar, S. G., Chong, T. T.-J., Apps, M. A. J., Batla, A., Stamelou, M., Jarman, P. R., Bhatia, K. P., & Husain, M. (2015). Reward Pays the Cost of Noise Reduction in Motor and Cognitive Control. *Current Biology: CB, 25*(13), 1707–1716.

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.

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

Musslick, S., & Cohen, J. D. (2021). Rationalizing constraints on the capacity for cognitive control. *Trends in Cognitive Sciences, 0*(0). https://doi.org/10.1016/j.tics.2021.06.001

Nee, D. E., Wager, T. D., & Jonides, J. (2007). Interference resolution: insights from a meta-analysis of neuroimaging tasks. *Cognitive, Affective & Behavioral Neuroscience, 7*(1), 1–17.

Nubar, Y., & Contini, R. (1961). A minimal principle in biomechanics. *The Bulletin of Mathematical Biophysics, 23*(4), 377–391.

Piray, P., & Daw, N. D. (2021). Linear reinforcement learning in planning, grid fields, and cognitive control. *Nature Communications, 12*(1), 4942.

Poggio, T., Torre, V., & Koch, C. (1985). Computational vision and regularization theory. *Nature, 317*(6035), 314–319.

Posner, M. I., & Petersen, S. E. (1990). The attention system of the human brain. *Annual Review of Neuroscience, 13*, 25–42.

Powers, W. T. (1973). *Behavior: The control of perception*. Aldine Chicago.

Putnam, H. (1967). Psychological predicates. *Art, Mind, and Religion, 1*, 37–48.

Pylyshyn, Z. W., & Storm, R. W. (1988). Tracking multiple independent targets: evidence for a parallel tracking mechanism. *Spatial Vision, 3*(3), 179–197.

Ritz, H., Frömer, R., & Shenhav, A. (2020). Bridging Motor and Cognitive Control: It’s About Time! *Trends in Cognitive Sciences*.

https://www.sciencedirect.com/science/article/pii/S1364661319302773?casa_token=lQOhjm4WW9AAAAAA:2xsNe0BxkGw_Tf8J-bpHROiRk44GQKTfBAJSIUJKZd4qf4jHVCoX8xerQS1JluNRJMFNzK9m4Ko

Rosenblueth, A., Wiener, N., & Bigelow, J. (1943). Behavior, Purpose and Teleology. In *Philosophy of Science* (Vol. 10, Issue 1, pp. 18–24). https://doi.org/10.1086/286788

Rust, N. C., & Cohen, M. R. (2022). Priority coding in the visual system. *Nature Reviews. Neuroscience*, 1–13.

Shadmehr, R., & Mussa-Ivaldi, F. A. (1994). Adaptive representation of dynamics during learning of a motor task. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *14*(5 Pt 2), 3208–3224.

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.

Silvetti, M., Vassena, E., Abrahamse, E., & Verguts, T. (2018). Dorsal anterior cingulate-brainstem ensemble as a reinforcement meta-learner. *PLoS Computational Biology*, *14*(8), e1006370.

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*. https://doi.org/10.1073/pnas.1906788116

Stringer, C., Pachitariu, M., Steinmetz, N., Reddy, C. B., Carandini, M., & Harris, K. D. (2019). Spontaneous behaviors drive multidimensional, brainwide activity. *Science*, *364*(6437), 255.

Tang, E., & Bassett, D. S. (2018). Colloquium: Control of dynamics in brain networks. *Reviews of Modern Physics*. https://journals.aps.org/rmp/abstract/10.1103/RevModPhys.90.031003

Tenenbaum, J. B., Kemp, C., Griffiths, T. L., & Goodman, N. D. (2011). How to grow a mind: statistics, structure, and abstraction. *Science*, *331*(6022), 1279–1285.

Tikhonov, N. (1943). On the stability of inverse problems. *Doklady Akademii Nauk SSSR*, *39*, 195–198.

Todorov, E. (2004). Optimality principles in sensorimotor control. *Nature Neuroscience*, *7*(9), 907–915.

Todorov, E., & Jordan, M. I. (2002). Optimal feedback control as a theory of motor coordination. *Nature Neuroscience*, *5*(11), 1226–1235.

Uno, Y., Kawato, M., & Suzuki, R. (1989). Formation and control of optimal trajectory in human multijoint arm movement. Minimum torque-change model. *Biological Cybernetics*, *61*(2), 89–101.

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*, 222.

Van Rooij, I. (2008). The tractable cognition thesis. *Cognitive Science*, *32*(6), 939–984.

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.

Yantis, S., & Serences, J. T. (2003). Cortical mechanisms of space-based and object-based attentional control. *Current Opinion in Neurobiology, 13*(2), 187–193.

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.

Zénon, A., Solopchuk, O., & Pezzulo, G. (2019). An information-theoretic perspective on the costs of cognition. *Neuropsychologia, 123*, 5–18.