Conditioning the Speed-Accuracy-Tradeoff:

Bert and Bob's take on the control problem

Bachelor's thesis

Radovan Vodila

Student number: 11930141

Word count: 17.591

Academic year: 2022/23

Keywords:
Cognitive control, numerical decision modeling, Speed-Accuracy-Tradeoff, reinforcement learning, drift-diffusion model

Supervisors:
Univ.-Prof. Dr. Senne Braem[1, 2]
Univ.-Prof. Dr. Tom Verguts[2]
Univ.-Prof. Dr. Pierre Sachse[3]

1: Department of Experimental Clinical and Health Psychology, Gent University
2: Department of Experimental Psychology, Gent University
3: Department of Experimental Psychology, University of Innsbruck

Contents

Acknowledgments

This project would not have been possible without the time and trust Senne and Tom put in me. Thank you for entrusting me with this study and dedicating your time to it, as well as for the guidance you provided throughout.

I want to thank Jonas for being a great colleague, mentor and a grand office mate.

Then I'd like to thank all Cocoflexers, as well as all members of Senne's and Tom's lab for making me feel welcome and appreciated, especially during lockdown in the early months of my stay.

A huge thanks goes out to Laura without whom I wouldn't have been able to present such a grammatically and syntactically sound thesis.

Next, I want to thank Pierre for his helpful thoughts and comments on each building block of this thesis.

Lastly, I'd like to thank the whole PP02 department, for introducing me to the grand game of academia.

Whenever I speak of 'We' in the upcoming chapters, I'm referring to the research team consisting of Jonas, Tom, Senne and I. Again, thank you for the knowledge you provided and the important concepts you taught me on this journey.

Abstract

Recent advances in cognitive neuroscience have brought forth a novel perspective on the control problem, grounding it in associative learning – a mechanism largely seen as the dichotomous counterpart of cognitive control. An important implication following this theory is that higher order functions are subject to the same reinforcement learning principles as lower-level behavior. Following this notion, the prediction can be made that humans adjust their control parameters based on learned association with contextual cues.

The presents study was designed to explore this prediction by employing a fast-paced visual discrimination task featuring two contexts, wherein participants were nudged to assume high, and low caution respectively in their decision making, which was quantified by the drift-diffusion model.

Data analysis points towards a null effect, which we attribute mainly to flawed design elements and conclude that these need to be catered for before a conclusion can be made.

Furthermore, a simulation-based approach will be proposed, which affords the visual investigation of performance of a simulated system informed by a particular set of DDM parameters. This was applied to the design and yielded valuable insights on the observed null effect, as well as on avenues for optimization.

Chapter 1:
Theoretical Foundation

Introduction

Imagine the following scenario: You spend your morning in a foggy forest, hunting for mushrooms. Doing so, your attention is set to detecting the generic ovalities of caps, as well as the color and shapes of their stems. Upon detection, your focus fixates on the target, confirming it as a true positive, and further classifying its kind according to size, color and shape. Once you're sure it's an edible boletus, you continue your inspection on whether it's a 'keeper'. To this end, you focus on the details of the fungus: the quality of the lamellae, as well as abnormalities indicating an infestation.

Within several seconds, your attentional mechanism shifted its focal points from detecting shapes within foliage to classification guided by its attributes up to quality grading informed by its detailed state. At each stage of your decision, different goals are active and hence lead your attentional mechanism to seek for different cues. These shifts of attentional focus occur periodically within your hike, as each pick resets the cycle to you scanning the foliage. You've flexibly adjusted your attention dozens of times to align with the currently active goal.

Let's consider a second scenario. You stand in front of your fridge and sigh at the emptiness of your compartment. Your gaze swings to your flatmate's area, as a full tray of Tiramisu lays there. Enthusiastically you take it out, energized with anticipation you lift your fork but then hesitate; you stow both instrument and object of desire, put on your running gear and go for a jog: You've remembered the dietary plan you swore to commit to and moreover, you couldn't (yet again) discard your moral values and (yet again) dig into your flatmate's belongings. Although admittedly the latter scenario might be just tangentially

scraping the bounds of realism for most of us, it illustrates the human capability to marshal behavior to align with higher-order goals. Both scenarios, albeit seemingly unrelated at first, illustrate the effects of a construct called cognitive control: the capability to regulate behavior adaptively and flexibly in order to achieve higher-order goals.

Conceptualizing Cognitive Control

Cognitive control can be defined as the capability to orchestrate behavior adaptively and flexibly in order to achieve higher-order goals. (Botvinick, Braver, Barch, Carter, & Cohen, 2001). The term 'goals' in this context refers to abstract goal representations, such as within-task micro-objectives in respect to which information is classified as relevant. The attentional system informed by cognitive control affords a filtering of sensory perception distinctly and solely for relevant cues. (Luck & Ford, 1998). You filter your visual input for different informational cues trying to locate mushrooms, as opposed to during quality-grading. Further, it informs under which circumstances not to exhibit learned behavior.

To illustrate, consider you burned yourself grabbing the blazing hot brass handle of your skillet. One painful experience was sufficient to create an aversion, making you hesitate the next time you reach out for it. Cognitive control affords the ability to override such inhibition – conditional to the certainty of it being cool. It is crucial to note that these goals are far from rigid: Goal representations need to flexibly adapt in ever-changing circumstances. In one moment, it is important to scan for stems within foliage, and in the next it's about classification, informed by color, girth and size of the fungi. And again, in the very next moment, you're checking the map to remind you of the route, analyzing the finicky lines and colors. Such flexibility of swinging seamlessly between goal representations also constitutes an important feature of cognitive control. (Braver, 2005).

Another attribute associated with the exertion of cognitive control is its subjective cost. This cost can be traced down to the cerebral energy-source glutamate. This 'fuel' is a valuable and notably depletable resource within the central nervous system. Upon depletion of this resource, the efficacy of cognitive control substantially weakens. (Westbrook & Braver, 2015).

The human ability to exert control over our behavior, inhibit urges and delay gratification has inspired generations of researchers. By now, it is largely well understood how control influences our behavior and which factors affect it. However, its underlying mechanisms are still left unclear. (Botvinick & Braver, 2015). Incumbent theories put a domain-general executive system in charge of control processes such as attentional filtering, action-inhibition, task switching, conflict adaptation, the Exploration-Exploitation-, and Speed-Accuracy-Tradeoff (SAT). This control system is conceptualized as being top-down-operating, as well as conscious.

Furthermore, this view posits the attribute of modularity. In such, the domain-general control system, control is imposed by a central unit specializing in this very function. Applied to the brain, one would be able to delineate an area which is serving solely this regulatory purpose from regions processing more rudimentary stimulus-response-mappings.

This means that the controlling cortex would be able to impose control onto - but not participate in 'basal' stimulus-response processes.

Hence, imposing control often manifests as the override of well-learned and habitual actions, consequently leading to control being set in contrast, and even in dichotomy to the concept of learning. To elaborate, associative learning is considered to be operating bottom-up: creating associations between perceived stimuli and following behavioral responses automatically. As illustrated, one does not need to contemplate the pain inflicted by the blazing brass handle to conclude that it ought to be avoided in the future, as to create that aversive conditioning.

Closely related to associative learning, reinforcement mechanisms also work based on associative links. Given a reward signal, a phasic upshoot in dopamine levels throughout the cortex is triggered. This state associates currently active contextual stimuli – informative, as well as uninformative ones – with the behavioral response the system carried out. Such stimulus-response associations form under the adhesion of reward signals. (Law & Gold, 2009; Saddoris et al., 2015). This soar of dopamine raises the likelihood of the system again performing the reinforced behavior when met with the associated stimuli. The system has learned to anticipate a state of high dopamine tonus following that particular action; hence it is keen on again reaching that state, and a strong predictor for it is performing the reinforced action. This principle is a well-established pillar of behaviorism and dates back to Thorndike's prominent law of effect. (Thorndike, 1898; 1911)

To illustrate, pressing a button can be conditioned as an action predictive of the arrival of a rewarding snack. Or to provide a more contemporary example, refreshing your social media feed acts as a strong predictor for the surge of that precious dopamine induced by the appearance of novel colorful images of attractive faces. Although being blessed with the capability of wishfully thinking us not to be susceptible to this basal mechanism, our reward system succumbs to the same archaic reinforcement principles as rodents and pigeons do.

In conclusion, cognitive control is regarded to be domain-general, conscious, top-down-operating, and an effortful mechanism, consisting of a set of supervisory processes, which are unidirectionally influencing stimulus-response mappings.

Control Problem

Conceptualizing both mechanisms in this manner serves to delineate, as well as to provide a metaphor to work along with. However, such modular perspectives grant little insight into the workings of that

'supervisory agent'. Keeping this homunculus in charge merely circumnavigates the question of its underpinning mechanisms, as little explanatory value is added by attributing all responsibility of behavioral control to a general-purpose agent. The quest to disentangle the underpinnings of cognitive control is referred to as the control problem. It is established that cognitive control oversees filtering our environment for relevant information as well as facilitating the override of habitual behavior. As mentioned, there is rich literature on the effects of cognitive control as well as its underlying factors. (Verbruggen, McLaren & Chambers, 2014).

Yet, it is not clear what underlies this phenomenon. It is not clear what guides this behavior-informing system. To disentangle this problem, a major question needs to be conquered:

'What informs this informer?'

Further broken down into primo: 'What informs it *when* to act?' and secondo: 'What informs it *how exactly* to act?

Umemoto and Halroyd (2015) for instance approached this problem by researching within an environment with multiple task options available. They posed the following question:
a. how does the control system decide what task to execute, and **b.** how vigorously to carry it out? Interestingly, they reported reward signals to have a modulatory effect on the exertion of control. This finding contradicts the traditional view of the control agent, as reward per definition is not supposed to play a role in behavioral control.

Furthermore, a recent meta-analysis reports that cognitive training rarely finds transfer. For one, they report extensive evidence that cognitive training improves performance on the trained tasks. However, they declare little evidence corroborating the notion of transferability: the less related the task in question is to the one subjects have been trained on, the less of an effect is observable. Crucially, they report that barely

any improvement transfers into everyday cognitive performance. (Simons et al., 2016). These findings suggest that training effects stay bound to the task-sets they were trained in, challenging the presumption of generalizability of cognitive control. Interestingly, such domain dependence or context specificity is an attribute of associative learning.

Not only does this challenge the feasibility of cognitive training, but also the very core of the incumbent modular view on cognitive control.

To make ground for an argument on the feasibility and context-specificity of cognitive control, imagine the following scenario:

On your way to work there is an ill-designed intersection where too much happens at the same time: the road lines are not intuitive and there is this one billboard which hinders your view to your right.

The first times you were stressed while attentively creeping forward to make your left turn; checking your mirrors every other second until you passed. Fast forward a couple months: By now you have adopted a more cautious state without really thinking about it. You happen to shift towards a more attentive state the moment you recognize the area of the intersection. It seems like you learned to adopt a higher attentive state in the critical environment. As if a high alert state was stamped onto that environment, mediated by the reoccurring need of attentiveness.

Following the traditional conceptualization of these two mechanisms, the system would have to consciously recruit a state of heightened awareness every time it enters the intersection. Such a state would always have to be preceded by the aware need for a state of heightened attentiveness, resulting in an effortful recruitment of same.

In 2016, Abrahamse and colleagues posited a new approach for taking on the control problem. In their report, they review an array of studies which violate the predictions of a modular perspective: Control being susceptible to reward signals, being context specific and being manipulatable in the absence of awareness. (Abrahamse et. al., 2016)

Curiously, these are the very features underlying associative learning. Moreover, these are the very attributes cognitive control was set in contradiction to. Building upon these findings, Abrahamse et al. inferred that the nature of cognitive control and associative learning might not be as distinctly separated as assumed and rather have the very same mechanisms constituting their core – positing the hypothesis that cognitive control might be embedded into associative learning.

An important prediction that follows this notion is that higher level control functions are subject to the very same reinforcement principles as lower-level behaviors and furthermore, that people regulate their control parameters based on learned associations with contextual cues. To illustrate, let's again picture that busy intersection. You reacted with high attentiveness and caution to a stressful situation and experienced the pleasurable outcome of passing unharmed: The control (network/ system) subsequently associated its parameter configuration (high caution) and the resulting behavioral implications (frequent checking of the mirrors, slow pace, etc.) with the context it occurred in. This results in an associative network binding context, response, and the overarching strategy. Subsequently, every time you approach the intersection, contextual cues trigger this network. Upon activation, it retrieves the embedded cognitive strategy, subsequently aligning your behavioral response to correspond to the overarching control parameter: being cautious.

In general, this network consists of three elements: perceptual, motor and goal representations. Contextual features embed as perceptual representations, actions taken as motor representations and active cognitive strategies as goal representations. Upon a favorable outcome of an action taken, momentarily active contextual features – informative as well as uninformative ones – are embedded as the base of the control network. The active goal representation can be reconceptualized as the

present cognitive strategy – may it be pronounced caution or a high readiness to switch tasks.

This strategy is embedded next to the contextual features and the executed motor representation. The reward signal remains in its traditional role of the adhesive component which induces as well as facilitates the construction of this network. In this vein, the learning perspective maintains the very same views on control representations. But most crucially, it provides explanation for the domain-dependence and lack of transfer via the context-bound embedding of control functions, rather than with a multitude of specific control processes acting solely in their respective competence. Similar arguments for a distributed view on cognitive processes were made by Eisenreich (2017) in his paper scrutinizing the modular conceptualization of the brain.

One important prediction that follows this control model is that people regulate their control parameters based on learned associations with contextual cues. To our knowledge, Braem (2017) was the first to provide behavioral evidence in favor of this hypothesis. The author reports having conditioned subjects to express a higher tendency of task switching behavior after disproportionally rewarding alternating tasks versus task repetitions.

In a free-choice testing phase, subjects were more likely to alternate tasks when this strategy was reinforced in the previous phase of the experiment. He concludes that reward indeed exerts a modulatory effect on task switching. Several studies conditioned stimuli to act as control-informing markers. Verbruggen and colleagues for instance associated stimuli to act as inhibitory markers, and as markers for raised attentional control. (Verbruggen & Logan, 2008).

Furthermore, Braem and colleagues reported that novel task instructions were more easily adopted in an environment, which was previously associated with a higher occurrence thereof. (Braem et al., 2020).

In order to adopt a new task instruction, a suitable goal representation needs to be generated, consequently replacing the former one. The new behavioral response subsequently shifts to being informed by this novel set of goals. He concludes that a higher readiness of running this integrative process was associated with a contextual cue, as subjects required less time to adopt new instructions in the trained context as opposed to controls.

More recently, Prével and colleagues (2021) reported a modulating effect of reward signals on conflict processing. This function of resolving conflict is essential to shield us from distracting information or prepotent response options, therefore contributing to the maintenance of goal-directed behavior.

Prasad and Mishra (2020) reported reinforcement playing a mediating role on control on the masked priming effect. Prior reward association of a given stimulus modulated the perceptual saliency of same in a non-reward testing phase.

Contemporary approaches to investigate the extent to which cognitive control can be conditioned mainly relied on blocked designs: throughout dozens of trials the subjects gradually learned to associate the stimulus with a particular expression of control. These learned associations were then again tested in a blocked manner.

The present study explores the possibility/attempt of conditioning control parameters beyond the realm of stable learning environments. To this end, an environment has been designed where the stimuli are presented in one of two locations, which we'll further refer to as contexts, or condition. Each context has its own reward policy, and the presentation of contexts fluctuate trial-by-trial / on a trial level, with the policies remaining stable within -context.

The reward policies incentivize different cognitive strategies. Therefore, it was advantageous to exert a particular strategy in context 1 and to

apply the converse approach in the other. Rephrased, upon context swing it was optimal to also adapt the strategy and parallelly shift the expression of control. It is crucial to understand that the same response can be evaluated differently in each context, therefore – to behave optimally, one had to pick up this difference in reward signals and adapt the strategy accordingly. This study first ought to investigate whether humans are capable of detecting this delta in reward signals within such a volatile environment. Building upon this, we aim to explore whether humans are able to adopt these optimal strategies and swing between them trial-by-trial. Thirdly, we aim to answer the question, whether such goal-representations get associated with the context, and if they transfer and remain stable in the absence of reward.

To this end, we engineered a task design incorporating three attributes of associative learning: reward-sensitivity, contextual dependence, and the absence of awareness. The strategy we chose to investigate is the expression of a strategy called Speed-Accuracy-Trade-off or rephrased as cautiousness.

The Speed-Accuracy-Tradeoff

The Speed-Accuracy-Tradeoff – further referred to as SAT – describes the cognitive strategy to either prioritize accuracy or speed in decision making. On one end of the spectrum – emphasizing accuracy over speed – a higher accuracy across multiple decisions is generated but speed sacrificed, entailing slower response times. This is counterposed by emphasizing speed, yielding quicker response times with the tradeoff of having a higher probability of erring. Gaining pace in decision speed entails the sacrifice of accuracy across trials, resulting in a higher error rate. To illustrate, let's regard a scenario way back in Paleolithic times. Our protagonists are Bert and Bob, two Neanderthals working as gatherers for their tribe.

One day, as usual, Bert takes his basket and sets out to forage. While collecting fungi he is cautious in his decisions. Each fungus needs to be inspected thoroughly to ensure its edibility, as erring, viz. returning with inedible ones would lead to physical unease at best and poisoning at worst. While foraging, he begins to traverse into the territory of a rival clan, unmistakably marked by a peculiar kind of pointy trees. His mindset changes into wanting to quickly fill up his basket, as it is advantageous to spend as little time as possible in such a dangerous area. Finally, Bert returns to his home cave and hands over his yield: The cook is upset, as he had to filter out way more wrongly picked mushrooms than usual.

How can that be? The larger number of erroneously picked mushrooms in Bert's basket can be explained by a shift of control parameters informing his decision making. By being in an uncertain and possibly dangerous environment, he adjusted his expression of SAT towards the speed pole, emphasizing speed over accuracy. Simply due to the rule of thumb 'the more time you spend in a dangerous area, the higher the probability of finding out why it is known as such'.

This unaware adjustment of his caution parameter enabled him to optimize his behavior in alignment to the situation he was in as well as the goal under which he was operating: Emphasizing a quickly filled up basket, taking the risk of lower accuracy doing so.

However, his cousin Bob sadly did not evolve to express such flexibility in his cognitive control parameters. Bob set out to forage and similarly ended up in an area under the control of a rival clan. As opposed to his cognitively flexible cousin, Bob did not adjust his SAT. He thoroughly inspected every fungus, which led to him spending additional time in this area. Consequently, Bob was spotted by a hostile scout and ambushed on his way back, never returning.

Mourning, Bert promises vengeance against his cousin's assassins. Again, he sets out to forage, but this time he remains in the clan's area, avoiding the path he once took. As he further wanders, he notices the same kind of peculiar trees which marked the enemy's territory, however still being in allied space. This thought soon fades while he continues filling up his basket.

To his surprise, he gets scolded by the cooks as he returns: Again, they found more inedible mushrooms than usual. Why didn't he spend more time selecting? He might as well have spent the normal time doing this task, probably coming back with more utilizable ones. Instead, he was back early, his yield speckled with inedible fungi. In disbelief, Bert looks towards the sky and realizes that in fact he returned much earlier than usual. Why didn't he take more time in his decisions, he wonders.

Following the associative learning perspective, Bert associated the goal of expressing a lower degree of caution with the contextual cue of that peculiar kind of tree. The tree was encoded as a contextual stimulus for triggering a low caution during decision making.

Then, without him being aware, this control network was triggered while executing the same task, being surrounded by the same contextual stimuli. This subsequently informed his control parameters and led to him swinging from expressing a high degree of cautiousness to the converse. With this goal active, he took less time deciding and raised his error rate.

Fig. 1. Bert's yield in three scenarios. The plot codes accuracy on the y-axis and speed on the x-axis. Each triangle is characterized by a different position on the Speed-Accuracy-Tradeoff.

Bert associated the context he was in with the goal representation informing his behavior. This led to him responding appropriately to the environment at hand. The attribute of domain dependence activates the network when encountering the embedded contextual stimulus. In Bert's case, a particular type of tree triggered the previously created network for lowering his caution in picking fungi (Fig. 1).

Presumably, Bob would not be underperforming in this novel setting as he would have remained cautious. Being cognitively inflexible, he would have kept foraging as usual – slowly and accurately. Unlike in Bert's case, no control network would have been formed and consequently couldn't have subliminally driven control parameters based on contextual markers. Sadly, one wouldn't be able to test this notion, as – crucially – Bob is dead.

Computational Modeling

The Speed-Accuracy-Tradeoff can be further characterized not only by response times (RTs), but also the variance thereof – both for correct and erroneous decisions taken. From this decomposition, a peculiar pattern emerges: Fast decisions result in a lower variance of their response times, contrasted by a wider spread of RTs in slow decisions. (Heitz, 2014 for review).

To explain such peculiarities, cognitive researchers employ computational tools which model the process in question and thereby provide insights into its underpinnings. This model consists of an algorithm which formalizes the problem in computational terms by removing all irrelevant features of the decision process and only incorporating the presumably relevant ones. This reduction in complexity is simply due to the notion that one cannot capture the working of each synapse within the control network; one does not have insight into the cognitive system, so its workings need to be approximated by reverse engineering.

Each model includes several parameters, which are assumed to approximate cognitive components relevant for the decision process.

These parameters ought to capture the influence of real word factors substantial for the process in question.

In the following section, I will first introduce unidirectional race models. These systems model processes for instance, wherein one ought to recognize an object varying in informational quality through a reduction of same. This reduction can be created by the injection of Gaussian noise, or by simply removing parts of the image. Essentially, the decision is made when enough evidence for the nature of a distorted image is accumulated.

Subsequently, build upon this concept to introduce a numeric tool used for computational modeling of bi-optional decision making processes.

Unidirectional Recognition Models

As mentioned, the informational quality of the presented stimulus is assumed to hold relevance for the decision making process. To illustrate, picture a task wherein you are instructed to hit a button the very moment you recognize the nature of the two stimuli of figure 2:

Both depict the same object but differ in informational quality. Stimulus 1 can be classified easily.

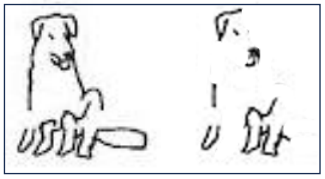

Fig. 2. Stimulus set 1.

However, stimulus 2 – providing less information – offers lower quality of evidence and hence demands more time to be recognized. A model using only this parameter would predict the average response time for high evidence images to be substantially faster than for lower quality stimuli. level of integration rises until a threshold is reached, initiating the decision. The amount of evidence needed to reach that bound represents the second parameter of this model – alpha (α). A higher bound means that more evidence is needed – the system takes longer to reach certainty.

This 'race' towards the decision threshold is the eponymous attribute of this model. Lastly, the time between stimulus onset and start of the decision process is called non-decision time – encoding period – and will also be referred to as theta (θ).

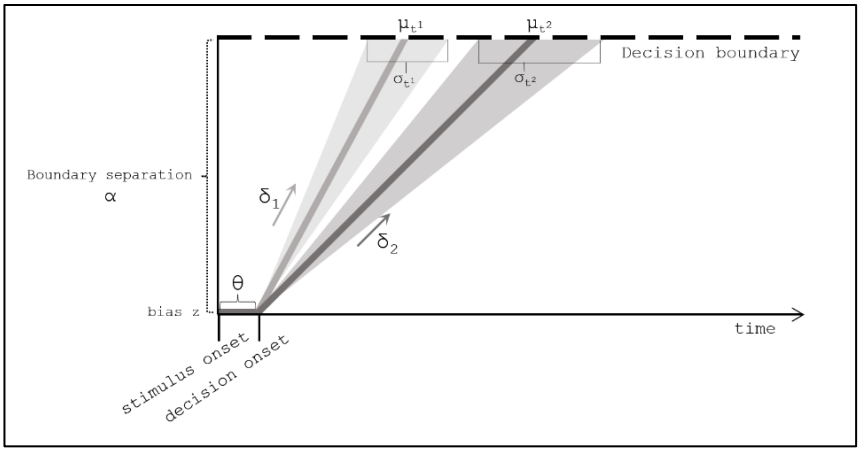

Fig. 3. Schematic of a rise-to-threshold model applied to a recognition tasks.

To summarize, our model decomposes 2afc decision making into the following parameters:

a) bias *z*, b) non-decision time *θ*, c) rate of evidence accumulation *δ* and d) vertical distance from starting point to the decision threshold *α*.

Integration level is coded on the y-axis and accumulates over time, which is depicted on the x-axis (Fig. 3). The accumulation of evidence over time draws from the notion of sequential sampling: at each time step t, the system integrates presented evidence.

Upon stimulus presentation, your visual system claims some time to sensorially process and encode the presented stimulus – accounted for by the parameter theta. From that point on, the decision process commences at starting point z.

The evidential quality of the stimulus informs the rate of deliberation, visualized as the accumulation of evidence 'racing' towards the decision bound. The higher the evidential quality, the more information can be extracted per time step, which results in a steeper integration-slope.

Keeping all other parameters constant, both increasing the slope δ and raising the starting point z decreases the response time, as the former leads to a steeper integration and the latter reduces the size of the decision space α and hence shortens the distance needing to be covered.

Additionally, increasing α results in the need for more evidence having to be integrated to initiate a response. Consequently, the decision process demands more time. We can formalize response time t as a function of the distance α in respect to slope δ and the nondecision time θ:

$$t = \frac{\alpha}{\delta} + \theta.$$

Crucially, the equation above holds true only in the absence of noise.

In a noise-free world, deliberating the same stimulus under a fixed set of parameters would always take the same amount of time.

Unfortunately, we happen to exist in a noise-riddled environment. Hence it is assumed that all sensory input we process is noisy, consisting of noise and signal. Signal, or information, is all sensory input relevant for our decision – in our case the 'isolated' representation of the stimulus. Noise on the other hand is the umbrella term for essentially everything else effecting the process. This can be internal physiological noise influencing the efficiency of our visual processing system, or intrusive thoughts – regardless of valence – surfacing, pulling attention away from the decision making process. Furthermore, bodily factors such as even slight hunger or completely external ones such as the audible weeping of a toddler nearby. Also, everything perceived in peripherical vision can constitute distractors.

Crucially, one has no way of knowing the individual influence these distractors entail. Noisy constituents are too vast to capture or to model, hence one reverts to subsume all these factors under one term, which then again can be incorporated into a model: statistical noise.

As we've seen in the introduction, response time distributions deviate more for slow decisions as opposed to fast ones, so the model needs to be further modified to account for this phenomenon. This is achieved by introducing noise into the model.

To illustrate, regard again Fig. 3: Both slopes/integrators start from the same point and must accumulate the same amount of evidence (a) to reach the decision boundary. Both parameters a and z stay constant, varying only in their integration rate delta:

System 1 integrates at a higher rate, depicted as a steeper slope. Due to its delta being higher compared to system 2, it reaches its decision bound faster on average. System 2 operates with a lower delta and consequently requires more time to reach its boundary – deciding slower. Due to its decision taking longer, it is likewise exposed to the influence of noise for longer. The longer the exposure, the stronger response times deviate and the larger the variance across trials will be.

Noise is implemented as a constant factor and its influence stands in direct proportion to time passing.

This relation is visualized in figure 3 through the ribbons around each integrator. The slower a system integrates, the wider the ribbon becomes.

Therefore, race models predict a higher variance of response times for slower responses, operating with a flatter integration slope. Conversely, decisions steered by a steep integrator slope occur quicker, entailing a lower deviation of response times.

With this relation in mind, our model offers an explanation for the observed variance patterns by decomposing a complex process into several subsets of it. The outlined race model attributes the occurrence of deviating variances to noisy interference modeled using a linear buildup of statistical distortion within each decision.

Crucially, such conceptualization of noisy interference implies that said pattern is not set in stone, as the depicted ribbon represents one standard

deviation from the mean – meaning that only 68.3% of the observations are expected to occur within the respective ribbon. Consequently, the model predicts 31.7% of all observations to occur outside of it.

Therefore, it is quite plausible to observe a slower outlier of system 1, as well as a faster outlier of system 2, resulting in the 'slower in nature' system 2 responding faster than system 1. Under the assumption of Gaussian noise, individually unlikely observations become expected across a vast number of trials.

By now, I've outlined the application of race models within recognition tasks. The response was initiated upon stimulus classification. Within the decision process, evidence was sequentially sampled – and incrementally accumulated until a threshold was reached which led to decision onset.

Building upon this foundation, race models can be extended to model decision processes between two choices – these architectures are referred to as drift-diffusion models.

Drift-diffusion model

Bioptional modeling adopts many characteristics of its unidirectional counterpart, as it also assumes the integration of evidence towards a constant threshold. Drift-diffusion models double the decision space each decision occurs in by introducing a second bound. Consequently the integrator activity can race towards either: its operational space now spans between the two bounds and its starting point being located centrally at time point 0.

Depending on the model configuration, bias z can shift towards either of the bounds, hence 'biasing' the decision process by reducing the distance to the respective bound. This property is the namesake of z.

Bidirectional models operate in the realm of tasks which offer two response options and presume a decision at every trial. The umbrella term for this task paradigm is "two-alternative forced-choice task" (2afc).

A bidirectional race model models a decision taken between two choices, formalized as upper and lower bound.

A common 2afc task is the visual discrimination task, consisting of a stimulus which must be classified as belonging to one of two categories.

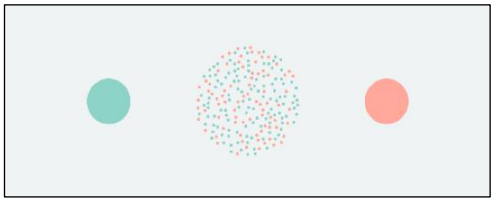

Fig. 4. Target stimulus 1 with a color coherence of 70% (blue).

The present study uses such a paradigm, wherein one ought to decide on the dominant color represented in a dot cloud (Fig. 4).

The difficulty of such visual discrimination is defined by the color coherence. High coherence represents a low difficulty as opposed to low coherence, which is substantially harder to differentiate as the ratio of the colors converges toward 1:1.

Let's model a trial of this task using the DDM framework: Succeeding the stimuli onset, a stimulus encoding period theta precedes the decision process. Subsequently, the integrator starts the deliberation of the stimuli: Depending on the informational quality of the stimulus, evidence toward one or the other choice is accumulated with rate delta. The better the evidential quality, the easier the stimulus deliberation. This leads to a higher integration rate and ultimately results in a faster decision. To use different wording: evidence for each bound competes with each other. When evidence towards one bound substantially overpowers its competitor, the integrator has an easy time deliberating the stimulus which allows for a rapid decision.

Again, this process is subject to noise, so that decisions informed by the same drift rate do not always terminate at the same time

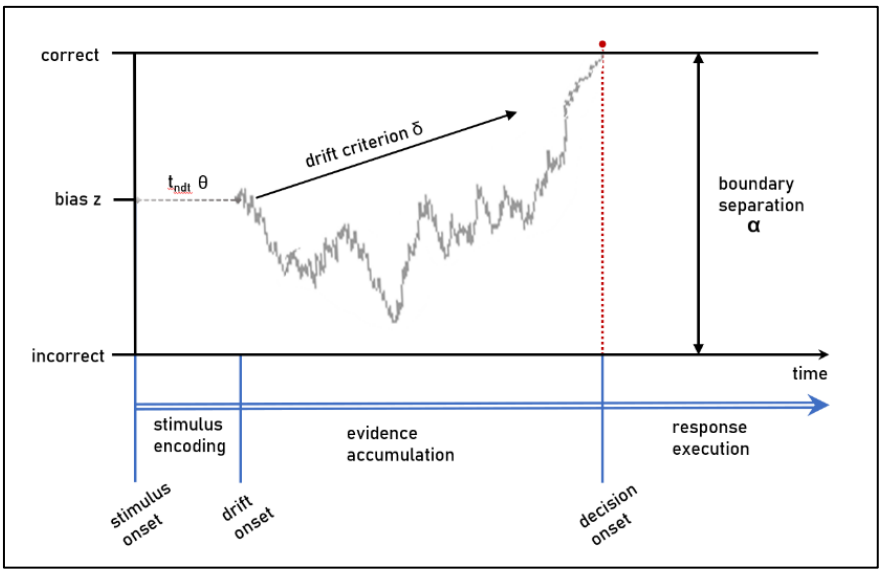

Fig. 5. Schematic of a Drift Diffusion Model decomposing a decision into non decision time and noisy evidence accumulation.

(producing RT distributions) and do not always terminate at the same boundary (producing errors). (Ratcliff & McKoon, 2008).

In this paradigm, the level of integration is subject to forces towards both bounds, resulting in upward, as well as downward directed movements. Visualized, the integrator performs a wiggly drift reminiscent of a Markovian random walk (Fig. 5). This drift is the name giving attribute of bidirectional Rise-to-Threshold models: drift-diffusion models.

The parameter delta is reconceptualized into drift rate: i.e., the rate/vector at which the drift strives towards the correct boundary. This again is determined by the quality of the sensory evidence with its lower bound being at null: Such a null-drift-rate is present while trying to deliberate an indifferentiable stimulus. I.e., a fully coherent stimulus consisting of an equal representation of either color.

Importantly, drift-diffusion models do not operate with one fixed drift rate. Rather, drift rates vary randomly across trials in a stable pattern or

probability distribution. The shape of this distribution is informed by the stimulus' evidential quality.

Each trial, a delta value is sampled from its distribution. DDM accounts for noise by this implementation of parameter sampling.

As mentioned, the drift does not always terminate at the same boundary. This occurs when noise is perceived as information and the integrator accumulates evidence toward the incorrect bound. The harder the task difficulty (the lower the coherence), the higher the probability of erring. To illustrate, consider a novel stimulus:

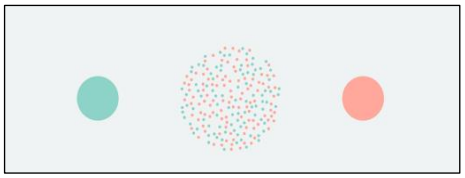

Fig. 6. Target stimulus 2 with a color coherence of 52% (blue).

Regard again stimulus 1 (Fig. 4) and stimulus 2 (Fig. 6). The former stimulus is easier to discriminate, as it provides stronger sensory evidence, whereas discriminating stimulus 2 can be considered to be more difficult, because it provides less evidential quality (a lower difference in colored dots). Hence a system would integrate the evidence of stimulus 1 faster (operating with a higher drift rate), resulting in quicker response times. Conversely, increases in task difficulty lower the drift rate, which leads to increases of average RT and a decline of accuracy rates.

Again, the quicker a decision is taken, the less noisy interference occurs to the deliberation process, producing a lower variance.

Stimulus 2 is less evident by nature, which consequently lowers the drift rate steering its deliberation process. A lower drift rate again results in more time required to reach the necessary level of evidence, leaving a larger temporal window for noise to exert influence: Response times deviate greater.

Boundary separation within drift-diffusion models define the vertical distance separating both boundaries (Fig. 5), therefore constituting the space the drift can occur in. Reducing α means that less evidence is needed to initiate a decision. Consequently, this lowers average response time, while also increasing error rate across trials.

This is a product of variability in drift rates induced by noise, which can drag the drift towards the wrong boundary and – given a low enough threshold – result in erring. Increasing α on the other hand increases the evidence needed to initiate a decision, resulting in slower response times, as well as higher accuracies.

Does this tradeoff sound familiar? The alpha parameter models the Speed-Accuracy-Tradeoff (SAT) or caution in decision making. A cautious decision will be modeled as operating with a relatively higher alpha than a careless one, reflecting the position within the SAT in decision making. Cautious decisions operate with a higher α, careless ones under a lower α.

Figure 7 depicts the same decision as figure 4 did: all parameters besides α were kept constant.

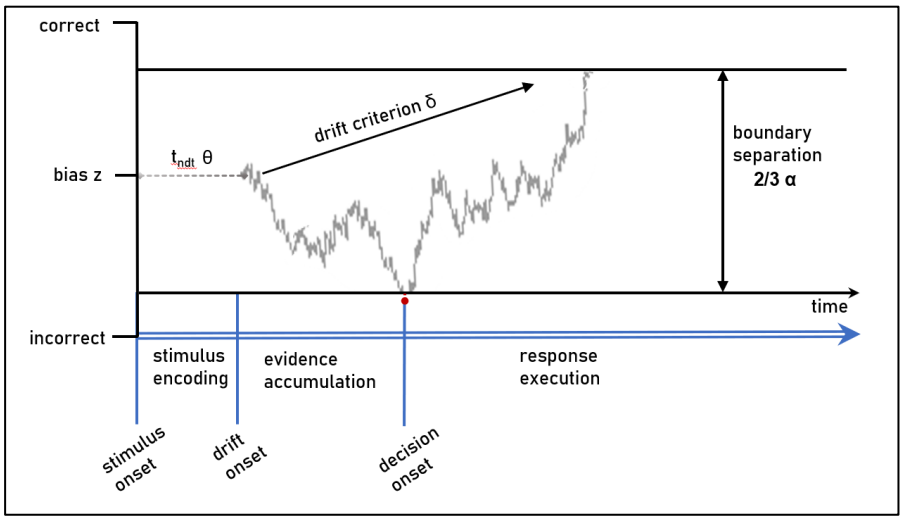

Fig. 7. Schematic of a lower alpha parameter applied to a Drift Diffusion Model. The lowered bounds facilitate the creation of errors by shortening the distance between the starting point and the decision threshold.

The reduced a value is visualized as lowered bounds – a smaller distance separating the starting point and its respective decision thresholds. In the early stage of both processes, noise substantially interferes with the deliberation process, dragging the integrator towards the wrong boundary. The system of figure 5 avoids erring, as its boundaries are high enough to continue sampling evidence and discriminate noise from actual information, eventually resulting in a correct response.

Conversely, the system depicted in figure 7 operates carelessly with a lowered boundary separation. It needs less evidence to initiate a decision: The noise-driven swing satisfies the threshold – visualized via the red marker.

It is important to note that the erroneous response was initiated substantially faster than the correct one. The decision onset is depicted as red markers set at the point where the drift reaches a boundary. This illustrates the Speed-Accuracy-Tradeoff in decision making in the drift-diffusion model. As previously outlined, this expression of caution is a cognitive strategy in decision making and belongs to the parameters of cognitive control.

Using this methodology of computational modeling, we can infer the expression caution across as well as within participants. The DDM outputs an estimate of the underlying parameters of the behavioral data. Therefore, we can not only compare RTs and accuracy rates, but also caution – drift rates of participants. Further, we can statistically test differences between parameters in both contexts.

Sign	Parameter
δ	Drift Rate/ Quality of sensory evidence
α	Boundary Separation / SAT / Cautiousness
z	Bias / Starting point
θ	Non-decision time

Fig. 8. Summary of DDM parameters.

Summary

To this end, we designed a bipartite environment, wherein reward signals differ between contexts, albeit remaining stable within context. For an optimal performance, subjects ought to raise their expressed degree of caution in one context, and lower it in the other. Contexts altered trial-by-trial, hence the window of picking up the difference in reward signals was extremely short for each exposure. Moreover, we test whether this divergence remains stable in the absence of reward signals.

The associative learning perspective on cognitive control predicts that these reward signals can be picked up via unaware mechanisms also underlying associative learning: The applied cognitive strategy would gradually become associated with its respective context via reward signals, hence resulting in a contextually conditioned expression of control parameters.

By observing that the divergence of caution remains stable in the absence of reward, we can conclude on evidence that cognitive control is indeed susceptible to the same basal mechanisms originally attributed only to associative learning.

Our study aims to provide further clarity towards the question whether cognitive control can be subject to the principles of associative learning.

Chapter 2:
Method

The present study ought to explore the extent to which conditioning of control parameters is possible within a volatile learning environment.
To this end, two contexts were introduced within our environment, namely the top and bottom half of the screen (±2° vert. visual angle, figure 8). In each context, different reward policies were used to evaluate the subject's performance. It is crucial to understand that identical task performance was evaluated differently in each context.

Materials

The experiment was performed on a computer monitor with a diameter of 17 inches, a resolution of 1920 x 1080 pixels, and a refresh rate of 60 hertz. Participants were positioned at a distance of 70 cm from the monitor. The experiment was designed in PsychoPy (Peirce 2007; 2009) and featured a visual discrimination decision making task. After an individual difficulty calibration, the task remained consistent throughout the experiment.

The survey was conducted in a dimly lit cubicle. To account for precise recording of response times in the range of µsek, a Cedrus Response Pad RB-740 was used as the input device. Participants rested their fingertips comfortably on one response key.

Participants

53 participants (37 female, aged 18–34, M=22, SD=2.42) took part in the experiment. All participants had normal or corrected-to-normal visual acuity. All subjects were students at Ghent University. They signed informed consent prior and received one participation credit in return for their participation. One participant also received a coupon for the online marketplace bol.com with a value of EUR100,- for having

acquired the highest score in the experiment. This reward was communicated in the introductory part of the survey.

Reward scheme

Each decision was classified as a. correct or incorrect, b. fast or slow and c. which context it was taken in. It follows a 2x4 matrix of possible feedback conditions (Fig. 9).

	Correct & Fast	Correct & Slow	Wrong & Fast	Wrong & Slow
Accuracy	+1	+0	-0.5	-1
Speed	+1	-0.5	+0	-1

Fig. 9. The applied reward schemes.

This resulted in correct & fast (CF), correct & slow (CS), wrong & fast (WF), and wrong & slow (WS) – brackets for each context. Both CF and WS were evaluated in the same way across contexts: +1 and -1, respectively.

The manipulation is located within the inner columns. In the accuracy condition – the one supposed to increase caution – it was optimal to prioritize correct, though slow decisions as opposed to erroneous and fast ones. However, the speed condition, which ought to decrease caution, and shift toward a speed emphasis on the Speed-Accuracy-Tradeoff, had these values inverted. Within this context, it was optimal to sacrifice accuracy to gain in speed, as the penalty was given in the correct and slow bracket (CS).

Given the same distribution of responses within an experiment, a more speed-focused strategy accumulated more bonus points in the speed-condition and adopting a more cautious strategy amassed more bonus points in the accuracy condition.

This delta in performance evaluation ought to nudge the subjects into adapting their SAT to optimize performance and maximize the gained bonus. The rationale behind this scheme was adopted from Fitts (1966).

Trial Design

Now, to incorporate volatility, the presentation of these contexts fluctuated trial-by-trial. A context was consecutively presented three times at most.

The task at hand was a perceptual discrimination task, wherein subjects had to classify stimuli into one of two categories. Subjects had to choose between two response options each trial, constituting a two alternative forced choice task paradigm (2afc).

Namely, participants ought to decide which color is majorly represented within a cloud of dots as mentioned before. Within this environment, we ought to condition subjects to adopt a cautious strategy when the stimulus is presented in one context, and conversely a less cautious one in the opposing one by varying the reward contingencies between contexts.

Stimuli

Every trial started with the presentation of a fixation cross on a lightly gray background (PsychoPy rgb = [0.88, 0.91, 0.91]) for a fixed duration of 500ms. Its location varied between three levels (top, center, bottom) across the experiment and was informative of the location the following stimuli would appear in. After 500ms, the cross was replaced by the task stimuli, consisting of two-colored flankers and the target stimulus, which was to be classified. The target consists of 200 colored dots. Each dot is colored in either cyan (PsychoPy rgb = [0.11, 0.67, 0.56]) or orange (PsychoPy rgb = [1, 0.32, 0.22]).

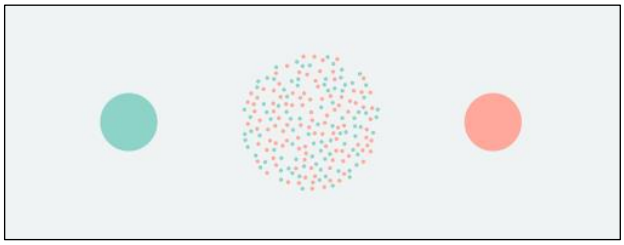

Fig. 10. Target stimulus with a coherence of 55% (blue).

The relative distribution of the colors represents the difficulty of the task and will be further referred to as stimuli coherence. Participants ought to visually analyze the stimuli and decide on the predominantly represented color. A low coherence, say 55%, entails a narrow delta between the amounts of dots and hence raises the difficulty to discriminate the dominant color. In this example, the dots are colored representing a 110:90 ratio (Fig. 10). Vice versa, a high coherence, say 70% represents a low degree of difficulty to discriminate the dominant color.

The target appears along two horizontally flanking stimuli (±6.5° visual angle respectively). These remained fixed on their respective horizontal position during the whole experiment, so the participant could habituate to their positions and focus on classifying the target. The participant ought to visually analyze the stimuli and decide on the dominant color of dots.

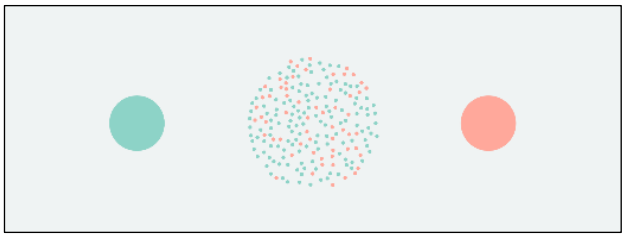

Fig. 11. A target with a coherence of 70% (blue).

One then presses the button, which stands for the majority-flanker. A time window of 5000ms was provided for each decision. Upon response, or after the deadline was reached, the feedback was presented in place of the target. Depending on the stage of the experiment, the feedback was either masked (‘###’), numeric (+1/ -1/ +0/ -0.5) or a string (‘correct’/ ’incorrect’), as well as ‘te laat’ (Dutch ‘too late’) appearing when the deadline was reached.

Procedure and Block Design

The present study ought to explore the possibilities of conditioning control parameters within a volatile environment. To this end, an

experiment consisting of three major epochs was engineered. These can be further grouped into two periods of interest (Fig. 12).

Participants were randomly, though in equal numbers, assigned to one of four groups, counterbalancing a) the color arrangement of the flankers, and b) the assignment of reward policies to context-locations.

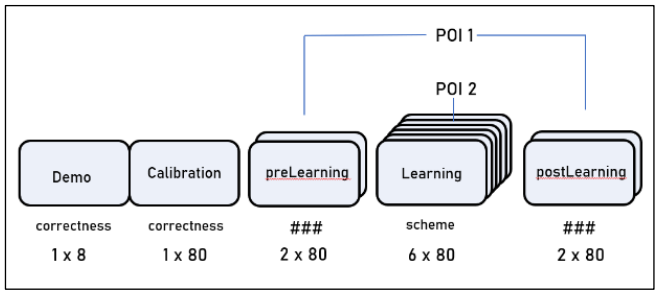

Fig. 12. Block-level design of the experimental procedure.

Demonstration Block

An introductory text followed informing the participants about the workings of the task, the temporal expense of the study as well as the possibility to have a self-timed break between blocks.

To get to know the task procedure, subjects completed a short demonstration block of 8 trials with moderate difficulty (randomly sampled coherence values around 0.6). Feedback of correct/incorrect was displayed and the stimuli remained vertically centered.

After the demo another instruction block followed stating that after the following block, the accumulation of points will commence, followed by an emphasis on a gift card worth EUR100,-. Furthermore, the competitive aspect was emphasized by stating that only the best player will earn that reward. Additionally, it was recommended to already take the following block seriously, as it may provide a competitive advantage. Subsequently, the procedure of the following blocks was explained.

Throughout 10 blocks of 80 trials each one may accumulate a score with the goal being to give correct answers as quickly as possible to maximize one's score.

Training Block: Difficulty Calibration

To ensure a comparable strain at decision making between all subjects, the task difficulty was individually calibrated. This was done using a staircase method (Wetherill & Levitt, 1965), specifically an adaptive fixed-step-size calibrated according to García-Pérez (1998). The rationale of this method is to let all participants go through a common task difficulty calibration in order to let them converge towards a cross-participant comparable difficulty level.

Within this block, performance was solely evaluated as correct or incorrect. All stimuli were presented in the vertically centered position throughout trials. A goal of 75% accuracy was set for all participants to converge on. The coherence of the target represented the task difficulty: a high coherence constitutes an easy difficulty and vice versa.

Pre-Learning Phase

After the difficulty calibration, participants completed two blocks of 80 trials each with their individual difficulty level. This epoch introduces the trial-by-trial fluctuation of the contexts, as well as masked feedback display: '###' (Fig. 13). No true reward was assigned, as the necessary values for evaluation were not yet calibrated. However, participants were told that they were being rewarded, albeit their reward being hidden. This setup is further referred to as masked feedback. Within these two blocks, an RT-threshold was calibrated, which delineated fast from slow responses. Therefore, all correct RTs were logged, sorted and the 60th percentile of this array was used as the discriminatory value, further referred to as response time criterion (RT criterion). This epoch is of high importance for the latter

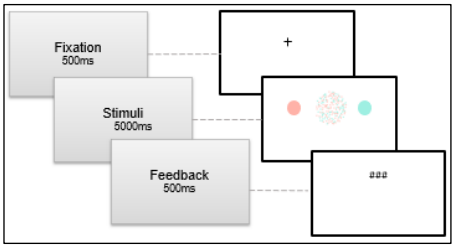

Fig. 13. Schematic of the task set up during pre-learning.

analyses, as it logged the subject's performance before the onset of the learning process.

Learning Phase

The subsequent learning phase then implemented both calibrated values and fluctuating contexts: the difficulty as well as the rtc to evaluate each decision according to the rewards scheme. Within 6 blocks, participants completed 480 trials in total in both contexts, which ought to nudge the participant into adopting a different reward schedule in each.

In accordance with the associative learning perspective on cognitive control, the delta in reward signals was supposed to be picked up by the control system informing caution within the decision. By registering reinforcement exerting caution in one context, the control system seeks to optimize its performance. Therefore, it adopts a cautious strategy – an accuracy emphasis – in one context. Conversely, it registers inverse reward signals in the opposing context, and adapts strategy accordingly toward a speed-emphasis.

This drive to maximize bonus and consequently maximize the chance of winning EUR100,- was the incentive for the control system to find a way to optimize performance.

It's important to keep in mind that one aim of this study was to find out whether this difference in reward could be picked up by the subjects as well as the extent of their adaptation to it. These 480 trials constituted the window in which this was supposed to take place. Furthermore, these trials serve the purpose of forming associations between the cognitive strategy and the context it was exerted in. This network would iteratively strengthen, as it yields reward signals for using the right strategy – high or low caution, respectively. Conversely, the absence of reward signals e.g., using the 'wrong' contextual strategy would incentivize the system to find a way to optimize its bonus, therefore randomly adapting, until

picking up on the right strategy and from this point, converging toward the right end on the pole, towards the right strategy to implement.

Post-Learning Phase

Given the creation of a network associating the cognitive strategy with its respective context, the post learning phase ought to answer the question of stability after reward signals cease.

The post learning epoch again consisted of two blocks à 80 trials with masked feedback display and fluctuating contexts. The difference to the pre-learning epoch is that subjects were still rewarded in the background. This was not possible in the early phase as the values for the reward policies had not been calibrated yet. This change happened in the back end only, subjects had the same experience in the third epoch as they did in the first.

Awareness test

Lastly, a questionnaire was implemented to check for subject awareness of the study design. This questionnaire consisted of an open-question part, followed by a multiple-choice section. Former inquired whether a difference in reward schemes was noticed in the two contexts and if so, to briefly outline it.

The latter showed three options to choose from: first was that some participants were rewarded at the top of the screen for being careful / giving mostly correct answers and were rewarded in the bottom half for being careless / giving especially quick answers. The second option stated exactly the opposite as statement 1 and indifference in reward schemes constituted the third option.

Procedure Summary

Regard again figure 12. The learning epoch, period of interest 2 (POI-2, 6 x 80 trials) covers the development of control parameters within the learning phase. This period tracks a. *if* and b. *how* strategy

distributions develop upon onset of reward signals. (Orange highlight in Fig. 14).

Period of interest 1 (POI-1, 2 blocks à 2 x 80 trials) compares the parameter distributions of epoch 1 and 3, to test a. *if*, and b. *how* strategy distributions have changed from their baseline. Figure 14 illustrates the two POIs. Here, 2 blocks were condensed into epochs (E) for readability.

We hypothesize that each subject exhibits the same level of cautiousness for both contexts within epoch 1 – the baseline.

Progressing through the learning phase, the subject's expression of caution gradually starts to diverge between contexts, reaching its peak distance at the very end of epoch 2. In epoch 3, this difference in cautiousness ought to shrink due to absence of reward signals, but nonetheless remain observable.

Data analysis

Both periods of interest were separated and analyzed individually. Period of interest 1 enveloped the pre- and post-learning phase (blocks 2,3,10 & 11). This period was characterized by masked feedback and compares the development of drift-diffusion model parameters before, and after the learning phase.

Period of interest 2 captures the trajectory of parameter evolution throughout the learning phase (blocks 4-9). Within this period, reward feedback was displayed. This reward ought to condition the participants to adopt diverging expressions of caution via reward signals. Epoch concatenates two blocks and hence an epoch featuring 80 trials in each condition is created. This transformation was

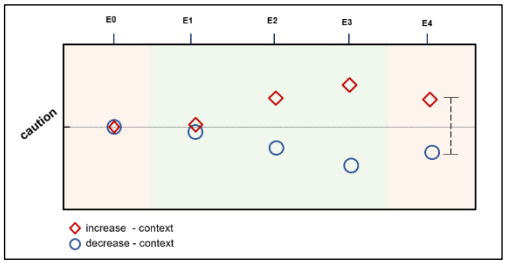

Fig. 14. Anticipated divergence in assumed caution across the experiment. Two blocks were collapsed into one estimation block (E) for the sake of readability.

crucial to have a block large enough for the subsequent parameter estimation. (Lerche, 2016). In this notation POI-1 envelops epochs 0 & 4, and POI-2 includes epochs 1, 2 & 3.

Statistical analysis was conducted with the statsmodels 0.13.5 module and always featured $\alpha = 0.05$. Visualizations were created using Matplotlib (Hunter, 2007), as well as the Seaborn library (Waskom, 2021) in Python 3.8.8.

Data preparation

RTs deviating from the mean by more than 3σ (absolute z-score ≥ 3; RT > 1.885) were identified as outliers. (see Berger, 2021 for discussion). No negative z-scores passed the threshold of z = -3. Furthermore, over half of the outliers occurred in the pre-learning phase (57%), suggesting that they are mainly due to learning effects within the new task.

RTs of outliers as well as trials featuring null-RTs (n=6) were identified. An exclusion of these trials wasn't feasible, as we needed 80 trials per block and participant for the upcoming analysis. In total, 795 RTs were corrected at individual level by adding 3σ to the mean of the respective block. This clipping was done for 1.8% of total observations. Demonstration and calibration block were omitted from the analysis.

Period of Interest 1

The nonparametric Wilcox signed rank test was employed to analyze RT-distributions within POI-1. This method was preferred to paired T-tests because of the violation of the normality-assumption. Wilcox does not assume normality, but rather similarity of distribution shape across groups as well as the given attribute of sphericity.

First, the between-context-variance of the pre- and post-learning epochs (E0, E4) was analyzed, followed by a between-context analysis of block 4 (B4) (last pre-learning block) and B10 (first post-learning block). This way we intended to capture the effect of the rewarded learning phase on post-learning blocks in relation to the pre-learning baseline.

In the beginning epoch (0), the distributions should not vary, as no learning was applied. Hence, epoch 1 was tested bidirectionally (two-sided): H0: $\mu^{E0}_{increase} \neq \mu^{E0}_{decrease}, |$ and respectively: H1: $\mu^{E0}_{increase} = \mu^{E0}_{decrease}$

For epoch 4, we hypothesized that the increase-condition will have distribution shifted to the right, characterized by slower values and conversely, the decrease-condition a higher density of faster values. Therefore, we applied a one-sided test: H0: $\mu^{E4}_{increase} \leq \mu^{E4}_{decrease}$. And respectively: H1: $\mu^{E4}_{increase} > \mu^{E4}_{decrease.}$

Additionally, a subject-level 2x2 rm-ANOVA (Girden, 1992) was applied to dissect the factors condition and the block on RT distribution between the two epochs.

Period of Interest 2

Dealing with correlated datapoints, a repeated measure 2x6 ANOVA was conducted on response times with condition (increase/decrease) and block as within-subject factors. Given a large enough n, ANOVAs are known to be robust against violations of normality (Lix, 1996). In this case, each of the 2x6 conditions consists of 2120 samples.

Due to the absence of an effect, employing post-hoc tests to determine directionality was refrained from.

Parameter Estimation with the Drift-Diffusion Model

To estimate DDM parameters, we used hierarchical Bayesian estimation. This has the advantage of individual fits being bound by group-level distributions (Wiecki, 2013). Hierarchical DDM was chosen because the estimates of parameters are allowed to vary trial-by-trial, affording the capability to model fluctuations of neural or psychological variables within a process. This attribute makes the HDDM (HDDM 0.6.0, pyMC 2.3.6) package incredibly useful for the modeling of decision-making processes. The HDDM uses Markov-Chain-Monte-Carlo-sampling (MCMC) for generating posterior distributions over

model parameters. The incorporated Bayesian statistics in the HDDM-backend allows the quantification of not only the most likely parameter-value, but also its distribution. Such an approach generates valuable knowledge about the associated uncertainty of a parameter-estimate through its variance. Due to the hierarchical nature of the HDDM-architecture, estimates for individual subjects are constrained by group-level prior distributions. Subsequently, individual parameter estimates lose statistical independence and inference is only meaningful at group-level. In particular, the model was specified such that on each trial t, the bias remained fixed at alpha/2. Alpha (boundary separation), delta (drift rate) and theta (non-decision time) were set to be based on the categorical estimator coded to be dependent on condition.

By convention, the expected outlier percentage was set to 5%. For the estimation, the original outliers were again introduced to the dataset, as the HDDM removes them by default. Besides, no manipulations were applied. 4000 samples were drawn from this model, discarding the first 1000 samples as 'burn-in'. Hypotheses in this case followed the same logic as the ones underlying the Wilcox paired rank test: The baseline epoch 0 was hypothesized to have no variance across conditions and was tested bidirectionally.

H0: $\mu^{E0}_{increase} \neq \mu^{E0}_{decrease,}$ and respectively: H1: $\mu^{E0}_{increase} = \mu^{E0}_{decrease}$

RT distributions of epoch 4 were expected to have a higher mean in the increase group a.o.t. the decrease group.

H0: $\mu^{E4}_{increase} \leq \mu^{E4}_{decrease,}$ and respectively H1: $\mu^{E4}_{increase} > \mu^{E4}_{decrease}$

One problem of DDM fitting occurs when the respective chains do not converge to the same stationary distribution and the MCMC algorithm subsequently does not sample from the actual posterior. The R-hat statistic compares between-chain variance to within-chain variance to control for this issue. Throughout all chains (parameters of both POIs for each participant), R-hat values of ~1 ($\bar{x}$ = 1.002) were observed.

Additionally, the Geweke statistic comparing means and variances of sequences from both head and tails of chains returned True, further indicating successful convergence.

SAT-Graphs

Recall figure 1 from chapter 1, depicting the SAT of Bert the neanderthal illustrating the proportion of edible goods in his basket in relation to the time spent – a so called SAT-graph.

On SAT-graphs, time is coded on x, whereas y codes the proportional correctness or accuracy. Black annotations depict data points, which were taken in temporal steps. It is important to understand that the graph does not provide information about these step sizes, as it depreciates them to ordinality.

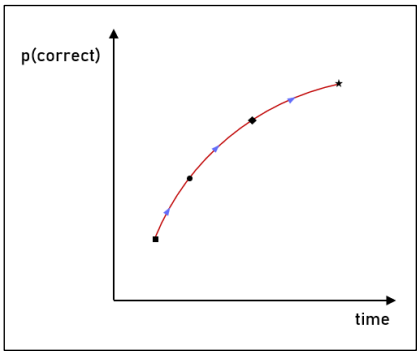

Fig. 15. Generic SAT-Plot coding accuracy on the y-axis and time on the x-axis.

The green sample was followed by the blue sample, as denoted by the blue arrow. We know that between these samples a shift towards a more accurate behavior occurred. The same behavior applies to the next samples, resulting in the starred sample which has the highest expression of caution – taking a lot of time and yielding high accuracies. This introductory figure is smoothed to an increasing function for clarity, but SAT-graphs in real world data are seldom so docile. They often change direction and cross each other; hence it is important to keep in mind to follow the order of the connections (depicted with arrows).

Chapter 3:
Results

Participant exclusion

No participants were excluded from the analysis, leaving the sample size at 53 subjects. Although detecting four significant outliers for calibrated coherence and response time criterions, no participants overlapped. Furthermore, no noteworthy deviations in mean RT or accuracy rates across participants were observed. Additionally, RT-distributions of each participant were inspected, which all showed the expected right-skewed shapes that response times characteristically follow. These findings suggest that all participants took the experiment seriously and performed accordingly.

Descriptive Analysis

The relation of both coherence (task difficulty) and response time criterion (slow-fast-threshold) in the participants' total score was first analyzed.

As expected, there was a significant positive correlation between RT criterion and total score ($r=0.53^{***}$). This is sound, as a high criterion raises the chances of reaching the +1 reward, which is independent of the currently active scheme.

However – and interestingly, there was a non-significant negative correlation between the coherence and total score ($r = -0.18$). Upon exclusion of subjects with outlier criterions ($r > 1.56$sec, $n=4$), this correlation turned out to be driven by same and corrected to a marginal $r = 0.03$. This fortifies the notion

that the coherence calibration worked as intended and successfully normed the task to individual skill-level. Importantly, the winner was no outlier in either value.

Fig. 16 shows a 3D-scatterplot illustrating the relation of both coherence and response time criterion to the total score. The red circle highlights the winner (RT criterion = 1.19 sec, coherence = 56%, score = 523).

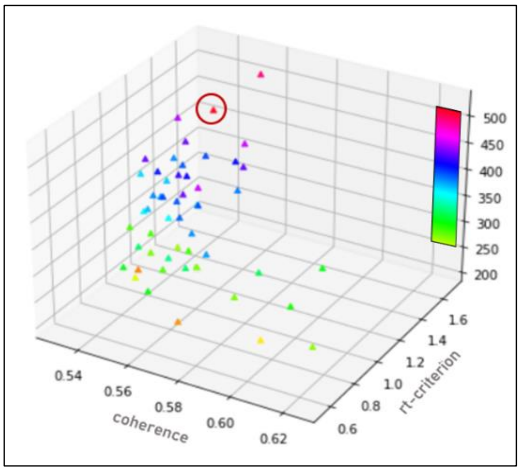

Fig. 16. 3D Scatterplot depicting the relation of RT criterion, coherence and the attained score.

On average, the response time criterion was calibrated at 960ms (std=230ms), and coherence at 56% (std = 0.2%).

From visual inspection, RT-distributions did not seem to vary across groups and only marginally across blocks, (Fig. 17). Importantly, no divergence in response behavior is observable across conditions. When the RTs shifted, they did so equally in both conditions.

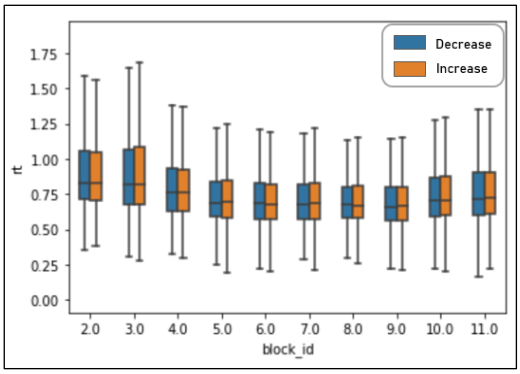

Fig. 17. Boxplots depicting RT distributions across the experiment for both conditions.

Awareness Test

Subjects were randomly, although in equal numbers, assigned to a context mapping, which assigned reward policies contexts. Mapping 1 (N=26) indicates that the increase reward scheme was attributed to the upper location. Vice versa, mapping 2 (N=27) indicates the bottom location being rewarded by the increase scheme.

At the end of the experiment subjects were asked whether they noticed any pattern in the reward contingencies, followed by a multiple-choice questionnaire consisting of three choices. Subjects' responses are depicted in Fig. 18. Correct answers are highlighted in green.

Mapping	Statement
1	
10	Accuracy in upper location
5	Accuracy in lower location
11	Indifference
2	
14	Accuracy in upper location
3	Accuracy in lower location
10	Indifference

Fig. 18. Results of the awareness test for each condition mapping across subjects. Green highlight indicates the subjects' mapping.

Across both mappings, the same pattern was observed. Namely, both statement 1 and 3 were preferred a.o.t. statement 2. Exhibiting the same tendency across mappings suggests that the mapping-counterbalancing did not influence the statement-choice or awareness of subjects. Furthermore, this indicates that subjects giving a correct answer were not aware of the nature of the rewards schemes, but supposedly subject to sequence effects of the questionnaire.

Period of Interest 1

POI-1: Inferential Statistics

In general, RTs were characterized by significant divergence from normality. Moreover log, square root or cubic-root transformations did not achieve normality according to Shapiro-Wilk. (Mishra et. al., 2019). QQ-Plots further exemplify this by visually displaying deviance from the normal distribution. The closest approximation to normality was achieved by log transformations, albeit remaining significantly deviant from it.

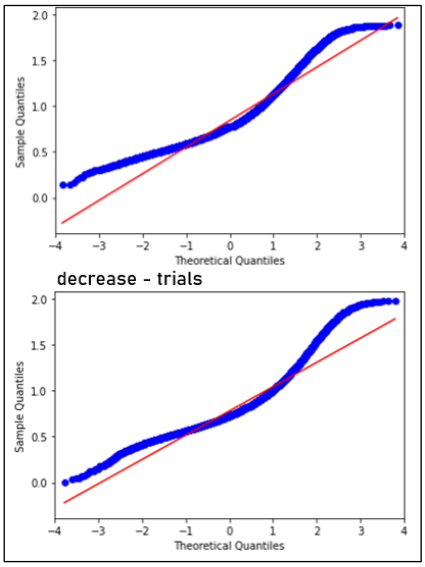

Fig. 19. QQ-Plots for increase (upper) and decrease (lower) condition. Both suggest the violation of normality.

A violation of the assumption of sphericity makes the ANOVA, as well as Wilcox test highly susceptible to type II error, hence

Mauchly's Test of Sphericity was applied to test the data. Results suggest that the assumption of sphericity is not violated, as $chi^2 = 3.67$, $W = 0.93$ and $p = 0.599$.

The comparison of conditions within epoch 0 (pre-learning) via the Wilcox signed rank test shows the expected non-significance ($z = 4368133$, $p = 0.48$).

It was hypothesized for epoch 4 to be characterized by a right-shifted distribution in RTs resulting in a higher mean. This hypothesis was falsified, as $z = 4554729.0$ and $p = 0.18$.

The 2x2 rm-ANOVA likewise resulted in no significance regarding the factor condition. Factor estimation block (estim) turned out to be highly significant (***), but this was expected, as it models between-epoch effects. Our interest lay in the within-block factor condition, which remained absent.

Anova

	F Value	Num DF	Den DF	Pr > F
condition	0.1854	1.0000	52.0000	0.6685
estim	169.1429	1.0000	52.0000	0.0000
condition:estim	0.0294	1.0000	52.0000	0.8646

Fig. 20. rmANOVA for Period of Interest 1.

POI-1: SAT-Plot

Figure 21 visualizes the behavior of each POI-1 block and condition within SAT-space. This figure was scaled for the sake of readability. Blocks of the pre-learning phase are situated on the right, and post learning on the left side.

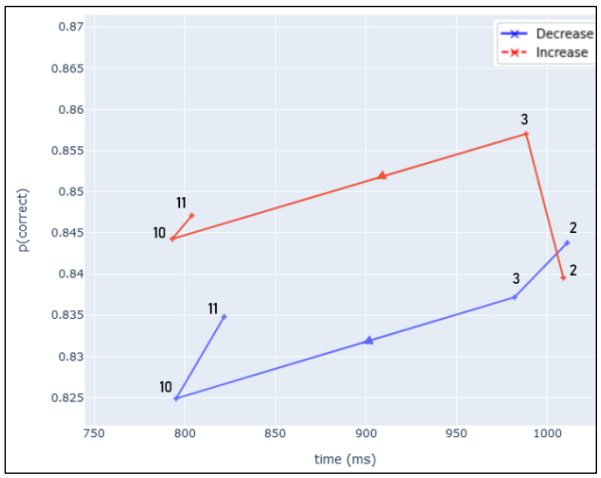

Fig. 21. SAT-Plot for blocks pre- (2,3) and post- (10,11) learning phase.

Importantly, six learning blocks separate points 3 and 10, so the connecting line must be interpreted with caution. Rather, it depicts the

vector with which both epochs traveled across SAT space, modulated by the learning phase.

Across the POI, a speedup of 195ms occurred, while generally losing accuracy. The decrease group lost more than the increase group, albeit this difference is marginal (delta= 0.9%). Increase group's initial increase in accuracy (+1.7%) is not interpretable, as it occurred before learning commenced. This graph ought to visualize the shift within the SAT space induced by the learning block. We anticipated a divergence between groups to be perceived with the increase condition moving to a locus of higher accuracy and slower RTs, and the decrease condition sacrificing accuracy to gain speed, moving to a locus in the lower left.

While it holds true that the increase group generally exhibits a higher accuracy, both conditions begin in a similar area (as anticipated), but also end up in the same general space, only separated by 1% in accuracy with identical pace. No divergence occurred. This again illustrates that the learning phase failed to diverge conditions within SAT space.

POI-1: DDM Parameter Estimation

Posterior distributions as output by the DDM tell the same story: No divergence in boundary separation occurred. In both conditions, subjects shifted to a less cautious strategy throughout the experiment. The Bayesian nature of this model allows for statistical analysis, testing whether the distributions behaved as hypothesized. As stated in the analysis section, we hypothesized that alpha values increase in the accuracy condition, whereas decrease in the speed condition. Latter hypothesis was accepted (***) – however, the same behavior

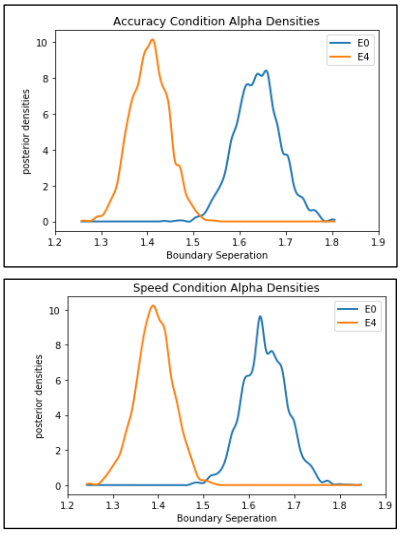

Fig. 22. Estimated alpha posterior densisites prior (E0), and post (E4) learning phase for increase (upper) and deacrease (lower) condition.

occurred in the accuracy condition, which ought to increase alpha values. Hence, the former hypothesis is rejected (p=0.99). Likewise, the interaction between block and condition was also insignificant (p=0.54).

Period of Interest 2

POI-2: Inferential Statistics

Again, Mauchly's Test of Sphericity was applied to the data, resulting in the upkeep of the assumption of sphericity (chi^2 = -370, W = 1626, p =1). Statistical analysis within the learning phase was conducted via repeated measure ANOVA modeling condition and block as within-subject factors. Factor *block*, (=progress throughout the experiment) turned out significant (***) in explaining variance of the response times. However, this is most likely due to effects unrelated to the nature of the study. We ought to test for the influence of the factor *condition*, which remains vastly insignificant.

Anova

	F Value	Num DF	Den DF	Pr > F
condition	0.1431	1.0000	52.0000	0.7068
block_id	21.5490	5.0000	260.0000	0.0000
condition:block_id	1.4227	5.0000	260.0000	0.2163

Fig. 23. rmANOVA results for learning blocks 4-9 of Period of Interest 2.

POI-2: SAT-Plot

Within the learning phase, the subjects generally gain 133ms in speed, while sacrificing 2% in accuracy. The same moving pattern is mirrored across conditions. The averaged transition from b4 to b5 (86ms) is most pronounced, accounting for 65% of total shift in response time, while sacrificing merely 0.8% of accuracy. This acceleration of RTs while keeping the same accuracy is most likely due to learning effects, which drive a more efficient performance. Progressing to b6,

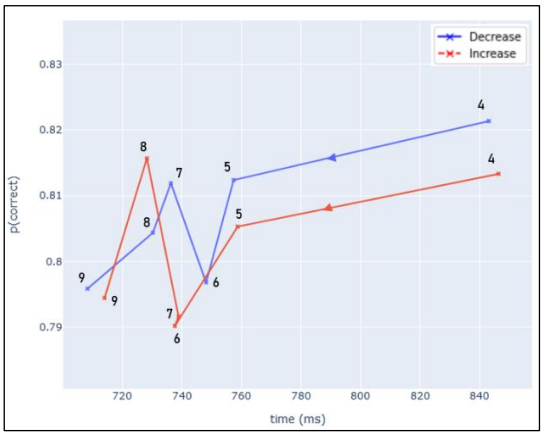

Fig. 24. SAT-Plot of learning blocks 4-9 for both conditions.

both conditions keep the same pace while losing accuracy, followed by a divergence in b7, although in the 'wrong' direction: The decrease group shoots up in accuracy (+1.4%), while the increase group performs identically to the previous block, resulting in a delta of 2%. Subsequently, the increase group performs the largest jump in accuracy from b7 to b8 (+2.4%), while becoming marginally faster (-10ms). Finally, both groups converge to a similar position for b9.

In general, both conditions exhibited the same pattern within the learning phase, becoming less cautious while gaining speed in the process. The aforementioned divergence cancels out by plotting POI-2 as epochs instead of blocks (Fig. 24). Notably, the movement within SAT space only occurs to the left: The learning phase is characterized only by acceleration, no slowing occurred between blocks.

POI-2 Feedback Distributions.

Feedback has only been given from block 4 on, wherein the necessary values for the reward scheme were calibrated. Hence, only POI-2 is subject to this analysis, as the post-learning block lacks its reference. Figure 25 depicts histograms of feedback distributions across, and within-blocks. Both +1 and -1 feedback categories were excluded, as they were

Fig. 25. Feedback distribution for POI-2 considering only the ambivalent classifications. The labels p.q are to be read as: the first integer (p) representing the block index, whereas the second one coding the condition (0 ~ increase; 1 ~ decrease).

assigned identically across conditions. It was expected to observe an increase of trials classified as wrong & fast (WF) in the decrease scheme and an increase of correct & slow (CS) trials in the increase scheme. For the decrease scheme, the blue WF category is advantageous. Conversely, trials evaluated by the increase scheme benefitted from the orange CS feedback.

Two neighboring bars represent a block, indicated by the first number of the index (1.x). The second number following the dot represents the condition: x.0 reads as increase (accuracy) trials, whereas x.1 reads as decrease (speed) trials. This way, an inter- as well as intra-block comparison of feedback distributions was possible.

It was hypothesized to observe a divergence in these values as the system progressively begins to adapt. The increase condition of block 6 was expected to consist of a respectively larger proportion of SC, a.o.t. the decrease condition, which ought to consist of more WF trials in respect to SC. Again, no such effect can be observed.

The only noteworthy change in feedback is from block 4 to block 5, reducing the percentage of SC trials across conditions, keeping WF constant. Notably, only marginal changes within blocks can be observed. This again fortifies the notion that adaptation did not occur. Ratios remained the same across the learning phase instead of diverging within blocks.

POI-2 DDM Parameter Estimation

Like POI-1, a similar pattern of parameter shifts across conditions can be observed. From e1 to e2, participants reduced their level of caution, which remained at the same level for e3.

Considering the unambiguous absence of an effect, statistical analysis via rm-ANOVA was refrained from.

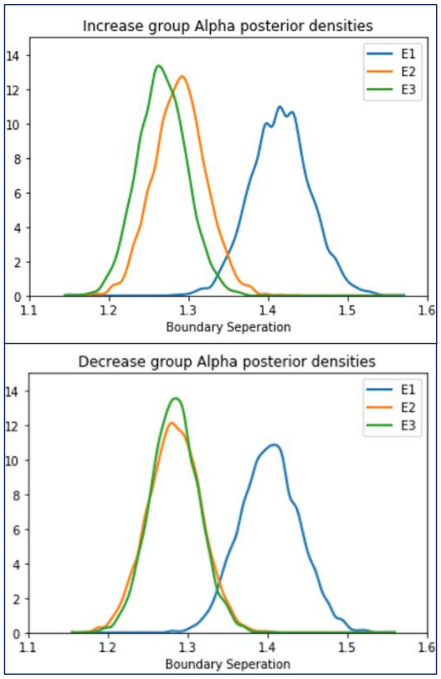

Fig. 26. Estimated alpha posterior densisites throughout the learning phase for increase (upper) and decrease (lower) condition.

Relation of SAT Graphs to Alpha Densities

I'd like to dedicate the last section of this chapter to the elegant relation of Speed-Accuracy-Tradeoff graphs and DDM posterior estimates. Figure 27 depicts the POI-2 epochs (6 blocks condensed into 3 epochs), starting from e1 (first learning epoch) to e3 (last learning epoch).

This has the same connotation as in the posterior density plots of figure 26. Viewing both side by side illustrates the relation of SAT graph and alpha density: An epochs position within the SAT space defined by two coordinates (accuracy & response time) directly translates to the posterior density of alpha. Alpha posterior density seems to combine the information given out of average RT & accuracy into one metric. Importantly, as mentioned in the DDM section, HDDM parameter estimates are only meaningful in group level

comparison. Likewise, it would be impossible to approximate the alpha density of a single point on SAT space without reference samples.

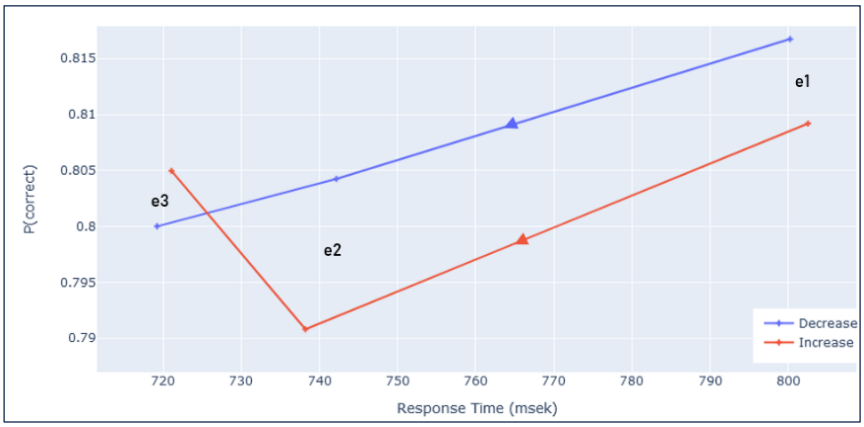

Fig. 27. SAT Graph of POI-2 estimation blocks (e) of the learning phase for both conditions. For readability, blocks 4 & 6, 7 & 8 and 9 & 10 were collapsed into e1, e2 & e3, respectively. Color encodes the condition.

The distance traveled within SAT space between e1 and e2 is larger in respect to the distance separating e2 and e3, which suggests that a larger shift in caution occurred moving from e1 to e2 compared to the shift from e2 to e3. This is the pattern across groups. Figure 27 depicts precisely the same pattern of shifting caution: The posterior density shifts stronger from e1 to e2, a.o.t. shifting from e2 to e3. The distance crossed mirrors the shift of alpha distributions. Alpha is a unidimensional metric, conveying the information of two-dimensional positioning on the SAT graph.

Now, regarding the movement between conditions, one can observe that within the increase condition, a relatively larger shift occurs from e2 to e3 (+1.5% accuracy & -17ms), than within the decrease condition (-0.43% accuracy & -23ms). From e2 to e3, a larger distance is covered in the increase condition, a.o.t the decrease condition. Now, regarding the alpha densities for e2 & e3 for both groups, one detects this very same pattern: Alpha density of e2 and e3 in the increase group deviate more than in the decrease condition.

Summary

Figure 28 visualizes the movement of epochs and conditions across the whole experiment.

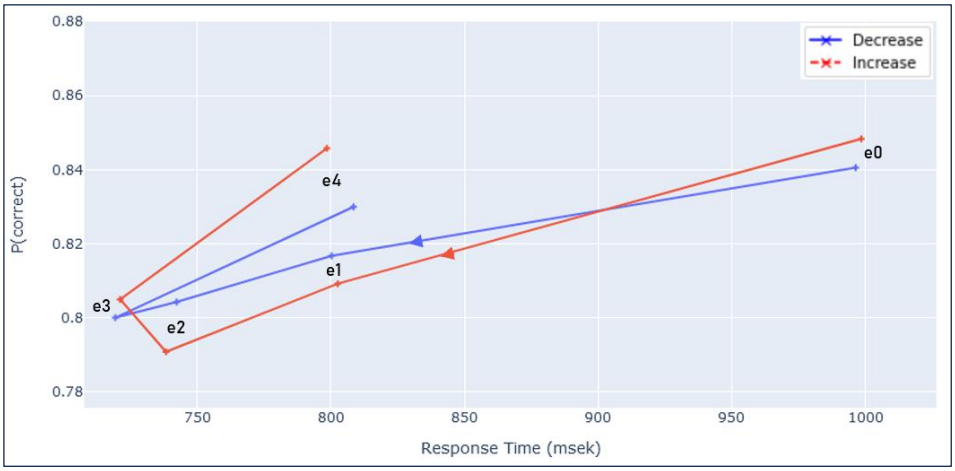

Fig. 28. Graph depicting all estimation blocks (e) across both conditions within SAT space. Color codes the condition. Both conditions trace the same general pattern. Little variance across conditions can be observed.

Throughout, the same pattern is mirrored across conditions. Epochs 0, 1, 2 & 3 continually accelerated in response time (-277ms), while sacrificing accuracy (-4.2%) in the process. This trend was reversed in the transition from e3 to e4, returning to almost baseline accuracy (+3.7%) while slowing 83ms in relation to e3. Notably, e4 converges to almost baseline (e0) accuracy (delta = 0.6%) while responding almost 194ms quicker. The transition to masked feedback seems to drive up RTs which entail higher accuracy rates, generally shifting towards more cautious behavior. This suggests that the masking of feedback – or the absence of a numerical one – reduces urgency to respond and drives decisions to higher caution via the slowing of response times. Although this study was not designed to test this notion, this finding might point to an interesting avenue to explore further. Importantly, epochs exhibited the same movement across conditions.

To conclude, conditions did not diverge in caution and all findings unequivocally point to the absence of an effect.

Chapter 4:
Limitations

Limitations of our study can be clustered into issues regarding the frontend, namely the reward policies and the presentation of reward to the subject and the calibration of the response time criterion in the backend.

Form of reward

Notably, participants in the present study were not subject to reward signals per se, but rather to a proxy of it. Amassed score raised the chance of winning a gift card but the score/ or feedback itself was not informative to the participant on their chances of winning. Neither trial-by-trial, nor at the end of the experiment, where the score was displayed. The score only becomes meaningful for the analyst comparing all participants – outside the scope of each singular participant. This way, the subject itself did not have insight in the determinant metric of his performance, and his feedback becomes a distant proxy of the anticipated chance of payoff, rather than actual reward.

Presumably, this rather indirect relation between displayed value and its meaning might have been too vague for the reinforcement system to interpret as reward. This argument refers to block 1 of the discussion, covering the problem of salience. In future approaches, one could transform the reward presentation into absolute units where the displayed value directly translates into financial gain or loss: the feedback '+1' would actually gain the participants a penny, and a -1 would take one away.

This way one could provide more tangible reward feedback, which could arguably create a more salient signal to be picked up by the reinforcement system.

In conclusion, the modus of reward presentation might have been a substantial factor driving the observed null effect. As we ought to manipulate

the archaic system of reinforcement learning, the implemented reward signals might have been too vague to be picked up by reinforcement mechanisms.

Reward Scheme

As will be discussed in section General Discussion, the reward scheme constituted the very core of our experiment design. Participants ought to be manipulated into adopting different expressions of caution by being exposed to two contexts, each associated with an individual reward policy. The results of our simulated incentive maps suggest that these did not provide sufficient incentive for the system to adapt diverging strategies between contexts. A model outlining hypothesized reasons can be found in chapter 5. The proposed framework of incentive maps provided important insights into the workings of various reward policies and helped us sharpen our understanding of our study design. Further, it points out promising directions to take in creating a reward scheme which satisfies the balance between salience and unaware processing.

Awareness Test Bias

As outlined in section General Discussion, awareness check responses might have been subject to recency or primacy effects, nudging participant to choose the first or last option. To ensure that the reward scheme really remained undetected, this effect needs to be controlled for: future studies ought to cycle the statement options, i.e., via balanced Latin squares (Lewis, 1989).

Response time Criterion

Lastly, I'll address the implementation of the response time criterion. This metric was calibrated in blocks 2 & 3 of the experiment (chapter 2: Pre-Learning) and was used to determine whether a response was classified as slow or fast via our reward scheme. We hypothesized that | fatigue and learning effects would cancel out while progressing through the experiment, but this turned out to be highly variable between participants. As outlined in chapter 3, we observe a rather large variance between participants in their RT criterion (std = 225ms by $\bar{x}$ = 960ms)

a.o.t. the coherence metric (std = 2%, $\bar{x}$ = 56%). This circumstance becomes visible when plotting the percentage of trials per block which have passed the threshold (= were classified as slow). Neither the mean, nor standard deviation varied significantly across conditions and were hence aggregated in figure 29 for the sake of readability.

The graph depicted in figure 29 generally follows the same inverse quadratic pattern as the response time plots do (Fig. 17): declining during learning and rising in the post-learning phase.

The calibration pipeline shall be re-engineered to drive the variance induced by block factor beneath the variance induced by subject factors. Moreover, the RT criterion ought to control the effects occurring in transitions from feedback to masked feedback. Ultimately, this eradication would lead to between-condition effects becoming more salient. To this end, one could implement a dynamic threshold (sliding window), which is informed by the average RT of the last *k* response times and weighted by the participants accuracy. This could account for training, as well as fatigue effects on an individual level, as the threshold adapts to the subject's performance and presumably ensuring the comparability of response time criterion passings across subjects.

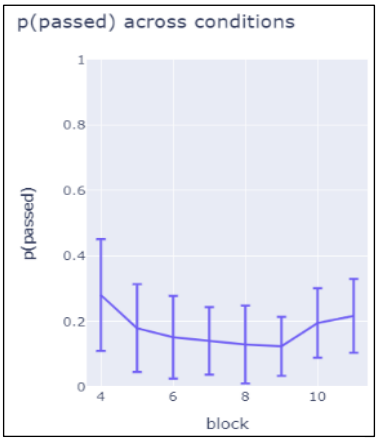

Fig. 29. Averaged proportion of trials reaching the RT criterion across blocks (subject average).

Chapter 5:
General Discussion

The following section explores possible explanations for the observed absence of an effect. Three hierarchical steps will be discussed, which are hypothesized to be necessary for an adaptation to occur.

The present study ought to investigate whether humans are capable of picking up a subtle difference in reward policies unwittingly while alternating between two contexts. The contexts incentivized cautious and reckless decision making, respectively via differing reward policies. Crucially, these contexts fluctuated trial-by-trial, keeping the temporal window of exposure to each policy very narrow. Building upon this notion of reward difference detection, the study ought to explore whether humans are able to adapt cognitive strategies 'on the fly' – informed merely by the presented context. Ultimately, the study ought to test whether such divergence in cautiousness a. occurs and b. remains stable in the absence of reward signals.

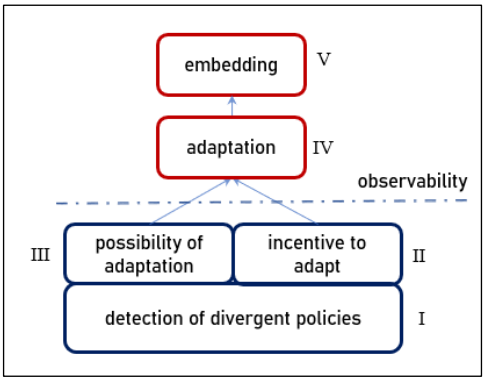

Fig. 30 Proposed hierarchy of prerequisites needed to be met for an adaptation to occur.

The task design was centered around three attributes of associative learning: reward-sensitivity, contextual dependence, and unawareness. These were implemented by a bonus system, exposure to two contexts and a fast-paced task, respectively.

I identified three mechanisms with hierarchical dependence, which presumably constitute the prerequisites for adaptation to occur. From now on the subject of the investigation will be referred to as "system". The term represents the reinforcement mechanisms exerting influence over the control system informing the subjects' behavior and therefore performance.

To achieve the desired outcome of embedding control parameters into a context (V), the system had to be conditioned to adapt different strategies for either of the contexts (IV). This adaptation requires two processes to take place:

In order to adapt, the possibility of reward optimization by a modification of strategies must be detected/realized (III). Such detection is made substantially more difficult if the system starts in a 'hybrid' parameter configuration amidst the two poles of the SAT, not receiving the benefit of neither a purely cautious, nor incautious strategy.

Building upon this detection, the swing in control parameters must be sufficiently incentivized (II). The subjective cost of adaptation must be outweighed by the respective anticipated reward – in our case the chance of winning a gift card, as well as its value.

Lastly and fundamentally, the divergent reward policies must be salient enough to be picked up by reinforcement mechanisms (I), albeit still being kept subtle enough to remain in the absence of awareness.

Only if prerequisites I, II & III (blue outlines in Fig. 30) are fulfilled the reward sensitivity of the system activates and sets the foundation for contextual adaptation to occur. Crucially, we have no insight into the fulfillment of each. The desired outcome would only become observable by adaptation, which needs all requirements to be satisfied. This circumstance is made even more complex by the unwitting nature of this adaptation process. We cannot ask the participants what would incentivize them to adapt, as it is no aware process per definition. In order to disentangle the problem of adaptation in such a fluctuating environment, each requirement must be explored individually. Ultimately, the challenge is to investigate, as well as to initiate a mechanism you only get feedback on if it works. Beneath that threshold of adaptation, our most powerful tools are informed guesses, or tinkering, as one could say.

Notion 1 – no detection of differing reward contingencies.

Adaptation can only occur when the system recognizes an advantage in pursuing it. Foundational to such incentive is the detection of differing reward signals between contexts.

The major problem at this stage is to engineer a reward scheme salient enough to be picked up by reinforcement mechanisms, while remaining subtle enough to stay 'under the radar' of conscious processing.

Our data suggests that the latter requirement was met, as subjects did not display awareness of the reward policies according to a self report. Participants across mappings displayed the same pattern in awareness-statements: The upper context seemed to have been associated intuitively with a more cautious approach in both mappings, regardless of the actual mapping.

These results suggest that the manipulation remained in the absence of awareness. However, it remains unclear whether the difference in reward policies was salient enough to be picked up.

Notion 2 – No opportunity for optimization detected.

Moving upstream, the next hurdle is to showcase room for reward optimization. A system can start the experiment with an alpha-configuration, which can be described as hybrid or a central position amidst both poles. No incentive of adapting strategies manifests, as no advantageous strategy is promoted. Consequently, the system does not detect room for optimization, resulting in the continuance of the incumbent hybrid strategy. In order to

engage optimization, the system has to detect a disadvantage in its current environment and the possibility to optimize it.

Let's assume that a system commences the experiment with an instruction-informed high degree of caution. In the increase context, it behaves optimally and gradually it'll detect a disadvantage of its current strategy when deciding in the opposing context. This would trigger an exploration of possible

strategies in the suboptimal context resulting in the gradual convergence to decreased caution in same.

Essentially, the system must detect a disadvantage to explore for optimization. Once optimized by diverging the expressions of caution, the system would shift into exploiting this configuration. Consequently, this would lead to contextual parameter embedding mediated by reoccurring reward signals. In this vein, this subset of the adaptation problem verges into the realm of the Exploration-Exploitation-Tradeoff.

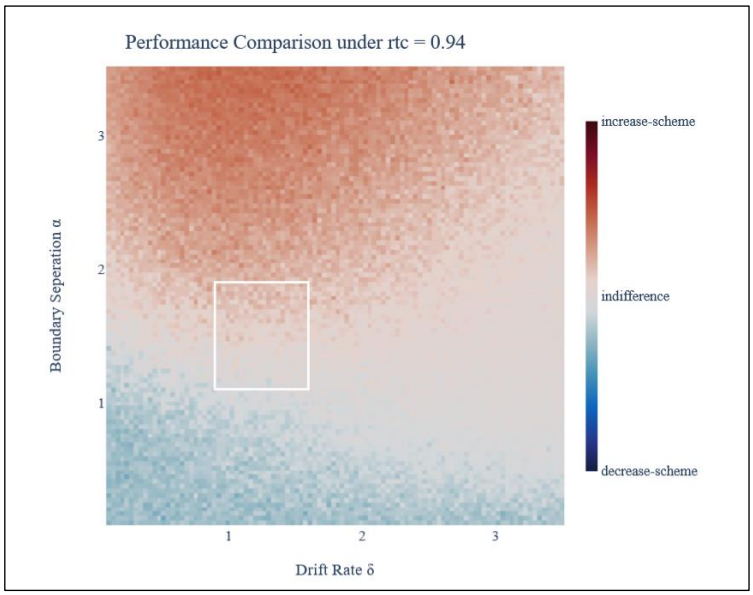

Fig. 31 Incentive map reflecting the experimentally applied reward policies. Red color reads as the parameter set being more profitable for the increase scheme and vice versa for blue color. The white rectangle frames the parameter space participants operated in.

Figure 31 shows a heatmap capturing performance differences of simulated trials evaluated by both policies. (This is outlined in detail in chapter 6). In this context, the term 'performance' is used as the amassed bonus across many trials, not the speed or accuracy of the decisions taken (i.e., within decision set performance).

The x and y axis code boundary separation and drift rate, respectively. Each unit of the heatmap contains the normalized reward difference of the same decision, evaluated by opposing schemes.

As these simulated decisions also account for noise, trials characterized by differing RTs as well as accuracy rates are created.

A negative value of this difference is coded as blue, which interprets as the decrease-scheme turning out dominant in that particular set of parameters. Vice versa, a red coding (a positive value) reads as the increase scheme being more advantageous. White and close to white tiles depict areas of indifference: no reward scheme turned out to be dominant; or put differently, one configuration of parameters led to comparable bonus outcomes in both schemes.

In this way, a metric was created to compare which policy is more advantageous to a system when taking a decision informed by a particular set of parameters. Importantly, the evaluation parameters (policies) used in this simulation are identical to the ones used in the main body of the experiment. The RT criterion of 0.94 constitutes the average value of our subjects.

The upper section of figure 31 covers an area characterized by high values of boundary separation/caution, which is rewarded by the increase scheme. This results in a stronger red coloring, as the system accumulates more points by being evaluated according to the increase scheme as opposed to the decrease scheme. A delta in performance of up to 40% emerges. The increase scheme ought to reward highly cautious decision making and the upper area captures decisions informed by this very parameter expression of high alpha. Conversely, the lower part of the map displays the area of operating under a low degree of caution, resulting in the decrease scheme being more rewarding: A low degree of boundary separation amasses more points and hence dominates the performance-delta by up to 30%. However, these values are unattainable by participants.

Notably, the white rectangle depicts the area which our subjects operated in. This observed space is mainly populated by white tiles, depicting indifference between schemes. Furthermore, the main body of participants started off with an alpha of 1.6 and shifted further down to 1.3 across the experiment. This

ribbon of alpha values is characterized by a performance difference ranging between one and 15%.

Incentive maps suggest that building block 2 of the pipeline was not satisfied. In our configuration, the reinforcement system would have had to pick up a marginal incentive, which I deem unlikely to have occurred. Building upon these results, we conclude on the notion that too little incentive was given to have initiated an exploration, which renders the subsequent adaptation of SAT parameters unattainable. The used reward policies provided marginal incentive to initiate a reinforcement-driven adaptation of cautiousness.

The introduced simulation approach constitutes a powerful tool for exploring possible reward schemes, as well as their utility when governed by different RT criterions. Furthermore, it enables us to extend our investigation beyond the bounds of observed expressions of parameter configurations. This notion as well as the modeling of various reward schemes will be continued in chapter 6. In this case, the simulations provided rich insights into our experimental design by showing too small a difference in reward signals between our policies. This suggests that subjects likely did not pick up a difference in schemes and subsequently did not detect the possibility of optimizing their performance.

Notion 3 – Adaptation Cost outweighs Payoff.

Even if I and II were satisfied, a third prerequisite building block needs to be catered for, which is subject to the notion of inherent costliness of cognitive control and its component of flexible adaptation in particular. We theorize that mainly two costs are to be considered: firstly, the cost of resources expensed in the shift and secondly, the temporal cost of shifting parameter configuration.

The former one consists of economic consideration regarding the subjective cost of adaptation: the exultance of mental effort is weighed against the subjectively perceived likelihood and quality of the anticipated payoff.

Payoff in this case depends on a. the likelihood of winning and b. the sum of money anticipated to be won. Both factors combined constitute the provided incentive, which is faced by the costliness of adaptation.

The second factor of costliness – opportunity cost of time – is characterized by the surplus in RT, which is needed to shift parameter settings. This surplus would mount onto the response time additively. These additional milliseconds can be determinant for the categorization into slow or fast. To further investigate this notion, computational models of statistical optimization within decision making can be applied, as reviewed by Bogasz (2007).

Again, regarding our reward policies, such crossing of the RT criterion is detrimental in the decrease scheme (possible feedback [-.5, -1], speed condition), whereas less so in the increase scheme (possible feedback [0, +1], accuracy condition). Paradoxically, the decrease context would likely penalize an adaptation because of the temporal cost the switch in strategy would entail.

In summary, the time-, as well as effort-related cost of rapid adaptation must be compensated by the perceived payoff. Only then will the system initiate the effortful adaptation process.

This third stage constitutes the last building block of our model characterizing the underpinnings of rapid control adaptation. I hypothesize, as mentioned, that all three must be satisfied for the reinforcement system to pick up the exploration process towards optimization.

Conclusion

Despite inconclusive results this study provided valuable insights for a further approach of the research question. The applied experimental design did not initiate a context dependent divergence in caution. However, we attribute this absence to a malfunctioning experimental design, rather than to the impossibility of achieving said divergence.

Interestingly, an effect of masked feedback was observed, which seemed to have enlarged both RTs and accuracy rates. Likewise, a shift to the top-right

(accuracy-domain) within the SAT-space was observed when transitioning from the last feedback block to masked feedback phase. This might open an interesting alley for future research investigating the role of an active reward display vs. a masked one while instructing equal evaluation in the backend.

The main contribution of this study is an array of valuable insights for the design of studies targeting this archaic reinforcement mechanism and its activation for regulating control parameters. With information sampled along every step of this journey I ideated a model aiming to explain the observed null-effect by decomposing the problem into hierarchically structured building blocks. Doing so allows for the individual investigation of each block, which will lead to the engineering of the optimal study design. Notably, this operates under the assumption that trial-by-trial control conditioning is possible. This approach of optimizing the study design is only one avenue to take for future research.

However, one ought to keep in mind that this whole problem might not be solvable. Potentially, cognitive control cannot be conditioned with such short windows of exposure to reward signals. Following this more pessimistic notion, another avenue opens up for approaching upcoming research, namely manipulating the windows of exposure to reward signals. Studies using blocked designs have managed to achieve the anticipated conditioning of control. Subsequently, it would be interesting to explore that space between block-, and trial-wise shift of reward policy. A mini-block of say 10–20 trials for each policy could be implemented to explore at what block size threshold the possibility of conditioning ceases. This second avenue would take a functioning blocked experimental design and iteratively lower the block size until no conditioning can be observed.

These avenues are by no means mutually exclusive but constitute two approaches needed to be taken to sharpen the understanding of the adaptation problem within volatile environments.

It is established that a block-wise adaptation occurs, which begs the question of where the threshold lies. The present study communicates that a trial-by-trial adaptation did not occur. If after thorough experimental re-engineering this absence of adaptation remains, one can assume that it indeed is not possible.

Now, I pose that these findings represent two poles on a spectrum of reward exposure. The present study using window sizes of ≤ 3 represents one pole. Opposing are studies using blocked design (nt~40, i.e., Braem, 2017) which achieved control conditioning. The space between 3 and 40 trials of reward exposure constitutes the spectrum. Now, how far can one decrease window size while still observing an effect? It would be interesting to see where that line is drawn.

Further I believe that this point does not constitute the end of the journey but rather a beginning from which one can embark on the exploration of the underpinnings of control conditioning. Regarding the complexity of the cognitive system, it will most likely not be as simple as decreasing block size of the same experimental design up until the point of no adaptation and concluding to have found the answer. Furthermore, it would be even more interesting to explore which modifications to study design could drive this threshold even lower. Presumably, more refined measures need to be taken from that point on to further decrease block size while keeping the adaptation in the absence of awareness. Subsequently, the outlined building blocks will once again become relevant, as each needs to be optimized for the respective task design. I hypothesize the role of incentive to gain more weight the lower the block size becomes.

In the upcoming section, a computational framework will be introduced which targets the incentive provided for switching control configurations via opposing reward policies.

Chapter 6:
Incentive Mapping

We recognized that the incentive to shift along the SAT axis is hard to capture. A substantial factor to this problem is the fact that we are trying to influence an unaware mechanism. This problem essentially translates to investigating the role of incentive in reward sensitivity within associative learning. Applying this to our case, I theorized that this problem can be decomposed into I. creating detectability, II. displaying the possibility of optimization and III. providing the incentive to adapt accordingly.

To provide insights into this cluster, a framework was engineered, which formalizes 'provided incentive' as the performance difference resulting from comparable decisions being evaluated by opposing schemes (Fig. 32).

	Correct & Fast	Correct & Slow	Wrong & Fast	Wrong & Slow
Accuracy	+1	R	P	-1
Speed	+1	P	R	-1

Fig. 32. Abstracted reward policies.

If decisions informed by a given set of parameters result in more bonus in scheme 1 as opposed to scheme 2, this parameter set is reinforced by the former. However, an advantage in one scheme constitutes an equal disadvantage in the opposing one. We argue that such difference in performance relates positively with the provided incentive to switch between parameter constellations.

Method

This approach utilizes the simulation library of the HDDM module (Wiecki, 2013; Python 2.7), which allows us to simulate decisions informed by arbitrary parameters. Namely, a set of n samples of decisions taken under a fixed set of parameters is generated, which is then evaluated by two reward schemes.

Notably, these schemes are variable in both ambivalent brackets (CS & WF). The ambivalence lies within the fact that the correct & slow (CS) category yields a reward in the increase scheme, and conversely a penalty in the decrease scheme. The same logic applies to the wrong & fast (WF) category. Penalty and reward values range between 0 and +-0.7. Both CF and WS remain fixed at +1 and -1, respectively.

As the ambivalent category values are mirrored in the opposing scheme, I hypothesize the provided incentive to diverge alpha values to be captured by varying the weight of their contribution towards performance. This results in prioritizing accuracy (prefer CS over WF), or vice versa, prioritizing speed (prefer WF over CS), simply due to these respective strategies offering more reward.

Furthermore, a custom RT criterion was implemented as well. When this criterion is reached, the possible feedback cuts down to either correct & slow (CS) or wrong & slow (WS). WS always translates to -1, but the CS category can be either a penalty or a bonus – depending on the applied scheme. Within the increase scheme, it is advantageous to pass that threshold, unlike so in the decrease scheme.

For a particular set of parameters, an averaged difference in performance (quantified by total score) was generated by the evaluation of the performance of trials via two schemes. These simulated decisions were informed by the same parameters (alpha = 1.8, delta = 1.2).

Each decision yields a binary correct/incorrect, as well as the response time. These metrics are subsequently fed into the reward scheme, which outputs a feedback value ranging from -1 to +1. If the decision was either correct & fast or wrong & slow, the feedback is identical. However, if the output is classified as either of the middle brackets, the schemes evaluate differently. Say, the decision fell into the slow correct category (RT > 1.12sec): the increase scheme outputs a reward (R) of +0.7 and the decrease scheme penalizes (P) with -0.58.

Notably, this framework allows us to test arbitrary values for the classifications in question. We let this decision occur k times respectively, which are then evaluated by the increase and decrease scheme. In this case, k represents 500 decisions. Again, the parameters informing these decisions remain fixed. We then simply determine the difference of the respective average bonus per decision of both schemes via subtraction.

If the difference in the average bonus between schemes is positive, the increase scheme granted a larger bonus a.o.t. the decrease scheme. Conversely, a negative value is interpreted as a higher reward when evaluated by the decrease scheme.

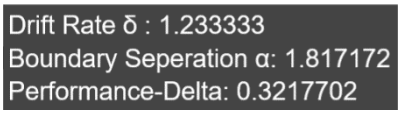

Fig. 33. Encoding of one unit within the incentive map.

Now regard figure 33. The outlined scenario of 1000 decisions taken under a fixed set of parameters result in a 32% higher bonus in the increase scheme a.o.t. the decrease scheme. Figure 34 depicts the parameter space ranging from 0.1 to 3.5 for both boundary separation and drift rate. Blue areas are to be interpreted as a larger utility provided by the decrease scheme and conversely, red areas depict the space wherein the increase scheme was more rewarding.

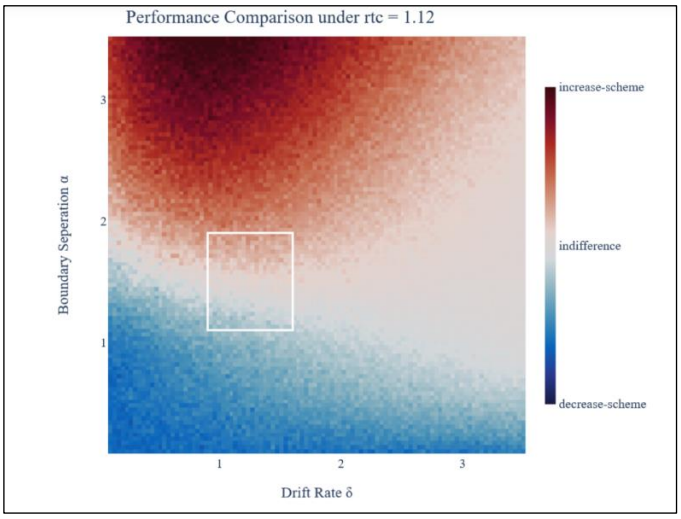

Fig. 34. Incentive map evaluated by reward = +0.7, penalty = -0.58 and classified by a RT criterion of 1.12. The white rectangle frames the observed parameter space participants operated in.

The parameter set informing Fig. 34 gives rise to a landscape which saturates in incentive along the upper and lower bound of the observed parameter space (white rectangle, Fig.34). Such an increase in color saturation points towards a divergence in utility for the respective strategies. Red areas point towards an advantage for deciding cautiously, and vice versa recklessly in blue areas.

Generally – and cycling back to the workings of the DDM – both a lower alpha and a higher delta lead to decisions being taken faster and result in a higher probability of the trial succeeding the RT criterion and thus being classified as fast. Regardless of correctness, a faster decision is classified as either correct & fast (CF) or wrong & fast (WF). The latter yields a reward in the decrease scheme, whereas a penalty in the increase scheme.

This results in a blue dominance in the low-alpha space, enveloping all delta values. Conversely, slower decisions have a higher chance of benefitting from the CS (correct & slow, rewarding) – category of the increase scheme, resulting in a strong dominance of red in the upper value space of alpha. Decisions take longer but have a higher likelihood of being correct. The saturation (respective dominance of a particular scheme) is to be seen as the correlate for the provided incentive.

One of the advantages of this approach is to capture humanly unattainable parameter space. Importantly, the observable parameter space, i.e., the range of parameters stemming from the data is depicted as the white rectangle both in Fig. 31 and Fig. 34.

Within this 'observable' space, the difference between-schemes ranges from -20% to +20% but is separated by a larger large ribbon of white indifference-space (abs(delta) <= 10%).

I theorize that an optimal scheme would be characterized with a narrow ribbon of indifference, neighbored by strong (highly saturated) scheme-dominated spaces.

The narrow ribbon would entail a relatively small adjustment of alpha being needed for moving to the opposing pole of the SAT, which again leads to profitability for this particular trial. Furthermore, the larger the needed adjustment of control parameter is, the costlier this swing arguably becomes. Hence, keeping the swing small would reduce the cost of adaptation (III).

Fig. 35. A desirable demarcation of opposing strategies characterized by strongly saturated incentives separated by a narrow white ribbon of coding indifference.

Figure 35 relates back to block II outlined in chapter 5, which discusses the importance of detecting the possibility of optimization, initiating the exploration process.

If a system starts within a broad, white ribbon (Fig. 31) the likelihood of detecting the opportunity of optimization remains quite low. However, if a system starts in a highly saturated – say red – region, the likelihood of optimizing parameters for the opposing context rises, as an advantage of operating in the increase context is mirrored as disadvantage for the decrease context. Consequently, it seems logical that the detection of such disadvantage would initiate exploration and subsequently lead to the gradual divergence in caution between contexts.

Limitations

A substantial downside of the current implementation lies in its backend, which generates 500 samples twice, which are then evaluated by both schemes. One set of decisions is fed through the evaluation pipeline which yields the average bonus. This process is not only redundant but introduces noise while not providing any additional information. This problem is visualized by generating an incentive map, which codes null for the ambivalent brackets.

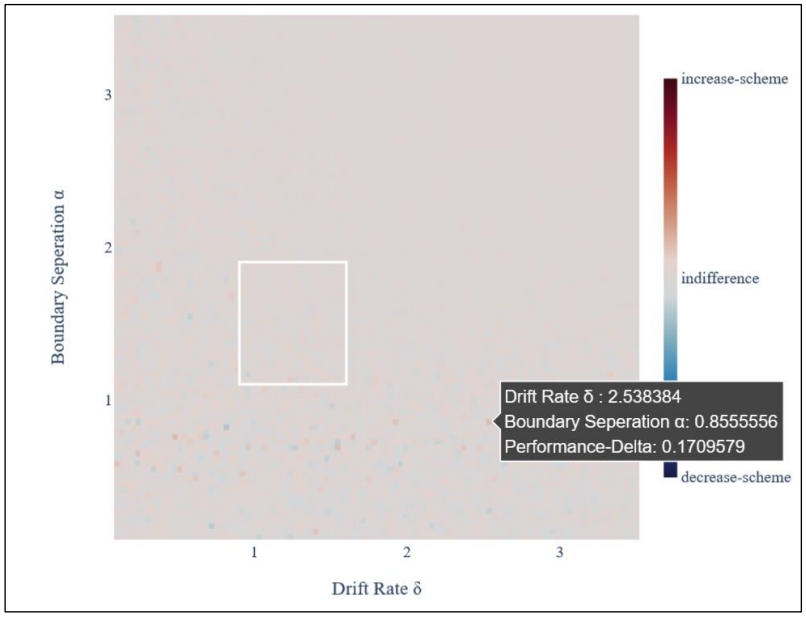

Fig. 36. Incentive map informed by null-value ambivalent brackets. As the utility of both wrong & slow and correct & fast trials (-1, +1 respectively) cancel out, no advantageous area for any strategy manifests. The subject has no utility in changing decision parameters. All color in this map stems from noise in the simulation algorithm.

The performance metric only considers the +1 and -1 feedback, which is mirrored in the scheme. Doing so, all scheme-induced difference in performance is eliminated and consequently all variance in performance is due to noise adding unnecessary and avoidable confounds to the data.

In Fig. 36, performance differences of up to 17% are observable – again, all differences in performance are purely noise-driven. This variance is

pronounced in the low-alpha space, which is characterized by rapid responses and a high probability of erring but spans across the entire plane.

Outlook

Despite these implementational teething issues, I regard the logic and functions of this framework as sound – and more importantly scalable through its modularity.

For instance, additional parameters can be implemented, such as collapsing bounds (Cisek, Puskas, & El-Murr, 2009) accounting for the rise of urgency with time passing, implemented as gradually lowering alpha-bounds. This would result in less integrated evidence needed to initiate a decision at timepoint t, a.o.t. timepoint t-1.

Furthermore, one could integrate the z (bias) parameter into the model, investigating the incentive provided to adjust the starting point of the decision process. In conclusion, this framework constitutes an approach of visualizing the incentive provided by opposing reward schemes to explore and consequently exploit respective parameter configurations. I used this method to validate the results stemming from real data, as well as to ideate possible approaches to work around the limitations found in the main body of the study. Further, I created a model which decomposes the process of adaptation into three sub-components, allowing for an individual investigation of each. Incentive mapping constituted a substantial role in building intuition behind the workings of said model.

I believe that the presented incentive map grants its greatest value in building intuition for the workings of reward in DDM modeling. This educational aspect is further amplified by its customizability.

As Samuel Karlin said, the purpose of models is not to fit the data, but to sharpen the questions.

Likewise, the incentive map serves this purpose while further providing visual guidance for building intuition on the workings of reward in decision making.

Bibliography

Abrahamse, E., Braem, S., Notebaert, W., & Verguts, T. (2016). Grounding cognitive control in associative learning. *Psychological Bulletin*, 142(7) 693-728. doi:10.1037/bul0000047

Bates, Mächler, Bolker, & Walker. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, 67(1), 1–48. doi:https://doi.org/10.18637/jss.v067.i01

Baumeister, Bratslavsky, Muraven, & Tice. (1998). Ego depletion: Is the self a limited resource? *Journal of Personality and Social Psychology, 74(5)*, 1252-1265.

Berger, & Kiefer. (2021). Comparison of Different Response Time Outlier Exclusion Methods: A Simulation Study. *Frontiers in psychology*. doi:doi.org/10.3389/fpsyg.2021.675558

Bogacz R. (2007). Optimal decision-making theories: linking neurobiology with behaviour. Trends in cognitive sciences, 11(3), 118–125. https://doi.org/10.1016/j.tics.2006.12.006

Bogacz, Hu, Holmes, & Cohen. (2010). Do humans produce the speed-accuracy tradeoff that maximizes reward rate? *Quarterly Journal of Experimental Psychology*, 63(5), 863-891. doi:10.1080/17470210903091643

Botvinick, & Braver. (2015). Motivation and cognitive control: from behavior to neural mechanism. *Annual Review of Psychology*, 66, 83-113.

Botvinick, & Cohen. (2014). The computational and neural basis of cognitive control: Charted territory and new frontiers. *Cognitive Science*, 38(6), 1249-1285. doi:10.1111/cogs.12126

Botvinick, Braver, Barch, Carter, & Cohen. (2001). Conflict monitoring and cognitive control. *Psychological Review*, 108(3), 624-652. doi:10.1037/0033-295X.108.3.624

Braem, & Egner. (2018). Getting a grip on cognitive flexibility. *Current Directions in Psychological Science*, 27(6), 470-476. doi:10.1177/0963721418787475

Braem, & Hommel. (2019). Executive functions are cognitive gadgets. *Behavioral and Brain Sciences*, 42, e173. doi:10.1017/S0140525X19001043

Braem, Hickey, Duthoo, & Notebaert. (2014). Reward determines the context-sensitivity of cognitive control. *Journal of Experimental Psychology: Human Perception and Performance*, 40(5), 1769-1778. doi:10.1037/a0037554

Braem, S. (2017). Conditioning task switching behavior. *Cognition*, 166, 272-276. doi:10.1016/j.cognition.2017.05.037

Braem, Xu, Liefooghe, & Abrahamse. (n.d.). Learning when to learn: Context-specific instruction encoding. doi:10.31234/osf.io/7tvx3

Braver. (2012). The variable nature of cognitive control: a dual mechanisms framework. *Trends in Cognitive Sciences, 16*(2), 106-113. doi:doi.org/10.1016/j.tics.2011.12.010

Cisek, Puskas, & El-Murr. (2009). Decisions in Changing Conditions: The Urgency-Gating Model. *Journal of Neuroscience, 29*(37), 11560-11571. doi:10.1523/JNEUROSCI.1844-09.2009

Dambacher, Hübner, & Schlösser. (2011). Monetary incentives in speeded perceptual decision: effects of penalizing errors versus slow responses. *Frontiers in Psychology*. doi:10.3389/fpsyg.2011.00248

Eisenreich, Akaishi, & Hayden. (2017). Control without Controllers: Toward a Distributed Neuroscience of Executive Control. *Journal of Cognitive Neuroscience*, 29(10), 1684–1698.

Fiedler, McCaughey, Prager, Eichberger, & Schnell. (2021). Speed-Accuracy Trade-Offs in Sample-Based Decisions. *Journal of Experimental Psychology: General*, 150 (6), 1203-1224. doi:10.1037/xge0000986

Fitts. (1966). Cognitive Aspects of Information processing: Set for Speed vs. Accuracy. *Journal of Experimental Psychology, 71*(6), 849-857.

Gershman, Pesaran, & Daw. (2009). Human reinforcement learning subdivides structured action spaces by learning effector-specific values. *The Journal of Neuroscience*, 29(43), 13524–13531.

Giesen, & Rothermund. (2014). You better stop! Binding "stop" tags. *The Quarterly Journal of Experimental Psychology, 67*, 809-832. doi:http://dx.doi.org/10.1080/17470218.2013

Girden, E. R. (1992). ANOVA: Repeated Measures. Newbury Park, CA: Sage. https://doi.org/10.4135/9781412983419

Heitz. (2014). The speed-accuracy tradeoff: history, physiology, methodology, and behavior. *Front. Neurosci.* doi:10.3389/fnins.2014.00150

Hübner, & Schlösser. (2010). Monetary reward increases attentional effort. *Psychonomic Bulletin & Review*, 17 (6), 821-826.

Hunter. (2007). Matplotlib: A 2D graphics environment. *Computing in Science \& Engineering, 9*(3), 90-95.

Kimberly, & Braver. (2011). Monetary incentives improve performance, sometimes: speed and accuracy matter, and so might preparation. *Frontiers in Psychology*. doi:10.3389/fpsyg.2011.00325

Lamy, Alon, Carmel, & Shalev. (2015). The role of conscious perception in attentional capture and object-file updating. *Psychological*, 5805–5811.

Law, & Gold. (2009). Reinforcement learning can account for associative and perceptual learning on a visual-decision task. *Nature Neuroscience*, 655–663. doi:10.1038/nn.2304

Lerche, & Voss. (2016). Model Complexity in Diffusion Modeling: Benefits of Making the Model More Parsimonious. *Frontiers in Psychology*, (7), 1126 - 1142. doi:10.3389/fpsyg.2016.01324

Lix, Keselman, H. J., & Keselman, J. C. (1996). Consequences of assumption violations revisited: A quantitative review of alternatives to the one-way analysis of variance F test. *Review of Educational Research*, 66(4), 579–619.

Luck, & Ford. (1998). On the role of selective attention in visual perception. *Proceedings of the National Academy of Sciences, 95*(3), 825 - 830.

Mishra P, Pandey CM, Singh U, Gupta A, Sahu C, Keshri A. (2019) Descriptive statistics and normality tests for statistical data. *Ann Card Anaesth. 22*(1). doi: 10.4103/aca.ACA_157_18.

Muraven, & Baumeister. (2000). Self-regulation and depletion of limited resources: Does self-control resemble a muscle. *Psychological Buletin, 126(2)*, 247-259.

Musslick, Cohen, & Shenav. (2018). Estimating the costs of cognitive control from task performance:. *Proceedings of the 40th Annual Meeting of the Cognitive Science Society* (pp. 800 - 805). Wisconsin: Madison.

Peirce, J. W. (2007). PsychoPy - Psychophysics software in Python. *Journal of Neuroscience Methods*, 162(1-2), 8-13. doi:10.1016/j.jneumeth.2006.11.017

Peirce, J. W. (2009). Generating stimuli for neuroscience using PsychoPy. *Frontiers in Neuroinformatics*, 2, 10. doi:10.3389/neuro.11.010.2008

Prasad, & Mishra. (2020). Reward Influences Masked Free-Choice Priming. *Front. Psychol*, 824-839. doi:10.3389/fpsyg.2020.576430

Prével, Krebs, Kukkonen, & Braem. (2021). Selective reinforcement of conflict processing in the Stroop task. *PLoS ONE*, 16 (7). doi:10.1371/journal.pone.0255430

Ratcliff, & McKoon. (2008). The Diffusion Decision Model: Theory and Data for Two-Choice Decision Tasks. *Neural Comput*, 873-922. doi:10.1162/neco.2008.12-06-420

Ratcliff, & Rouder. (1998). Modeling Response Times for Two-Choice Decisions. *Psychological Science, 9*(5), 347-356. doi:10.1111/1467-9280.00067

Ratcliff, & Rouder. (2000). A diffusion model account of masking in two-choice letter identification. *Journal of Experimental Psychology: Human Perception and Performance, 26*(1), 127-140.

Ratcliff, & Smith. (2004). A comparison of sequential sampling models for two-choice reaction time. *Psychological Review*, 111(2), 333–367.

Ratcliff, & Tuerlinckx. (2002). Estimating parameters of the diffusion model: Approaches to dealing with contaminant reaction times and parameter variability. *Psychonomic Bulletin & Review*, 9(3), 438-481. doi:10.3758/BF03196302

Saddoris, Cacciapagalia, Wightman, & Carelli. (2015). Differential dopamine release dynamics in the nucleus accumbens core and shell reveal complementary signals for error prediction and incentive motivation. *The Journal of Neuroscience, 35*(33), 11572-11582.

Simons, Boot, Gathercole, Chabris, Hambrick, & Stime-Morrow. (2016). Do "Brain-Training" Programs Work? *Psychological Science in the Public Interest*, 17(3), 103–186.

Skinner, B. F. (1953). *Science and human behavior.* New York: The Macmillan Company. doi:10.1002/sce.37303805120

Spapé, & Hommel. (2008). He said, she said: Episodic retrieval induces conflict adaptation in an auditory Stroop task. *Psychonomic Bulletin & Review, 15*(12), 1117-1121.

Thorndike. (1898). A proof of the law of effect. *Science*, 173-175. doi:10.1126/science.77.1989.173-a

Tolman. (1925). Purpose and cognition: The determiners of animal learning. *Psychological Review*, 32(4), 285-297. doi:10.1037/ h0072784

Umemoto, & Holroyd. (2015). Task-specific effects of reward on task switching. *Psychological Research*, 79, 698–707. doi:10.1007/s00426-014-0595-z

van den Berg, Krebs, Lorist, & Woldorff. (2014). Utilization of reward-prospect enhances preparatory attention and reduces stimulus conflict. *Cognitive, Affective & Behavioral Neuroscience*, 14(2), 561-577. doi:10.3758/s13415-014-0

van Gaal, Ridderinkhof, van den Wildenberg, & Lamme. (2009). Dissociating consciousness from inhibitory control: Evidence. *Journal of Experimental Psychology: Human Perception and Performance*, 1129–1139.

van Gaal, Ridderinkhof, vand en Wildenberg, & Lamme. (2009). Dissociating consciousness from inhibitory control: Evidence. *Journal of Experimental Psychology: Human Perception and Performance, 35*, 1129-1139. doi:http://dx.doi.org/10.1037/a0013551

Verbruggen, & Logan. (2008). Automatic and controlled response inhibition: Associative learning in the go/no-go and stop-signal paradigms. *Journal of Experimental Psychology: General, 137*(4), 649–672. doi:10.1037/a0013170

Verbruggen, McLaren, & Chambers. (2014). Banishing the control homunculi in studies of action control and behavior change. *Perspectives in Psychological Science*, 9(5), 497-524. doi:10.1177/1745691614526414

Waskom. (2021). seaborn: statistical data visualization. *The Open Journal*, 3021. doi:10.21105/joss.03021

Westbrook A, Braver TS. Cognitive effort: A neuroeconomic approach. Cogn Affect Behav Neurosci. 2015 Jun;15(2):395-415. doi: 10.3758/s13415-015-0334-y.

Wiecki, Sofer, & Frank. (2013). HDDM: Hierarchical Bayesian estimation of the Drift-Diffusion Model in Python. *Frontiers in neuroinformatics, 7*(14). doi:https://doi.org/10.3389/fninf.2013.00014