

## **The interaction between arousal and cognitive control** Tromp, J.J.

### Citation

Tromp, J. J. (2025, April 9). *The interaction between arousal and cognitive control*. Retrieved from https://hdl.handle.net/1887/4211609

Version:	Publisher's Version
License:	<u>Licence agreement concerning inclusion of doctoral</u> <u>thesis in the Institutional Repository of the University</u> <u>of Leiden</u>
Downloaded from:	https://hdl.handle.net/1887/4211609

**Note:** To cite this publication please use the final published version (if applicable).



# **Chapter 1** Introduction



## Introduction

Imagine yourself in a library attempting to write this introduction. In front of you are a laptop, your phone, and some snacks you're saving for later. To your left, a few students are quietly whispering to each other. To your right, people walk by and occasionally stop to browse the bookshelves. Outside, the rain softly taps on the windows.

It is remarkable that most humans, most of the time, can focus on writing amidst this myriad of possible distractions. How do we manage to override millions of years of evolutionary tendencies (registering novel movements in our environment) for an internally set goal that didn't even exist until a few hundred years ago (writing a dissertation) on a device invented only a few decades ago (the laptop)? And, equally intriguingly, why does this ability to pursue these goals fluctuate so much? At times, you might find yourself in a state of flow, impervious to distractions, while at other times, every little sound pulls you away from your goal.

The human capacity for flexible goal-directed behavior described here is called 'cognitive control,' although this concept serves more as an umbrella term for many distinct processes and mechanisms that enable this unique ability. This dissertation will focus on understanding the neural mechanisms that give rise to variability in cognitive control, specifically focusing on the interaction between the ascending arousal system and cognitive control processes.

#### **Cognitive control**

Decades of neurophysiological research have sought to understand the capacity for cognitive control. I define cognitive control as the collection of mechanisms responsible for flexibly adapting thought and action in the service of task goals, in contrast to automatic or reflexive thought and behavior. The controlled versus automatic distinction is fundamental in cognitive psychology (and was popularized by the late Daniel Kahneman (2011), and much effort has been put into characterizing cognitive control both on the cognitive and the neurobiological levels of analysis (Botvinick & Braver, 2015).

On the cognitive level, cognitive control is often assessed through cognitive tasks that measure an individual's ability to manage and adapt their thought processes and actions in response to changing task demands and conflicting information. The Stroop task (Chapter 2), arrow flanker task (Chapter 3), and task-switching task (Chapter 4) are three historically rich cognitive control tasks used in the work presented here that each emphasize a different aspect of the capacity for cognitive control.

In the Stroop task, participants are asked to name the color of the font of a word that may spell out a different color (e.g., the word "red" printed in blue ink) (MacLeod, 1991). The task measures cognitive control by requiring participants to suppress the automatic process of reading the word in favor of identifying the ink color. The performance difference between congruent trials (where the word and ink color match) and incongruent trials (where they do not) is known as the incongruence cost, and this measure indexes the effort to override the automatic reading response. The arrow flanker task involves participants responding to a central target stimulus (an arrow pointing left or right) flanked by distractor stimuli that can either be congruent (pointing in the same direction) or incongruent (pointing in the opposite direction) (Ridderinkhof, Wylie, van den Wildenberg, Bashore, & van der Molen, 2021). The incongruence cost here is reflected in the slower response times and increased error rates when the flankers are incongruent, highlighting the effort to narrowly focus on the central target and overcome the conflicting information from the flankers.

In task-switching experiments, participants switch between different tasks (e.g., categorizing shapes by color or type) based on a given cue (Kiesel et al., 2010). The cognitive control required for this task is evident in the switch cost, which is the increased response time and error rate when switching tasks compared to repeating the same task. This cost arises because switching tasks requires reconfiguring mental processes to accommodate new rules and suppress previously relevant ones.

Cognitive control tasks like the ones described above have also played a vital role in examining the neural underpinnings of cognitive control so far. What is clear is that the frontal lobe plays a crucial role in the capacity for cognitive control (Friedman & Robbins, 2022; Miller & Cohen, 2001). Furthermore, neurophysiological studies using EEG have identified specific brain oscillations linked to cognitive control. Theta-band (4–8 Hz) oscillations in the midfrontal cortex, particularly in and around the anterior cingulate cortex, are strongly associated with conflict monitoring and resolution during these tasks (Cavanagh & Frank, 2014). For instance, during the Stroop and flanker tasks, increased theta activity is observed in response to incongruent stimuli, reflecting the heightened demand for cognitive control to manage conflicting information (Botvinick, Braver, Barch, Carter, & Cohen, 2001).

In addition to theta oscillations, alpha-band (8–12 Hz) activity is also implicated in cognitive control, particularly inhibiting irrelevant information and maintaining task goals (Jensen & Mazaheri, 2010). For example, during task-switching tasks, which require rapid shifts between different cognitive operations, the switch costs—reflected in increased reaction times and error rates—correlate with variations in both theta and alpha rhythms (Cooper, Darriba, Karayanidis, & Barceló, 2016). These EEG markers provide a temporal and spectral window into the dynamic processes underlying cognitive control, highlighting how different brain regions coordinate to adaptively regulate behavior in the face of varying cognitive demands (Miller & Cohen, 2001).

While cognitive control is renowned as enabling the uniquely human capacity for flexible behavior, it is also notorious for its limitations: humans are limited in how many control-demanding tasks they can perform simultaneously (e.g., playing a new piano piece and parsing difficult mathematical equations), how intensely they can focus on one single task (e.g. studying in the library while your neighbors are whispering to each other) and how consistently it can be deployed (e.g., resisting a cookie in the morning versus after a long day of work) (Musslick & Cohen, 2021). Why is it that such an important capacity has such glaring limitations? Predominantly, researchers have attempted to explain the limitations as weaknesses, for example, viewing cognitive control as dependent on a central resource that is depleted if used too much (Hagger, Wood, Stiff, & Chatzisarantis, 2010; Lavie, Hirst, de Fockert, & Viding, 2004). While compelling, decades of research have failed to provide strong

evidence for the depletion account (Friese, Loschelder, Gieseler, Frankenbach, & Inzlicht, 2019) or other, and comparable accounts.

Meanwhile, another approach has been more successful in addressing the limitations of control and serves as the backbone for the work in this dissertation: the resource-rational analysis approach (Lieder & Griffiths, 2019). This approach seeks to identify the optimal computations necessary for a given function and uses these to generate hypotheses about the underlying mechanisms and observed behavior. This method has driven significant progress in many scientific domains, including psychology (Anderson, 1990; Geisler, 2003; Tenenbaum, Griffiths, & Kemp, 2006), but until recently has been surprisingly scarce in cognitive control research.

A critical step in the resource-rational analysis approach is defining the 'objective function' or goal of cognitive control, which is challenging given the broad scope of processes it encompasses. So far, several levels of optimization have been tried out. Task-level optimization focuses on optimizing performance in specific tasks. Meta-level optimization addresses broader control systems. An example of a theory on this level is the expected value of control theory, which theorizes how cognitive control limitations arise through balancing control investment and rewards (Shenhav, Botvinick, & Cohen, 2013). Optimization under constraints, or bounded rationality, addresses the practical constraints on cognitive control, emphasizing that optimization must account for the limited mechanistic resources available.

In contrast to the weakness accounts, the resource-rational analysis approach views the limitations of cognitive control not as flaws but as pointers to optimizations we perhaps have not understood yet (Musslick & Cohen, 2021; Shenhav et al., 2017). In this vein, Musslick and Cohen (2021) take the bounded rationality approach and propose that the limitations in cognitive control result from optimizing trade-offs inherent in neural systems. For example, the trade-off between learning efficiency and processing efficiency: sharing representations in neural networks enables rapid learning and generalization but at the cost of multitasking performance, as you cannot use the same representation simultaneously for two different tasks. Another trade-off is between cognitive stability and flexibility: being more stable on one task, by definition, results in being worse at switching from task to task. Being stable is beneficial when writing your dissertation in the library, but not so much when you are a doctor on duty, having to respond quickly to emergency calls.

As mentioned, the level to which one exerts control fluctuates over time, both on larger (e.g., circadian) timescales and from moment to moment. The resource-rational analysis approach suits our goal well to generate hypotheses about this variance, and we explicitly work with the approach in Chapter 4. However, while the resource-rational approach constrains the search space to hypothetical mechanisms that facilitate (near) optimality, there could still be countless possible mechanisms in theory.

#### Ascending arousal system

Parallel to the classic work in cognitive control, a separate stream of research has gradually conducted a paradigm shift in cognitive neuroscience. This research shows that the brain is not a passive receiver and processor of inputs but a dynamically active organ that continuously changes states, sometimes through external inputs, but often without environmental changes

(Fox & Raichle, 2007; McGinley, Vinck, et al., 2015). Changing brain states allows the brain to change how it responds to the environment without having to go through the laborious process of changing the wiring of the brain, thus providing the brain with a vital tool to adapt to the ever-changing demands of the world and body over several timescales – from moment-to-moment to circadian to even longer timeframes.

The ascending arousal system, a set of interconnected brainstem nuclei with axonal projections to the thalamus and the cortex, underlies many of these changes in brain states. As the name suggests, these nuclei regulate the level of arousal, the global state of activation of our nervous system (Calderon, Kilinc, Maritan, Banavar, & Pfaff, 2016; Munn, Müller, Wainstein, & Shine, 2021). The ascending arousal system regulates arousal levels by projecting neuromodulators through their diffuse projections, thereby changing how neurons subsequently respond to incoming signals. The activity of the ascending arousal systems can be best measured invasively through direct neural recordings in the brainstem, but (with a stroke of luck for researchers that work with humans) research has shown that pupil size is a reliable non-invasive alternative to estimate the activity of the ascending arousal system (Aston-Jones & Cohen, 2005; McGinley, Vinck, et al., 2015; Munn et al., 2021; Reimer et al., 2014). On top, different aspects of pupil size, for example, the distinction between pupil size between trials (baseline pupil size) and change in pupil size during tasks (pupil derivative), can even capture subtleties in how the ascending arousal systems shape arousal (Reimer et al., 2016). This reliability and accessibility have led to a surge in the use of pupillometry as a tool in cognitive control research (van der Wel & van Steenbergen, 2018).

How exactly each unique cell population and neuromodulator in the ascending arousal system sculpts brain states is still the topic of ongoing research (with pupillometry playing an important role). What is already clear is that the locus coeruleus (LC) plays a central role in influencing cognitive processes through its noradrenergic projections (Sara, 2009; Sara & Bouret, 2012). The LC is a small nucleus in the pons responsible for the bulk of the brain's norepinephrine (NE) production. Despite its small size, the LC sends extensive and divergent projections throughout the brain, supplying NE to the cerebral cortex, cerebellum, and hippocampus. Slow fluctuations of activity of the LC are modulated through, for example, inputs from hypothalamic hypocretin/orexin inputs (España, Reis, Valentino, & Berridge, 2005) which align with circadian rhythm. However, moment-to-moment fluctuations are also possible through cortical connections from the anterior cingulate cortex (ACC) and the orbitofrontal cortex (OFC) to the LC (Aston-Jones & Cohen, 2005). These connections are especially interesting in the context of this dissertation, as the ACC and OFC are both heavily implicated in cognitive control processes (Botvinick et al., 2001). Recent findings by Jordan et al. (2023) further support this cortex-LC-cortex loop by demonstrating that the frontal brain regions detect visuomotor prediction errors and then use the LC to broadcast these visuomotor prediction errors across the cortex, thereby facilitating sensorimotor plasticity (Jordan & Keller, 2023).

Aston-Jones and Cohen (2005) propose that the LC exhibits two distinct modes of activity: phasic and tonic. Phasic activation, characterized by bursts of activity in response to task-related events, is associated with enhanced performance and accurate behavioral responses, promoting the exploitation of known rewards. Conversely, tonic activity, marked by elevated baseline activity without phasic bursts, is linked to increased distractibility and exploration of

new opportunities when the utility of the current task diminishes. Furthermore, Aston-Jones and Cohen (2005) hypothesize that NE released from LC neurons modulates the activity of target neurons by enhancing the contrast between synaptically evoked activity and spontaneous activity. In other words, NE increases the signal-to-noise ratio by enhancing the responsiveness of target neurons to incoming signals while simultaneously reducing their background noise. This modulates neural gain is particularly well-suited to test in computational models – the topic of the next two sections.

#### **Computational Models in Cognitive Neuroscience**

Many theories on arousal and cognitive control have been expressed solely verbally, particularly in psychology. While verbal theories and hypotheses are undeniably useful, they have significant pitfalls. One of these pitfalls is that verbal theories can be opaque or ambiguous. With computational modeling, it is not necessary to abandon verbal theory per se; you also express it in purely mathematical terms. In this section, I will lay out several arguments for using computational models in cognitive neuroscience, specifically focusing on understanding neural mechanisms of cognitive phenomena.

One of the primary benefits of computational models is that they force researchers to formalize their verbal theories quantitatively. This process of formalization requires specifying the exact relationships between variables, making the implicit assumptions of verbal theories explicit and testable. However, merely adopting computational models does not automatically resolve all ambiguities; there are still implicit virtues that guide the choices researchers make when designing these models. As Heijnen et al. (2024) emphasize, these choices often remain unspoken in the literature, leading to potential biases and locked-in research traditions. Therefore, it's crucial for researchers to also be transparent about the underlying motivations and criteria that inform their modeling decisions to ensure that the models serve their intended purpose effectively.

Following the formalization of their theory, the researcher can simulate and predict the outcomes of different experimental conditions, thereby serving as a computational testbed for the implications of that theory. These predictions allow researchers to test their theories' validity in silico before conducting costly and time-consuming experiments. By running simulations, researchers can explore a wide range of hypothetical scenarios, identify critical variables, and refine their hypotheses, making subsequent empirical studies more focused and efficient.

Computational modeling also offers a way out of what has been termed the "Flatland fallacy" the tendency to oversimplify complex, high-dimensional psychological phenomena into lowdimensional explanations. As discussed by Jolly and Chang (2019), this fallacy arises because of our cognitive limitations and the fact that traditional methodological approaches often lead us to favor simpler, more tractable theories, even when the phenomena we study are inherently complex and multidimensional. Traditional experimental designs and statistical analyses often limit researchers to examining a few variables at a time, leading to potentially oversimplified conclusions. In contrast, computational models can handle high-dimensional data, capturing the intricate interactions between numerous factors that shape, for example, cognitive control variability. By embracing this complexity instead of turning to reductionist approaches, researchers can gain a deeper and more accurate understanding of the phenomena under study.

Especially in research as conducted in this dissertation, computational models are invaluable. In essence, the brain is an information-processing machine, and its cognitive capacities can be mechanistically explained by laying bare how neural computations over neural representations give rise to those capacities (Barack & Krakauer, 2021; Piccinini, 2020). These computations are real, causally effective properties of the nervous system, not just metaphorical descriptions. One step further, some philosophers (e.g., Piccinini, 2020) argue that these computations are multiply realizable: on a certain level of analysis, the computations are identical whether implemented in silicon or biological networks. While I am unconvinced by the multiple realizability argument (I think there are biochemical properties within cells that cannot be replicated in non-biological networks; Sterling & Laughlin, 2015), I do support the idea that the neural networks implemented in silicon can capture real properties of the neural networks implemented in biological systems (Stinson, 2020). Therefore, when we try to elucidate the neural mechanisms underlying cognitive control, replicating the properties of those mechanisms in computational models implemented in silicon is a promising way to make progress in understanding the biological neural mechanisms.

Next to these ontological arguments for using computational models, there is an important epistemic point as well: computational models promote cumulative science by providing a framework for systematically building on previous research. Unlike verbal theories, which can be ambiguous and open to multiple interpretations, computational models are precise and reproducible. They can be shared, tested, and extended by other researchers, facilitating collaboration and the gradual accumulation of knowledge (Doerig et al., 2023; Jolly & Chang, 2019). This iterative process of model refinement and validation helps develop robust theories that better capture the mechanisms that instantiate cognitive control. A good example of how this can play out is with the class of models that stem from the tradition of connectionism, which is the topic of the next section.

#### **Connectionist Models of Control**

The models used throughout this dissertation fall under the category of parallel distributed processing (PDP) models or connectionist models. These models utilize highly interconnected, neuron-like processing units to simulate cognitive processes, closely mirroring the structure and function of biological neural networks. In the past four decades since the start of the research program (Rumelhart, Hinton, & McClelland, 1986), connectionist models have been able to successfully emulate how humans solve cognitive tasks, thereby showing that there is a possible answer to one of the biggest questions in cognitive neuroscience: how can complex cognition arise from "simple" neurons? According to connectionism, the answer lies in distributed processing rather than centralized master units.

Especially in the field of cognitive control, there is a rich history of successful connectionist models. A good example of this history is the subclass of connectionist models that iteratively attempt to capture how humans solve the Stroop task (Botvinick et al., 2001; J. D. Cohen, Dunbar, & McClelland, 1990a; J. D. Cohen & Huston, 1994; J. D. Cohen & Servan-Schreiber, 1992a, 1992b; Kalanthroff, Davelaar, Henik, Goldfarb, & Usher, 2018; Miller & Cohen, 2001). A history which we supplement in Chapter 2. This iterative improvement of computational

models that attempt to model the same cognitive phenomenon is a good example of the epistemic advantage of computational modeling mentioned in the previous section.

#### Approach

In the work reported below, I aimed to explore the interplay between arousal and cognitive control, thereby examining how fluctuations in arousal contribute to variability in cognitive performance. By combining empirical tests, electrophysiology, pupillometry, and computational modeling, the research seeks to advance our understanding of the neural and computational mechanisms underpinning cognitive control and its variability.

The hypothesis I pursued is that the variance in the human ability to exert control can be mechanistically explained (at least partly) by arousal dynamics.

#### Overview

The dissertation is divided into five chapters, with this introduction being the first. The second chapter examines how fluctuations in global neural gain influence cognitive control, both empirically and in computational models. The third chapter explores the effects of phasic alertness on cognitive control, presenting a novel account of how phasic alertness boosts arousal and generates an urgency signal that impacts performance. The fourth chapter addresses the stability-flexibility trade-off, investigating how individuals regulate this trade-off in response to different contextual demands through task-switching paradigms and neurophysiological measures. In the fifth chapter, I outline the next steps I think are essential for further research. What follows now is a more detailed summary of the following three chapters.

Chapter Two investigates how fluctuations in global neural gain affect cognitive control. Using behavioral methods, pupillometry, and computational modeling, we examine the impact of neural gain on the Stroop task, a classic measure of cognitive control. First, simulations of a comprehensive connectionist model of the Stroop task, which includes task conflict and both proactive and reactive forms of control, predict accurately that increasing global gain leads to an overall speeding of reaction times, increased Stroop interference, and decreased Stroop facilitation. Pupillometric analyses reveal that the pre-trial pupil derivative, a non-invasive index of global gain, correlates with Stroop task performance as predicted in the model simulations: increased pupil derivative correlates with increased interference and decreased facilitation. An analysis of the internal model dynamics suggests that a gain-related increase in task conflict and corresponding (within-trial) increase in reactive control are vital for understanding this pattern of behavioral results. Indeed, a similar connectionist model without this task-conflict-control loop could not account for the results. Thus, this chapter shows how spontaneous fluctuations in neural gain significantly impact reactive cognitive control and that we can predict the impact of these fluctuations in neural gain on behavior through computational modeling.

**Chapter Three** explores a decade-old mystery in the arrow flanker task. Why does a short tone before the stimulus presentation speed up response times while also increasing incongruence costs? Here, we present a novel, biologically informed account suggesting that phasic alertness generates an evidence-independent urgency signal. This urgency signal shortens overall response times and amplifies competition between evidence accumulators, impairing cognitive control. By using a combination of the arrow flanker task, pupil measurements, and

electrophysiological data, the study demonstrates that phasic alertness increases conflict in incongruent trials. A connectionist model of the flanker task incorporating time-varying urgency and lateral inhibition among evidence accumulators successfully reproduces the behavioral effects of phasic alertness, highlighting the interplay between dynamic changes in urgency, cognitive control, and evidence accumulation. This chapter provides insights into the complex relationship between phasic alertness, urgency, and cognitive control.

**Chapter Four** addresses the stability-flexibility trade-off in cognitive control, which is essential for adapting to varying contextual demands in daily life. Using a cued task-switching paradigm, the study reported in this chapter investigates whether humans can efficiently regulate this trade-off by adjusting their cognitive flexibility in response to different contextual demands and the neural correlates of this adaptation. Therefore, participants performed a parity-magnitude task-switching task under varying switch-frequency conditions while we measured their pupil size and alpha/theta oscillatory power. The study replicates the trade-off between switch and congruency costs empirically and in computational simulations. Results show that participants exhibit decreased switch costs in high-switch contexts, indicating greater flexibility, but increased incongruence costs, indicating reduced stability. Furthermore, baseline pupil size and pre-cue alpha power levels correlate with switch frequency, suggesting these markers track stability-flexibility adaptations. This chapter bridges the gap between empirical and theoretical research by combining neurophysiological measures with computational models, providing a comprehensive understanding of the mechanisms underlying adaptations to the stability-flexibility trade-off.

In summary, the chapters in this dissertation collectively advance our understanding of how ascending arousal systems, particularly the locus coeruleus-norepinephrine (LC-NE) system, influence various aspects of cognitive control. By integrating empirical tests, computational modeling, and neurophysiological measurements, this research sheds light on the dynamic regulation of cognitive control and offers potential pathways for enhancing cognitive performance through targeted interventions.