FACULTY OF PSYCHOLOGY
AND EDUCATIONAL SCIENCES

Motivation and cognitive control in depression

Ivan Grahek

Supervisor: Prof. Dr. Ernst Koster
Co-supervisor: Prof. Dr. Ruth Krebs

A dissertation submitted to Ghent University in partial fulfilment of the requirements for the degree of Doctor of Psychology

Academic year 2019 – 2020

GHENT
UNIVERSITY

Chair of the Examination Board

Prof. dr. Bart Soenens
Ghent University

Examination board

Prof. dr. Eva Van den Bussche
KU Leuven

Prof. dr. Roshan Cools
Donders Institute for Brain, Cognition and Behaviour

Prof. dr. Marie-Anne Vanderhasselt
Ghent University

Prof. dr. Tom Verguts
Ghent University

Guidance committee

Prof. dr. Ernst H. W. Koster (promotor)
Ghent University

Prof. dr. Ruth M. Krebs (co-promotor)
Ghent University

Prof. dr. Marie-Anne Vanderhasselt
Ghent University

Prof. dr. Alvaro Sanchez-Lopez
Complutense University of Madrid

Table of contents

Acknowledgements 5

Chapter 1: General introduction 7

Chapter 2: Cognitive control in depression 15

Chapter 3: Motivation and cognitive control in depression 47

Chapter 4: Reward in the visual cortex 85

Chapter 5: Learning efficacy 121

Chapter 6: Expecting cognitive effort 149

Chapter 7: General discussion 175

English summary 193

Nederlandstalinge Samenvatting 195

Data storage fact sheets 197

Acknowledgements

Over the past four years I have been very lucky to work with a great number of excellent scientists who have inspired me and made this period very fun and exciting. Being able to spend my time discussing and developing ideas in such a stimulating environment has been an enormous privilege which I am very grateful for.

Ernst Koster has provided me with full support since proposed to write a PhD project with him, and this thesis owes most to him. Ernst, I couldn't have asked for a better mentor. You have always been supportive and you have given me an incredible amount of freedom to develop my own ideas, while always being present and guiding me. I am so grateful for this.

Ruth Krebs has enthusiastically joined this project from the beginning and supported me throughout it. She has introduced me to the field of motivation and cognition and inspired me to continue my research in this field. Ruth, designing experiments with you and discussing results has been one of the most intellectually rich parts of the past four years.

Amitai Shenhav was central in developing parts of this thesis and has provided me with a great amount of support over the last two years. Collaborating with him and Sebastian Musslick has brought a great amount of joy and inspiration.

Gilles Pourtois has inspired me with his enthusiasm for electrophysiology and science. He has supported me with kindness and was always very open for discussion. I am very grateful for that.

Antonio Schettino has been with me from the start of this project and I have learned so much from him. His knowledge and friendship have been strong driving forces behind this project. Most importantly, working with Antonio has thought me what true collaboration should be: fun, engaging, and full of joy.

Jonas Everaert has been central at the beginning of this project. Without Jonas and his support I would not have ended up in Belgium. His kindness and enthusiasm for science have inspired me to start this thesis.

Ladislas Nalborczyk has inspired me with his passion for statistics and science. He was central in the development of my statistical thinking. I am grateful for his support, friendship, and enthusiasm.

Martin Finn has been with me from the start of this thesis and discussions with him have always pushed me to deeply examine my theoretical positions. Discussions with him have enormously improved my thinking. His friendship has been one of the main driving forces of this project.

Davide Gheza has been the central part of this thesis. We have shared all of our designs, results, and worries over the past four years. Davide's deep knowledge of electrophysiology has been crucial in my understanding of the subject and his comments have been central to multiple studies in this thesis. This thesis would not have been possible without his friendship. Thank you Davide.

I have received continuous financial support from Ghent University and Research Foundation – Flanders. I am thankful for the opportunities these institutions have provided me with.

I wish to thank hundreds of participants in my experiments. Your patience and good will has been an inspiration.

I have encountered and collaborated with so many great scientists over the past years. While I cannot name all of them here, I wish to thank all of the people from the Department of Experimental-Clinical and Health Psychology and the Department of Experimental Psychology at Ghent University. I also wish to thank all of the people at the Department of Cognitive, Linguistic, and Psychological Sciences at Brown University.

Finally, my deepest thank you goes to my mother Ana and my sister Nela. Thank you for all of your support and understanding.

Ivan
October 2019

Chapter 1

General introduction

Depression[1] is a very common mental disorder and is ranked as the single leading cause of disability worldwide (World Health Organization, 2017). At least one in ten individuals is likely to experience depression as the lifetime prevalence estimates are 14% in high-income and 11% in low-income countries (Kessler & Bromet, 2013). The disorder is linked to severe suffering and high societal costs (Kessler, 2012). Importantly, depression is a highly recurrent disorder and individuals who suffer from the first episode are likely to experience another one (40-60% probability) with this probability rising to 90% for the individuals who have experienced three episodes (Bockting, Hollon, Jarrett, Kuyken, & Dobson, 2015). Despite the obvious importance of treating depression, the psychological and neural causes of the disorder are not well understood.

Depression profoundly influences how we experience ourselves and the world around us. It impacts our thoughts, feelings, and desires. Such changes in cognitive, emotional, and motivational processes are key components in both diagnosing and treating depression (DSM-V; American Psychiatric Association, 2013). Cognitive theories of depression have focused on the maladaptive changes in the way depressed individuals process information (Clark & Beck, 1999; Gotlib & Joormann, 2010). These theories have led to the development of successful treatments for depression and to the development of a research program which seeks to map depression-related impairments in specific cognitive processes. This research program has relied on the insights from cognitive psychology and, more recently, cognitive neuroscience, to understand key impairments in depression (Clark & Beck, 2010; DeRubeis, Siegle, & Hollon, 2008).

[1] Throughout this thesis we use the term depression to denote the Major Depressive Disorder. Unless otherwise specified, we refer to unipolar depression.

Cognitive control in depression

Following the increasingly fine-grained specifications of cognitive processes in cognitive science, research on depression has focused on impairments in processes such as attention, memory, and cognitive control (Gotlib & Joormann, 2010). This line of research has pointed to impairments in cognitive control as the central cognitive deficit in depression (Disner, Beevers, Haigh, & Beck, 2011; Joormann & Tanovic, 2015; Snyder, 2013), as well as many other mental disorders (Millan et al., 2012). Cognitive control[2] represents a set of processes which enable us to flexibly adapt our cognition and behavior in accordance with our goals, and it is thus crucial for goal-directed behavior (Botvinick & Cohen, 2014; Diamond, 2012; Friedman & Miyake, 2017). For example, cognitive control is needed in order to focus our attention on the current task and inhibit all other distracting thoughts and events. Existing research on cognitive control in depression has focused on investigating the specific components of cognitive control such as inhibition, task-switching, and working memory updating, following the tripartite model of cognitive control (Miyake et al., 2000). This research has also focused on emotion-specific impairments in these processes. For example, focusing on the ability to inhibit negatively valenced material (Joormann & Gotlib, 2010), or discard negative stimuli from working memory (Joormann, Levens, & Gotlib, 2011).

In sum, much of the research on cognition in depression has been focused on characterizing the topography of specific impairments in different cognitive processes. This approach follows the developments in cognitive psychology which has divided cognitive processes into smaller units. Both fields have addressed the questions of *what* are the crucial components of control and *how* do individuals differ in control processes. However, the question of *why* there are individual differences in cognitive control has received far less attention. Recent theoretical developments in cognitive science allow us to start asking this question.

[2] In this thesis we consider the terms cognitive control, executive control, and executive functions to refer to the same set of processes. While the terms executive functions and executive control are mainly used in individual differences research, the term cognitive control is often used in cognitive neuroscience.

Cognitive control as a decision-making process based on motivation

One of the main problems in cognitive control research is how we decide when to exert cognitive control and when to engage in more automatic processing. Recent theoretical and computational models of cognitive control propose that control can be seen as a decision-making process (Kool, Shenhav, & Botvinick, 2017; Shenhav, Botvinick, & Cohen, 2013). In this view the decision about how much control to allocate based on costs and benefits of control allocation.

Cognitive control is an effortful process and individuals tend to avoid exerting control whenever possible (Shenhav et al., 2017; Westbrook & Braver, 2015). However, more control will be exerted if the rewards for doing so are sufficiently large (Botvinick & Braver, 2015). Thus control allocation can be seen as a form of cost-benefit decision-making in which the costs of exerting control are weighted against the benefits of doing so (Shenhav et al., 2013). An increasing number of formal models of cognitive control rely on such components of motivation to understand cognitive control (Brown & Alexander, 2017; Holroyd & McClure, 2015; Manohar et al., 2015; Shenhav et al., 2013; Silvetti, Alexander, Verguts, & Brown, 2014). These approaches have led to successful computational implementations of control theories which are able to simulate a wide range of existing empirical data (Lieder, Shenhav, Musslick, & Griffiths, 2018; Musslick, Shenhav, Botvinick, & Cohen, 2015). Thus this field of research is beginning to provide potential reasons for *why* individuals differ in how much control they allocate in specific situations. Going beyond an impairment in the ability to allocate control, this novel view suggests that exerting control is based on estimates of costs and benefits related to exerting control. Importantly, these estimates need to be learned from previous experiences, highlighting the role of learning in cognitive control (Bhandari, Badre, & Frank, 2017; Chiu & Egner, 2019). Thus, the observed depression-related deficits in control can be caused by the changes in this cost-benefit decision making process about control allocation.

Thesis outline

The main goal of this thesis is to provide a bridge between the recent developments in cognitive science and depression research. In this way, the thesis stands at the intersection between cognitive and clinical research, with the aim to translate cognitive science insights into improved understanding of cognitive control deficits in depression. In order to do so, we have divided our work into two sections. First, we have focused on theoretical work aimed at providing a new framework for understanding cognitive deficits in depression. Second, guided

by this framework, we conducted empirical work aimed at developing measures which can in future be used to test the specific predictions of our theoretical work.

Novel theoretical and empirical developments in the field of motivation and cognitive control provide a new avenue for understanding the causes of cognitive control impairments in depression. Instead of postulating the impairments in the ability to exert control, this view provides an understanding based on components of motivation, as well as learning from previous experiences. In this way, it shifts both the theoretical and empirical focus in addressing cognitive impairments in depression. The first goal of this thesis is to provide a new theoretical framework for understanding cognitive impairments in depression. In *Chapter 1* we analyze the existing theoretical models aimed at understanding cognitive control impairments in depression. We outline the main conceptual issues in the field and propose the ways in which they can be addressed. *Chapter 2* builds on this work and proposes a new theoretical framework for understanding cognitive control in depression. We review how crucial components of motivation (outcome value, outcome controllability, and effort costs) are impaired in depression and then propose how these impairments can impact cognitive control. This theoretical work relies on a formal model of cognitive control in order to provide behavioral predictions based on simulations from the computational model we use. The first two chapters also introduce all of the relevant theoretical concepts which are used in the empirical chapters.

Based on the framework we have proposed, the second goal of the thesis is to develop behavioral and electrophysiological procedures to assess the three components of motivation on which the framework relies. Since our theoretical work provides a novel view on cognitive control in depression, there were no existing experimental paradigms which could be readily used to test the predictions of our framework. To address this issue, we developed paradigms which assess the influence of value, controllability, and effort on cognitive control. In *Chapter 3* we investigate how value and attentional control interact to determine how stimuli of different values are processed in the visual cortex. This study provides a novel way to assess the separate influence of value and attentional control on sensory processing. In *Chapter 4* we investigate the neural correlates of learned action-outcome contingencies and how they influence cognitive control. We develop a novel method to address the influence of both learning and one of the main components of our model – outcome controllability – on cognitive control. In *Chapter 5* we present a new behavioral approach to assess effort-based allocation of control. In this chapter we explore how the expectations about cognitive effort modulate cognitive control. All

of the empirical chapters of the thesis include links to the Open Science Framework where the data and analysis scripts are stored.

We conclude the thesis with the general discussion in which we focus on the implications of our framework for future clinical research on cognition in depression. Within this thesis, we focused on devising a new theoretical approach to cognitive control impairments in depression, as well as novel procedures which can be used to test the predictions of our model. We discuss how these experimental procedures can be used in clinical populations in order to start directly testing the proposed framework. Finally, we discuss how cognitive and clinical science can work together to improve the understanding and development of treatments for depression.

References

American Psychiatric Association. (2013). *Diagnostic and Statistical Manual of Mental Disorders* (5th ed.). Washington DC: American Psychiatric Publishing.

Bhandari, A., Badre, D., & Frank, M. J. (2017). Learning cognitive control. *The Wiley Handbook of Cognitive Control*, 376–391. https://doi.org/10.1002/9781118920497.ch21

Bockting, C. L., Hollon, S. D., Jarrett, R. B., Kuyken, W., & Dobson, K. (2015). A lifetime approach to major depressive disorder: The contributions of psychological interventions in preventing relapse and recurrence. *Clinical Psychology Review*, *41*, 16–26. https://doi.org/10.1016/j.cpr.2015.02.003

Botvinick, M. M., & Braver, T. S. (2015). Motivation and Cognitive Control: From Behavior to Neural Mechanism. *Annual Review of Psychology*, *66*(September 2014), 83–113. https://doi.org/10.1146/annurev-psych-010814-015044

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

Brown, J. W., & Alexander, W. H. (2017). Foraging Value, Risk Avoidance, and Multiple Control Signals: How the Anterior Cingulate Cortex Controls Value-based Decision-making. *Journal of Cognitive Neuroscience*, *29*(10), 1656–1673. https://doi.org/10.1162/jocn_a_01140

Chiu, Y. C., & Egner, T. (2019). Cortical and subcortical contributions to context-control learning. *Neuroscience and Biobehavioral Reviews*, *99*(December 2018), 33–41. https://doi.org/10.1016/j.neubiorev.2019.01.019

Clark, D. A., & Beck, A. T. (1999). *Scientific foundations of cognitive theory and therapy of depression*. John Wiley & Sons.

Clark, David A., & Beck, A. T. (2010). Cognitive theory and therapy of anxiety and depression: Convergence with neurobiological findings. *Trends in Cognitive Sciences*, *14*(9), 418–424. https://doi.org/10.1016/j.tics.2010.06.007

DeRubeis, R. J., Siegle, G. J., & Hollon, S. D. (2008). Cognitive therapy versus medication for

depression: treatment outcomes and neural mechanisms. *Nature Reviews Neuroscience*.

Diamond, A. (2012). Executive functions. *Annual Review of Psychology*, *64*, 135–168. https://doi.org/10.1146/annurev-psych-113011-143750

Disner, S. G. S., Beevers, C. G. C., Haigh, E. A. P., & Beck, A. T. (2011). Neural mechanisms of the cognitive model of depression. *Nature Reviews Neuroscience*, *12*(8), 467–477. https://doi.org/10.1038/nrn3027

Friedman, N. P., & Miyake, A. (2017). Unity and diversity of executive functions: Individual differences as a window on cognitive structure. *Cortex*, *86*, 186–204. https://doi.org/10.1016/j.cortex.2016.04.023

Gotlib, I. H., & Joormann, J. (2010). Cognition and depression: current status and future directions. *Annual Review of Clinical Psychology*, *6*, 285–312. https://doi.org/10.1146/annurev.clinpsy.121208.131305

Holroyd, C. B., & McClure, S. M. (2015). Hierarchical control over effortful behavior by rodent medial frontal cortex: A computational model. *Psychological Review*, *122*(1), 54–83. https://doi.org/10.1037/a0038339

Joormann, J., & Gotlib, I. H. (2010). Emotion regulation in depression: relation to cognitive inhibition. *Cognition & Emotion*, *24*(2), 281–298. https://doi.org/10.1080/02699930903407948

Joormann, J., Levens, S. M., & Gotlib, I. H. (2011). Sticky thoughts: Depression and rumination are associated with difficulties manipulating emotional material in working memory. *Psychological Science*, *22*(8), 979–983. https://doi.org/10.1177/0956797611415539

Joormann, J., & Tanovic, E. (2015). Cognitive vulnerability to depression: Examining cognitive control and emotion regulation. *Current Opinion in Psychology*, *4*, 86–92. https://doi.org/10.1016/j.copsyc.2014.12.006

Kessler, R. C. (2012). The costs of depression. *Psychiatric Clinics of North America*, *35*(1), 1–14. https://doi.org/10.1016/j.psc.2011.11.005

Kessler, R. C., & Bromet, E. J. (2013). The epidemiology of depression across cultures. *Annual Review of Public Health*, *9*(34), 119. https://doi.org/10.1146/annurev-publhealth-031912-114409

Kool, W., Shenhav, A., & Botvinick, M. M. (2017). Cognitive Control as Cost-Benefit Decision Making. *The Wiley Handbook of Cognitive Control*, 167–189. https://doi.org/10.1002/9781118920497.ch10

Lieder, F., Shenhav, A., Musslick, S., & Griffiths, T. L. (2018). Rational metareasoning and the plasticity of cognitive control. *Plos Computational Biology*, *14*(4). https://doi.org/10.13140/RG.2.2.24500.14721

Manohar, S. G., Chong, T. T. J., Apps, M. A. J., Batla, A., Stamelou, M., Jarman, P. R., … Husain, M. (2015). Reward Pays the Cost of Noise Reduction in Motor and Cognitive Control. *Current Biology*, *25*(13), 1707–1716. https://doi.org/10.1016/j.cub.2015.05.038

Millan, M. J., Agid, Y., Brüne, M., Bullmore, E. T., Carter, C. S., Clayton, N. S., … Young, L. J. (2012). Cognitive dysfunction in psychiatric disorders: characteristics, causes and the quest for improved therapy. *Nature Reviews Drug Discovery*, *11*(2), 141–168. https://doi.org/10.1038/nrd3628

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, a H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex "Frontal Lobe" tasks: a latent variable analysis. *Cognitive Psychology*, *41*(1), 49–100. https://doi.org/10.1006/cogp.1999.0734

Musslick, S., Shenhav, A., Botvinick, M. M., & Cohen, J. D. (2015). A computational model of control allocation based on the Expected Value of Control. *Reinforcement Learning and Decision Making Conference*, *59*(1978), 2014.

Shenhav, A., Botvinick, M., & Cohen, J. (2013). The expected value of control: An integrative theory of anterior cingulate cortex function. *Neuron*, *79*(2), 217–240. https://doi.org/10.1016/j.neuron.2013.07.007

Shenhav, A., Musslick, S., Falk, L., Kool, W., Griffits, T. L., Cohen, J. D., & Botvinick, M. M. (2017). Toward a rational and mechanistic account of mental effort. *Annual Review of Neuroscience*, *40*. https://doi.org/10.1146/annurev-neuro-072116-031526

Silvetti, M., Alexander, W., Verguts, T., & Brown, J. W. (2014). From conflict management to reward-based decision making: Actors and critics in primate medial frontal cortex. *Neuroscience and Biobehavioral Reviews*, *46*, 44–57. https://doi.org/10.1016/j.neubiorev.2013.11.003

Snyder, H. R. (2013). Major depressive disorder is associated with broad impairments on neuropsychological measures of executive function: a meta-analysis and review. *Psychological Bulletin*, *139*(1), 81–132. https://doi.org/10.1037/a0028727

Westbrook, A., & Braver, T. S. (2015). Cognitive effort: A neuroeconomic approach. *Cognitive, Affective and Behavioral Neuroscience*, *15*(2), 395–415. https://doi.org/10.3758/s13415-015-0334-y

World Health Organization. (2017). Depression and other common mental disorders: global health estimates, 1–24. https://doi.org/CC BY-NC-SA 3.0 IGO

Chapter 2

Cognitive Control in Depression: Toward Clinical Models Informed by Cognitive Neuroscience[1]

Abstract

Cognitive control dysfunctions are thought to contribute to the onset and maintenance of depression. However, the causes and nature of these dysfunctions remain unknown. Here, we critically review contemporary research on cognitive control in depression. We identify three main conceptual issues in this field: 1) uncritical use of the tripartite model; 2) reliance on descriptive explanations; and 3) lack of integration with emotional and motivational impairments. Recent advances in cognitive neuroscience offer possibilities to resolve these issues. We review this progress focusing on the ability to detect the need for control, the role of motivation, and the flexibility-stability balance. We propose that depression-related dysfunctions arise from issues in detecting when, how, and for how long to engage in goal-oriented processing. In conclusion, we argue that integrating advances in cognitive neuroscience into clinical research can help to move from a descriptive towards a more mechanistic understanding of cognitive dysfunctions in depression.

[1] Based on: Grahek, I., Everaert, J., Krebs, R. M., & Koster, E. H. W. (2018). Cognitive control in depression: Toward clinical models informed by cognitive neuroscience. *Clinical Psychological Science*, 6(4), 464-480.

Introduction

Depression represents one of the leading causes of disability worldwide (Kessler & Bromet, 2013). This highly common disorder is linked to severe individual suffering and high societal costs (Kessler, 2012). Notably, the relapse and recurrence rates of depression remain high (Bockting, Hollon, Jarrett, Kuyken, & Dobson, 2015; Vittengl, Clark, Dunn, & Jarrett, 2007), indicating that improvement of existing treatments is urgently needed. To advance both theory and treatment of depression, it is crucial to uncover the mechanisms underlying hallmark features of depression. These mechanisms may include abnormalities in cognitive, emotional, and motivational processes which contribute to the onset and maintenance of depression (for recent reviews see: Admon & Pizzagalli, 2015; Crocker et al., 2013; Joormann & Vanderlind, 2014). One of the central cognitive abnormalities investigated in depression are cognitive control dysfunctions. These dysfunctions have become the focus of a rapidly increasing number of empirical and theoretical studies over more than a decade.

Cognitive control refers to a set of mental processes that allow flexible adaptation of cognition and behavior in accordance with an individual's current goals (Braver, 2012; Friedman & Miyake, 2017; Shenhav, Botvinick, & Cohen, 2013). These processes are critical for goal-directed, non-automatic behavior and are found to be disturbed in a wide range of psychiatric disorders including depression (Millan et al., 2012; Snyder, Miyake, & Hankin, 2015). Three cognitive control processes that are most often investigated in relation to depression are: *inhibition* (overriding dominant or prepotent responses), *shifting* (switching between mental sets or tasks), and *updating* (adding or discarding of working memory contents) (Miyake et al., 2000; Miyake & Friedman, 2012). Cognitive control dysfunctions are purported to represent a key vulnerability factor for depression (Joormann et al., 2007; Siegle et al., 2007). Dysfunctions in these processes have been observed in clinically depressed individuals (for a meta-analysis see Snyder, 2013), individuals with self-reported elevated levels of depressive symptoms (Derakshan, Salt, & Koster, 2009; Owens, Koster, & Derakshan, 2012), and patients with depression in remission (Demeyer, De Lissnyder, Koster, & De Raedt, 2012; Levens & Gotlib, 2015). These results from cross-sectional studies suggest that such impairments are not merely an epiphenomenon of a depressive mood or episode. Indeed, there is evidence that cognitive control may have a causal influence on depressive symptoms. Research indicates that improving cognitive control through training in depressed and at-risk populations helps to reduce depressive symptoms (Koster, Hoorelbeke, Onraedt, Owens, & Derakshan, 2017; Siegle et al., 2007). Also, longitudinal studies have provided initial evidence for the importance of

cognitive control dysfunctions in predicting depressive symptoms (Demeyer et al., 2012; Pe, Brose, Gotlib, & Kuppens, 2015). Finally, cognitive control dysfunctions have also been linked to information-processing biases, increased emotional reactivity to stress, and difficulties to downregulate negative emotions (Joormann & Vanderlind, 2014; Koster et al., 2011; Siegle et al., 2007). Taken together, cognitive control represents an important construct in understanding vulnerability to depression.

Currently there is increasing recognition of the importance of cognitive control dysfunctions in depression, accompanied by a fast development of cognitive training procedures aimed at reducing depressive symptoms. We believe that now is the right time to take a step back and critically examine the current understanding of cognitive control in depression. Models and conceptualizations of cognitive control in depression have guided important discoveries of critical cognitive abnormalities. However, these models do not adequately address questions about causes and mechanisms of the depression-related cognitive control impairments. For example, it remains unclear if depression is related to the general reduction of the ability to exert control, or it is more related to the problems in detecting when to engage in controlled processing and goal-oriented behavior, and how intensely to do so. Also, it is not clear if the cognitive control dysfunctions are general or specific to the processing of emotionally negative material. In short, the mechanisms behind these cognitive impairments remain unknown.

Recent advances in cognitive neuroscience have led to important improvements in understanding cognitive control mechanisms (for a broad overview see: Cohen, 2017). These advances offer a mechanistic view and plausible neurobiological substrates of cognitive control. The progress in research on depression-linked cognitive control dysfunctions notwithstanding, we advocate the view that future advances in clinical research should be informed by these novel developments in cognitive neuroscience. While current research and theorizing offer explanation at the descriptive level, there is a strong need for a mechanistic approach to dysfunctional cognitive control in depression. This will be a crucial next step in developing a more sophisticated understanding of cognitive control dysfunctions and may provide novel directions for treatment strategies. Hence, the aim of this paper is to bridge the current gap between clinical research and theories of cognitive control in depression and novel developments in cognitive neuroscience. In the following sections, we critically review the current state of theory and research on cognitive control in depression and provide an overview of recent developments in understanding cognitive control. By doing so, we aim to demonstrate

how progress in cognitive neuroscience can be applied to research on depression to advance the understanding of cognitive control dysfunctions.

Cognitive Control and Depression

Theoretical Models of Cognitive Control in Depression

Contemporary cognitive models of depression (Disner et al., 2011) propose that genetic vulnerability in combination with adverse early experiences and stressful events (e.g., experiences of loss) can result in depressogenic ways of processing emotional information. For instance, if someone has experienced low parental warmth in childhood, this person may develop beliefs that he/she is unlovable. Such core beliefs may determine how someone interacts with other people and stressful situations. This person may focus excessively on cues signaling that their current partner does not love them. This may result in difficulties to disengage from this type of information and regulating of the elicited thoughts and emotions. Indeed, research has documented specific emotional biases in cognitive processes of attention and memory in depressed and dysphoric individuals (for a review see: Everaert et al., 2012). Cognitive models have been put forward to explain such depressogenic information processing biases in terms of impaired disengagement of attention from negative stimuli (Koster et al., 2011) and cognitive control dysfunctions (Joormann, 2010; Joormann et al., 2007; Joormann & Vanderlind, 2014). These models are mainly descriptive and focused on detecting the processes that may be involved in depressogenic information processing.

The Impaired Cognitive Control account (Joormann, 2010; Joormann & Vanderlind, 2014; Joormann et al., 2007) is the most elaborate account linking cognitive control dysfunctions to depressive symptoms. Within this framework, cognitive control dysfunctions are defined as specific difficulties in controlling the contents of working memory (WM) (Joormann & Vanderlind, 2014). Following the tripartite model of cognitive control (Miyake & Friedman, 2012; Miyake et al., 2000), it is proposed that depression is related to decreased ability to limit the access of irrelevant negative information into working memory (*inhibition*) and a decreased ability to remove negative content that is no longer relevant from WM (*updating*). For example, negative cognitions about past failures that remain active in WM (e.g., "I failed my previous exam, why would I pass now") may interfere with current performance on a task (e.g., when one is preparing for the next exam). Reduced ability to perform inhibition and updating are theorized to further skew information processing resulting in exaggerated processing of negative material and interrupting effective emotion regulation.

Several theoretical models have also elaborated on how cognitive control dysfunctions and depressogenic information processing biases contribute to the onset and maintenance of depressive symptoms. It is proposed that a decreased ability to exert cognitive control, particularly when processing negative information, underlies ineffective use of emotion regulation strategies to increase depressive symptoms (Joormann, 2010; Koster et al., 2011; Whitmer & Gotlib, 2013). Of particular importance in this context is the emotion regulation strategy called *rumination,* which refers to the tendency to engage in preservative negative thinking about the past and present and is closely linked to depressive symptoms (Nolen-Hoeksema, Wisco, & Lyubomirsky, 2008). The indirect influence of cognitive control dysfunctions in processing negative information on depressive symptoms via changes in emotion regulation strategies has gained initial empirical support (Everaert et al., 2017; for a review see: Mor & Daches, 2015).

Theoretical models have also included neurobiological substrates of cognitive control dysfunctions in depression (Disner, Beevers, Haigh, & Beck, 2011). While cognitive control processes are supported by a wide range of interacting brain regions and circuitries, most research in the field of depression has been focused on two regions in particular. That is, depression has been linked to abnormal activation of the dorsolateral prefrontal cortex (dlPFC) and the anterior cingulate cortex (ACC). These two regions are strongly associated with cognitive control (Davidson, Pizzagalli, Nitschke, & Putnam, 2002; Gotlib & Hamilton, 2008; Pizzagalli, 2011). Current theoretical models frequently relate diminished cognitive control in depression to hypoactivity of the dlPFC (but also ventral and medial PFC) and to hypoactivity of the ACC (Disner et al., 2011; Joormann et al., 2007). The main hypothesis is that depression-related hypoactivity in these areas is related to the inability to effectively regulate negative affect.

In summary, cognitive control dysfunctions are assigned a central role in depressogenic information processing and emotion regulation difficulties. These dysfunctions are currently understood as an impaired (i.e. attenuated) ability to engage cognitive control processes such as inhibition or WM updating. In the next section, we describe the empirical research that was guided by aforementioned theoretical models.

Research Findings on Cognitive Control in Depression

Behavioral studies

Depression-linked dysfunctions in cognitive control processes have been examined in the case of processing both neutral and emotional material. Standard cognitive control tasks

present stimuli that are neutral in emotional tone. For example, the Stroop task presents a series of words (e.g. red, green, blue) printed in different colors (e.g. red, yellow, blue) and prompts participants to name the ink color of the printed word. This task requires participants to override an automatic response (i.e., processing the content of the word) and to execute a controlled behavior (i.e., naming the color of the printed word). Emotional variants of standard cognitive control tasks present emotionally salient stimuli. For example, in the emotional 2-back task, participants view streams of emotional faces and decide for each face if the current emotional expression is the same as the one presented two faces before. This task requires participants to update the WM contents continuously. The use of emotional stimuli allows for the investigation of valence-specific difficulties in WM operations. Depression research has been conducted using both single tasks which tap into a specific cognitive control component (e.g. using the Stroop task as a measure of inhibition) and multiple tasks tapping into the same component (i.e., the latent-variable approach; Miyake et al., 2000).

Meta-analytic work has provided robust evidence for dysfunctions in cognitive control *in response to neutral information*. A meta-analysis by Snyder (2013) integrated 113 studies comparing the performance of participants with depression to healthy controls on a wide range of cognitive control tasks measuring different components such as inhibition, shifting, and updating. Snyder adopted the latent-variable approach by analyzing multiple tasks thought to measure the same cognitive control component. The results revealed depression-related impairments with medium effect size on *neutral* measures for several cognitive control processes, namely inhibition, shifting, updating, and others. Also, this study has provided evidence that the magnitude of the impairments can be related to depression symptom severity, with more severe impairments characteristic for severely depressed individuals. Another meta-analysis (Rock, Roiser, Riedel, & Blackwell, 2014) included 24 studies which used the same neuropsychological battery of tasks measuring cognitive control (i.e., the *Cambridge Neuropsychological Test Automated Battery*; Fray et al., 1996). This approach ensured the inter-study homogeneity of the tasks used to assess cognitive control. The study revealed medium effect size for the difference between depressed and non-depressed control groups on emotionally neutral measures of cognitive control. In sum, results from these meta-analyses provide robust evidence for depression-related dysfunctions in cognitive control on emotionally neutral tasks.

Research has also provided evidence for dysfunctional cognitive control *in response to emotional information* (for detailed reviews see: Gotlib & Joormann, 2010; Joormann & Vanderlind, 2014). In this research field, a variety of experimental tasks that presumably tap

into different components of cognitive control have been used. Unfortunately, there are currently no meta-analyses to provide a systematic synthesis of previous work. This line of research has largely adopted the three-component model of cognitive control (Miyake & Friedman, 2012; Miyake et al., 2000). Several studies have investigated updating of working memory (WM) and found evidence of depression-related difficulties in manipulating material in WM, especially when processing negative material (Joormann et al., 2011). Depressed individuals are also slower to discard sad faces and faster to discard happy faces from WM compared to healthy controls (Levens & Gotlib, 2010). With regard to the shifting function in depression, evidence for both general and emotion-specific dysfunctions has been found. While subclinical levels of depression may be characterized by emotion-specific dysfunctions in the form of shifting away from negative material, clinical depression levels may be characterized by dysfunctions in shifting between both negative and positive mental representations (De Lissnyder et al., 2012; Demeyer et al., 2012). Moreover, there is some evidence for depression-related deficits in inhibition of negative material. Using a negative affective priming task, Joorman and Gotlib have found that depressed individuals, compared to healthy controls, exhibit difficulties in inhibiting negative words (Joormann & Gotlib, 2010). In sum, studies in this field provide some evidence for depression-related dysfunctions in exerting cognitive control over emotional material.

Neuroimaging studies

Neuroimaging studies have started to investigate the neural substrates of cognitive control dysfunctions in depression. A recent review has pointed to depression-related decreases in activation of the dlPFC and dACC during both emotional and neutral tasks that demand increased cognitive control (Pizzagalli, 2011). This review also emphasized an important role of the hypoactivation and/or reduced deactivation of the rostral ACC – a region of the ACC related to evaluating emotional and motivational significance of events. For example, one study has shown that depressed patients have an increased activity of the dACC and parietal and bilateral insular cortices when removing positive compared to removing negative words from WM (Foland-Ross et al., 2013). Also, difficulties in shifting attention away from emotional stimuli in participants with mild to moderate depression levels has been related to weaker activation of the lateral PFC and parietal regions, regions associated with cognitive control and attentional processes (Beevers et al., 2010). A number of studies have focused on the processing of emotional material in depression, and especially the processing of negative stimuli. Some of the meta-analyses in this domain have found evidence of abnormal activity in the dACC and

decreased activity in the dlPFC, regions strongly associated with cognitive control (Hamilton et al., 2012), while others have failed to replicate these results (Müller et al., 2016; for a discussion see: Barch & Pagliaccio, 2017).

However, research on neural correlates of cognitive control dysfunctions in depression is in early stages. The number of studies in this domain is limited and sample sizes are often small. Also, the small number of studies and the heterogeneity of tasks used (e.g., emotional and neutral variants of various cognitive control tasks) limit the possibility of drawing strong conclusions about neural substrates of dysfunctional cognitive control in depression. Another important problem in this field is related to the issue of reverse inference – inferring the engagement of a psychological process from patterns of brain activity (Poldrack, 2006, 2011). For example, the observed difference between depressed and healthy individuals in the dlPFC activity when processing negative material cannot be interpreted in terms of the reduction in the ability of the depressed individuals to exert control over negative material. This type of reverse inference is commonly found in the interpretation of the neuroimaging results in depression. Finally, heterogeneity of depression leaves the possibility that different depression subtypes are related to different neurobiological changes, which could be one of the reasons for the lack of consistency of the neuroimaging studies (for a recent example of the work on neurophysiological subtypes of depression see: Drysdale et al., 2016).

State of the Art on Cognitive Control in Depression

Meta-analyses of behavioral studies reveal reliable depression-related dysfunctions in cognitive control processes when measured with neutral tasks. Neuroimaging studies have started to provide initial evidence for the neurobiological substrates of these dysfunctions, but this research is in early stages. A smaller number of studies offers initial evidence for emotion-specific dysfunctions in different cognitive control components such as inhibition and WM updating. These findings provide support for theoretical models that ascribe an important role to reduced cognitive control in the onset and maintenance of depressive symptoms.

While the relationship between dysfunctional cognitive control and depressive symptoms is empirically supported, the causes and nature of these dysfunctions remain unclear. Questions about the relative importance of specific control dysfunctions in processing emotional material and the question of component-specific vs. general control dysfunctions remain unanswered. Crucially, current research is describing cognitive control dysfunctions, but not proposing mechanisms through which these dysfunctions originate and are maintained. We propose that current conceptualizations of cognitive control used in depression research

contribute to the state of research in this field with lingering major research questions. In the next section, we outline some of the problems with views on cognitive control in this field and analyze how they are hampering further theoretical and empirical progress. We identify three main problems: (1) problems with the way in which the three-component model is used; (2) the problem of the depressed homunculus; and (3) lack of theoretical integration of cognitive control with emotional and motivational processes. We elaborate on each of these issues in the following sections.

Conceptual Problems with Cognitive Control in Depression Research

Problems with the Three-Component Model

The three-component model of cognitive control has been used to guide a large portion of empirical and theoretical work on depression, where most of the studies and frameworks have adopted the division of cognitive control into *inhibition*, *shifting*, and *updating* (Miyake et al., 2000). This three-component model of cognitive control is based on correlational and individual differences research and is primarily descriptive. The three components were extracted through confirmatory factor analysis of multiple (non-emotional) tasks known to engage cognitive control processes and activate the PFC. The resulting components show both unity and diversity, which means that they represent correlated but separable facets of cognitive control. This approach to cognitive control has generated a wealth of research focusing on the relations between specific cognitive control components and other factors such as intelligence, genetic factors, and psychopathology (Braver, Cole, & Yarkoni, 2010; Diamond, 2012; Friedman & Miyake, 2017). In research on depression, studies on cognitive control have investigated depression-related deficits in one of the three components of cognitive control by selecting experimental tasks that are thought to primarily tap into a specific component. While this approach has led to some advances in charting cognitive control dysfunctions in depression, there are several issues with this approach that seem problematic from a methodological and conceptual point of view.

The task-impurity problem

One important methodological problem is related to specificity of the measurement of cognitive control processes. All tasks that have been used to assess cognitive control components involve multiple cognitive processes such that every task reflects an impure measure of the process (Miyake & Friedman, 2012). Apart from the processes targeted to address a particular research question, each cognitive control task also involves low-level visual

processes (e.g., color processing in a Stroop task) and other non-targeted cognitive control processes (e.g., shifting between mental representations often involves updating the contents of WM). This problem is called the *task-impurity problem* (Miyake & Friedman, 2012). This issue is even more problematic in psychopathology research where complex tasks are often used to assess multiple cognitive control processes at once (for a detailed discussion see: Snyder et al., 2015). One of the solutions to address this problem is the use of the latent-variable approach. This approach involves measuring one component with multiple tasks tapping into the same process in order to extract a more pure measure of the cognitive control process involved in the tasks. This approach is rarely applied in clinical research because it requires larger sample sizes and longer study protocols. Rather, most of the subclinical and clinical studies include only one task that is assumed to measure a single cognitive control component in relatively small samples. This imposes significant limitations on depression research. Specifically, the task-impurity problem and the lack of solutions to tackle this problem, challenge the claimed specificity of reported research findings in terms of cognitive control components. This hampers current understanding of the nature of cognitive control dysfunctions involved in depression.

Generalizability of the original component structure

Another problem related to the three-component structure of cognitive control is that this structure was extracted from data collected with tasks presenting emotionally neutral material. It remains unclear to what extent this structure can be replicated in the context of emotional stimulus materials. To date, research has yet to examine similarities and differences in the structure of cognitive control components in response to emotional vs. neutral material. Some indications of potential differences from the original component structure come from a recent study in which no correlations were found between inhibition, shifting, and updating when using emotional tasks (Everaert et al., 2017). In this respect, however, it is important to note that there is an increasing number of studies using control tasks with neutral material that did not replicate the initial three-component structure (Miyake & Friedman, 2012). As a result of these recent insights, changes in the initial conceptualization have recently been proposed (Friedman & Miyake, 2017; Miyake & Friedman, 2012).

The need for controlled processing

As we have stated earlier, the three-component model of cognitive control is primarily descriptive and reflects the factor structure of cognitive control components that provides the

best fit with the dataset at hand. While this approach to cognitive control is very useful for investigating possible dysfunctions in specific components, it is less useful when it comes to other questions relevant for depression research. A very important issue that has attracted limited attention in depression research concerns the question of when, with which intensity, and for how long individuals employ cognitive control. Cognitive control is regarded as a goal-directed and controlled process in opposition to more automatic and habitual processes (for a discussion of the concept of automaticity see: Moors & De Houwer, 2006). While the distinction between automatic and controlled processes has been implied as an important one in depression research (Beevers, 2005; Teachman et al., 2012), dysfunctional cognitive control has rarely been investigated from this perspective. Potential dysfunctions in switching from automatic to more controlled processing (i.e. engaging cognitive control), as well as in detecting the need for such a switch, are hard to address within the framework of the three-component model as it is currently used in depression research.

Interim summary

While the division of cognitive control into three components has led to some advances in clinical research, the use of the three-component model entails problems which cannot be resolved easily. This seems especially true for clinical research. The issues of measuring cognitive control components and the question of processing emotional material significantly contribute to the current state of depression research on cognitive control dysfunctions. Importantly, strictly adhering to the three-component model prevents depression research from posing questions about when and how individuals employ cognitive control.

The Depressed Homunculus

One of the long-standing issues of cognitive control research is the homunculus problem. The problem refers to the tendency of cognitive theories to attribute the ability of control over cognitive processes to a unitary controller – the homunculus (Verbruggen, McLaren, & Chambers, 2014). This problem has been long recognized and fractioning of the controller into more basic processes has been proposed as solution (Monsell & Driver, 2000). In this context, the division of control processes into inhibition, shifting, and updating is commonly understood as partitioning of the controller (A. D. Baddeley, 2012). Still, the homunculus is often merely replaced by multiple homunculi surviving in each of those processes. This problem can be tackled in several ways. In order to use cognitive control processes as explanatory concepts, further work is needed on understanding the simple sub-

processes and their interactions which lead to what is termed *inhibition, shifting*, or *updating* (Verbruggen et al., 2014). Moreover, a complementary approach to tackling the homunculus problem comes from cognitive neuroscience. Models in this domain are trying to replace the homunculus by proposing explicit computational and neural mechanisms which underlie cognitive control processes (Hazy, Frank, & O'Reilly, 2006).

The homunculus problem is particularly visible in clinical psychology where there is a tendency to explain clinical symptoms by the malfunction of the homunculus. As we have previously stated, dysfunctional cognitive control is often used to explain a wide range of depressive symptoms as well as abnormalities in other processes such as emotion regulation. For example, the tendency to ruminate on negative aspects of an event can be explained by the inability to shift attention away from negative thoughts. Although this may seem like an explanation relying on a mechanism, it is just a re-description of the observed behavior in different terms, if no theoretical explanation is proposed for these inabilities (Verbruggen et al., 2014). In the case of cognitive control, there is a strong need for a more mechanistic explanation which could replace the one relying on the malfunction of the homunculus. This explanation should rely on a specific, mechanistic understanding of the dysfunctional cognitive control processes and well specified neural substrates of these processes.

Modern theories of depression that include cognitive control dysfunctions should aim to go beyond the explanations relying on general dysfunctions of one or multiple components of cognitive control. Instead, they should rely on mechanistic explanations and address the causes of the dysfunctions observed in cognitive control tasks. In this context, the questions of why and how individuals employ cognitive control are important because current explanations are largely homunculus-based. In this paper, we provide some of the building blocks for developing such a framework. We point to specific processes that are crucial for cognitive control and are likely altered in depression.

Cognitive Control is not an Isolated Mechanism: The Role of Motivation and Emotion

Models of cognitive control in depression rarely take into account crucial links between cognitive, emotional, and motivational processes. A growing number of researchers propose that emotion and motivation are crucial parts of cognitive control processes, and that the strong distinction between cognitive, emotional, and motivational processes is not theoretically and practically useful (Inzlicht, Bartholow, & Hirsh, 2015; Pessoa, 2008, 2009). Current views on cognitive control in depression research do not account for the interface between emotion,

motivation, and cognitive control. Consequently, most research has focused on the influence of emotion on cognitive control in depression and conceptualized them as separate processes.

Relatedly, depression research has largely focused on cognitive processing of negative material, whereas the processing of motivationally salient material has been picked up only recently in this research domain. This is remarkable in light of the significant role assigned to motivational impairments (i.e., anhedonia) in depression (Pizzagalli, 2014; Treadway & Zald, 2013). Current theoretical models and empirical research have neglected the role of motivation to employ cognitive control processes in the context of depression. The link between motivational processes and cognitive control is crucial given the importance of control for goal-directed behavior. Indeed, motivation has been shown to have strong effects on cognitive control and is of great importance in contemporary models (Botvinick & Braver, 2015). The lack of motivation to engage in controlled processing could play a pivotal role in cognitive control dysfunctions observed in depression, especially in individuals with anhedonic symptoms. However, the importance of motivational factors in this context is currently impossible to estimate given the lack of empirical studies in this domain. The potential role of motivation in cognitive control dysfunctions in depression is an important new avenue for both theoretical and empirical work. By taking into account the potentially different contributions of anhedonia and prolonged negative affect to dysfunctional cognitive control in depression this research field can start to investigate how different depression symptoms affect cognitive processes. This will be a crucial step towards taking into account the heterogeneity of depression which has been largely neglected in this field (Fried & Nesse, 2015).

In conclusion, an important challenge for future cognitive frameworks of depression is to integrate observed dysfunctions in emotional, motivational, and cognitive control processes rather than to investigate and conceptualize them as separable, but interacting processes. This will help to understand the complex nature of cognitive dysfunctions in depression, and relate them to emotional and motivational deficits – potentially revealing mechanistic relationships.

Summary

Our current knowledge about how cognitive control processes may be altered in depression is limited in several important ways. Some of the main problems in this domain are related to the conceptualization and measurement of cognitive control in depression research, reliance on descriptive and homunculus-based explanations, and the lack of integration between cognitive, motivational, and emotional processes. These problems are hindering progress in understanding the causes of dysfunctional cognitive control in depression. Below, we discuss

how several advances in contemporary models of cognitive control can provide novel insights and overcome some of the challenges we have described. We will focus on three big topics in cognitive control research, namely the problem of switching from automatic to more controlled processing, the role of motivation in cognitive control, and the flexibility of cognitive control.

Cognitive Control in Cognitive (Neuro)Science

When to Engage in Controlled Processing?

The distinction between automatic and controlled processing is one of the central topics of research in cognitive psychology (Posner & Snyder, 1975; Shiffrin & Schneider, 1977). While some situations demand control, in other situations, behavior can be automatic with no negative consequences. The ability to overcome habitual actions and engage in more strategic and goal-driven behavior is one of the hallmarks of human behavior. Here, a key question is how the need to switch from more automatic to more controlled processing and behavior is determined. In other words, how do individuals "know" when to engage in controlled processing or when to "turn on cognitive control"?

In this context, there is a prevailing notion that the cognitive system will stay in an automatic mode of processing until a need for cognitive control is detected via changes in current goals or via performance monitoring. Several theoretical models have adopted this approach and have posited the ACC as a key region responsible for detecting the need for control. Some of the proposals for the mechanism through which the need for control is detected include: the presence of response conflict (i.e., Conflict Monitoring Theory: Botvinick et al., 2001), commission of errors followed by omission of rewards (Holroyd & Coles, 2002), and the discrepancy between predicted and obtained outcomes (Alexander & Brown, 2011). The exact way in which the need for cognitive control is detected is still a subject of intense investigation (Brown, 2017; Holroyd & Yeung, 2012; Shenhav et al., 2013; Silvetti et al., 2014). Most theorists agree that the need for control is first detected based on certain changes in the environment, which in turn signal the need to implement control. The role of detecting the need for cognitive control and the intensity of control is assigned to the ACC. The ACC transmits signals to other regions (such as the dlPFC) that in turn implement control.

Motivation and Cognitive Control

The importance of motivation in cognitive control processes is inherent in the definition of cognitive control as a set of processes that support goal-directed, non-automatic behavior. While it may seem that it would be most adaptive to be constantly engaged in controlled

processing and behavior, new research and theoretical advances indicate that engaging in this type of processing carries an intrinsic cost which is named mental effort (Shenhav et al., 2017). It has been shown that engaging in tasks high in cognitive demand, such as cognitive control tasks, is inherently costly and that individuals tend to avoid it, even if this is tied to forgoing substantial rewards (Kool, McGuire, Rosen, & Botvinick, 2010; Westbrook, Kester, & Braver, 2013). In order to pursue a task that involves mental effort, individuals need to be sufficiently motivated. Indeed, a growing number of studies have demonstrated that in non-depressed individuals motivation can enhance a number of cognitive control processes such as response inhibition, task-switching, updating, etc. (Botvinick & Braver, 2015; Braver et al., 2014; Krebs & Woldorff, 2017). Research demonstrating motivational enhancements in cognitive control and costs of control (in terms of mental effort) has led to the development of novel theoretical approaches.

One of the novel theoretical developments which emphasizes the role of motivation in cognitive control comes from the authors of the three-component model (Miyake et al., 2000). Recently, the authors of this model have shifted their focus from diversity towards unity of cognitive control components. These authors have proposed the unity/diversity framework in which a common factor represents the unity of cognitive control processes, while updating-specific and shifting-specific factors represent diversity (Friedman & Miyake, 2017; Miyake & Friedman, 2012). Importantly, the common factor shared by all control components is defined as the "ability to actively maintain task goals and goal-related information and use this information to effectively bias lower-level processing" (Miyake & Friedman, 2012, p. 11).

The open question in this field centers on the problem of how motivation (i.e., current goals) is translated to the allocation of cognitive control. One of the possible mechanisms that may determine when to allocate cognitive control (or effort) is the process reliant on cost-benefit decision-making including information on how much reward is expected from engaging control and how effortful this will be (for a review of these models see: Kool et al., 2017). One of the prominent theories taking this approach is the Expected Value of Control (EVC) theory (Shenhav et al., 2013). EVC theory proposes that a cost-benefit analysis underlies the decision about the amount, timing, and strength of control allocation. This cost-benefit decision-making process is based on three types of information: the expected payoff from the allocation of control, the amount of control needed, and the cost of the control in terms of cognitive effort. EVC theory proposes that the dorsal ACC integrates this information, calculates the expected value of control, and then signals this to regions which implement control such as the dlPFC.

The Balance Between Flexibility and Stability

The idea that information retained in WM can be used to control and guide subordinate cognitive processes and behavior is an enduring principle formulated in early models of WM such as the model of Baddeley and Hitch (1974). Flexible cognitive control poses conflicting functional demands on WM. On the one hand, it is important to maintain stable goal representations in order to guide attentional and decision making processes. In order to do this, the contents of WM have to be protected and any interfering representations must be inhibited. On the other hand, it has to be possible to flexibly update contents of WM in case of salient changes in the environment or changes in one's goals. This poses conflicting demands on working memory and this problem has been termed "flexibility vs. stability paradox" (Bhandari, Badre, & Frank, 2017). While too much flexibility can promote distraction, too little flexibility can lead to rigidity. Hence, a balance between flexibility and stability is crucial for optimal allocation of cognitive control. Two important questions in this context are which mechanisms support flexibility and stability, and how individuals manage to obtain the optimal balance between these processes.

Cognitive neuroscience and computational models of working memory and cognitive control have proposed WM gating as the mechanism underlying flexibility and stability (Braver & Cohen, 1999; Chatham & Badre, 2015; Frank, Loughry, & O'Reilly, 2001; Hazy et al., 2006). If the gate is open WM is sensitive to external input, while when the gate is closed the existing representations are stably maintained. In this way, input gating serves as a selection mechanism that determines the time at which contents of WM can be updated. This gating mechanism is thought to rely on dopamine signaling within the PFC and the striatum (Braver & Cohen, 1999; Westbrook & Braver, 2016) and on fronto-striatal loops involving the PFC and basal ganglia (Frank et al., 2001; Hazy et al., 2006). More recently, a similar output gating mechanism has been proposed (Chatham & Badre, 2015; Chatham, Frank, & Badre, 2014). This mechanism determines which of the representations currently held in WM will control further processing (e.g. biasing attention towards goal-relevant stimuli).

While the described gating models provide an explicit mechanism that underlies the ability to flexibly update or stably maintain WM representations which underlie goal-directed behavior, an important open question is how the gating system learns when to be flexible or stable. The consensus in the field is that it is necessary to avoid potential homunculus-based explanation by relying on learning processes. One of the promising possibilities is that the system is trained by reward prediction errors, i.e. reinforcement learning (Bhandari et al., 2017).

Another perspective proposes that control can be grounded in associative learning and conceptualized as a process reliant on associative networks including perceptual, motor, and goal representations (Abrahamse, Braem, Notebaert, & Verguts, 2016; Verguts & Notebaert, 2009). These perspectives are not mutually exclusive and they offer a new way of circumventing the homunculus problem by linking cognitive control to learning processes while providing a neurobiologically plausible mechanism to support these proposals.

Summary

Cognitive neuroscience research has provided increasingly specific mechanistic views on cognitive control. The homunculus-based views are being replaced by fine-grained processes, such as detecting the need for controlled processing, motivation as an integral component of control, and learning-based flexibility of cognitive control. These novel advances are offering a detailed view of cognitive control as a high-order cognitive process emerging from the interactions between multiple processes with a strong neurobiological foundation. Although these models are still being developed, they provide a wide range of possibilities to solve some of the previously discussed issues in clinical research on cognitive control dysfunctions in depression. Current empirical work and theoretical models have begun to characterize cognitive control dysfunctions in depression and have provided ways in which these dysfunctions contribute to the onset and maintenance of symptoms via emotion regulation strategies. An important next step is to provide an explanation for the causes of these dysfunctions and to investigate their precise nature. In the next section, we discuss the ways in which the described developments in basic cognitive control research can be integrated into depression research and used to advance clinical science.

Updating Current Views on Cognitive Control in Depression

Engaging in Controlled Processing and Depression

The distinction between automatic and controlled processing has been addressed in depression research (Beevers, 2005; Teachman et al., 2012), but an integration with cognitive control processes is lacking. The question of the potential depression-related impairments in detecting when to switch from more automatic to more controlled processing modes (i.e., cognitive control processes) remains largely unaddressed. From a neurobiological point of view, there is evidence that depression is related to disrupted activity of the ACC, a region involved in detecting the need for cognitive control. For example, several studies have revealed abnormal error processing in depression leading to decreased cognitive control recruitment

(Holmes & Pizzagalli, 2009; Holmes & Pizzagalli, 2007). This evidence of reduced ACC activation in depressed individuals coming from both neuroimaging and electrophysiological studies has inspired models which propose disrupted activity of the ACC as a possible biomarker for depression (Holroyd & Umemoto, 2016; Olvet & Hajcak, 2008; Pizzagalli, 2011).

Many of the cognitive models of depression include dysfunctional cognitive control, but do not specify the origin of these dysfunctions. The described models of cognitive control, which deal with the question of how switching between automatic and controlled processing occurs, offer a new avenue that could advance both cognitive research and theorizing in the domain of depression. Starting from the described models of cognitive control, it could be argued that failures to detect the need for increased control, which can be observed as disruptions in ACC activity, cause signals indicating the need to switch to more controlled behavior to be weaker. This is then observed as the decreased tendency of depressed individuals to engage cognitive control. This point of view offers a mechanism for observed cognitive control dysfunctions in depression and ties them closely to neurobiological mechanisms. Still, how to explain a disruption in the ability to detect the need for control? This is where some of the novel models that link cognitive control to motivation may play an important role.

Motivation and Cognitive Control in Depression

Sustained negative affect and anhedonia – the loss of interest or pleasure in previously pleasant activities, are regarded as cardinal symptoms of depression (Gotlib & Furman, 2015). While negative affect has received a lot of attention in cognitive models of depression, anhedonia has been somewhat disregarded. Recently, there is an upsurge of research interest in anhedonia in depression revealing impairments in reward processing and willingness to exert effort.

Research on anhedonia in depression has led to recent proposals to re-conceptualize anhedonia. While there is mixed evidence that depression is linked to reductions in consummatory value (i.e., loss of subjective pleasure coming from obtaining rewards), deficits in the ability to change behaviors in order to maximize reward attainment, in implicit reinforcement learning, and in reward-based decision making have been reliably demonstrated (Admon & Pizzagalli, 2015; Pizzagalli, 2014; Treadway & Zald, 2013). A recent meta-analysis has demonstrated that these depression-linked impairments in reinforcement learning are more related to reduced reward sensitivity then to the learning rates (Huys, Pizzagalli, Bogdan, & Dayan, 2013). Also, electrophysiological studies have demonstrated depression-related reduced

responses to reward attainment (termed reward sensitivity), measured as reward prediction errors (Olvet & Hajcak, 2008) and reward positivity (Proudfit, 2015; Proudfit, Bress, Foti, Kujawa, & Klein, 2015). Another recent research line has demonstrated that depressed individuals are less willing to invest effort into obtaining rewards compared to controls (Treadway, Bossaller, Shelton, & Zald, 2012), and that their perceived level of invested effort differs from objective measures of actual effort invested (Cléry-Melin et al., 2011). These results suggest that depression is not necessarily related to decreased experience of subjective pleasure, but there is electrophysiological evidence for reduced sensitivity to obtaining rewards. Importantly, increased depression levels seem to be linked to a decreased willingness to modify behavior to obtain rewards, an impaired ability to learn from obtaining rewards, and to a dissociation among experienced pleasure and willingness to invest effort into achieving pleasure.

Taking into account the importance of motivation and effort for cognitive control, the investigation of contributions of impairments in these processes to depression-linked dysfunctions in cognitive control seems crucial. Starting from the EVC theory of cognitive control and applying it to depression there are several important things to notice. The EVC theory proposes that decisions about timing and intensity of cognitive control are based on reward prospect related to engaging in controlled processing and expected amount of effort related to this. Both reward processing and effort expenditure are known to be changed in depression. It seems plausible to hypothesize that this could be one of the reasons underlying the depression-related cognitive control dysfunctions. In simplified terms, the idea could be that for depressed individuals the perceived gain from engaging in costly control processes seems small, while the effort seems too big, which could in turn lead to reduced exertion of control. This idea offers the possibility to reinterpret the neuroimaging results on the depression-related changes in the brain regions related to cognitive control. From the perspective of the EVC theory, changes in the perceived values of rewards and the willingness to exert effort would affect the cost-benefit calculation in the dACC which would decrease the intensity of control implemented by the dlPFC.

The contribution of anhedonic symptoms to dysfunctional cognitive control is a research field that – based on our analysis – could be highly informative. While there are theory-based indications that this link could be important, there is almost no empirical research on the topic. Future research should investigate the potential differences in the levels of cognitive control dysfunctions related to different symptoms of depression. The first step in this domain will be to compare cognitive control dysfunctions between individuals with elevated negative mood

and individuals with increased anhedonic symptoms. This will be a crucial step towards uncovering the potential heterogeneity within cognitive control dysfunctions in different subtypes of depression. Further on, research on the role of motivation in cognitive control dysfunctions should elucidate the potentially separable contributions of lack of motivation (e.g. "Is this reward worthy enough?") and the lack of willingness to exert effort (e.g. "How much effort am I willing to invest to attain this reward"). We are currently working on developing paradigms in which we can manipulate both reward prospect and the need to invest effort during a cognitive control task.

Flexibility and Stability in Depression

Psychological flexibility, largely reliant on flexible cognitive control, has been theorized as a fundamental aspect of mental health and may be disrupted in various forms of psychopathology (Kashdan & Rottenberg, 2010). There is evidence that depression is related to both problems in flexibly updating the working memory and in maintaining WM contents, which may guide further information processing. Most of the research and theoretical models in this field are directed towards finding precise impairments in these processes and how they are related to the valence of WM content. Yet, research evidence is still mixed. The cognitive neuroscience models we have described above allow to rephrase the initial question from searching for impairments towards examining *how* depressive symptoms are related to the balance between flexibility and stability. While there is no direct research on the topic yet, there is some relevant indirect evidence about the influence of mood on cognitive flexibility, showing that positive mood promotes flexibility at the cost of stability (Dreisbach, 2006; Dreisbach & Goschke, 2004). One of the possibilities is that prolonged negative mood could promote stable maintenance of WM representations at the cost of the ability to flexibly update content of WM with new information relevant for goal-directed behavior.

Another line of indirect evidence for a relation between flexibility-stability balance and depressive symptoms comes from neuroscience research on depression-related changes in the activity of the striatum – a region central in the gating models of WM. As we have previously described, fronto-striatal loops are implicated in WM gating and neurobiological models of cognitive control. Importantly, dysfunctions of this circuitry are found in a wide range of psychiatric disorders (Gunaydin & Kreitzer, 2016). To date there is evidence that depression is associated with changes in activity of both the dorsal and the ventral striatum in response to pleasant and rewarding stimuli (Pizzagalli, 2014). Also, there is some evidence for depression-related changes in striatal-dACC connectivity in response to losses and rewards (Admon et al.,

2015). However, it remains unclear if dysfunction of the striatal regions in depression can be observed only in the context of processing motivationally salient stimuli, or whether these dysfunctions are broader and can also be observed in neutral cognitive control tasks. An interesting avenue for further research would be to examine striatal activation in depression during neutral cognitive control tasks. The possibility of depression-related changes in the striatum – a region important for gating of information in or out of the WM – in neutral cognitive control tasks in depressed individuals would offer a plausible neurobiological mechanism that can account for some of the cognitive control dysfunctions in depression.

Finally, the theoretical accounts that relate cognitive control to learning mechanisms offer an opportunity to replace the depressed homunculus by well-defined basic mechanisms. For example, one of the proposals which we have described links reward prediction errors and reinforcement learning, known to be altered in depression (Gradin et al., 2011; Kumar et al., 2008; Pizzagalli, 2014), to adaptation of gating policies underlying cognitive control processes. Also, there is evidence for depression-related changes in the dopaminergic system which is crucial in gating models of cognitive control (Dunlop & Nemeroff, 2007; Nestler & Carlezon, 2006). Another proposal relies on associative learning and offers an interesting possibility to study potential depression-related impairments in this type of learning and explore whether these are related to cognitive control dysfunctions (Abrahamse et al., 2016; Verguts & Notebaert, 2009).

Summary

The use of the three-component model has led to research on depression-related impairments in inhibition, shifting, and WM updating. While this approach has contributed to understanding of dysfunctional cognitive control, it has led to multiple problems as well. First of all, the important next step is to avoid homunculus-based explanations at the level of impaired cognitive control processes (e.g., depression-related impairments in updating). While these explanations are a good first step in guiding research, current progress in cognitive neuroscience provides useful avenues by which further research and theorizing can advance. By shifting the focus away from merely establishing specific dysfunctions, we can pose more specific questions such as: how and why do depressed individuals fail to detect the need to engage in controlled processing? Moreover, understanding cognitive control as a process reliant on reinforcement or associative learning will provide the field with the basic mechanisms through which the dysfunctions can occur. Importantly, by raising the questions of how "cognitive control is learned" depression research can investigate more direct links between

stressful life events, which necessarily involve learning, with dysfunctional cognitive control. Finally, the important role of motivation and the willingness to exert effort in cognitive control processes has been largely overlooked in depression research. Novel models of cognitive control offer an exciting possibility of linking anhedonic symptoms with dysfunctional cognitive control in depression. Further integration of cognitive and neuroscience models of cognitive control into the clinical research on depression is a next step to take. Doing so carries the potential to elucidate causes of cognitive control dysfunctions and to enable the understanding of more fine-grained depression-related dysfunctions in cognitive processes.

Conclusions

Current cognitive models of depression consider cognitive control dysfunctions as an important risk factor for depression and an important hub in the depressogenic information processing. After more than a decade of establishing the presence and influence of dysfunctions in cognitive control, it is timely that research goes beyond this descriptive level. Current cognitive models of depression assume cognitive control dysfunctions, either general or more specific ones (e.g. inhibition of irrelevant negative material), without offering a cause and a mechanistic explanation for these dysfunctions. Moreover, most current cognitive models of depression assume that cognitive control is attenuated in depression, which is a very simplistic view of a complex set of dysfunctions. Starting from cognitive neuroscience models of cognitive control, we propose novel avenues for research and theorizing which offer solutions to current limitations of models of dysfunctional cognitive control. We offer a perspective in which deficits in cognitive control observed in depressed individuals do not stem from attenuated ability to exert cognitive control per se. Rather, we argue that dysfunctions arise due to the impaired ability to detect when, with which intensity, and for how long to engage in controlled processing. This process can rely on changes in the environment that signal the need for control (e.g. making a lot of errors) or a cost-benefit decision making process weighing between the prospect of reward related to switching to controlled processing mode and the costs of doing so. Alternatively, it can also rely on previous learning and knowing when to flexibly engage with the environment vs. when to stably maintain a goal. In this way, cognitive control dysfunctions can be viewed not as a structural problem arising from the malfunction of a particular brain region or a homunculus-based process, but rather as a systematic failure to effectively engage in controlled processing that supports goal-directed behavior.

This reconceptualization offers new perspectives on key depressive symptoms. For example, some of the maladaptive emotion regulation strategies can become an automatic mode

of processing negative thoughts and emotions (e.g. rumination). In order to employ a more adaptive emotion regulation strategy, cognitive control might be needed to switch from ruminating and to start reappraising. If there is a failure to detect the need for a switch when ruminating (conflict detection), if there is a lack of motivation (anhedonia), and if switching is perceived as very effortful (effort expenditure problems), this can lead to a failure to employ cognitive control and switch to a different processing mode. In turn, this can then lead to further rumination. While this can be explained as a failure to switch from one task set to another (e.g. switching impairment), but we believe that a focus on potential causes of this process offers a much more in-depth perspective on depressed cognition. While these processes can be generally viewed as a failure to, for example, switch from one task set to another, we believe that the perspective which we provide here is much more specific.

From a clinical perspective, breaking down complex cognitive dysfunctions in depression into multiple smaller-scale problems offers new opportunities with regard to developing targeted treatments and cognitive control trainings. Importantly, cognitive control dysfunctions are also proposed to be a transdiagnostic mechanism as these dysfunctions are observed in a wide range of mental disorders (Goschke, 2014). Advances in the mechanistic understanding of cognitive control dysfunctions in psychopathology will lead to a better understanding of the relations between this and other transdiagnostic mechanisms which will be crucial in moving from the symptom-based towards the mechanism-based view of psychiatric disorders (Cuthbert & Insel, 2013; Insel et al., 2010). Finally, depression research offers an exciting field for testing cognitive neuroscience models of cognitive control in a more applied manner. In this way, the intersection between cognitive and clinical science offers productive avenues for advancing both fields.

Acknowledgements

This work was supported by the Special Research Fund (BOF) of Ghent University [grant number 01D02415 awarded to IG], the starting grant of the European Research Council (ERC) under the Horizon 2020 framework [grant number 636110 awarded to RK], the Concerted Research Action Grant of Ghent University [grant number BOF16/GOA/017 awarded to EHWK], the Special Research Fund (BOF) of Ghent University [grant number BOF16/PDO/046 awarded to JE], and the Belgian American Educational Foundation (BAEF) [postdoctoral fellowship awarded to JE]. We thank Igor Marchetti for his comments on earlier versions of the manuscript and Martin Finn for proofreading the final version of the manuscript.

References

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

Admon, R., Nickerson, L. D., Dillon, D. G., Holmes, A. J., Bogdan, R., Kumar, P., … Pizzagalli, D. a. (2015). Dissociable cortico-striatal connectivity abnormalities in major depression in response to monetary gains and penalties. *Psychological Medicine*, *45*(1), 121–31. http://doi.org/10.1017/S0033291714001123

Admon, R., & Pizzagalli, D. A. (2015). Dysfunctional Reward Processing in Depression. *Current Opinion in Psychology*, *4*(January), 114–118. http://doi.org/10.1016/j.copsyc.2014.12.011

Alexander, W. H., & Brown, J. W. (2011). Medial Prefrontal Cortex as an action-outcome predictor. *Nature Neurosci.*, *14*(10), 1338–1344. http://doi.org/10.1038/nn.2921.Medial

Baddeley, A. D. (2012). Working Memory: Theories, models, and controversies. *Annual Review of Psychology*, *63*(1), 1–29. http://doi.org/10.1146/annurev-psych-120710-100422

Baddeley, A., & Hitch, G. (1974). Working memory. *Psychology of Learning and Motivation*, *8*, 47–89. http://doi.org/10.1016/j.cub.2009.12.014

Barch, D. M., & Pagliaccio, D. (2017). Consistency, Replication, and Meta-analyses of Altered Brain Activity in Unipolar Depression. *JAMA Psychiatry*, *74*(1), 56. http://doi.org/10.1001/jamapsychiatry.2016.2844

Beevers, C. G. (2005). Cognitive vulnerability to depression: A dual process model. *Clinical Psychology Review*, *25*(7), 975–1002. http://doi.org/10.1016/j.cpr.2005.03.003

Beevers, C. G., Clasen, P., Stice, E., & Schnyer, D. (2010). Depression symptoms and cognitive control of emotion cues: A functional magnetic resonance imaging study. *Neuroscience*, *167*(1), 97–103. http://doi.org/10.1016/j.neuroscience.2010.01.047

Bhandari, A., Badre, D., & Frank, M. J. (2017). Learning Cognitive Control. In T. Egner (Ed.), *Wiley Handbook of Cognitive Control* (pp. 376–391). Wiley-Blackwell.

Bockting, C. L., Hollon, S. D., Jarrett, R. B., Kuyken, W., & Dobson, K. (2015). A lifetime approach to major depressive disorder: The contributions of psychological interventions in preventing relapse and recurrence. *Clinical Psychology Review*, *41*, 16–26. http://doi.org/10.1016/j.cpr.2015.02.003

Botvinick, Matthew, M., & Braver, T. (2015). Motivation and Cognitive Control: From Behavior to Neural Mechanism. *Annual Review of Psychology*, *66*, 83–113. http://doi.org/10.1146/annurev-psych-010814-015044

Botvinick, M. M., Braver, T. S., Barch, D. M., Carter, C. S., & Cohen, J. D. (2001). Conflict monitoring and cognitive control. *Psychological Review*, *108*(3), 624–652. http://doi.org/10.1037/0033-295X.108.3.624

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

Braver, T. S., & Cohen, J. D. (1999). Dopamine, cognitive control, and schizophrenia: the gating model. *Progress in Brain Research*, *121*, 327–349.

Braver, T. S., Cole, M. W., & Yarkoni, T. (2010). Vive les differences! Individual variation in neural mechanisms of executive control. *Current Opinion in Neurobiology*, *20*(2), 242–250. http://doi.org/10.1016/j.conb.2010.03.002

Braver, T. S., Krug, M. K., Chiew, K. S., Kool, W., Westbrook, J. A., Clement, N. J., … Somerville, L. H. (2014). Mechanisms of motivation-cognition interaction: challenges and opportunities. *Cognitive, Affective & Behavioral Neuroscience*, *14*(2), 443–72. http://doi.org/10.3758/s13415-014-0300-0

Brown, J. W. (2017). Models of Anterior Cingulate Cortex Function in Cognitive Control. In T. Egner (Ed.), *The Wiley Handbook of Cogntive Control*. Wiley-Blackwell.

Chatham, C. H., & Badre, D. (2015). Multiple gates on working memory. *Current Opinion in Behavioral Sciences*, *1*, 23–31. http://doi.org/10.1016/j.cobeha.2014.08.001

Chatham, C. H., Frank, M. J., & Badre, D. (2014). Corticostriatal output gating during selection from working memory. *Neuron*, *81*(4), 930–942. http://doi.org/10.1016/j.neuron.2014.01.002

Cléry-Melin, M. L., Schmidt, L., Lafargue, G., Baup, N., Fossati, P., & Pessiglione, M. (2011). Why don't you try harder? an investigation of effort production in major depression. *PLoS ONE*, *6*(8), 1–8. http://doi.org/10.1371/journal.pone.0023178

Cohen, J. D. (2017). Cognitive Control: Core Constructs and Current Considerations. In T. Egner (Ed.), *The Wiley Handbook of Cogntive Control*. Wiley-Blackwell.

Crocker, L. D., Heller, W., Warren, S. L., O'Hare, A. J., Infantolino, Z. P., & Miller, G. a. (2013). Relationships among cognition, emotion, and motivation: implications for intervention and neuroplasticity in psychopathology. *Frontiers in Human Neuroscience*, *7*, 261. http://doi.org/10.3389/fnhum.2013.00261

Cuthbert, B. N., & Insel, T. R. (2013). Toward the future of psychiatric diagnosis: the seven pillars of RDoC. *BMC Medicine*, *11*(1), 126. http://doi.org/10.1186/1741-7015-11-126

Davidson, R. J., Pizzagalli, D., Nitschke, J. B., & Putnam, K. (2002). Depression: perspectives from affective neuroscience. *Annual Review of Psychology*, *53*(1), 545–574. http://doi.org/10.1146/annurev.psych.53.100901.135148

De Lissnyder, E., Koster, E. H. W., Everaert, J., Schacht, R., Van den Abeele, D., & De Raedt, R. (2012). Internal cognitive control in clinical depression: general but no emotion-specific impairments. *Psychiatry Research*, *199*(2), 124–30. http://doi.org/10.1016/j.psychres.2012.04.019

Demeyer, I., De Lissnyder, E., Koster, E. H. W., & De Raedt, R. (2012). Rumination mediates the relationship between impaired cognitive control for emotional information and depressive symptoms: A prospective study in remitted depressed adults. *Behaviour Research and Therapy*, *50*(5), 292–7. http://doi.org/10.1016/j.brat.2012.02.012

Derakshan, N., Salt, M., & Koster, E. H. W. (2009). Attentional control in dysphoria: An investigation using the antisaccade task. *Biological Psychology*, *80*(2), 251–255. http://doi.org/10.1016/j.biopsycho.2008.09.005

Diamond, A. (2012). Executive functions. *Annual Review of Psychology*, *64*, 135–68. http://doi.org/10.1146/annurev-psych-113011-143750

Disner, S. G. S., Beevers, C. G. C., Haigh, E. A. P., & Beck, A. T. (2011). Neural mechanisms of the cognitive model of depression. *Nature Reviews Neuroscience*, *12*(8), 467–77. http://doi.org/10.1038/nrn3027

Dreisbach, G. (2006). How positive affect modulates cognitive control: the costs and benefits of reduced maintenance capability. *Brain and Cognition*, *60*(1), 11–9. http://doi.org/10.1016/j.bandc.2005.08.003

Dreisbach, G., & Goschke, T. (2004). How positive affect modulates cognitive control: reduced perseveration at the cost of increased distractibility. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *30*(2), 343–353. http://doi.org/10.1037/0278-7393.30.2.343

Drysdale, A. T., Grosenick, L., Downar, J., Dunlop, K., Mansouri, F., Meng, Y., … Liston, C. (2016). Resting-state connectivity biomarkers define neurophysiological subtypes of depression. *Nature Publishing Group*, *23*(1). http://doi.org/10.1038/nm.4246

Dunlop, B. W., & Nemeroff, C. B. (2007). The role of dopamine in the pathophysiology of depression. *Archives of General Psychiatry*, *64*(3), 327–337. http://doi.org/10.1001/archpsyc.64.3.327

Everaert, J., Grahek, I., & Koster, E. H. W. (2017). Individual differences in cognitive control modulate cognitive biases linked to depressive. *Cognition & Emotion*, *31*(4), 736–746. http://doi.org/10.1080/02699931.2016.1144562

Everaert, J., Koster, E. H. W., & Derakshan, N. (2012). The combined cognitive bias hypothesis in depression. *Clinical Psychology Review*, *32*(5), 413–24. http://doi.org/10.1016/j.cpr.2012.04.003

Foland-Ross, L. C., Hamilton, J. P., Joormann, J., Berman, M. G., Jonides, J., & Gotlib, I. H. (2013). The neural basis of difficulties disengaging from negative irrelevant material in major depression. *Psychological Science*, *24*(3), 334–44. http://doi.org/10.1177/0956797612457380

Frank, M. J., Loughry, B., & O'Reilly, R. C. (2001). Interactions between frontal cortex and basal ganglia in working memory: A computational model. *Cognitive, Affective, & Behavioral Neuroscience*, *1*(2), 137–160. http://doi.org/10.3758/CABN.1.2.137

Fray, P. J., Robbins, T. W., & Sahakian, B. J. (1996). Neuorpsychiatyric applications of CANTAB. *International Journal of Geriatric Psychiatry*, *11*, 329–336.

Fried, E. I., & Nesse, R. M. (2015). Depression sum-scores don't add up: why analyzing specific depression symptoms is essential. *BMC Medicine*, *13*(1), 72. http://doi.org/10.1186/s12916-015-0325-4

Friedman, N. P., & Miyake, A. (2017). Unity and diversity of executive functions: Individual differences as a window on cognitive structure. *Cortex*, *86*, 186–204. http://doi.org/10.1016/j.cortex.2016.04.023

Goschke, T. (2014). Dysfunctions of decision-making and cognitive control as transdiagnostic mechanisms of mental disorders: advances, gaps, and needs in current research. *International Journal of Methods in Psychiatric Research*, *23*(41–57). http://doi.org/10.1002/mpr.1410

Gotlib, I. H., & Furman, D. J. (2015). Depression. *Emerging Trends in the Social and Behavioral Sciences: An Interdisciplinary, Searchable, and Linkable Resource*, 1–15. http://doi.org/10.1002/9781118900772.etrds0074

Gotlib, I. H., & Hamilton, J. P. (2008). Neuroimaging and depression: Current status and unresolved

issues. *Current Directions in Psychological Science*, *17*(2), 159–163. http://doi.org/10.1111/j.1467-8721.2008.00567.x

Gotlib, I. H., & Joormann, J. (2010). Cognition and depression: current status and future directions. *Annual Review of Clinical Psychology*, *6*, 285–312. http://doi.org/10.1146/annurev.clinpsy.121208.131305

Gradin, V. B., Kumar, P., Waiter, G., Ahearn, T., Stickle, C., Milders, M., … Steele, J. D. (2011). Expected value and prediction error abnormalities in depression and schizophrenia. *Brain*, *134*(6), 1751–1764. http://doi.org/10.1093/brain/awr059

Gunaydin, L. A., & Kreitzer, A. C. (2016). Cortico–Basal Ganglia Circuit Function in Psychiatric Disease. *Annual Review of Physiology*, *78*(1), 327–350. http://doi.org/10.1146/annurev-physiol-021115-105355

Hamilton, P. J., Etkin, A., Furman, D. J., Lemus, M. G., Johnson, R. F., & Gotlib, I. H. (2012). Functional neuroimaging of major depressive disorder: a meta-analysis and new integration of baseline activation and neural response data. *American Journal of Psychiatry*, *169*(7), 693–704.

Hazy, T. E., Frank, M. J., & O'Reilly, R. C. (2006). Banishing the homunculus: Making working memory work. *Neuroscience*, *139*(1), 105–118. http://doi.org/10.1016/j.neuroscience.2005.04.067

Holmes, A. J., & Pizzagalli, D. a. (2009). Spatio-temporal Dynamics of Error Processing Dysfunctions in Major Depressive Disorder. *Archives of General Psychiatry*, *65*(2), 179–188. http://doi.org/10.1001/archgenpsychiatry.2007.19.Spatio-temporal

Holmes, A., & Pizzagalli, D. (2007). Task feedback effects on conflict monitoring and executive control: relationship to subclinical measures of depression. *Emotion*, *7*(1), 68–76. http://doi.org/10.1037/1528-3542.7.1.68

Holroyd, C. B., & Umemoto, A. (2016). The Research Domain Criteria Framework: The Case for Anterior Cingulate Cortex. *Neuroscience and Biobehavioral Reviews*, *71*, 418–443. http://doi.org/10.1016/j.neubiorev.2016.09.021

Holroyd, C. B., & Yeung, N. (2012). Motivation of extended behaviors by anterior cingulate cortex. *Trends in Cognitive Sciences*, *16*(2), 122–128. http://doi.org/10.1016/j.tics.2011.12.008

Holroyd, C., & Coles, M. (2002). The neural basis of human error processing: Reinforcement learning, dopamine, and the error-related negativity. *Psychological Review*, *109*(4), 679–709. http://doi.org/10.1037//0033-295X.109.4.679

Huys, Q. J., Pizzagalli, D. A., Bogdan, R., & Dayan, P. (2013). Mapping anhedonia onto reinforcement learning: a behavioural meta-analysis. *Biology of Mood & Anxiety Disorders*, *3*(1), 12. http://doi.org/10.1186/2045-5380-3-12

Insel, T., Cuthbert, B., Garvie, M., Heinssen, R., Pine, D. S., Quinn, K., … Wang, P. (2010). Research Domain Criteria (RDoC): Toward a New Classification Framework for Research on Mental Disorders. *The American Journal of Psychiatry*, *167*(7), 748–751. http://doi.org/10.1176/appi.ajp.2010.09091379

Inzlicht, M., Bartholow, B. D., & Hirsh, J. B. (2015). Emotional foundations of cognitive control. *Trends in Cognitive Sciences*, *19*(3), 126–132. http://doi.org/10.1016/j.tics.2015.01.004

Joormann, J. (2010). Cognitive Inhibition and Emotion Regulation in Depression. *Current Directions in Psychological Science*, *19*(3), 161–166. http://doi.org/10.1177/0963721410370293

Joormann, J., & Gotlib, I. H. (2010). Emotion regulation in depression: relation to cognitive inhibition. *Cognition & Emotion*, *24*(2), 281–98. http://doi.org/10.1080/02699930903407948

Joormann, J., Levens, S. M., & Gotlib, I. H. (2011). Sticky thoughts: Depression and rumination are associated with difficulties manipulating emotional material in working memory. *Psychological Science*, *22*(8), 979–83. http://doi.org/10.1177/0956797611415539

Joormann, J., & Vanderlind, W. M. (2014). Emotion Regulation in Depression: The Role of Biased Cognition and Reduced Cognitive Control. *Clinical Psychological Science*, *2*(4), 402–421. http://doi.org/10.1177/2167702614536163

Joormann, J., Yoon, K. L., & Zetsche, U. (2007). Cognitive inhibition in depression. *Applied and Preventive Psychology*, *12*(3), 128–139. http://doi.org/10.1016/j.appsy.2007.09.002

Kashdan, T. B., & Rottenberg, J. (2010). Psychological flexibility as a fundamental aspect of health. *Clinical Psychology Review*, *30*(7), 865–878. http://doi.org/10.1016/j.cpr.2010.03.001

Kessler, R. C. (2012). The costs of depression. *Psychiatric Clinics of North America*, *35*(1), 1–14. http://doi.org/10.1016/j.psc.2011.11.005

Kessler, R. C., & Bromet, E. J. (2013). The epidemiology of depression across cultures. *Annual Review of Public Health*, *9*(34), 119. http://doi.org/10.1146/annurev-publhealth-031912-114409

Kool, W., McGuire, J. T., Rosen, Z. B., & Botvinick, M. M. (2010). Decision making and the avoidance of cognitive demand. *Journal of Experimental Psychology: General*, *139*(4), 665–682. http://doi.org/10.1037/a0020198

Kool, W., Shenhav, A., & Botvinick, M. M. (2017). Cognitive Control as Cost-Benefit Decision Making. In T. Egner (Ed.), *The Wiley Handbook of Cogntive Control*. Wiley-Blackwell.

Koster, E. H. W., De Lissnyder, E., Derakshan, N., & De Raedt, R. (2011). Understanding depressive rumination from a cognitive science perspective: the impaired disengagement hypothesis. *Clinical Psychology Review*, *31*(1), 138–45. http://doi.org/10.1016/j.cpr.2010.08.005

Koster, E. H. W., Hoorelbeke, K., Onraedt, T., Owens, M., & Derakshan, N. (2017). Cognitive control interventions for depression: A systematic review of findings from training studies. *Clinical Psychology Review*, *53*, 79–92. http://doi.org/10.1016/j.cpr.2017.02.002

Krebs, R. M., & Woldorff, M. G. (2017). Cognitive control and reward. In T. Egner (Ed.), *Wiley Handbook of Cognitive Control* (pp. 422–439). Wiley-Blackwell.

Kumar, P., Waiter, G., Ahearn, T., Milders, M., Reid, I., & Steele, J. D. (2008). Abnormal temporal difference reward-learning signals in major depression. *Brain*, *131*(8), 2084–2093. http://doi.org/10.1093/brain/awn136

Levens, S. M., & Gotlib, I. H. (2010). Updating positive and negative stimuli in working memory in depression. *Journal of Experimental Psychology. General*, *139*(4), 654–64. http://doi.org/10.1037/a0020283

Levens, S. M., & Gotlib, I. H. (2015). Updating emotional content in recovered depressed individuals: Evaluating deficits in emotion processing following a depressive episode. *Journal of Behavior Therapy and Experimental Psychiatry*, *48*, 156–163. http://doi.org/10.1016/j.jbtep.2015.03.009

Millan, M. J., Agid, Y., Brüne, M., Bullmore, E. T., Carter, C. S., Clayton, N. S., … Young, L. J. (2012). Cognitive dysfunction in psychiatric disorders: characteristics, causes and the quest for improved therapy. *Nature Reviews Drug Discovery*, *11*(2), 141–168. http://doi.org/10.1038/nrd3628

Miyake, A., & Friedman, N. P. (2012). The Nature and Organization of Individual Differences in Executive Functions: Four General Conclusions. *Current Directions in Psychological Science*, *21*, 8–14. http://doi.org/10.1177/0963721411429458

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, a H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex "Frontal Lobe" tasks: a latent variable analysis. *Cognitive Psychology*, *41*(1), 49–100. http://doi.org/10.1006/cogp.1999.0734

Monsell, S., & Driver, J. (2000). Banishing the control homunculus. In S. Monsell & J. Driver (Eds.), *Control of cognitive processes: Attention and performance XVIII* (pp. 3–32). Cambridge, MA: MIT Press.

Moors, A., & De Houwer, J. (2006). Automaticity: A Theoretical and Conceptual Analysis. *Psychological Bulletin*, *132*(2), 297–326. http://doi.org/10.1037/0033-2909.132.2.297

Mor, N., & Daches, S. (2015). Ruminative Thinking. *Clinical Psychological Science*, *3*(4), 574–592. http://doi.org/10.1177/2167702615578130

Müller, V. I., Cieslik, E. C., Serbanescu, I., Laird, A. R., Fox, P. T., & Eickhoff, S. B. (2016). Altered Brain Activity in Unipolar Depression Revisited. *JAMA Psychiatry*, 1–9. http://doi.org/10.1001/jamapsychiatry.2016.2783

Nestler, E. J., & Carlezon, W. A. (2006). The Mesolimbic Dopamine Reward Circuit in Depression. *Biological Psychiatry*, *59*(12), 1151–1159. http://doi.org/10.1016/j.biopsych.2005.09.018

Nolen-Hoeksema, S., Wisco, B. E., & Lyubomirsky, S. (2008). Rethinking Rumination. *Perspectives on Psychological Science*, *3*(5), 400–424. http://doi.org/10.1111/j.1745-6924.2008.00088.x

Olvet, D. M., & Hajcak, G. (2008). The error-related negativity (ERN) and psychopathology: Toward an endophenotype. *Clinical Psychology Review*, *28*(8), 1343–1354. http://doi.org/10.1016/j.cpr.2008.07.003

Owens, M., Koster, E. H. W., & Derakshan, N. (2012). Impaired filtering of irrelevant information in dysphoria: An ERP study. *Social Cognitive and Affective Neuroscience*, *7*(7), 752–763. http://doi.org/10.1093/scan/nsr050

Pe, M. L., Brose, A., Gotlib, I. H., & Kuppens, P. (2015). Emotional updating ability and stressful events interact to prospectively predict increased depressive symptoms. *Emotion*, *16*(1), 73–82. http://doi.org/10.1037/emo0000097

Pessoa, L. (2008). On the relationship between emotion and cognition. *Nature Reviews. Neuroscience*, *9*(2), 148–58. http://doi.org/10.1038/nrn2317

Pessoa, L. (2009). How do emotion and motivation direct executive control? *Trends in Cognitive Sciences*, *13*(4), 160–166. http://doi.org/10.1016/j.tics.2009.01.006

Pizzagalli, D. A. (2011). Frontocingulate Dysfunction in Depression: Toward Biomarkers of Treatment Response. *Neuropsychopharmacology*, *36*(1), 183–206.

http://doi.org/10.1038/npp.2010.166

Pizzagalli, D. A. (2014). Depression, Stress, and Anhedonia: Toward a Synthesis and Integrated Model. *Annual Review of Clinical Psychology*, *10*(1), 393–423. http://doi.org/10.1146/annurev-clinpsy-050212-185606

Poldrack, R. A. (2006). Can cognitive processes be inferred from neuroimaging data? *Trends in Cognitive Sciences*, *10*(2), 59–63. http://doi.org/10.1016/j.tics.2005.12.004

Poldrack, R. A. (2011). Inferring mental states from neuroimaging data: From reverse inference to large-scale decoding. *Neuron*, *72*(5), 692–697. http://doi.org/10.1016/j.neuron.2011.11.001

Posner, M. I., & Snyder, C. R. R. (1975). Facilitation and inhibition in the processing of signals. In *Attention and performance V* (pp. 669–682).

Proudfit, G. H. (2015). The reward positivity: From basic research on reward to a biomarker for depression. *Psychophysiology*, *52*(4), 449–459. http://doi.org/10.1111/psyp.12370

Proudfit, G. H., Bress, J. N., Foti, D., Kujawa, A., & Klein, D. N. (2015). Depression and event-related potentials: Emotional disengagement and reward insensitivity. *Current Opinion in Psychology*, *4*, 110–113. http://doi.org/10.1016/j.copsyc.2014.12.018

Rock, P. L., Roiser, J. P., Riedel, W. J., & Blackwell, A. D. (2014). Cognitive impairment in depression: a systematic review and meta-analysis. *Psychological Medicine*, *44*(10), 2029–40. http://doi.org/10.1017/S0033291713002535

Shenhav, A., Botvinick, M., & Cohen, J. (2013). The expected value of control: An integrative theory of anterior cingulate cortex function. *Neuron*, *79*(2), 217–240. http://doi.org/10.1016/j.neuron.2013.07.007

Shenhav, A., Musslick, S., Falk, L., Kool, W., Griffits, T. L., Cohen, J. D., & Botvinick, M. M. (2017). Toward a rational and mechanistic account of mental effort. *Annual Review of Neuroscience*, *40*. http://doi.org/10.1146/annurev-neuro-072116-031526

Shiffrin, R. M., & Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending and a general theory. *Psychological Review*, *84*(2), 127–190. http://doi.org/10.1037/0033-295X.84.2.127

Siegle, G. J., Ghinassi, F., & Thase, M. E. (2007). Neurobehavioral therapies in the 21st century: Summary of an emerging field and an extended example of cognitive control training for depression. *Cognitive Therapy and Research*, *31*(2), 235–262. http://doi.org/10.1007/s10608-006-9118-6

Silvetti, M., Alexander, W., Verguts, T., & Brown, J. W. (2014). From conflict management to reward-based decision making: Actors and critics in primate medial frontal cortex. *Neuroscience and Biobehavioral Reviews*, *46*, 44–57. http://doi.org/10.1016/j.neubiorev.2013.11.003

Snyder, H. R. (2013). Major depressive disorder is associated with broad impairments on neuropsychological measures of executive function: a meta-analysis and review. *Psychological Bulletin*, *139*(1), 81–132. http://doi.org/10.1037/a0028727

Snyder, H. R., Miyake, A., & Hankin, B. L. (2015). Advancing understanding of executive function impairments and psychopathology: Bridging the gap between clinical and cognitive approaches. *Frontiers in Psychology*, *6*. http://doi.org/10.3389/fpsyg.2015.00328

Teachman, B. A., Joormann, J., Steinman, S. A., & Gotlib, I. H. (2012). Automaticity in anxiety disorders and major depressive disorder. *Clinical Psychology Review*, *32*(6), 575–603. http://doi.org/10.1016/j.cpr.2012.06.004

Treadway, M. T., Bossaller, N. A., Shelton, R. C., & Zald, D. H. (2012). Effort-based decision-making in major depressive disorder: A translational model of motivational anhedonia. *Journal of Abnormal Psychology*, *121*(3), 553–558. http://doi.org/10.1037/a0028813

Treadway, M. T., & Zald, D. H. (2013). Parsing Anhedonia: Translational Models of Reward-Processing Deficits in Psychopathology. *Current Directions in Psychological Science*, *22*(3), 244–249. http://doi.org/10.1177/0963721412474460

Verbruggen, F., McLaren, I. P. L., & Chambers, C. D. (2014). Banishing the Control Homunculi in Studies of Action Control and Behavior Change. *Perspectives on Psychological Science*, *9*(5), 497–524. http://doi.org/10.1177/1745691614526414

Verguts, T., & Notebaert, W. (2009). Adaptation by binding: a learning account of cognitive control. *Trends in Cognitive Sciences*, *13*(6), 252–257. http://doi.org/10.1016/j.tics.2009.02.007

Vittengl, J. R., Clark, L. A., Dunn, T. W., & Jarrett, R. B. (2007). Reducing relapse and recurrence in unipolar depression: A comparative meta-analysis of cognitive-behavioral therapy's effects. *Journal of Consulting and Clinical Psychology*, *75*(3), 475–488. http://doi.org/10.1037/0022-006X.75.3.475.Reducing

Westbrook, A., & Braver, T. S. (2016). Dopamine Does Double Duty in Motivating Cognitive Effort. *Neuron*, *89*(4), 695–710. http://doi.org/10.1016/j.neuron.2015.12.029

Westbrook, A., Kester, D., & Braver, T. S. (2013). What Is the Subjective Cost of Cognitive Effort? Load, Trait, and Aging Effects Revealed by Economic Preference. *PLoS ONE*, *8*(7), 1–8. http://doi.org/10.1371/journal.pone.0068210

Whitmer, A. J., & Gotlib, I. H. (2013). An attentional scope model of rumination. *Psychological Bulletin*, *139*(5), 1036–61. http://doi.org/10.1037/a0030923

Chapter 3

Motivation and Cognitive Control in Depression[1]

Abstract

Depression is linked to deficits in cognitive control and a host of other cognitive impairments arise as a consequence of these deficits. Despite of their important role in depression, there are no mechanistic models of cognitive control deficits in depression. In this paper we propose how these deficits can emerge from the interaction between motivational and cognitive processes. We review depression-related impairments in key components of motivation along with new cognitive neuroscience models that focus on the role of motivation in the decision-making about cognitive control allocation. Based on this review we propose a unifying framework which connects motivational and cognitive control deficits in depression. This framework is rooted in computational models of cognitive control and offers a mechanistic understanding of cognitive control deficits in depression.

[1] Based on: Grahek, I., Shenhav, A., Musslick, S., Krebs, R. M., & Koster, E. H. (2019). Motivation and cognitive control in depression. *Neuroscience & Biobehavioral Reviews*, *102*, 371-381.

Introduction

Depression[2] profoundly influences the way in which we process information and think about ourselves, others, and the world around us. An individual suffering from depression will take a longer time to disengage from the processing of negative information and will experience difficulties in suppressing irrelevant thoughts, or shifting attention from one task to another in order to reach a goal. Such issues will make it difficult for that individual to regulate emotions and adapt to the changing environment. This is why cognitive processes are a crucial target for understanding and treating depression (Clark and Beck, 2010; Kaser et al., 2017).

Depression is characterized by impairments in attention, memory, and cognitive control (Millan et al., 2012). Cognitive control deficits are related to central features of depression such as concentration and memory problems and a host of other cognitive impairments and biases arise as a consequence of these deficits (Disner et al., 2011; Gotlib and Joormann, 2010). Cognitive control is crucial in motivated, goal-directed behavior. It represents a set of processes that allow for the flexible adaptation of cognition and behavior in accordance with our current goals (Botvinick and Cohen, 2014; Friedman and Miyake, 2017; Shenhav et al., 2013). For example, cognitive control is necessary if we want to inhibit negative thoughts and shift our attention to a new task. Impairments in such control processes are found across a wide range of psychiatric disorders (Millan et al., 2012), and they have been consistently linked with depressive symptoms (Snyder, 2013). However, our understanding of cognitive control in depression is limited in several important ways.

Current understanding of cognitive control in depression is predominately descriptive and research is focused on detecting deficits in specific cognitive processes, such as the inhibition of negative material (for a review see: Grahek et al., 2018). In spite of the important progress in charting cognitive control deficits related to depressive symptoms, the origin of these deficits remains poorly understood. It is not known why deficits in cognitive control develop, or how they are maintained. Most of the existing models view cognitive control deficits in depression as the reduced ability to exert control and do not offer mechanisms through which these deficits emerge. Currently there is a strong need for the development of a more mechanistic account of cognitive control deficits in depression. A mechanistic account

[2] Throughout this paper we use the term depression to denote the Major Depressive Disorder.

moves beyond identification and description of a phenomenon. It does so by appealing to a mechanism: a structure defined by its components, their organization and interactions, which produce a phenomenon (Bechtel and Abrahamsen, 2005; Machamer et al., 2000). In this paper we argue that recent developments in research on motivation and cognitive control, as well as the development of computational theories of cognitive control, can contribute to such a mechanistic understanding of control deficits in depression. The use of such theories to explain cognitive control deficits in depression, instead of developing new models specific for psychopathology, holds the promise of advancing the understanding of cognitive control in both healthy and depressed individuals.

In this paper we first review disparate literatures on motivation and cognitive control in depression. Further, we describe computational models of cognitive control and demonstrate how they can be used to link motivation and cognition in depression. On the basis of this review, we rely on a computational model of cognitive control to propose a framework in which cognitive control deficits in depression arise from alterations in crucial components of motivation: reward anticipation, effort costs, and estimates of environment controllability. This view offers the possibility for re-conceptualizing depression-related cognitive control deficits. Instead of a *reduced ability* to employ control, we propose that control deficits can be viewed as changes in the decision-making process underlying cognitive control allocation. This decision-making process relies on crucial components of motivation: reward anticipation, effort, and estimates of the ability to control the environment. We use simulations of two cognitive tasks to demonstrate how this framework can be used to derive behavioral predictions about the impact of motivational impairments on cognitive control.

Cognitive control in depression

Deficits in cognitive control have not only been documented in clinically depressed individuals (Snyder, 2013), but also in patients in remission (Demeyer et al., 2012; Levens and Gotlib, 2015), and in at-risk populations (Derakshan et al., 2009; Owens et al., 2012). Meta-analytic evidence from behavioral studies suggests that depression is reliably linked to deficits in cognitive control (Rock et al., 2014; Snyder, 2013). There is also emerging evidence that cognitive remediation training aimed at improving cognitive control processes reduces depressive symptoms (for a review see: Koster, Hoorelbeke, Onraedt, Owens, & Derakshan, 2017). At the neural level, depressive symptoms have been linked to changes in the activity of the dorsolateral prefrontal cortex (dlPFC) and anterior cingulate cortex (ACC; Davidson, Pizzagalli, Nitschke, & Putnam, 2002; Gotlib & Hamilton, 2008; Pizzagalli, 2011). Meta-

analyses of the neuroimaging studies also point to the differences between healthy and depressed individuals, both in activation in these two regions during cognitive tasks (McTeague et al., 2017), as well as in the gray matter volume (Goodkind et al., 2015). However, the neuroimaging studies have often been conducted on very small samples, and there is considerable heterogeneity in their results (e.g., Müller et al., 2016; for a discussion see: Barch & Pagliaccio, 2017). Multiple authors have proposed that the reduced activity in the dlPFC and the ACC is related to the diminished ability of depressed individuals to employ cognitive control (Disner et al., 2011; Joormann, Yoon, & Zetsche, 2007).

Cognitive control processes are considered to be an important vulnerability factor for depression. Cognitive impairments in attention, interpretation, and memory may arise as a consequence of control deficits (Gotlib & Joormann, 2010; Millan et al., 2012; Siegle et al., 2007). For example, Gotlib and Joorman (2010) have suggested that depressed individuals' difficulties in disengaging attention from negative stimuli, or forgetting such stimuli, could be caused by cognitive control deficits. This proposal has recently received empirical support (Everaert et al., 2017). Lowered levels of cognitive control increase and sustain depressive symptoms via their proximal links with emotion regulation strategies such as rumination (Joormann & Vanderlind, 2014; Whitmer & Gotlib, 2013). Research on cognitive control in depression has been focused on charting deficits in different cognitive control processes. Specific deficits in processes such as inhibition, shifting, and updating have been documented (Joormann & Tanovic, 2015). These deficits are commonly thought of as the *lowered ability* to inhibit certain thoughts or stimuli, shift attention away from them, or update the contents of working memory. However, it is important to note that not all accounts of cognitive impairments in depression postulate a reduced ability. For example, the cognitive-initiative account of memory in depression focuses on changes in initiative – a concept close to motivation – to explain memory impairments in depression (Hertel, 2000, 1994).

In a recent analysis of theoretical models of cognitive control in depression (Grahek et al., 2018) we identified three main conceptual problems in the field: (1) the use of descriptive models of cognitive control, (2) the reliance on describing the impairments instead of searching for mechanisms, and (3) the lack of integration between cognitive, motivational, and emotional impairments. These issues are hindering further progress in understanding how and why cognitive control is impaired in depression. In order to overcome some of the problems that we have outlined in our earlier work, in this paper we propose an integrated framework that links alterations in motivational processes with cognitive control deficits. This framework allows us

to move away from the view that cognitive control deficits in depression stem from a reduced ability to exert control. Instead, we will argue for a view in which the deficits arise as a result of altered expectations about the value of exerting control.

Components of motivation in depression

Cognitive impairments in depression are closely linked to impairments in emotional and motivational processes (Crocker et al., 2013). A wealth of depression research has focused on the relationship between cognitive and emotional processes. This approach has led to insights in key deficits related to the disengagement from emotionally negative material (Koster et al., 2011) and the ability to deploy cognitive control over emotional material (Joormann, 2010). While the processing of negative material and the presence of negative affect have been studied in relation to cognitive control impairments in depression, motivational deficits remain largely unexplored in this context. This is why we focus this paper on the link between motivation and cognition in depression. The links between motivation and cognitive processing of emotional material our out of the scope of the current paper.

Motivation is goal-directed when effort is invested in order to bring about desired outcomes (Braver et al., 2014). Here we will focus on components of motivation that are relevant for goal-directed behavior because of their relevance for cognitive control processes. Motivated goal-directed behavior is flexible and sensitive to the current state of the individual and the environment. Two types of representations are crucial in driving this type of behavior: 1) action-outcome contingencies, and 2) the value of potential outcomes (Balleine and O'Doherty, 2010; Dickinson, 1985; Wood and Rünger, 2016). Action-outcome contingencies represent the probability that an action will result in a desired outcome. We will refer to these representations as *outcome controllability* (one's estimate of their ability to control outcomes in an environment) and *outcome value* (the expected reinforcement - total reward and/or punishment - for reaching an outcome). The third concept that we will consider is effort, a variable that is central to the study of motivation. Effort represents the intensification of physical or mental activity needed to reach a goal (Inzlicht et al., 2018; Kurzban et al., 2013). This intensification comes with a cost, and we will refer to the effort requirements for reaching an outcome as *effort costs.*

Goal-directed motivated behavior emerges with the integration of these three components. For example, imagine a person working long hours in order to get a promotion at work. This behavior is motivated and goal-directed because this person believes that working

hard (high effort) will lead to the promotion (high outcome controllability), and the promotion is desired (high outcome value). As we review next, there is evidence that each of these three components of motivation can be impaired in depression (Barch et al., 2015; Griffiths et al., 2014).

Outcome controllability

The classic paradigm for investigating the role of outcome controllability in response to stressors was developed by Seligman and Maier (1967). They demonstrated that animals who were subjected to uncontrollable stressors (inescapable shocks) exhibited passivity, a phenomenon they referred to as learned helplessness. They found that controllable stressors (escapable shocks) did not induce learned helplessness. Overall, uncontrollable stressors were found to induce responses that resemble some of the symptoms of depression (Maier, 1984; Maier & Watkins, 2005). Further work by Maier and colleagues (2006) revealed that animals were able to detect the possibility of control over their environment. While animals have a default reaction of passivity when experiencing stress, this default response can be overcome by learning that the stressors are controllable. The medial prefrontal cortex (mPFC) in rodents detects whether a stimulus is controllable and inhibits the default passivity in response to shocks. In this way, the ability to exert control over the environment serves as a protective factor against negative behavioral and physiological effects of stress (Maier & Seligman, 2016). Seligman and colleagues demonstrated that learned helplessness is also observed in humans, and suggested that it has significant relevance for understanding depression and related disorders (Abramson, Seligman, & Teasdale, 1978; Maier & Seligman, 2016; Seligman, 1972).

The idea that environmental controllability is crucial for one's health, security, and well-being is supported by several other lines of research (Leotti et al., 2010). For instance, Moscarello and Hartley (2017) have recently proposed that goal-directed behavior is strongly influenced by estimates of agency. The authors propose that animals and humans infer their ability to control their environment by ascertaining the relationship between their actions and motivationally significant outcomes. These estimates are generalized and determine the probability of being able to exert control in a novel environment. Estimates of agency are used to calibrate ongoing behaviors. If the estimates of agency are high (i.e., high probability of being able to control the environment), goal-directed behavior is promoted. If the estimates are low, behavior is more likely to rely on an innate reactive repertoire. In this way, controllability of the environment, inferred from previous learning, is crucial for promoting either goal-directed or habitual behavior (Liljeholm, Tricomi, O'Doherty, & Balleine, 2011; Miller,

Shenhav, & Ludvig, in press). Lowered estimates of outcome controllability, resulting in alterations of goal-directed behavior, could be an important factor in depression.

Recent theoretical frameworks of depression also emphasize the importance of outcome controllability. For example, Pizzagalli (2014) has proposed a model of anhedonia in depression in which stress plays a central role in the development of anhedonic symptoms. According to Pizzagalli, certain stressors, especially if they are uncontrollable, induce anhedonic behavior by causing dysfunctions in mesolimbic dopamine pathways crucial for motivated behavior. De Raedt and Hooley (2016) have argued that individual's expectancies about their ability to cope with future negative events play a crucial role in depression. These expectancies are proposed to be formed based on previous coping experiences and are one factor that determines the actual ability to cope with stressors when they occur. Notably, these authors suggest that expectancies influence the ability to cope with stressors by modulating the proactive allocation of cognitive control prior to encountering a stressful situation. In sum, altered estimates of outcome controllability seem to be an important factor contributing to the levels of depressive symptoms.

Outcome value

Representations of outcome value are also strongly altered in depression. In recent years, motivational impairments in depression have received significant attention and a more fine-grained view of these impairments is emerging (Pizzagalli, 2014; Treadway and Zald, 2013). Anhedonia is one of the two core symptoms of depression and is defined as a loss of pleasure in previously enjoyable activities or a loss of interest in pursuing them (DSM-5; American Psychiatric Association, 2013). Anhedonia is a good predictor of antidepressant treatment success (Uher et al., 2012), the course of depression (Spijker et al., 2001; Wardenaar et al., 2012), and the time to remission and number of depression-free days after antidepressant treatment (McMakin et al., 2012). In spite of the importance of anhedonia in depression, the conceptualization and measurement of anhedonia has been heterogeneous and inconsistent (for a discussion see: Rizvi, Pizzagalli, Sproule, & Kennedy, 2016). Traditionally, anhedonia has been primarily viewed as an impairment in consummatory pleasure (i.e. liking rewards when they are obtained). However, the evidence for impairments in consummatory pleasure in depression is mixed (Barch et al., 2015; Treadway and Zald, 2013).

Animal models suggest that the mesolimbic dopamine system is selectively involved in reward motivation, but not in hedonic responses when rewards are gained (Haber and Knutson, 2010; Salamone and Correa, 2012). These insights have stimulated research on anhedonia in

depression and schizophrenia, revealing impairments in motivation that are not related to hedonic responses. As a result, an emerging, more nuanced account of motivation in anhedonia, emphasizes multiple reward processing deficits such as anticipation of rewards, reinforcement learning, effort expenditure, and value-based decision making (Barch and Dowd, 2010; Der-Avakian and Markou, 2012; Pizzagalli, 2014; Romer Thomsen et al., 2015; Strauss and Gold, 2012; Zald and Treadway, 2017). These deficits point toward changes in outcome value. They influence how outcome values are learned, anticipated, and translated into behavior.

Depression is related to a number of reward processing deficits (for reviews see: Barch et al., 2015; Pizzagalli, 2014; Treadway & Zald, 2011). Converging evidence from self-report studies, behavioral tasks, physiological, and neuroimaging experiments suggests that depression, especially in the presence of anhedonia, is linked to reduced anticipation of rewards and impaired implicit reinforcement learning. For example, monetary incentive delay tasks (Knutson et al., 2000) were developed in order to decompose anticipatory (e.g., the period after notifying participants about a possible reward) and consummatory (e.g., receipt of a monetary reward) aspects of reward processing. In these tasks depressed individuals, relative to healthy controls, display reduced behavioral and neural responses in anticipation of rewards (Pizzagalli, 2014). The literature on motivational impairments in depression suggests that depressed individuals, especially in the presence of anhedonic symptoms, assign value outcomes in a manner that differs from healthy controls. This can be caused by anticipating lower payoffs and/or impaired reinforcement learning.

Effort costs

Another important factor influencing motivated behavior is effort that needs to be expended in order to reach a desired outcome. Effort involves the expected cost necessary to reach an outcome. This cost is weighed against expected benefits in order to choose which actions to pursue (Wallis and Rushworth, 2014). Reduced effort exertion is associated with multiple mental disorders (Culbreth et al., 2018; Salamone et al., 2016). Clinical studies have focused on the exertion of physical effort in order to obtain rewards, demonstrating that anhedonia in depression is related to reduced effort exertion (for an excellent review see: Zald & Treadway, 2017). In order to investigate this process, Treadway and colleagues developed the Effort Expenditure for Rewards Task (EEfRT; Treadway, Bossaller, Shelton, & Zald, 2012), which involves having participants choose how much effort they want to expend in order to gain varying amounts of reward. Using this task, the authors have demonstrated that depressed individuals are less willing to exert effort than healthy controls (see also: Cléry-Melin et al.,

2011; Yang et al., 2014). However, the research on effort exertion in anhedonia and depression has largely been focused on physical effort while ignoring cognitive effort.

Although cognitive effort has been an important topic of research for a long time (Kahneman, 1973), in recent years there has been an upsurge in cognitive and neuroscience research on this topic (Kool and Botvinick, 2018; Westbrook and Braver, 2015). More specifically, research on cognitive control has focused on the role of effort costs in decision-making about control allocation (for a recent review see: Shenhav et al., 2017). Cognitive control processes require more effort than automatic ones, and effort needs to be expended in order to override automatic processes in order to reach a goal. Research is starting to demonstrate that there are individual differences in the exertion of this type of effort (Westbrook et al., 2013). To date, there are not a lot of studies on cognitive effort in depression. However, a recent study has demonstrated the inverse relationship between depressive symptoms and a questionnaire-based measure of the willingness to exert cognitive effort (Marchetti et al., 2018). Also, the first experimental study of cognitive effort in depression revealed similar results to the results obtained with physical effort (Hershenberg et al., 2016).

Cognitive control as a process reliant on motivation

Controllability, value, and effort of a desired outcome depend on previous learning. They have been studied extensively in the context of animal and human learning at both behavioral and neural levels (for an overview see: Daw & O'Doherty, 2014). The importance of these components of motivated behavior has also been recognized in other fields of psychology and neuroscience. For example, in social psychology the concepts of self-efficacy (Bandura, 2001, 1977) and locus of control (Rotter, 1966) are closely linked to controllability. Also, concepts of feasibility and desirability of goals (Oettingen & Gollwitzer, 2001) correspond to the concepts of controllability and outcome value. Similar concepts can be found in psychology of motivation (Atkinson, 1957; Wigfield & Eccles, 2000) and work psychology (Bonner & Sprinkle, 2002). The role of motivation is now increasingly recognized in cognitive psychology and cognitive neuroscience.

Automatic or habitual responding can be suitable for a number of everyday situations, but for tasks that are more novel, uncertain, or complex, individuals need to engage cognitive control. This set of processes helps to overcome automatic response tendencies in favor of controlled modes of information processing and behavior. This allows for the coordination of our thoughts and actions in accordance with our goals. Recent research has started to emphasize

the role of motivation in the timing, intensity, and direction of cognitive control allocation. It is now recognized that the allocation of cognitive control is driven by goals and therefore closely linked to motivation (Braver et al., 2014).

Reward prospect enhances cognitive control processes (for reviews see: Botvinick & Braver, 2015; Krebs & Woldorff, 2017). This effect has been observed on various tasks that tap into different components of control such as: attentional control (Padmala and Pessoa, 2011), response inhibition (Leotti & Wager, 2010), conflict adaptation (Braem et al., 2012), and task-switching (Aarts et al., 2010). The effect is not restricted to situations in which reward is signaled by advance cues augmenting preparatory control processes. Comparable results are found when reward is signaled simultaneously with the target thereby promoting fast control adjustments (Krebs et al., 2010). But why is there a need to adapt control in the first place? Is it not the most optimal strategy to always exert maximal control over our thoughts and actions? The emerging answer is that exerting cognitive control carries intrinsic effort-related costs (Shenhav et al., 2017). It has been shown that exerting control is effortful and that individuals tend to avoid it (Kool et al., 2010). Both reward prospect associated with an outcome, and the effort needed to reach that outcome are important factors in determining the way in which cognitive control is allocated.

Another important factor in determining how cognitive control is allocated is the learned contingency between actions and outcomes. Although this factor has received less empirical attention, several computational models of cognitive control emphasize its importance. For example, the model of Alexander and Brown (2011) stresses the importance of predicting mappings between responses and outcomes in cognitive control. The model proposed by Shenhav and colleagues (2013) posits that the allocation of cognitive control relies, among other factors, on the probability of an outcome conditional on the type and intensity of control. In this way, allocation of cognitive control depends on predictions about probabilities of certain actions (e.g., intense control allocation or a certain response) producing desired outcomes (e.g., solving a task correctly).

Computational models of cognitive control

The insight that cognitive control is closely related to motivation is formalized in a growing number of computational models of cognitive control. At present there are multiple theoretical and computational models which deal with the problem of how cognitive control is allocated (for recent reviews see: Botvinick & Cohen, 2014; Verguts, 2017). These models

specify a set of components needed for optimal decision-making about control allocation and they are not mutually exclusive. Many of them include the representations of outcome value and/or outcome controllability (Brown and Alexander, 2017; Holroyd and McClure, 2015; Shenhav et al., 2013; Verguts et al., 2015). Also, many of these models assume that learning is crucial in forming these representations and emphasize the importance of learning for cognitive control (Abrahamse et al., 2016; Bhandari et al., 2017). Our framework relies on the current formulation of the Expected Value of Control (EVC) theory (Shenhav et al., 2013). This theory includes representations of outcome controllability, value, and cost in the decision-making process about cognitive control allocation.

The EVC theory provides a normative account of cognitive control allocation as resulting from a decision-making process in which potential gains of control allocation are weighted against their costs (Shenhav et al., 2013). The theory postulates that these decisions determine how control is allocated across candidate control signals. These control signals vary on two dimensions: *signal identity*, which determines the processes which are engaged (e.g., paying attention to the ink color in a Stroop task), and *signal intensity* which determines the amount of control allocated (e.g., how intensely to pay attention to the ink color). The theory proposes that control signals are selected in such a way that maximizes the expected value of control at any given moment.

The EVC for a given control signal within a given state is determined by three components: efficacy, value, and cost. *Efficacy* is defined as the probability of a certain outcome given a signal of a particular identity and intensity, and the current state. In other words, the efficacy component can be described as the probability that certain outcomes will occur (i.e., that our actions lead to what we desire) or simply as action-outcome contingencies inferred from previous experience. In standard cognitive control paradigms, such as the Stroop task, an outcome can be a correct or an incorrect response on a particular trial. In this context, we can assign a probability to each of the two possible outcomes given each of the possible signal intensities and identities. *Value* is assigned to each of the possible outcomes and it represents the value of an outcome in terms of possible rewards or punishments associated with the outcome. The rewards can be both intrinsic (i.e., motivation to do the task well) and extrinsic (i.e., monetary rewards for good performance). Punishments can also come from multiple sources such as monetary loss or poor task performance. Outcome values are also modulated by the time it takes for the outcomes to occur, as individuals try to maximize reward rates (rewards per unit of time) and not rewards per se. *Cost* is defined as the expected cost associated

with the specified intensity of control allocation. Within the EVC theory, cost arises as the property of the neural system. This means that as the intensity of a signal to allocate control becomes stronger, so does the cost of allocating control. In sum, the EVC is determined by the probability of an outcome for a given control signal (efficacy), the value of that outcome, and the cost associated with control allocation.

At the neural level, the EVC theory was developed to explain the function of the dorsal anterior cingulate cortex (dACC). The theory assigns the critical role in decision-making about control allocation to this region. It is proposed that the dACC plays a key role in calculating the EVC and that the outcome of this process is then signaled to other regions which implement control, such as the dlPFC (Shenhav et al., 2016, 2013). The dACC is assigned a role in both monitoring for the changes in the environment relevant for control (rewards, punishments, errors, etc.), and, based on this, specifying which control signals maximize the EVC.

The EVC theory provides a detailed and formal account of how cognitive control is allocated. One of the main advantages of the theory is its ability to integrate a wide range of behavioral and neural findings. The computational implementation of the theory has been shown to account for various effects associated with the allocation of cognitive control such as the sequential adaptation effects and post-error slowing (Musslick et al., 2015), including how individuals can learn about features of their environment that predict incentives for and demands of control allocation (Lieder et al., 2018).

Proposed framework

We propose a framework that integrates disparate areas of research: cognitive control deficits and motivational impairments in depression. This framework encompasses the controllability, value, and effort of outcomes in order to provide a more mechanistic understanding of these deficits. Furthermore, the framework emphasizes the crucial role of learning in the process of allocating cognitive control. In this way the framework allows for the integration of motivational, learning, and cognitive control deficits in depression. We put forward direct links between components of motivation and reduced cognitive control in depression and propose how these links emerge through learning. In this way our framework offers a first step toward understanding how cognitive control deficits arise. In order to do so, we rely on the EVC theory which offers a computationally explicit model of cognitive control (Shenhav et al., 2013). This allows us to focus on the ways in which expected value of control might be reduced in depression, which results in allocating cognitive control with reduced

intensity. We propose that cognitive control deficits in depression arise as a consequence of changes in the decision-making process about control allocation. These deficits in depression occur due to lowered expected value of control. The expected value of control relies on three components of motivation: controllability of outcomes, their value, and the effort needed to attain them.

We argue that research on motivation in depression provides insights into mechanisms leading to a reduction in the expected value of control. The efficacy component in the EVC theory is closely linked to what we have termed outcome controllability - the estimates of the probability that actions will lead to desired outcomes. These representations are central in guiding motivated behavior (Moscarello and Hartley, 2017). As we have reviewed, there is emerging evidence that estimates of outcome controllability and beliefs related to controllability of the environment, are changed in depression. The value component in the EVC theory is related to what we have termed outcome value - the expected reinforcers following an outcome. We have outlined how reward anticipation and reinforcement learning are impaired in depression, especially in the presence of anhedonic symptoms (Treadway and Zald, 2013). The cost component in the EVC theory is linked to the cost of effort needed to obtain rewards. We have reviewed how this process is altered in depression. All of these components of motivation rely on previous learning to estimate the probability of being able to exert control in a new environment, the value of outcomes in that environment, and the effort needed in order to reach them. These insights from the study of motivation in depression offer evidence for the reduction in efficacy, value, and cost of control. This leads to lower expected value of control which in turn results in reduced allocation of cognitive control (Figure 1).

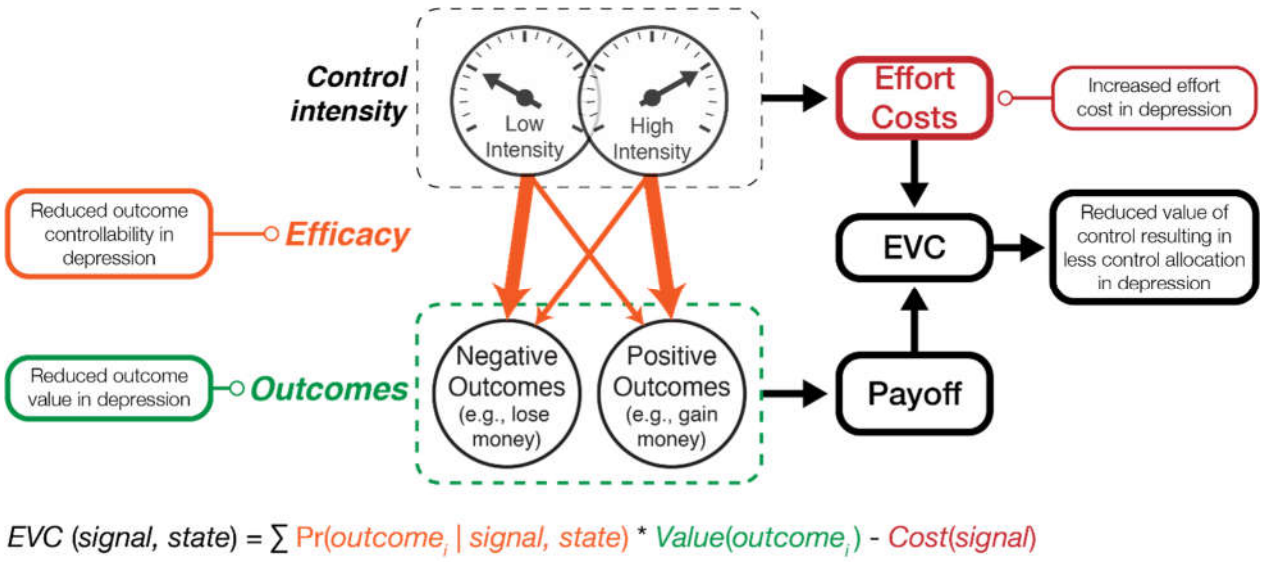


Figure 1. The schematic representation of the proposed framework. Depression-related changes in the outcome value (e.g. reduced reward anticipation), outcome controllability (e.g. lowered estimates of controllability), and effort costs (e.g. reduced effort exertion), lead to the reduced value of control. This leads to the lowered amounts of control being allocated. The expected value of control (*EVC*) for a signal of a given intensity is calculated as the sum of the values of each possible outcome weighted by the probability of reaching that outcome for the given signal. The cost of allocating control is subtracted from that sum. The figure was adapted from Shenhav et al., 2013 with permission from the authors.

Without a doubt, estimates of controllability, value, and effort of outcomes, are formed through prior learning. Estimates formed in one situation are generalized and used in novel similar environments. For example, imagine a person who developed low estimates of controllability in the work environment. Over the years that person may have learned that no matter what they do in the work environment, it is unlikely that desired outcomes occur. That person will tend to allocate less cognitive control in that environment and will develop beliefs about the inability to do the job well. This will lead to a reduction in motivated behavior and will further strengthen existing low controllability estimates and beliefs. Alternatively, another individual might estimate high controllability in the same work environment. However, that individual can anticipate low amounts of rewards associated with achieving good results in work, or high amounts of effort needed to do so. That individual does not anticipate pleasure in work anymore and will also tend to allocate less cognitive control during work hours.

Within our framework, depression-related deficits in cognitive control processes such

as inhibition, task switching, or working memory updating (Joormann & Tanovic, 2015), can be conceptualized as products of changes in the expected value of control. In this way our framework integrates cognitive research on depression with impairments in other domains. Our framework outlines the links between cognitive control deficits and the study of motivation in depression. The framework can account for both the existing research findings and further develop the field by connecting these findings to previously unrelated literatures. Within our framework, cognitive control deficits in depression (Gotlib and Joormann, 2010) can be studied in relation with reward processing impairments in depression (Admon and Pizzagalli, 2015) and changes in perceived controllability of an environment (Moscarello and Hartley, 2017).

This framework has several key implications. First, cognitive control deficits do not have a single cause and can be caused by impairments in different processes. Deficits can thus be present in depressed individuals with different clusters of symptoms (e.g., mainly depressive mood or mainly anhedonia). Second, cognitive control deficits are causally related to reward processing impairments, effort costs, and estimates of controllability of an environment. These impairments influence allocation of cognitive control through learned estimates of reward probability, action-outcome contingencies, and needed effort. Third, cognitive control deficits are not a product of the simple reduced ability to exert control. They are a result of lowered expectations about the value of exerting cognitive control.

The proposed framework goes beyond the current understanding of cognitive control deficits in depression. It does so by providing a more mechanistic understanding of such deficits. First, it identifies the three crucial components that give rise to these deficits. Second, it proposes that the three components play a crucial role in the lowered expected value of control which, in turn, leads to deficits in cognitive control. Further, we propose below how this mechanistic understanding can be used to derive model-based behavioral predictions, as well how it can be related to the neurobiology of depression.

Simulation-based behavioral predictions

The proposed framework outlines the three key motivational components which determine how cognitive control is allocated. By relying on a formalized model of cognitive control, the framework is able to generate precise predictions about the influence of each of these components on the behavioral performance in tasks which require cognitive control. In this section we describe how the computational implementation of the EVC theory can be used to make such predictions. This allows our framework to go beyond the currently existing data

and make testable predictions that can be falsified or corroborated in future studies.

After the original formulation of the EVC theory (Shenhav et al., 2013), the more recent work has developed a computational implementation of the theory (Lieder et al., 2018; Musslick et al., 2015). In the computational implementation, performance of each task (e.g., responding to the color or to the word of a Stroop stimulus) is implemented as a process of evidence accumulation toward a decision boundary. Allocation of control can modify the parameters of this decision process (e.g. the rate of evidence accumulation) to improve performance, depending on the current goal and environment. This implementation relies on the drift-diffusion model of decision making, which has been widely used to describe decision-making (Forstmann et al., 2016; Ratcliff et al., 2016), value-based choice (Krajbich and Rangel, 2011; Tajima et al., 2016), as well as cognitive control processes (Bogacz et al., 2006; Cohen et al., 1990; Dunovan et al., 2015; Dutilh and Vandekerckhove, 2013; Schmitz and Voss, 2012; Ueltzhöffer et al., 2015).

The computational implementation of the EVC theory (Lieder et al., 2018; Musslick et al., 2015) has been successfully used to simulate a wide range of the existing empirical results in the domain of cognitive control (e.g. congruency effects, congruency sequence effects, task-switching costs) and the influence of motivation on cognitive control (incentive-based enhancements in the processes such as inhibition, task-switching, and conflict adaptation). Here we use this model in order to provide behavioral predictions about cognitive control in depression. Specifically, we explore how the depression-related changes in the efficacy, cost, and value components should influence behavior on tasks that require cognitive control. These simulations, based on a model that predicts well the existing data in healthy individuals, provides clear predictions related to cognitive control allocation in depression.

To demonstrate how control-demanding behavior is affected by changes in the decision-making process about control allocation, we simulated the behavior of an agent across two paradigms while varying parameters of that agent's EVC-driven control valuation (Musslick et al., 2015; for a detailed description of these simulations, see Supplementary Materials). We first simulated performance in a Stroop task in which the simulated agent had to categorize a the ink color (e.g. red or green) of a color word (e.g. "RED" or "GREEN") on each trial (Stroop, 1935). We assessed the overall control allocation and the resulting performance cost associated with responding to a color that is response-incongruent with the word (e.g. "RED" displayed in green) compared to responding to a color that is response-congruent with a word (e.g. "RED" displayed in red). We also simulated behavior in a cognitive effort discounting (COGED) task,

in which the agent must choose between performing a difficult, high-demand task for $2 and an easy, low-demand (baseline) task for a variable amount of monetary reward on each trial. We assessed the subjective value of performing a high-demand task by measuring amount of monetary reward offered for the baseline task for which the EVC agent is indifferent between the two tasks, and by normalizing this amount by the reward offered for the high-demand task ($2). We evaluated how these complementary measures – of task performance vs. task preference – were influenced by changes in the agent's (a) cost of cognitive control (b) sensitivity to reward, and (c) expected efficacy of exerting cognitive control (control efficacy).

Consistent with observations in healthy human subjects, our simulated agents demonstrated an incongruency cost, generating more errors when performing incongruent trials compared to congruent trials (Stroop, 1935). Critically, agents chose to exert less control (Fig 1A-C), resulting in higher incongruency costs (Fig 1D-F) when they had (i) higher costs of control, (ii) lower reward sensitivity, and/or (iii) when they perceived their control as being less efficacious. The simulated agent also replicated the behavior of healthy human subjects in studies of cognitive effort selection, assigning lower subjective values to tasks that demanded higher amounts of cognitive control (Westbrook et al., 2013). However, the subjective value for a given task decreased as (i) the cost of cognitive control increased, (ii) reward sensitivity decreased, and/or (iii) the control efficacy decreases (Figure 2G-I). Interestingly, the influence of these changes in parameters are slightly magnified for high task difficulties compared to low

task difficulties.

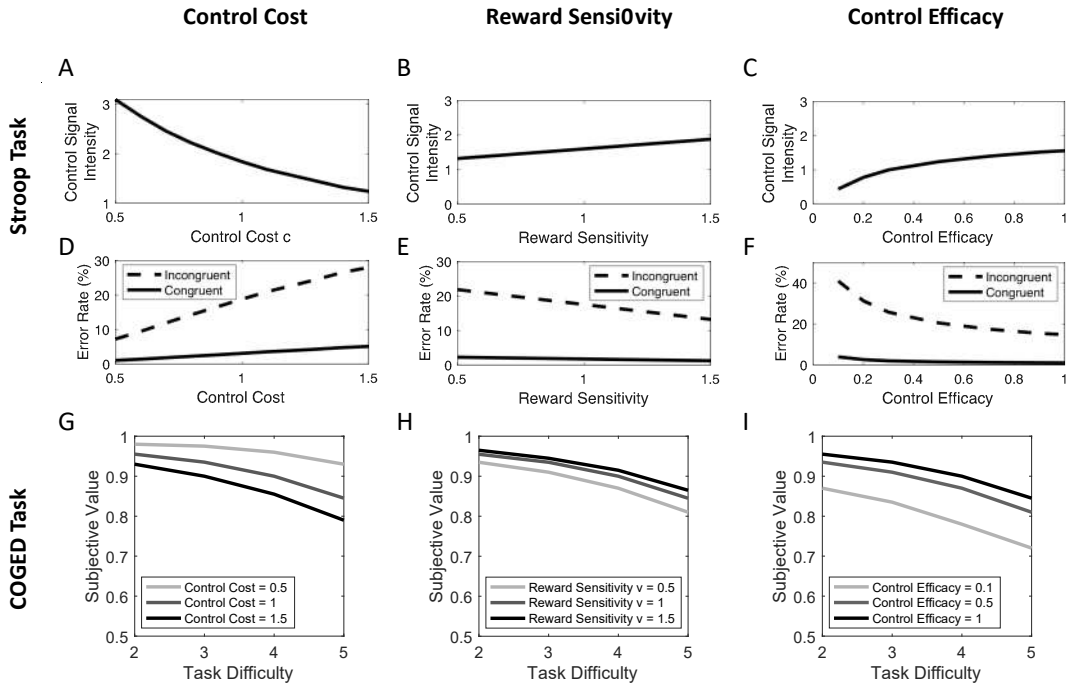


Figure 2. Effects of control cost, reward sensitivity and control efficacy on simulated behavior of the EVC model. Behavior of the model was simulated in the Stroop paradigm and COGED paradigm (see supplementary Materials for details). (A-C) The amount of cognitive control allocated in a Stroop task is shown as a function of control cost, sensitivity to reward and expected control efficacy. (D-F) The error rate on incongruent and congruent trials is shown as a function of the three model parameters. (G-I) The subjective value of a task in the COGED paradigm is plotted as a function of task difficulty for different values of control costs, sensitivity to reward and control efficacy.

The computational implementation of our framework offers a possibility of studying cognitive control deficits in depression within the developing framework of computational psychiatry (Huys et al., 2016; Montague et al., 2012). Several computational frameworks have already started to model behavioral control in depression (Huys et al., 2015; Huys and Dayan, 2009) and our framework offers the possibility of extending this research into the domain of cognitive control deficits in depression. Our framework provides a clear path forward by applying a normative theory of cognitive control to understand cognitive control impairments in depression. Further, it points to the crucial components which interact to produce control

deficits. We have described the advantages of the EVC theory above, but we do recognize the importance of other related computational models of cognitive control. Outcome value, outcome controllability, and effort costs are being recognized as crucial in allocation of cognitive control across different models of control. Because of this, our review of the relevant literature, and the proposed framework, will be useful in guiding future research beyond the boundaries of the EVC theory. The further use of the computational models developed in cognitive neuroscience to understand cognitive control deficits in depression will advance the field in several ways. First, it will allow for the more unified understanding of cognition in both healthy and in individuals suffering from mental illnesses. Second, it will avoid the creation of models that are tailored specifically to understand cognition in depression. Finally, the application of these models to depression, and other mental disorders, will be able to inform and help improve the existing computational models of cognitive control.

Neural Level

Cognitive control deficits in depression are proposed to be related to functional changes in the dlPFC and the ACC (Disner et al., 2011). This view is based on the conceptualization of cognitive control functions as reliant purely on the prefrontal cortex. However, this view has been challenged by the discovery of the role of other regions, such as the basal ganglia and corticostriatal loops in cognitive control and goal-directed behavior (for reviews see: Balleine & O'Doherty, 2010; Chatham & Badre, 2015; Haber, 2016). By building on the EVC theory, our framework offers a more specific view on the role of these two regions in depression. We propose that depression-related changes in the decision-making process about control allocation are related to the activity of the dACC. These changes can further result in lower levels of dlPFC activation which implements control based on the signals from the dACC. In this way, depression is not related to a lowered ability of the dlPFC to implement control, but related to changes in inputs to the dACC used in decision-making about control allocation. In line with the importance of the dACC in our framework, current research has pointed to the crucial role of the ACC in depression (Holroyd and Umemoto, 2016; McTeague et al., 2017), as well as to the specific role of the dACC (Goodkind et al., 2015). Our framework also points to the important role of efficacy and value representations which serve as inputs in the decision-making process. In this way, it can guide further research in connecting the role of the dACC and other regions and networks related to encoding value, reward processing, and efficacy (Haber and Knutson, 2010; Moscarello and Hartley, 2017; Treadway and Zald, 2011).

Relationship with other constructs relevant for depression

Lowered levels of cognitive control have a reciprocal relationship with negative beliefs and attributions. Negative beliefs about the self, the others, and the world represent a crucial cognitive vulnerability factor to depression (Beck, 1972). Attributions that are global, internal, and stable are also an important vulnerability factor (Abramson et al., 1989). Cognitive control deficits can strengthen those beliefs, but are also strengthened by them. Interestingly, there is already some progress on studying the maladaptive beliefs within the framework of computational psychiatry (Moutoussis et al., 2017). Finally, lowered levels of cognitive control are related to maladaptive use of emotion regulation strategies such as rumination (Nolen-Hoeksema et al., 2008) which contribute to the onset and maintenance of depressive symptoms (Joormann & Vanderlind, 2014). Impaired cognitive control also affects other cognitive processes thus producing cognitive biases in attention, interpretation, and memory, which further promote the maladaptive use of emotion regulation strategies (Everaert et al., 2012; Gotlib and Joormann, 2010).

An important research line has focused on the relationship between depressive symptoms and cognitive control over affectively negative material. These studies have demonstrated specific impairments in shifting attention away from negative material, removing negative material from the working memory, and inhibiting negative stimuli (Gotlib and Joormann, 2010; Joormann and Vanderlind, 2014; Koster et al., 2011). In its current form, our framework does not focus on these deficits. However, we hope to integrate this body of work in the framework in future studies. One of the interesting possibilities for such integration is the influence of negative material on efficacy estimates through prior learning.

Future directions and open questions

The proposed framework opens novel avenues for research, namely the links between cognitive control in depression and alterations in components of motivation. Future studies in this domain can more directly test some of the implications of our framework. To date there are not many studies that have investigated the links between components of motivation and cognitive control deficits in depression. We hope that our framework will inspire more studies in this direction. Here we outline some of the existing evidence and propose the paradigms that could be used in future research.

Several studies have already demonstrated the importance of motivation for cognitive functioning in depression (Moritz et al., 2017; Scheurich et al., 2008). Recently, studies have

focused on more specific components of motivation. For example, several studies have demonstrated that reward-based improvements in cognitive control and attention are related to depressive symptoms (Anderson et al., 2014; Jazbec et al., 2005; Ravizza and Delgado, 2014). Future studies should focus on precise distinctions between different types of impairments related to reward processing. In this context, we believe that reward anticipation and effort costs are the most interesting processes to investigate in relation to cognitive processes in depression. Depression research can rely on paradigms developed in cognitive science to study motivation and cognitive control (Botvinick and Braver, 2015). The study of effort expenditure is also gaining a great deal of attention and paradigms to study physical effort already exist (Treadway et al., 2012). For depression research, novel insights on cognitive effort (Shenhav et al., 2017) can be of particular relevance. Also, the relationship between anhedonic symptoms and cognitive control deficits in depression will be a crucial next step in better understanding cognition in depression. In this domain research should be guided by recent developments in measuring anhedonia through behavioral tasks and questionnaires (for a review on measures of anhedonia see: Rizvi et al., 2016).

In the domain of efficacy, there are no readily available paradigms that could directly inform us about its relationship with cognitive control. However, developments in the field of learned helplessness suggest that the interactions between stress and controllability are crucial for stress regulation (Maier and Seligman, 2016). Recent research is starting to provide paradigms to study stress controllability in humans (Bhanji et al., 2016; Hartley et al., 2014). The critical next step is to develop paradigms that would allow the investigation of cognitive control processes in relation to controllability. Recently, paradigms aimed to investigate different components of motivated action and learning have been developed in schizophrenia research. Using a novel paradigm, Morris and associates have demonstrated impairments in action-outcome learning in schizophrenia (Morris, Cyrzon, Green, Le Pelley, & Balleine, 2018; Morris, Quail, Griffiths, Green, & Balleine, 2015; see also: Liljeholm et al., 2011). The use of similar paradigms would allow for further work on computational models of motivation and cognition in depression. Recent computational work has demonstrated that generalization of learned action-outcome contingencies can account for a wide range of behavioral features associated with learned helplessness in animals and humans (Huys and Dayan, 2009; Lieder et al., 2013). However, there is a strong need for more empirical work which could precisely measure the importance of outcome controllability in depression.

Cognitive control deficits (as well as anhedonia) are transdiagnostic in nature and co-

occur in many disorders (Goschke, 2014; Whitton et al., 2015). Although our model is focused on depression, the links between cognitive control and reward processing are relevant for other psychiatric disorders as well. Future studies in this domain hold the promise of not only detecting transdiagnostic constructs such as the positive valence and cognitive systems (Cuthbert and Insel, 2013; Insel et al., 2010), but starting to investigate the relations between these systems. In this context, the relationship between reward processing (positive valence system) and cognitive control (cognitive system) across different clinical populations can be of great interest.

In current research, depression is often regarded and measured as a single, homogeneous disorder and we have treated it as such in this paper. Importantly, diagnosis of depression requires only presence of one out of the two core symptoms: prolonged negative affect and the loss of interest in previously pleasurable activities (DSM-5; American Psychiatric Association, 2013). A growing amount of evidence is pointing to the heterogeneous nature of depression and the need to analyze specific symptoms of depression (Fried and Nesse, 2015). Our framework emphasizes the possibility that cognitive control deficits in depression emerge from different causes. This allows for a study of how potential different causes of cognitive control deficits in depression can be related to different symptom clusters. For example, reward processing impairments, resulting in lowered levels of cognitive control allocation, can be related to anhedonic symptoms. In the same way, changes in efficacy estimates can produce control deficits, but they could be more related to the presence of negative affect.

The study of mechanisms through which cognitive control deficits in depression emerge is of great relevance for developing future cognitive treatments for depression. While there is no evidence for the effectiveness of cognitive training in healthy individuals (Simons et al., 2016), cognitive trainings in depressed individuals are showing considerable promise (Koster et al., 2017; Motter et al., 2016; Siegle et al., 2007). However, the mechanisms through which cognitive trainings improve depressive symptoms remain unknown. Moreover, studies on cognitive training in depression have not focused on motivation yet. Understanding the components of motivation that are affected by these trainings, as well as including incentives in the trainings, will allow for more precise and individualized treatments.

Conclusions

Cognitive control deficits represent an important vulnerability factor for depression and play a central role in cognitive impairments in depression. We propose a framework of

cognitive control in depression in which these deficits emerge as a consequence of changes in the decision-making process about control allocation. We argue that alterations in core components of motivated behavior – reward anticipation, effort costs, and estimates of environment controllability – are crucial mechanisms contributing to dysfunctional cognitive control in depression. This view provides a more mechanistic understanding of cognitive control in depression, offers a better computational and neural understanding of the deficits, and connects cognitive research on depression with other fields of study such as motivation and agency. We believe that our framework can guide future research on mechanisms underlying depressive cognition which can result in improving treatments for depression.

Acknowledgements

This work was supported by the Special Research Fund (BOF) of Ghent University [grant number 01D02415 awarded to IG], by a Center of Biomedical Research Excellence grant from the National Institute of General Medical Sciences [grant number P20GM103645 awarded to AS], by the starting grant of the European Research Council (ERC) under the Horizon 2020 framework [grant number 636116 awarded to RK], and by the Concerted Research Action Grant of Ghent University [grant number BOF16/GOA/017 awarded to EHWK], and a grant from the John Templeton Foundation [awarded to SM; The opinions expressed in this publication are those of the authors and do not necessarily reflect the views of the John Templeton Foundation].

Supplementary Materials

Simulations of control allocation

In order to simulate the effects of motivational variables on control allocation we generated behavior from an agent that performs cognitive control tasks[3] (Musslick et al., 2015). In these simulations, each trial is described as an interaction between the control system and the task environment. The agent specifies the optimal control signal based on an internal representation that is an estimate of the next trial (inferred state $\hat{S}$). The agent implements the specified control signal while performing the corresponding task in the actual task environment (actual state S). After each trial, the agent updates the inferred state for the next trial based on

[3] The source code for all simulations is available at https://github.com/musslick/EVCDepression

its observation of the current trial.

To generate response probabilities and reaction times we used a drift diffusion model (DDM, Bogacz et al., 2006; Ratcliff, 1978) that accumulates evidence towards one of two responses in a Stroop task (e.g. one response indicating the color green and the other response indicating the color red). The DDM simulates performance on a task as the accumulation of evidence about the stimulus towards one of two responses. The drift rate determines whether accumulated evidence leads to either the lower or the upper threshold (e.g. leading to either the left or right response button) and its magnitude determines how fast evidence is accumulated. The threshold determines the amount of evidence required to indicate a response, regulating the speed-accuracy tradeoff: higher thresholds lead to higher likelihood of reaching the correct response at the expense of slower reaction times (Ratcliff, 1985, 1978). We assume that the rate of accumulation toward one of the two response boundaries is governed by an automatic component and a controlled component

$$d = d_{\text{control}} + d_{\text{automatic}}$$

where the automatic component reflects automatic processing of each of the color and word of the stimulus that is unaffected by control:

$$d_{\text{automatic}} = a_{\text{color}} + a_{\text{word}}$$

The absolute magnitude of the color-response association a_{color}, as well as the magnitude of the word-response association a_{word} depends on the strength of the association of each stimulus feature with a given response and its sign depends on the response (e.g. $a_{\text{color}} < 0$ if the response is associated with the left button, $a_{\text{color}} > 0$ if response is associated with the right button). Thus, for congruent trials a_{color}, and a_{word} have the same sign, whereas the opposite sign for incongruent (conflict) trials. The controlled component of the drift rate is the sum of the two stimulus values, each weighted by the intensity of the corresponding control signal, one for processing the color dimension of the stimulus u_{color} and one for processing the word dimension of the stimulus u_{word}:

$$d_{\text{control}} = a_{\text{control}} \cdot u_{\text{control}} + a_{\text{word}} \cdot u_{\text{word}}$$

Thus, each control signal biases processing towards one of the stimulus dimensions. As a result, higher control signal intensity for processing the color dimension of the stimulus improves performance – speeds responses and lowers error rates – for the Stroop task. Mean reaction times (RTs) and response probabilities for a given parameterization of drift rate on trial *t* are

derived from an analytical solution to the DDM (Navarro and Fuss, 2009).

In order to specify the optimal set of control signals $U = \{u_{\text{color}}, u_{\text{word}}\}$ on a given trial, the model estimates the expected value for each configuration of control signal intensities based on its internal model of the next trial $\hat{S} = \{\hat{a}_{\text{color}}, \hat{a}_{\text{word}}\}$. This is done by weighting the expected reward for an outcome against the cost associated with the chosen control signal configuration:

$$EVC(U, \hat{S}) = \varepsilon \cdot P\left(\text{correct} \middle| U, \hat{S}\right) \cdot V(R) - Cost(U)$$

Where ε corresponds to the expected efficacy of exerting cognitive control, $P\left(\text{correct} \middle| U, \hat{S}\right)$ corresponds to the probability of reaching the decision threshold for the correct response and $V(R)$ corresponds to the subjective value of responding correctly. Here, the subjective value $V(R) = v \cdot R$ corresponds to the amount of reward offered for a correct response R weighted by the model's sensitivity to the reward v. The total cost of cognitive control $Cost(U)$ is computed as the sum of the costs for each control signal,

$$Cost(U) = e^{c \cdot u_{\text{color}}} + e^{c \cdot u_{\text{word}}}$$

where the cost for each control signal is an exponential function of the intensity of the control signal, scaled by the cost parameter c. The model selects the control signal configuration with the maximum EVC within the inferred next trial $\hat{S}$, out of all the configurations under consideration:

$$U^* \leftarrow argmax_i\left[EVC(U, \hat{S})\right]$$

Performance in the actual state S is determined by the influence of the chosen control signals on the true parameters of the DDM (e.g., by adjusting the drift rate).

Task environments

Stroop Paradigm. To illustrate effects of distractor interference in the EVC model, we simulated behavior on a Stroop task. In this task, the model is presented with a two-dimensional stimulus, one dimension representing an ink color and another dimension representing a color word. On each trial, the EVC model is required to indicate the response associated with the ink color. The trial sequence encompassed 200 trials, half of which were response congruent and half of which were response incongruent. To simulate congruent trials, we set $a_{\text{color}} = 0.1, a_{\text{word}} = 0.33$ such that both stimuli dimension promote the same response. On incongruent trials, we set $a_{\text{color}} = 0.1, a_{\text{word}} = -0.33$ such that the word dimension is associated with a different response than the color dimension. Note that the word-response

association has a higher absolute magnitude than the color-response association, reflecting the assumption that word reading is a more automatic process than color naming. To simulate an expected mixture of congruent and incongruent trials, we parameterized the expected state as the average of the two trial conditions, $\hat{S} = \{\hat{a}_{\text{color}} = 0.1, a_{\text{word}} = 0\}$. Control was implemented in the form of two control signals, one for processing the color dimension and one for processing the word dimension. The range of control signal intensities as varied from 0 to 4 in steps of 0.02 and the reward received for a correct response was set to $R = 40$. DDM parameters were set as follows: starting point = 0.0, noise coefficient = 1.0, non-decision time = 0.25s and threshold = 1.5. To demonstrate how control-demanding behavior is affected by changes in the decision-making process about control allocation we systematically varied the control cost parameter c from 0.5 to 1.5 in steps of 0.1, the reward sensitivity v from 0.5 to 1.5 in steps of 0.1 and the expected control efficacy ε from 0.1 to 1.0 in steps of 0.1 across simulations. Note that we varied only one parameter at a time while holding the other parameters constant at $c = 1, v = 1, \varepsilon = 1$. For each parameter setting, we assessed the mean amount of control that the model allocated for each trial type, as well as the mean error rate.

COGED Paradigm. To demonstrate the effects of changing model parameters on demand-avoidance, we simulated behavior in the cognitive effort discounting (COGED) experiment described by Westbrook & Braver (2015). In this paradigm, subjects can choose on each trial whether they want to perform a baseline low-demand task for a low reward or a higher-demand alternative task for a higher reward. The amount of reward offered for the baseline task is adjusted to identify the point of indifference, that is, the reward at which subjects are indifferent between performing the low-demand baseline task and performing the high-demand task. To simulate this paradigm, we modeled both tasks as different types of trials that the model can choose between. Each trial encompassed a stimulus with a color dimension that mapped to one of two responses with $a_{\text{color}} > 0$. However, unlike in the Stroop task described above there was no word dimension, $a_{\text{word}} = 0$. The difficulty of the high-demand task was manipulated across experiment blocks by varying the color-response association a_{color} from 0.2 to 0.4 in steps of 0.0667 and the difficulty of the baseline task was fixed to $a_{\text{color}} = 1$ (higher color-response associations may reflect higher saturation values for a color patch). For each set of simulations, we fixed the reward for the high-demand task to ($R = 200$) while steadily increasing the amount of reward offered for the low-demand task in steps of 1, beginning from an initial reward value of $R = 1$. On each trial, the EVC agent determined the

highest EVC separately for each task and chose the task with the highest predicted EVC. We then assessed the amount of reward offered for the low-demand task for which the model would be indifferent between performing the low-demand task and the (more rewarding) high-demand task and normalized this value by the amount of reward offered for the high-demand task. Following the notation by Westbrook & Braver (2015), we refer to this reward value as the subjective value of completing the high-demand task. For instance, if the model would switch to performing the low-demand task at an offered reward of 120 then the (discounted) subjective value of the high-demand task would be 120. The range of control signal intensities was varied from 0 to 6 in steps of 0.05 and DDM parameters were set as follows: starting point = 0.0, noise coefficient = 1, non-decision time = 0.25s and threshold = 1.5. To demonstrate how control-demanding behavior is affected by changes in the decision-making process about control allocation, we systematically varied the control cost parameter c from 0.5 to 1 to 1.5, the reward sensitivity v from 0.5 to 1 to 1.5 and the expected control efficacy ε from 0.1 to 0.5 to 1.0 across simulations. Note that we varied only one parameter at a time while holding the other parameters constant at $c = 1, v = 1, \varepsilon = 1$. For each parameter setting, we assessed the subjective value as a function of the task difficulty of the high-demand task.

References

Aarts, E., Roelofs, A., Franke, B., Rijpkema, M., Fernández, G., Helmich, R.C., Cools, R., 2010. Striatal dopamine mediates the interface between motivational and cognitive control in humans: Evidence from genetic imaging. Neuropsychopharmacology 35, 1943–1951. https://doi.org/10.1038/npp.2010.68

Abrahamse, E., Braem, S., Notebaert, W., Verguts, T., 2016. Grounding cognitive control in associative learning. Psychol. Bull. 142. https://doi.org/10.1037/bul0000047

Abramson, L.Y., Metalsky, G.I., Alloy, L.B., 1989. Hopelessness depression: A theory-based subtype of depression. Psychol. Rev. 96, 358–372. https://doi.org/10.1037/0033-295X.96.2.358

Abramson, L.Y., Seligman, M.E.P., Teasdale, J.D., 1978. Learned helplessness in humans and reformulation. J. Abnorm. Psychol. 87, 49–74. https://doi.org/10.1037/0021-843X.87.1.49

Admon, R., Pizzagalli, D.A., 2015. Dysfunctional Reward Processing in Depression. Curr. Opin. Psychol. 4, 114–118. https://doi.org/10.1016/j.copsyc.2014.12.011

Alexander, W.H., Brown, J.W., 2011. Medial Prefrontal Cortex as an action-outcome predictor. Nat. Neurosci. 14, 1338–1344. https://doi.org/10.1038/nn.2921.Medial

American Psychiatric Association, 2013. Diagnostic and Statistical Manual of Mental Disorders, 5th ed. American Psychiatric Publishing, Washington DC.

Anderson, B.A., Leal, S.L., Hall, M.G., Yassa, M.A., Yantis, S., 2014. The attribution of value-based

attentional priority in individuals with depressive symptoms. Cogn. Affect. Behav. Neurosci. 14, 1221–7. https://doi.org/10.3758/s13415-014-0301-z

Atkinson, J.W., 1957. Motivational determinants of risk taking behavior. Psychol. Rev. 64, 359–372. https://doi.org/10.1037/h0043445

Balleine, B.W., O'Doherty, J.P., 2010. Human and rodent homologies in action control: Corticostriatal determinants of goal-directed and habitual action. Neuropsychopharmacology 35, 48–69. https://doi.org/10.1038/npp.2009.131

Bandura, A., 2001. Social Cogntive Theory: An Agentic Perspective. Annu. Rev. Psychol. 52, 1–26. https://doi.org/10.1146/annurev.psych.52.1.1

Bandura, A., 1977. Self-efficacy: Toward a unifying theory of behavioral change. Psychol. Rev. 84, 191–215. https://doi.org/10.1037/0033-295x.84.2.191

Barch, D.M., Dowd, E.C., 2010. Goal representations and motivational drive in schizophrenia: The role of prefrontal-striatal interactions. Schizophr. Bull. 36, 919–934. https://doi.org/10.1093/schbul/sbq068

Barch, D.M., Pagliaccio, D., 2017. Consistency, Replication, and Meta-analyses of Altered Brain Activity in Unipolar Depression. JAMA Psychiatry 74, 56. https://doi.org/10.1001/jamapsychiatry.2016.2844

Barch, D.M., Pagliaccio, D., Luking, K.R., 2015. Mechanisms underlying motivational deficits in psychopathology: similarities and differences in depression and schizophrenia, in: Behavioral Neuroscience of Motivation. Springer International Publishing, pp. 411–449.

Bechtel, W., Abrahamsen, A., 2005. Explanation : a mechanist alternative. Stud. Hist. Philos. Biol. Biomed. Sci. 36, 421–441. https://doi.org/10.1016/j.shpsc.2005.03.010

Beck, A.T., 1972. Depression: Causes and treatment. University of Pennsylvania Press.

Bhandari, A., Badre, D., Frank, M.J., 2017. Learning Cognitive Control, in: Egner, T. (Ed.), Wiley Handbook of Cognitive Control. Wiley-Blackwell, pp. 376–391.

Bhanji, J.P., Kim, E.S., Delgado, M.R., 2016. Perceived Control Alters the Effect of Acute Stress on Persistence 145, 356–365.

Bogacz, R., Brown, E., Moehlis, J., Holmes, P., Cohen, J.D., 2006. The physics of optimal decision making: A formal analysis of models of performance in two-alternative forced-choice tasks. Psychol. Rev. 113, 700–765. https://doi.org/10.1037/0033-295X.113.4.700

Bonner, S.E., Sprinkle, G.B., 2002. The effects of monetary incentives on effort and task performance: theories, evidence, and a framework for research. Accountine, Organ. Soc. 27, 303–345. https://doi.org/10.1016/S0361-3682(01)00052-6

Botvinick, M.M., Braver, T., 2015. Motivation and Cognitive Control: From Behavior to Neural Mechanism. Annu. Rev. Psychol. 66, 83–113. https://doi.org/10.1146/annurev-psych-010814-015044

Botvinick, M.M., Braver, T.S., 2015. Motivation and Cognitive Control : From Behavior to Neural Mechanism. Annu. Rev. Psychol. 1–31. https://doi.org/10.1146/annurev-psych-010814-015044

Botvinick, M.M., Cohen, J.D., 2014. The Computational and Neural Basis of Cognitive Control :

Charted Territory and New Frontiers 38, 1249–1285. https://doi.org/10.1111/cogs.12126

Braem, S., Verguts, T., Roggeman, C., Notebaert, W., 2012. Reward modulates adaptations to conflict. Cognition 125, 324–332. https://doi.org/10.1016/j.cognition.2012.07.015

Braver, T.S., Krug, M.K., Chiew, K.S., Kool, W., Westbrook, J.A., Clement, N.J., Adcock, R.A., Barch, D.M., Botvinick, M.M., Carver, C.S., Cools, R., Custers, R., Dickinson, A., Dweck, C.S., Fishbach, A., Gollwitzer, P.M., Hess, T.M., Isaacowitz, D.M., Mather, M., Murayama, K., Pessoa, L., Samanez-Larkin, G.R., Somerville, L.H., 2014. Mechanisms of motivation-cognition interaction: challenges and opportunities. Cogn. Affect. Behav. Neurosci. 14, 443–72. https://doi.org/10.3758/s13415-014-0300-0

Brown, J.W., Alexander, W.H., 2017. Foraging Value, Risk Avoidance, and Multiple Control Signals: How the Anterior Cingulate Cortex Controls Value-based Decision-making. J. Cogn. Neurosci. 29, 1656–1673. https://doi.org/10.1162/jocn_a_01140

Chatham, C.H., Badre, D., 2015. Multiple gates on working memory. Curr. Opin. Behav. Sci. 1, 23–31. https://doi.org/10.1016/j.cobeha.2014.08.001

Clark, D.A., Beck, A.T., 2010. Cognitive theory and therapy of anxiety and depression: Convergence with neurobiological findings. Trends Cogn. Sci. 14, 418–424. https://doi.org/10.1016/j.tics.2010.06.007

Cléry-Melin, M.L., Schmidt, L., Lafargue, G., Baup, N., Fossati, P., Pessiglione, M., 2011. Why don't you try harder? an investigation of effort production in major depression. PLoS One 6, 1–8. https://doi.org/10.1371/journal.pone.0023178

Cohen, J.D., Dunbar, K., Mcclelland, J.L., 1990. On the Control of Automatic Processes : A Parallel Distributed Processing Account of the Stroop Effect. Psychol. Rev. 97, 332–361.

Crocker, L.D., Heller, W., Warren, S.L., O'Hare, A.J., Infantolino, Z.P., Miller, G. a, 2013. Relationships among cognition, emotion, and motivation: implications for intervention and neuroplasticity in psychopathology. Front. Hum. Neurosci. 7, 261. https://doi.org/10.3389/fnhum.2013.00261

Culbreth, A.J., Moran, E.K., Barch, D.M., 2018. Effort-cost decision-making in psychosis and depression: could a similar behavioral deficit arise from disparate psychological and neural mechanisms? Psychol. Med. 1–16. https://doi.org/10.1017/S0033291717002525

Cuthbert, B.N., Insel, T.R., 2013. Toward the future of psychiatric diagnosis: the seven pillars of RDoC. BMC Med. 11, 126. https://doi.org/10.1186/1741-7015-11-126

Davidson, R.J., Pizzagalli, D., Nitschke, J.B., Putnam, K., 2002. Depression: Perspectives from Affective Neuroscience. Annu. Rev. Psychol. 53, 545–574. https://doi.org/10.1146/annurev.psych.53.100901.135148

Daw, N.D., O'Doherty, J.P., 2014. Multiple Systems for Value Learning, in: Glimcher, P.W., Fehr, E. (Eds.), Neuroeconomics: Decision Making and the Brain: Second Edition. Academic Press, pp. 393–410. https://doi.org/10.1016/B978-0-12-416008-8.00021-8

De Raedt, R., Hooley, J.M., 2016. The role of expectancy and proactive control in stress regulation : A neurocognitive framework for regulation expectation. Clin. Psychol. Rev. 45, 45–55. https://doi.org/10.1016/j.cpr.2016.03.005

Demeyer, I., De Lissnyder, E., Koster, E.H.W., De Raedt, R., 2012. Rumination mediates the relationship between impaired cognitive control for emotional information and depressive symptoms: A prospective study in remitted depressed adults. Behav. Res. Ther. 50, 292–7. https://doi.org/10.1016/j.brat.2012.02.012

Der-Avakian, A., Markou, A., 2012. The neurobiology of anhedonia and other reward-related deficits. Trends Neurosci. 35, 68–77. https://doi.org/10.1016/j.tins.2011.11.005

Derakshan, N., Salt, M., Koster, E.H.W., 2009. Attentional control in dysphoria: An investigation using the antisaccade task. Biol. Psychol. 80, 251–255. https://doi.org/10.1016/j.biopsycho.2008.09.005

Dickinson, A., 1985. Actions and habits: the development of behavioural autonomy. Philos. Trans. R. Soc. Lond. B. Biol. Sci. 308, 67–78.

Disner, S.G.S., Beevers, C.G.C., Haigh, E.A.P., Beck, A.T., 2011. Neural mechanisms of the cognitive model of depression. Nat. Rev. Neurosci. 12, 467–77. https://doi.org/10.1038/nrn3027

Dunovan, K., Lynch, B., Molesworth, T., Verstynen, T., 2015. Competing basal ganglia pathways determine the difference between stopping and deciding not to go 1–24. https://doi.org/10.7554/eLife.08723

Dutilh, G., Vandekerckhove, J., 2013. A Diffusion Model Account of Age Differences in Posterror Slowing 28, 64–76. https://doi.org/10.1037/a0029875

Everaert, J., Grahek, I., Koster, E.H.W., 2017. Individual differences in cognitive control over emotional material modulate cognitive biases linked to depressive symptoms. Cogn. Emot. 31. https://doi.org/10.1080/02699931.2016.1144562

Everaert, J., Grahek, I., Koster, E.H.W., 2017. Individual differences in cognitive control modulate cognitive biases linked to depressive. Cogn. Emot. 31, 736–746. https://doi.org/10.1080/02699931.2016.1144562

Everaert, J., Koster, E.H.W., Derakshan, N., 2012. The combined cognitive bias hypothesis in depression. Clin. Psychol. Rev. 32, 413–24. https://doi.org/10.1016/j.cpr.2012.04.003

Forstmann, B.U., Ratcliff, R., Wagenmakers, E., 2016. Sequential Sampling Models in Cognitive Neuroscience : Advantages , Applications , and Extensions. https://doi.org/https://doi.org/10.1146/annurev-psych-122414-033645

Fried, E.I., Nesse, R.M., 2015. Depression sum-scores don't add up: why analyzing specific depression symptoms is essential. BMC Med. 13, 72. https://doi.org/10.1186/s12916-015-0325-4

Friedman, N.P., Miyake, A., 2017. Unity and diversity of executive functions: Individual differences as a window on cognitive structure. Cortex 86, 186–204. https://doi.org/10.1016/j.cortex.2016.04.023

Goodkind, M., Eickhoff, S.B., Oathes, D.J., Jiang, Y., Chang, A., Jones-Hagata, L.B., Ortega, B.N., Zaiko, Y. V., Roach, E.L., Korgaonkar, M.S., Grieve, S.M., Galatzer-Levy, I., Fox, P.T., Etkin, A., 2015. Identification of a common neurobiological substrate for mental Illness. JAMA Psychiatry 72, 305–315. https://doi.org/10.1001/jamapsychiatry.2014.2206

Goschke, T., 2014. Dysfunctions of decision-making and cognitive control as transdiagnostic mechanisms of mental disorders: advances, gaps, and needs in current research. Int. J. Methods

Psychiatr. Res. 23. https://doi.org/10.1002/mpr.1410

Gotlib, I.H., Hamilton, J.P., 2008. Neuroimaging and depression: Current status and unresolved issues. Curr. Dir. Psychol. Sci. 17, 159–163. https://doi.org/10.1111/j.1467-8721.2008.00567.x

Gotlib, I.H., Joormann, J., 2010. Cognition and depression: current status and future directions. Annu. Rev. Clin. Psychol. 6, 285–312. https://doi.org/10.1146/annurev.clinpsy.121208.131305

Grahek, I., Everaert, J., Krebs, R.M., Koster, E.H.W., 2018. Cognitive control in depression: Towards clinical models informed by cognitive neuroscience. Clin. Psychol. Sci. https://doi.org/10.1177/2167702618758969

Griffiths, K.R., Morris, R.W., Balleine, B.W., 2014. Translational studies of goal-directed action as a framework for classifying deficits across psychiatric disorders. Front. Syst. Neurosci. 8, 1–16. https://doi.org/10.3389/fnsys.2014.00101

Haber, S.N., 2016. Corticostriatal circuitry. Dialogues Clin. Neurosci. 18, 7–21. https://doi.org/10.1002/bit.260190307

Haber, S.N., Knutson, B., 2010. The reward circuit: Linking primate anatomy and human imaging. Neuropsychopharmacology 35, 4–26. https://doi.org/10.1038/npp.2009.129

Hartley, C.A., Gorun, A., Reddan, M.C., Ramirez, F., Phelps, E.A., 2014. Neurobiology of Learning and Memory Stressor controllability modulates fear extinction in humans. Neurobiol. Learn. Mem. 113, 149–156. https://doi.org/10.1016/j.nlm.2013.12.003

Hershenberg, R., Satterthwaite, T.D., Daldal, A., Katchmar, N., Moore, T.M., Kable, J.W., Wolf, D.H., 2016. Diminished effort on a progressive ratio task in both unipolar and bipolar depression. J. Affect. Disord. 196, 97–100. https://doi.org/10.1016/j.jad.2016.02.003

Hertel, P., 2000. The cognitive-initiative account of depression-related impairments in memory. Psychol. Learn. Motiv. 39, 47–71. https://doi.org/10.1016/S0079-7421(00)80030-4

Hertel, P.T., 1994. Depression and Memory: Are Impairments Remediable Through Attentional Control? Curr. Dir. Psychol. Sci. 3, 190–193. https://doi.org/10.1111/1467-8721.ep10770707

Holroyd, C.B., McClure, S.M., 2015. Hierarchical control over effortful behavior by rodent medial frontal cortex: A computational model. Psychol. Rev. 122, 54–83. https://doi.org/10.1037/a0038339

Holroyd, C.B., Umemoto, A., 2016. The research domain criteria framework: The case for anterior cingulate cortex. Neurosci. Biobehav. Rev. 71, 418–443. https://doi.org/10.1016/j.neubiorev.2016.09.021

Huys, Q.J.M., Daw, N.D., Dayan, P., 2015. Depression: A Decision-Theoretic Analysis. Annu. Rev. Neurosci. 38, 1–23. https://doi.org/10.1146/annurev-neuro-071714-033928

Huys, Q.J.M., Dayan, P., 2009. A Bayesian formulation of behavioral control. Cognition 113, 314–328. https://doi.org/10.1016/j.cognition.2009.01.008

Huys, Q.J.M., Maia, T. V, Frank, M.J., 2016. Computational psychiatry as a bridge from neuroscience to clinical applications. Nat. Neurosci. 19, 404–13. https://doi.org/10.1038/nn.4238

Insel, T., Cuthbert, B., Garvie, M., Heinssen, R., Pine, D.S., Quinn, K., Sanislow, C., Wang, P., 2010. Research Domain Criteria (RDoC): Toward a New Classification Framework for Research on

Mental Disorders. Am. J. Psychiatry 167, 748–751. https://doi.org/10.1176/appi.ajp.2010.09091379

Inzlicht, M., Shenhav, A., Olivola, C.Y., 2018. The Effort Paradox: Effort Is Both Costly and Valued. Trends Cogn. Sci. 22, 337–349. https://doi.org/10.1016/j.tics.2018.01.007

Jazbec, S., McClure, E., Hardin, M., Pine, D.S., Ernst, M., 2005. Cognitive control under contingencies in anxious and depressed adolescents: an antisaccade task. Biol. Psychiatry 58, 632–639. https://doi.org/10.1016/j.biopsych.2005.04.010

Joormann, J., 2010. Cognitive Inhibition and Emotion Regulation in Depression. Curr. Dir. Psychol. Sci. 19, 161–166. https://doi.org/10.1177/0963721410370293

Joormann, J., Tanovic, E., 2015. Cognitive vulnerability to depression: examining cognitive control and emotion regulation. Curr. Opin. Psychol. 4, 86–92. https://doi.org/10.1016/j.copsyc.2014.12.006

Joormann, J., Vanderlind, W.M., 2014. Emotion Regulation in Depression: The Role of Biased Cognition and Reduced Cognitive Control. Clin. Psychol. Sci. 2, 402–421. https://doi.org/10.1177/2167702614536163

Joormann, J., Yoon, K.L., Zetsche, U., 2007. Cognitive inhibition in depression. Appl. Prev. Psychol. 12, 128–139. https://doi.org/10.1016/j.appsy.2007.09.002

Kahneman, D., 1973. Attention and Effort. Prentice-Hall, Englewood Cliffs, NJ.

Kaser, M., Zaman, R., Sahakian, B.J., 2017. Cognition as a treatment target in depression. Psychol. Med. 47, 987–989. https://doi.org/10.1017/S0033291716003123

Knutson, B., Westdorp, A., Kaiser, E., Hommer, D., 2000. FMRI visualization of brain activity during a monetary incentive delay task. Neuroimage 12, 20–27. https://doi.org/10.1006/nimg.2000.0593

Kool, W., Botvinick, M., 2018. Mental labour. Nat. Hum. Behav. https://doi.org/10.1038/s41562-018-0401-9

Kool, W., McGuire, J.T., Rosen, Z.B., Botvinick, M.M., 2010. Decision making and the avoidance of cognitive demand. J. Exp. Psychol. Gen. 139, 665–682. https://doi.org/10.1037/a0020198

Koster, E.H.W., De Lissnyder, E., Derakshan, N., De Raedt, R., 2011. Understanding depressive rumination from a cognitive science perspective: the impaired disengagement hypothesis. Clin. Psychol. Rev. 31, 138–45. https://doi.org/10.1016/j.cpr.2010.08.005

Koster, E.H.W., Hoorelbeke, K., Onraedt, T., Owens, M., Derakshan, N., 2017. Cognitive control interventions for depression: A systematic review of findings from training studies. Clin. Psychol. Rev. 53, 79–92. https://doi.org/10.1016/j.cpr.2017.02.002

Krajbich, I., Rangel, A., 2011. Multialternative drift-diffusion model predicts the relationship between visual fi xations and choice in value-based decisions. Proc. Natl. Acad. Sci. U. S. A. 108. https://doi.org/10.1073/pnas.1101328108

Krebs, R.M., Boehler, C.N., Woldorff, M.G., 2010. The influence of reward associations on conflict processing in the Stroop task. Cognition 117, 341–347. https://doi.org/10.1016/j.cognition.2010.08.018

Krebs, R.M., Woldorff, M.G., 2017. Cognitive control and reward., in: Egner, T. (Ed.), Wiley

Handbook of Cognitive Control. Wiley-Blackwell, pp. 422–439.

Kurzban, R., Duckworth, A., Kable, J.W., Myers, J., 2013. An opportunity cost model of subjective effort and task performance. Behav. Brain Sci. 36, 661–679. https://doi.org/10.1017/S0140525X12003196

Leotti, L. a, Wager, T.D., 2010. Motivational influences on response inhibition measures. J. Exp. Psychol. Hum. Percept. Perform. 36, 430–47. https://doi.org/10.1037/a0016802

Leotti, L.A., Iyengar, S.S., Ochsner, K.N., 2010. Born to choose: The origins and value of the need for control. Trends Cogn. Sci. 14, 457–463. https://doi.org/10.1016/j.tics.2010.08.001

Levens, S.M., Gotlib, I.H., 2015. Updating emotional content in recovered depressed individuals: Evaluating deficits in emotion processing following a depressive episode. J. Behav. Ther. Exp. Psychiatry 48, 156–163. https://doi.org/10.1016/j.jbtep.2015.03.009

Lieder, F., Goodman, N.D., Huys, Q.J., 2013. Learned helplessness and generalization. Proc. Annu. Meet. Cogn. Sci. Soc. 35.

Lieder, F., Shenhav, A., Musslick, S., Griffiths, T.L., 2018. Rational metareasoning and the plasticity of cognitive control. Plos Comput. Biol. 14. https://doi.org/10.13140/RG.2.2.24500.14721

Liljeholm, M., Tricomi, E., O'Doherty, J.P., Balleine, B.W., 2011. Neural Correlates of Instrumental Contingency Learning: Differential Effects of Action-Reward Conjunction and Disjunction. J. Neurosci. 31, 2474–2480. https://doi.org/10.1523/JNEUROSCI.3354-10.2011

Machamer, P., Darden, L., Craver, C.F., 2000. Thinking About Mechanisms. Philos. Sci. 67, 1–25.

Maier, S.F., 1984. Learned helplessness and animal models of depression. Prog. Neuro-Psychopharmacology Biol. Psychiatry 8, 435–446. https://doi.org/10.1016/S0278-5846(84)80032-9

Maier, S.F., Amat, J., Baratta, M. V., Paul, E., Watkins, L.R., 2006. Behavioral control, the medial prefrontal cortex, and resilience. Dialogues Clin. Neurosci. 8, 397–406. https://doi.org/10.1016/j.ynstr.2014.09.003

Maier, S.F., Seligman, M.E.P., 2016. Learned helplessness at fifty: Insights from neuroscience. Psychol. Rev. 123, 1–19. https://doi.org/10.1037/rev0000033

Maier, S.F., Watkins, L.R., 2005. Stressor controllability and learned helplessness: The roles of the dorsal raphe nucleus, serotonin, and corticotropin-releasing factor. Neurosci. Biobehav. Rev. 29, 829–841. https://doi.org/10.1016/j.neubiorev.2005.03.021

Marchetti, I., Shumake, J., Grahek, I., Koster, E.H.W., 2018. Temperamental factors in remitted depression: The role of effortful control and attentional mechanisms. J. Affect. Disord. 235. https://doi.org/10.1016/j.jad.2018.04.064

McMakin, D.L., Olino, T.M., Porta, G., Dietz, L.J., Emslie, G., Clarke, G., Wagner, K.D., Asarnow, J.R., Ryan, N.D., Birmaher, B., Shamseddeen, W., Mayes, T., Kennard, B., Spirito, A., Keller, M., Lynch, F.L., Dickerson, J.F., Brent, D.A., 2012. Anhedonia predicts poorer recovery among youth with selective serotonin reuptake inhibitor treatmentresistant depression. J. Am. Acad. Child Adolesc. Psychiatry 51, 404–411. https://doi.org/10.1016/j.jaac.2012.01.011

McTeague, L.M., Huemer, J., Carreon, D.M., Jiang, Y., Eickhoff, S.B., Etkin, A., 2017. Identification of common neural circuit disruptions in cognitive control across psychiatric disorders. Am. J.

Psychiatry 174, 676–685. https://doi.org/10.1176/appi.ajp.2017.16040400

Millan, M.J., Agid, Y., Brüne, M., Bullmore, E.T., Carter, C.S., Clayton, N.S., Connor, R., Davis, S., Deakin, B., DeRubeis, R.J., Dubois, B., Geyer, M.A., Goodwin, G.M., Gorwood, P., Jay, T.M., Joëls, M., Mansuy, I.M., Meyer-Lindenberg, A., Murphy, D., Rolls, E., Saletu, B., Spedding, M., Sweeney, J., Whittington, M., Young, L.J., 2012. Cognitive dysfunction in psychiatric disorders: characteristics, causes and the quest for improved therapy. Nat. Rev. Drug Discov. 11, 141–168. https://doi.org/10.1038/nrd3628

Miller, K.J., Shenhav, A., Ludvig, E.A., n.d. Habits without values. Psychol. Rev.

Montague, P.R., Dolan, R.J., Friston, K.J., Dayan, P., 2012. Computational psychiatry. Trends Cogn. Sci. 16. https://doi.org/10.1016/j.tics.2011.11.018

Moritz, S., Stöckert, K., Hauschildt, M., Lill, H., Jelinek, L., Beblo, T., Diedrich, S., Arlt, S., 2017. Are we exaggerating neuropsychological impairment in depression? Reopening a closed chapter. Expert Rev. Neurother. 17, 839–846. https://doi.org/10.1080/14737175.2017.1347040

Morris, R.W., Cyrzon, C., Green, M.J., Le Pelley, M.E., Balleine, B.W., 2018. Impairments in action–outcome learning in schizophrenia. Transl. Psychiatry 8, 54. https://doi.org/10.1038/s41398-018-0103-0

Morris, R.W., Quail, S., Griffiths, K.R., Green, M.J., Balleine, B.W., 2015. Corticostriatal control of goal-directed action is impaired in schizophrenia. Biol. Psychiatry 77, 187–195. https://doi.org/10.1016/j.biopsych.2014.06.005

Moscarello, J.M., Hartley, C.A., 2017. Agency and the Calibration of Motivated Behavior. Trends Cogn. Sci. 21, 725–735. https://doi.org/10.1016/j.tics.2017.06.008

Motter, J.N., Pimontel, M.A., Rindskopf, D., Devanand, D.P., Doraiswamy, P.M., Sneed, J.R., 2016. Computerized cognitive training and functional recovery in major depressive disorder : A meta-analysis. J. Affect. Disord. 189, 184–191. https://doi.org/10.1016/j.jad.2015.09.022

Moutoussis, M., Shahar, N., Hauser, T.U., Dolan, R.J., 2017. Computation in Psychotherapy , or How Computational Psychiatry Can Aid Learning-Based Psychological Therapies. Comput. Psychiatry 2. https://doi.org/10.1162/CPSY_a_00014

Müller, V.I., Cieslik, E.C., Serbanescu, I., Laird, A.R., Fox, P.T., Eickhoff, S.B., 2016. Altered Brain Activity in Unipolar Depression Revisited. JAMA Psychiatry 1–9. https://doi.org/10.1001/jamapsychiatry.2016.2783

Musslick, S., Shenhav, A., Botvinick, M.M., Cohen, J.D., 2015. A computational model of control allocation based on the Expected Value of Control. Reinf. Learn. Decis. Mak. Conf. 59, 2014.

Navarro, D.J., Fuss, I.G., 2009. Fast and accurate calculations for first-passage times in Wiener diffusion models. J. Math. Psychol. 53, 222–230. https://doi.org/10.1016/j.jmp.2009.02.003

Nolen-Hoeksema, S., Wisco, B.E., Lyubomirsky, S., 2008. Rethinking Rumination. Perspect. Psychol. Sci. 3, 400–424. https://doi.org/10.1111/j.1745-6924.2008.00088.x

Oettingen, G., Gollwitzer, P.M., 2001. Goal setting and goal striving, in: Tesser, A., Schwarz, N. (Eds.), Blackwell Handbook of Social Psychology: Intraindividual Processes. Blackwell Publishers, pp. 329–348.

Owens, M., Koster, E.H.W., Derakshan, N., 2012. Impaired filtering of irrelevant information in

dysphoria: An ERP study. Soc. Cogn. Affect. Neurosci. 7, 752–763. https://doi.org/10.1093/scan/nsr050

Padmala, S., Pessoa, L., 2011. Reward reduces conflict by enhancing attentional control and biasing visual cortical processing. J. Cogn. Neurosci. 23, 3419–3432. https://doi.org/10.1162/jocn_a_00011

Pizzagalli, D.A., 2014. Depression, Stress, and Anhedonia: Toward a Synthesis and Integrated Model. Annu Rev Clin Psychol 393–423. https://doi.org/10.1146/annurev-clinpsy-050212-185606.Depression

Pizzagalli, D.A., 2011. Frontocingulate dysfunction in depression: Toward biomarkers of treatment response. Neuropsychopharmacology 36, 183–206. https://doi.org/10.1038/npp.2010.166

Ratcliff, R., 1985. Theoretical Interpretations of the Speed and Accuracy of Positive and Negative Responses. Psychol. Rev. 92, 212–225. https://doi.org/10.1037/0033-295X.92.2.212

Ratcliff, R., 1978. A theory of memory retrieval. Psychol. Rev. 85, 59–108. https://doi.org/10.1037/0033-295X.85.2.59

Ratcliff, R., Smith, P.L., Brown, S.D., Mckoon, G., 2016. Diffusion Decision Model : Current Issues and History. Trends Cogn. Sci. xx, 1–22. https://doi.org/10.1016/j.tics.2016.01.007

Ravizza, S.M., Delgado, M.R., 2014. Motivational enhancement of cognitive control depends on depressive symptoms. Emotion 14, 646–50. https://doi.org/10.1037/a0036754

Rizvi, S.J., Pizzagalli, D.A., Sproule, B.A., Kennedy, S.H., 2016. Assessing anhedonia in depression: Potentials and pitfalls. Neurosci. Biobehav. Rev. 65, 21–35. https://doi.org/10.1016/j.neubiorev.2016.03.004

Rock, P.L., Roiser, J.P., Riedel, W.J., Blackwell, A.D., 2014. Cognitive impairment in depression: a systematic review and meta-analysis. Psychol. Med. 44, 2029–40. https://doi.org/10.1017/S0033291713002535

Romer Thomsen, K., Whybrow, P.C., Kringelbach, M.L., 2015. Reconceptualizing anhedonia: novel perspectives on balancing the pleasure networks in the human brain. Front. Behav. Neurosci. 9, 1–23. https://doi.org/10.3389/fnbeh.2015.00049

Rotter, J.B., 1966. Generalized expectancies for internal versus external control of reinforcement. Psychol. Monogr. 80, 1–28. https://doi.org/10.1037/h0092976

Salamone, J.D., Correa, M., 2012. The Mysterious Motivational Functions of Mesolimbic Dopamine. Neuron 76, 470–485. https://doi.org/10.1016/j.neuron.2012.10.021

Salamone, J.D., Yohn, S.E., Lopez-Cruz, L., San Miguel, N., Correa, M., 2016. Activational and effort-related aspects of motivation: Neural mechanisms and implications for psychopathology. Brain 139, 1325–1347. https://doi.org/10.1093/brain/aww050

Scheurich, A., Fellgiebel, A., Schermuly, I., Bauer, S., Wölfges, R., Müller, M.J., 2008. Experimental evidence for a motivational origin of cognitive impairment in major depression. Psychol. Med. 38, 237–246. https://doi.org/10.1017/S0033291707002206

Schmitz, F., Voss, A., 2012. Decomposing Task-Switching Costs With the Diffusion Model. J. Exp. Psychol. Hum. Percept. Perform. 38, 222–250. https://doi.org/10.1037/a0026003

Seligman, M.E., 1972. Learned helplessness. Annu. Rev. Med. 23, 407–412. https://doi.org/10.1146/annurev.me.23.020172.002203

Seligman, M.E.P., Maier, S.F., 1967. Failure To Escape Traumatic Shock. J. Exp. Psychol. 74, 1–9. https://doi.org/10.1037/h0024514

Shenhav, A., Botvinick, M., Cohen, J., 2013. The expected value of control: An integrative theory of anterior cingulate cortex function. Neuron 79, 217–240. https://doi.org/10.1016/j.neuron.2013.07.007

Shenhav, A., Cohen, J.D., Botvinick, M.M., 2016. Dorsal anterior cingulate cortex and the value of control. Nat. Neurosci. 19, 1286–1291. https://doi.org/1038/nn.4382

Shenhav, A., Musslick, S., Lieder, F., Kool, W., Griffiths, T.L., Cohen, J.D., Botvinick, M.M., 2017. Toward a Rational and Mechanistic Account of Mental Effort. Annu. Rev. Neurosci. 40, 99–124. https://doi.org/10.1146/annurev-neuro-072116-031526

Siegle, G.J., Ghinassi, F., Thase, M.E., 2007. Neurobehavioral therapies in the 21st century: Summary of an emerging field and an extended example of cognitive control training for depression. Cognit. Ther. Res. 31, 235–262. https://doi.org/10.1007/s10608-006-9118-6

Simons, D.J., Boot, W.R., Charness, N., Gathercole, S.E., Chabris, C.F., Hambrick, D.Z., Stine-Morrow, E.A.L., 2016. Do " Brain-Training " Programs Work ? Psychol. Sci. Public Interes. 17. https://doi.org/10.1177/1529100616661983

Snyder, H.R., 2013. Major depressive disorder is associated with broad impairments on neuropsychological measures of executive function: a meta-analysis and review. Psychol. Bull. 139, 81–132. https://doi.org/10.1037/a0028727

Spijker, J., Bijl, R. V, de Graaf, R., Nolen, W. a, 2001. Determinants of poor 1-year outcome of DSM-III-R major depression in the general population: results of the Netherlands Mental Health Survey and Incidence Study (NEMESIS). Acta Psychiatr. Scand. 103, 122–30. https://doi.org/acp016 [pii]

Strauss, G.P., Gold, J.M., 2012. A new perspective on anhedonia in schizophrenia. Am. J. Psychiatry 169, 364–373. https://doi.org/10.1176/appi.ajp.2011.11030447

Stroop, J.R., 1935. Studies of interference in serial verbal reactions. J. Exp. Psychol. 18, 643–662. https://doi.org/10.1037/h0054651

Tajima, S., Drugowitsch, J., Pouget, A., 2016. Optimal policy for value-based decision-making. Nat. Commun. 7, 1–12. https://doi.org/10.1038/ncomms12400

Treadway, M.T., Bossaller, N.A., Shelton, R.C., Zald, D.H., 2012. Effort-based decision-making in major depressive disorder: A translational model of motivational anhedonia. J. Abnorm. Psychol. 121, 553–558. https://doi.org/10.1037/a0028813

Treadway, M.T., Zald, D.H., 2013. Parsing Anhedonia: Translational Models of Reward-Processing Deficits in Psychopathology. Curr. Dir. Psychol. Sci. 22, 244–249. https://doi.org/10.1177/0963721412474460

Treadway, M.T., Zald, D.H., 2011. Reconsidering anhedonia in depression: Lessons from translational neuroscience. Neurosci. Biobehav. Rev. 35, 537–555. https://doi.org/10.1016/j.neubiorev.2010.06.006

Ueltzhöffer, K., Armbruster-Genç, D.J.N., Fiebach, C.J., 2015. Stochastic Dynamics Underlying Cognitive Stability and Flexibility. Plos Comput. Biol. 11, 1–46. https://doi.org/10.1371/journal.pcbi.1004331

Uher, R., Perlis, R.H., Henigsberg, N., Zobel, A., Rietschel, M., Mors, O., Hauser, J., Dernovsek, M.Z., Souery, D., Bajs, M., Maier, W., Aitchison, K.J., Farmer, A., McGuffin, P., 2012. Depression symptom dimensions as predictors of antidepressant treatment outcome: Replicable evidence for interest-activity symptoms. Psychol. Med. 42, 967–980. https://doi.org/10.1017/S0033291711001905

Verguts, T., 2017. Computational models of cognitive control, in: Egner, T. (Ed.), The Wiley Handbook of Cognitive Control. Wiley-Blackwell, Hoboken, NJ, pp. 376–391.

Verguts, T., Vassena, E., Silvetti, M., 2015. Adaptive effort investment in cognitive and physical tasks: a neurocomputational model. Front. Behav. Neurosci. 9. https://doi.org/10.3389/fnbeh.2015.00057

Wallis, J.D., Rushworth, M.F.S., 2014. Integrating benefits and costs in decision making, in: Neuroeconomics (Second Edition). pp. 411–433.

Wardenaar, K.J., Giltay, E.J., Van Veen, T., Zitman, F.G., Penninx, B.W.J.H., 2012. Symptom dimensions as predictors of the two-year course of depressive and anxiety disorders. J. Affect. Disord. 136, 1198–1203. https://doi.org/10.1016/j.jad.2011.11.037

Westbrook, A., Braver, T.S., 2015. Cognitive effort: A neuroeconomic approach. Cogn. Affect. Behav. Neurosci. 15, 395–415. https://doi.org/10.3758/s13415-015-0334-y

Westbrook, A., Kester, D., Braver, T.S., 2013. What Is the Subjective Cost of Cognitive Effort? Load, Trait, and Aging Effects Revealed by Economic Preference. PLoS One 8, 1–8. https://doi.org/10.1371/journal.pone.0068210

Whitmer, A.J., Gotlib, I.H., 2013. An attentional scope model of rumination. Psychol. Bull. 139, 1036–61. https://doi.org/10.1037/a0030923

Whitton, A.E., Treadway, M.T., Pizzagalli, D.A., 2015. Reward processing dysfunction in major depression, bipolar disorder and schizophrenia. Curr. Opin. Psychiatry 28, 7–12.

Wigfield, A., Eccles, J.S., 2000. Expectancy–Value Theory of Achievement Motivation. Contemp. Educ. Psychol. 25, 68–81. https://doi.org/10.1006/ceps.1999.1015

Wood, W., Rünger, D., 2016. Psychology of Habit. Annu. Rev. Psychol. 67, 289–314. https://doi.org/10.1146/annurev-psych-122414-033417

Yang, X. hua, Huang, J., Zhu, C. ying, Wang, Y. fei, Cheung, E.F.C., Chan, R.C.K., Xie, G. rong, 2014. Motivational deficits in effort-based decision making in individuals with subsyndromal depression, first-episode and remitted depression patients. Psychiatry Res. 220, 874–882. https://doi.org/10.1016/j.psychres.2014.08.056

Zald, D.H., Treadway, M.T., 2017. Reward Processing, Neuroeconomics, and Psychopathology. Annu. Rev. Clin. Psychol 133, 1–325. https://doi.org/10.1146/annurev-clinpsy-032816-044957

Chapter 4

Reward enhances stimulus processing in the visual cortex independently of voluntary attention: Insights from Steady-State Visual Evoked Potentials[1]

Abstract

Rewards influence the processing of stimuli in the visual cortex, but the mechanisms through which this effect occurs remain unclear. One proposed mechanism is that the presence of potential rewards increases the deployment of voluntary attention. Another possibility is that rewards increase the saliency of previously neutral stimuli independently of voluntary attention. We contrasted these two theoretical proposals by having participants perform a global motion detection task in which they were instructed on each trial to pay attention to one of the two features which were linked to different reward schedules. We recorded steady-state visual evoked potentials (SSVEPs) in 43 human subjects of both genders to simultaneously measure processing of both features in visual cortex. This design allowed us to dissociate the effect of voluntary attention from the effect of reward. The introduction of rewards increased the processing of the feature linked to high rewards independently of voluntary attention. However, this enhancement disappeared when participants were aware that they could not earn any more rewards. This shows that rewards can directly influence sensory processing in the visual cortex, independently of voluntary attention. However, processing in visual cortex was readjusted following the change in the availability of rewards in the environment. Neither the voluntary attention, nor the saliency account can fully explain these results. We suggest that they need to be integrated to allow for the flexible balance between reward-driven increase in saliency and voluntary deployment of attention based on reward availability.

[1] Based on: Grahek, I., Schettino, A., Koster, E. H. W., & Andersen, S. K. Reward enhances stimulus processing in the visual cortex independently of voluntary attention: Insights from Steady-State Visual Evoked Potentials. (in preparation)

Introduction

Maximizing rewards and avoiding punishments is one of the main determinants of human behavior. In order to increase the probability of obtaining rewards, it is necessary to prioritize the information needed to earn rewards. Selective attention is crucial for adaptive behavior as it facilitates processing of relevant and suppression of irrelevant information in our environment (Chun, Golomb, & Turk-Browne, 2011; Desimone & Duncan, 1995). This process depends on our current goals (e.g., searching for our keys) and the physical salience of stimuli (e.g., a loud noise; Corbetta & Shulman, 2002; Posner, 1980; Theeuwes, 2010). Recent research has indicated that motivation can influence selective attention by impacting both of these factors. Reward expectation can enhance voluntary selective attention, and reward associations can increase the saliency of previously neutral stimuli. In most situations attention is guided by the combination of both voluntary allocation of attention and reward history of stimuli (Awh, Belopolsky, & Theeuwes, 2012). For example, while we are searching for keys (goal-relevant target) our attention can be captured by a cake (goal-irrelevant distractor). These two ways in which rewards influence selective attention have been commonly studied in isolation and the neural mechanisms through which they jointly guide attention remain unclear. Specifically, it remains unclear how voluntary selective attention and reward history interact to determine the processing of goal-relevant and irrelevant stimuli in the visual cortex.

Voluntary selective attention (as well as other cognitive control processes) is proposed to be enhanced when individuals anticipate that they can earn rewards for good task-performance (Botvinick & Braver, 2015; Krebs & Woldorff, 2017; Pessoa, 2015). A number of fMRI and EEG studies found reward-based enhancements in preparation for upcoming target stimuli driven by enhanced activity in frontoparietal regions involved in attentional control (Krebs, Boehler, Roberts, Song, & Woldorff, 2012; Pessoa & Engelmann, 2010; Schevernels, Krebs, Santens, Woldorff, & Boehler, 2014) and by enhanced task-set representations in these regions (Etzel, Cole, Zacks, Kay, & Braver, 2016; Wisniewski, Reverberi, Momennejad, Kahnt, & Haynes, 2015). While these studies suggest that reward influences attentional control via modulations in the frontoparietal network, it remains unclear how such modulations translate to affect the processing of attended and unattended stimuli in visual cortex.

Another set of studies has focused on the processing of stimuli associated with earning rewards. These studies have demonstrated that stimuli currently or previously associated with rewards capture attention in an automatic fashion, and even in cases when this conflicts with

the current goals (Anderson, 2016; Awh, Belopolsky, & Theeuwes, 2012; Chelazzi, Perlato, Santandrea, & Della Libera, 2013; Failing & Theeuwes, 2017). Behavioral studies have demonstrated that stimuli predictive of rewards capture attention, and that they can do so in subsequent trials when rewards are no longer present (Anderson, Laurent, & Yantis, 2011; Della Libera & Chelazzi, 2009; Failing & Theeuwes, 2014). At the neural level it has been shown that stimuli related to rewards receive increased sensory processing (Serences, 2008). Studies focusing specifically on attention have shown that the attentional capture by rewarding stimuli can be related to changes in the early processing of such stimuli in the visual cortex (i.e., increase in the P1 component; Donohue et al., 2016; Hickey, Chelazzi, & Theeuwes, 2010; Luque et al., 2017; MacLean & Giesbrecht, 2015). However, other studies have failed to find evidence for such early modulations in the visual cortex, and found changes at later stages of stimulus processing (increased N2pc component and improved decoding in later processing stages; Qi, Zeng, Ding, & Li, 2013; Tankelevitch, Spaak, Rushworth, & Stokes, 2019). One fMRI study (Hickey & Peelen, 2015) has provided evidence for the simultaneous enhancement in representation of reward-related stimuli and suppression of stimuli devoid of a specific motivational value. More specifically, using a multivoxel pattern analysis and decoding technique, these authors found a gain increase in object-selective visual cortex for stimuli paired with rewards, while those not associated with this incentive were suppressed.

Importantly, the effects of reward history and voluntary attention are often difficult to dissociate, and they are commonly confounded in cognitive tasks (Maunsell, 2004). Thus the open question is whether the reward-related changes in stimulus processing in visual cortex occur as a consequence of voluntary selective attention, changes in stimulus saliency, or a combination of both. One option (the *voluntary selective attention view*) is that rewards influence the processing of stimuli by increasing the amount of voluntary attention deployed toward these stimuli. This hypothesis comes from the models which are focused on the role of motivation in the allocation of attention and cognitive control (Brown & Alexander, 2017; Holroyd & McClure, 2015; Shenhav, Botvinick, & Cohen, 2013; Verguts, Vassena, & Silvetti, 2015). These models propose that the amount of attention allocated toward stimuli is dependent on the amount of rewards which can be earned by doing so. Another possibility (*the saliency view*) is that rewards increase the saliency of stimuli, and thus capture attention automatically, independently from voluntary attention. This view comes from theoretical models focused on the role of reward history in guiding selective attention (Anderson, 2016; Awh et al., 2012; Chelazzi et al., 2013; Failing & Theeuwes, 2017). These models propose that the processing of

stimuli related to high rewards is facilitated, while the processing of other stimuli is suppressed, and that this effect is long lasting, even when rewards are no longer available.

In this study, we orthogonally manipulated voluntary attention and reward probability in order to test the predictions of these models. To this end, participants performed a continuous global motion discrimination task (Andersen, Muller, & Hillyard, 2009; Andersen & Müller, 2010). On each trial, two superimposed random dot kinematograms (RDKs) of different color (red and blue), were presented concurrently and participants were cued on a trial by trial basis to attend to one of them. Thus, these two RDKs served as goal-relevant (attended) and goal-irrelevant (unattended) stimuli, respectively. Critically, after a baseline period used as control condition, these two colors were systematically associated with a low or high probability of earning a reward in a training phase. Subsequently we examined the influence of the previous reward history in the test phase in which rewards were no longer available. The two RDKs flickered at different frequencies, thereby driving separate steady-state visual evoked potentials (SSVEPs). SSVEPs offer the unique advantage of simultaneously tracking the processing of multiple stimuli as the specific oscillatory response of each stimulus can be extracted (frequency tagging), and the two resulting signals can be compared to each other (Andersen & Müller, 2010; Kashiwase, Matsumiya, Kuriki, & Shioiri, 2012; Müller, Teder-Sälejärvi, & Hillyard, 1998). Voluntary attention is known to increase SSVEP amplitudes of attended stimuli (Morgan, Hansen, & Hillyard, 1996). Further, SSVEP amplitudes are highly sensitive to changes in the physical salience of stimuli and are increased for more salient stimuli (Andersen, Muller, & Martinovic, 2012). Hence, analyzing SSVEPs in this design provided us with the ability to simultaneously track the visual processing of attended and unattended stimuli related to high or low rewards respectively. This design thus enabled us to dissociate the effects of voluntary attention and reward, which are often difficult to distinguish (Maunsell, 2004).

We tested predictions arising from the theoretical models developed to account for the effects of rewards on cognitive control (Brown & Alexander, 2017; Holroyd & McClure, 2015; Shenhav et al., 2013; Verguts et al., 2015) and the effects of reward history on attention (Anderson, 2016; Awh et al., 2012; Chelazzi et al., 2013; Failing & Theeuwes, 2017) respectively. The first class of models predict that reward influences sensory processing through voluntary attention (Figure 1A), and the second class of models predict that rewards directly influence the saliency of stimuli (Figure 1B). Both groups of models predict behavioral improvements and enhanced processing (higher SSVEP amplitudes) of the stimuli related to high rewards. However, the saliency view predicts that the processing of the high reward stimuli

will be enhanced irrespective of attention, while the voluntary attention view predicts that the processing of the high reward stimuli will be enhanced only when they are attended. The saliency view also predicts the suppressed processing of the low reward stimuli, while the voluntary attention view predicts that the processing of these stimuli will be suppressed only when they are not attended. Finally, the saliency view predicts that these effects will persist when rewards are no longer available (test phase), while the voluntary attention view predicts that the processing of both high and low reward stimuli will retreat to baseline levels.

Methods

Participants

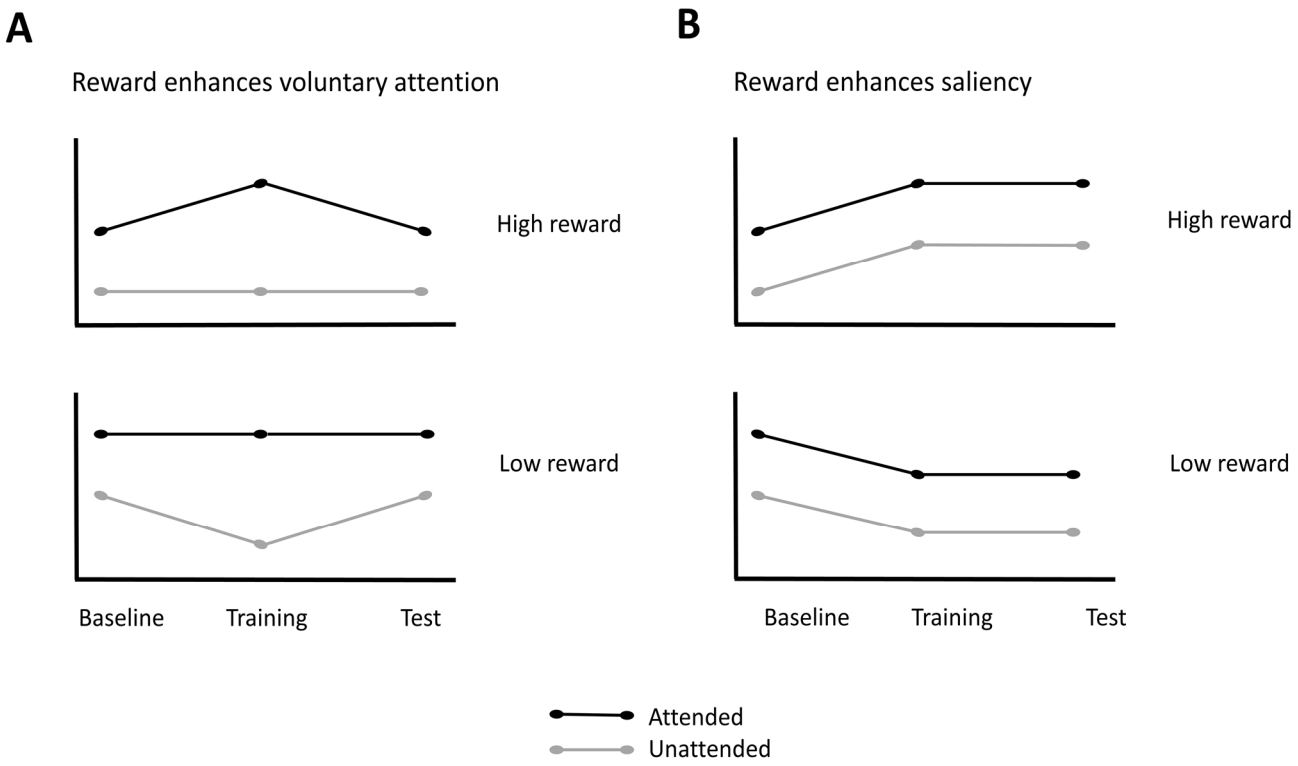


Figure 1. Predictions from the saliency and voluntary selective attention views. A. If rewards influence the processing of stimuli through changes in voluntary attention, the processing of the high reward stimulus will be enhanced when it is attended, while the processing of the low reward stimulus will be suppressed when it is unattended (training phase). The processing of the stimuli related to high and low rewards will retreat back to baseline once the rewards are no longer available (test phase). **B.** If rewards increase saliency of a stimulus independently from voluntary attention, the processing of the high reward stimulus will be enhanced and the processing of the low reward stimulus will be suppressed in both the attended and unattended condition (training phases). Further, this effect will persist even when rewards are no longer available (test phase).

We tested 48 participants with normal or corrected-to-normal vision and no history of psychiatric or neurological disorders. Four participants were excluded due to technical problems during EEG recording and one person was excluded due to noisy EEG data. Thus, the final data set consisted of 43 participants (39 females, 14 males; median age = 22). Participants

received a fixed payoff of 20 €, plus up to 6 € depending on task performance (on average 25.5 €). The study was approved by the ethics committee of Ghent University.

Stimuli and task

We used the coherent motion detection task (Andersen & Müller, 2010; *Figure 1A*), in which participants were presented with two overlapping circular RDKs of isoluminant colors (red and blue) on a grey background. Viewing distance was fixed with a chinrest at 55 cm from the 21-inch CRT screen (resolution of 1024 x 768 pixels, 120 Hz refresh rate). At the beginning of each trial, participants were instructed which of the two RDKs to attend by a verbal audio cue: "red" (241 ms) or "blue" (266 ms). The two RDKs had a diameter corresponding to 20.61 degrees of visual angle and consisted of 125 randomly and independently moving dots each (0.52 degrees of visual angle per dot). The two RDKs flickered at different frequencies (10 or 12 Hz). One-third of trials contained one, two, or three coherent motion intervals, occurring with equal probability in the attended and unattended color RDK. During these intervals, dots in one of the RDKs moved with 75% coherence in one of four cardinal directions (up, down, left, or right) for 300 ms. Participants had to detect the occurrence of coherent motion in the attended RDK as fast as possible by pressing the space key on a standard AZERTY USB keyboard while ignoring such coherent motion in the unattended RDK. Responses occurring between 200 ms and 1500 ms after coherent motion onset of the attended or unattended dots were counted as hits or false alarms, respectively. Correct responses were followed by a tone (200 ms sine wave of either 800 or 1,200 Hz, counterbalanced across participants). Late or incorrect responses were followed by an error sound (200 ms square wave tone of 400 Hz).

The experiment started with 4 practice blocks of 60 trials in each block. After each block, participants received feedback on their performance. After finishing the practice phase, participants completed 12 blocks (each consisting of 50 trials) divided into 3 phases (*baseline*, *training*, and *test*; *Figure 1B*) of 4 blocks each. Each phase contained 100 trials on which participants were instructed to attend to the red color and 100 trials in which they were instructed to attend to the blue color. Out of those 100 trials, 40 trials contained no dot motion, while 60 trials contained one, two, or three dot motions (20 of each). The trials in which participants attended one or the other color, and the trials with different number of motions were intermixed. Participants did the coherent motion detection task, as described above, throughout all three phases (baseline, training, and test). In the training phase, participants could earn additional monetary rewards (up to 6 €) based on their actual performance. They were

instructed that one of the colors would be paired with high probability (80%) and the other color with low probability (20%) of earning 10 extra cents for each correct motion detection. The mapping between color and reward probability was counterbalanced across participants. Receipt of the reward was signaled by a new tone that replaced the usual correct tone. If the correct tone was a sine wave of 800 Hz, the reward tone was a sine wave of 1,200 Hz and vice versa. At the end of each of 4 training blocks, participants received feedback regarding both their performance and the amount of reward earned within the block. The third phase (test) was identical to baseline (i.e., no monetary rewards assigned). The entire task lasted for approximately 50 minutes, including short breaks in between blocks. After finishing the task, participants completed two questionnaires aimed at assessing reward sensitivity (BIS-BAS; Franken et al., 2005) and depression levels (BDI-II; Van der Does, 2002). The collection of the questionnaire data is not reported here as it was collected for exploratory purposes in order to form a larger database of neural and self-report measures of reward processing. This experiment was realized using Cogent 2000 developed by the Cogent 2000 team at the FIL and the ICN and Cogent Graphics developed by John Romaya at the LON at the Wellcome Department of Imaging Neuroscience.

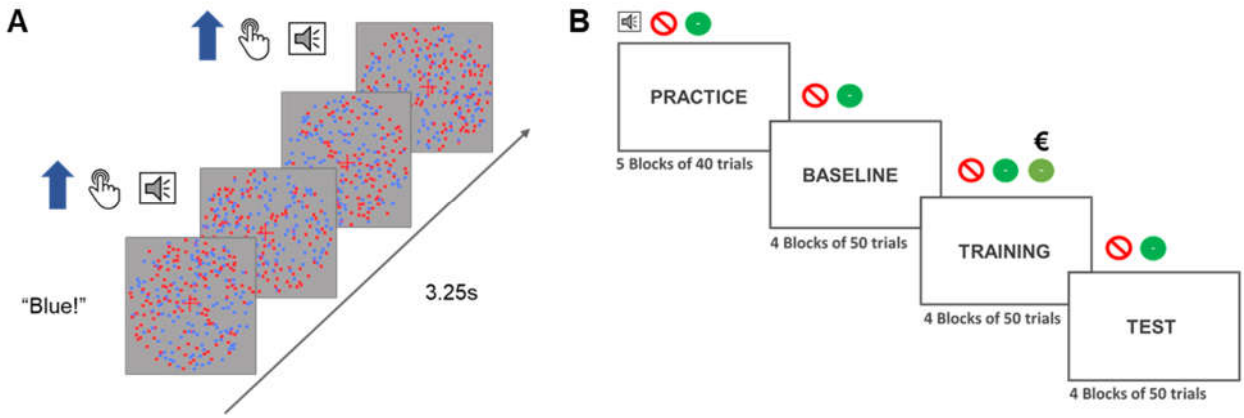


Figure 1. Depiction of a single trial and the phases of the experiment. A. Each trial started with an audio cue (”Blue” or ”Red”) which instructed participants which color to attend to in that trial. The trial lasted for 3.25 seconds during which dots of either of the colors could move from 0 to 3 times in total. If the participants were instructed to attend to the blue dots and the blue dots moved coherently, they had to press the response button. In that case they would hear the auditory feedback signaling the correct detection of the motions. **B.** The experiment started with a practice and a baseline block in which the participants heard an audio cue at the beginning of the trial and two types of feedback sounds (incorrect or correct). In the training block a third sound was introduced to signal that the participants were both correct and received a reward for that response. They would still at times hear the old correct feedback which would signal that they were correct, but not rewarded. The test phase was the same as the baseline phase.

EEG recording and pre-processing

Electroencephalographic activity (EEG) was recorded with an ActiveTwo amplifier (BioSemi, Inc., The Netherlands) at a sampling rate of 512 Hz. Sixty-four Ag/AgCl electrodes were fitted into an elastic cap, following the international 10/10 system (Chatrian, Lettich, & Nelson, 1985). The common mode sense (CMS) active electrode and the driven right leg (DRL) passive electrode were used as reference and ground electrodes, respectively. Additional external electrodes were applied to the left and right mastoids, as well as on the outer canthi of each eye and in the inferior and superior areas of the left orbit (to record horizontal and vertical electrooculogram, EOG).

Data pre-processing was performed offline with custom MATLAB scripts and functions included in EEGLAB v14.1.1b (Delorme & Makeig, 2004). After subtracting the mean value of the signal (DC offset), the continuous EEG data were epoched between 0 and 3,250 ms, corresponding to the beginning and end of the trial, respectively. After referencing to *Cz*, FASTER v1.2.3b (Nolan, Whelan, & Reilly, 2010) was used for artifact identification and rejection using the following settings: (i) over the whole normalized EEG signal, channels with variance, mean correlation, and Hurst exponent exceeding $z = \pm 3$ were interpolated via a spherical spline procedure (Perrin, Pernier, Bertrand, & Echallier, 1989); (ii) the mean across channels was computed for each epoch and, if amplitude range, variance, and channel deviation exceeded $z = \pm 3$, the whole epoch was removed; (iii) within each epoch, channels with variance, median gradient, amplitude range, and channel deviation exceeding $z = \pm 3$ were interpolated; (iv) grand-averages with amplitude range, variance, channel deviation, and maximum EOG value exceeding $z = \pm 3$ were removed; (v) epochs containing more than 12 interpolated channels were discarded. Subsequently, automated routines were used to reject all trials with blinks or horizontal eye-movements exceeding 25 microvolt. For details, see our commented code at https://osf.io/5hryf/. After pre-processing, the average number of interpolated channels was 3.61 (SD = 1.23, range 1 – 6) and the mean percentage of rejected epochs was 8.77% (SD = 6.71, range 0 – 27.78). After re-referencing to averaged mastoids, trials in each condition were averaged separately for each participant, resulting in the following averages: (i) baseline, red attended; (ii) baseline, blue attended; (iii) training, red attended; (iv) training, blue attended; (v) test, red attended; (vi) test, blue attended.

After removing linear trends, SSVEP amplitudes were computed as the absolute of the complex Fourier coefficients of the trial averaged EEG in a time-window from 500 ms (to

exclude the typically strong phasic visual evoked response to picture onset) to 3,250 ms after stimulus onset. Electrodes with maximum SSVEP amplitudes were identified by calculating isocontour voltage maps based on grand-averaged data collapsed across all conditions. Based on this, a cluster consisting of the four electrodes Oz, O2, POz, and Iz was chosen for further analysis. SSVEP amplitudes were normalized (rescaled) for each participant and frequency (10 and 12 Hz) separately by dividing amplitudes by the average amplitude of the two conditions in the baseline.

Statistical analyses

Behavioral and EEG data were analyzed using Bayesian multilevel regressions. We fitted and compared multiple models of varying complexity to predict observer sensitivity (d′, i.e. participants' ability to distinguish between target and distractor coherent motion), reaction times for correct responses, and SSVEP amplitudes. For the behavioral data, mean reaction times of correct detections (hits) and sensitivity (d′) were analyzed. Sensitivity index d′ (Macmillan & Creelman, 2004) was calculated with adjustments for extreme values (Hautus, 1995) using the *psycho* R package (for the method see: Pallier, 2002).

Each of the fitted models included both constant and varying effects (also known as fixed and random). Participant-specific characteristics are known to affect both behavioral performance (e.g., response speed) and EEG signal (e.g., skull thickness, skin conductance, hair); therefore, we decided to model this variability by adding varying intercepts in our models. Additionally, the studied effects (i.e., selective attention and reward sensitivity) are known to vary in magnitude over participants, so we opted for including varying slopes in our models. It should be noted that, because of the simultaneous estimation of group-level and participant-level parameters, multilevel models display a property called *shrinkage*[2].

Models were fitted in R using the *brms* package (Bürkner, 2016) which employs the probabilistic programming language *Stan* (Carpenter et al., 2016) to implement Markov Chain Monte Carlo (MCMC) algorithms in order to estimate posterior distributions of the parameters of interest (details about the fitted models can be found in the data analysis scripts). Each of the

[2] In brief, estimates that strongly deviate from the mean (e.g., a participant performing the task much worse than the average of the total sample) will be pulled toward the group mean (McElreath, 2016). This advantageous property prevents extreme values from having large effects on the results.

models were fitted using weakly informative prior distributions (described below) and Gaussian likelihood. Four MCMC simulations ("chains") with 6,000 iterations (3,000 warmup) and no thinning were run to estimate parameters in each of the fitted models. Further analyses were done following the recommendations for Bayesian multilevel modeling using *brms* (Bürkner, 2016, 2017; Nalborczyk & Bürkner, 2019). We confirmed that all models converged by examining trace plots, autocorrelation, and variance between chains (Gelman-Rubin statistic; Gelman & Rubin, 1992). We compared models based on their fit to the actual data using the Bayesian R^2 (Gelman, Goodrich, Gabry, & Ali, 2017), and their out-of-sample predictive performance using the Widely Applicable Information Criterion (WAIC; Watanabe, 2010). The best model was selected and the posterior distributions of conditions of interest were examined. Differences between conditions were assessed by computing the mean and the 95% highest density interval (HDI) of the difference between posterior distributions of the respective conditions (Kruschke, 2014). Additionally, we calculated the evidence ratios (ERs) for our hypotheses as the ratios between the percentage of posterior samples on each side of the zero of a difference distribution between two conditions. ERs can be interpreted as the probability of a hypothesis (e.g. "Condition A is larger than condition B") against its alternative ("Condition B is larger than condition A").

Behavioral data

We fitted three models to predict sensitivity (d′) and reaction times (in milliseconds) separately (*Figure 2* for the raw data and *Supplementary Table 1* for the descriptive statistics). First, we fitted the *Null model* with a constant and varying intercepts across participants. This model was fitted in order to explore the possibility that the data would be best explained by simple random variation between participants. To investigate the effect of reward phase (baseline, training, test), we fitted the *Reward phase* model which included only reward phase as the constant predictor, as well as varying intercepts and slopes across participants for this effect. To investigate the possible interaction between reward phase and reward, we fitted the *Reward phase* × *Reward Probability* model including the intercepts and slopes of these two effects and their interaction as both constant and varying effects. All models had a Gaussian distribution as the prior for the intercept (for sensitivity: centered at 1.8 with a standard deviation of 1; for reaction times: centered at 500 with a standard deviation of 200). The models with slopes also included a Gaussian distribution as prior for the slopes (for sensitivity: centered at 0 with a standard deviation of 2; for reaction times: centered at 0 with a standard deviation of 200). The means for the priors for the intercepts were selected based on the previous study

with the same task (Andersen & Müller, 2010). The standard deviations of all of the prior distributions were chosen so that the distributions are very wide and thus only weakly informative. Note that there are two additional models that, although possible to fit, are not plausible in the context of our experiment. Specifically, the model including only the effect of reward probability overlooks the fact that this effect would necessarily be most pronounced in the training phase, thus interacting with the effect of reward phase. The same logic applies to the model with additive effects of reward phase and probability (i.e., these effects could not act independently in our experimental design).

SSVEP amplitudes

We fitted seven models to predict the trial-averaged SSVEP amplitudes (in a.u. due to the normalization) across conditions (*Figure 2C, Figure 2D,* and *Supplementary Table 2*). The *Null model* included one constant and multiple varying intercepts across participants. The *Attention model* included the constant effect of attention; the *Reward Phase model* included the constant effect of reward phase; the *Reward Phase + Attention* model included the additive effects of reward phase and attention; and the *Reward Phase × Attention* model also included the interaction between reward phase and attention. The *Reward probability × Reward phase + Attention* model consisted of the constant effects of reward and phase, their interaction, and the independent effect of attention. The last model was the *Reward probability × Reward phase × Attention* model which included all constant effects and their interaction. All models, except for the *Null model*, included varying intercepts and slopes across participants for all of the constant effects. All models included a Gaussian distribution as the prior for the intercept (centered at 1 with a standard deviation of 1). The mean across both attended and unattended conditions is approximately 1 in this paradigm (Andersen & Müller, 2010), while the normalized amplitudes are in the 0-2 range (the normalized amplitude of 2 for the attended stimulus would equal the physical removal of the unattended stimulus) which is why we opted for the standard deviation of 1 for the prior distributions. In addition, the models with slopes included a Gaussian distribution as the prior for the slopes (centered at 0 with a standard deviation of 1). As was the case for the behavioral data, several models were not fitted because they were not plausible in the context of our experiment (e.g., models that include both reward phase and probability, but not their interaction, are implausible because reward probability cannot affect the baseline phase).

Results

Behavioral results

Sensitivity d′

Of all the tested models, the *Reward phase* × *Reward probability* model best predicted sensitivity (*Table 1*). The posterior distributions of the interaction model (*Figure 2A* and *Table 2*) revealed that sensitivity improved in the training phase compared to the baseline for both the low reward (M = 0.14; 95% HDI [0.01, 0.27]; ER = 57.82) and the high reward color (M = 0.04; 95% HDI [-0.08, 0.17]; ER = 3.10). This improvement was far more pronounced for low compared to high reward (M = 0.10; 95% HDI [-0.08, 0.27]; ER = 6.25). Conversely, there was no evidence for a difference between training and test phases in the low (M = 0.00; 95% HDI [-0.13, 0.13]; ER = 1.09), and a small reduction in sensitivity in the high reward condition (M = -0.08; 95% HDI [-0.20, 0.05]; ER = 8.52). These results suggest a higher sensitivity for coherent motion detection in the training phase compared to baseline, that was most pronounced for the low relative to the high reward color. There was also very little evidence of a change in sensitivity from the training to the test phase.

Table 1

Mean and standard errors (in parenthesis) of WAIC and Bayesian R^2 for each model predicting sensitivity and reaction times.

Model	*WAIC (SE)*	*Bayesian R^2 (SE)*
Sensitivity		
Null	533.3 (26.5)	0.27 (0.05)
Reward phase	541.0 (26.5)	0.27 (0.05)
Reward phase × Reward probability	202.7 (19.1)	0.84 (0.01)
Reaction times		
Null	2,500.2 (31.6)	0.50 (0.04)
Reward phase	2,483.0 (35.3)	0.56 (0.04)
Reward phase × Reward probability	2,322.5 (30.0)	0.82 (0.02)

Table 2

Means and 95% HDIs of the posterior distributions of reaction times and sensitivity in each condition.

Reward phase	Reward probability	Sensitivity (d′)	Reaction times (milliseconds)
Baseline	High	1.64 [1.39, 1.87]	546.54 [534.33, 559.30]
Baseline	Low	1.48 [1.25, 1.69]	551.13 [539.34, 563.50]
Training	High	1.69 [1.44, 1.93]	524.91 [512.94, 536.30]
Training	Low	1.62 [1.41, 1.84]	537.99 [526.48, 550.32]
Test	High	1.61 [1.36, 1.84]	528.97 [515.90, 541.99]
Test	Low	1.62 [1.41, 1.84]	539.85 [525.63, 554.34]

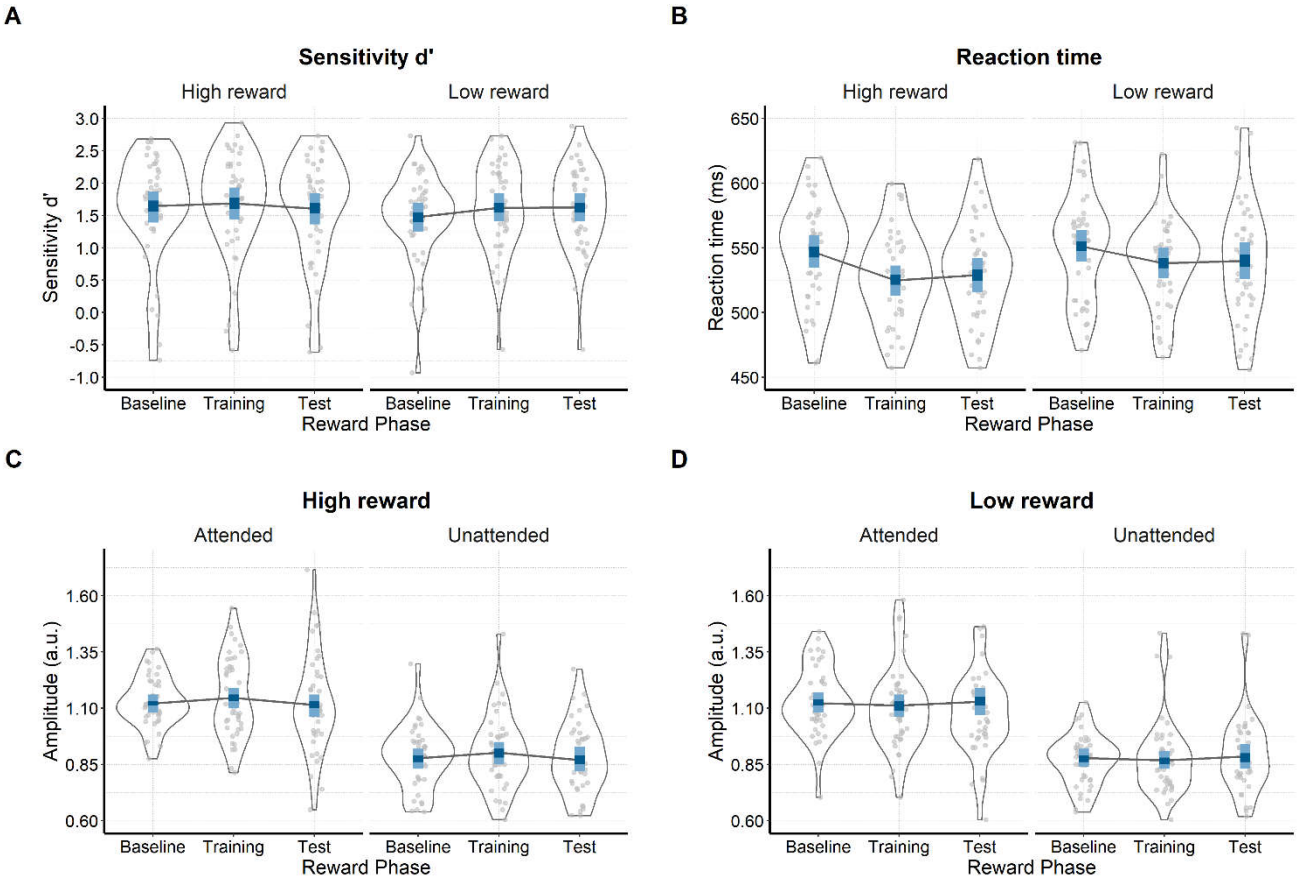


Figure 2. Raw and modelled data. Violin plots displaying raw data for each participant (grey dots), separately for each condition. Results from the winning models are presented in blue (dark blue – 50% HDIs and light blue – 95% HDIs). **A.** Sensitivity (d′) **B.** Reaction times (ms) **C.** SSVEP amplitudes (arbitrary units) in response to the color related to high reward on trials in which it is attended or unattended. **D.** SSVEP amplitudes for the color linked to low reward on trials when it was attended or unattended.

Reaction times

The *Reward phase* × *Reward probability* model also best predicted the reaction times (*Figure 2B* and *Table 1*). In the training, compared to the baseline phase, participants were reliably faster in detecting the motions of both the high (*M* = -21.60; 95% HDI [-29.90, -12.80]; ER > 12,000., i.e. the whole posterior distribution was below zero thus the ER is larger than the total number of posterior samples) and the low reward colors (*M* = -13.10; 95% HDI [-21.70, -4.69]; ER = 999). Moreover, this difference between baseline and training was larger for detecting motions of the high relative to low reward color (*M* = -8.49; 95% HDI [-18.60, 2.06]; ER = 17.18). We found less evidence for changes in reaction times between the training and the test phase. There was a very small increase in the reaction times in the test compared to training phase for the high reward color (*M* = 4.07; 95% HDI [-4.52, 13.10]; ER = 4.40), and no difference for the low reward (*M* = 1.87; 95% HDI [-6.93, 10.70]; ER = 1.98). These results indicate that participants were faster in detecting coherent motions in the condition in which they could earn rewards (training), and more so for high than low reward color. Also, there was a very small increase in the reaction times for the high reward condition and no difference in the low reward condition when the rewards were no longer available (test). Supplementary analyses carried out to assess possible training effects indicated some evidence for the presence of training effects in sensitivity and scant evidence for such effects in reaction times (*Supplementary materials*).

SSVEP amplitudes

As shown in *Figure 3* SSVEP amplitudes averaged over conditions peaked at central occipital channels (i.e., *Oz*, *POz*, *O2*, *Iz*). Also, the amplitude spectra showed the expected peaks at the frequencies of 10 and 12 Hz.

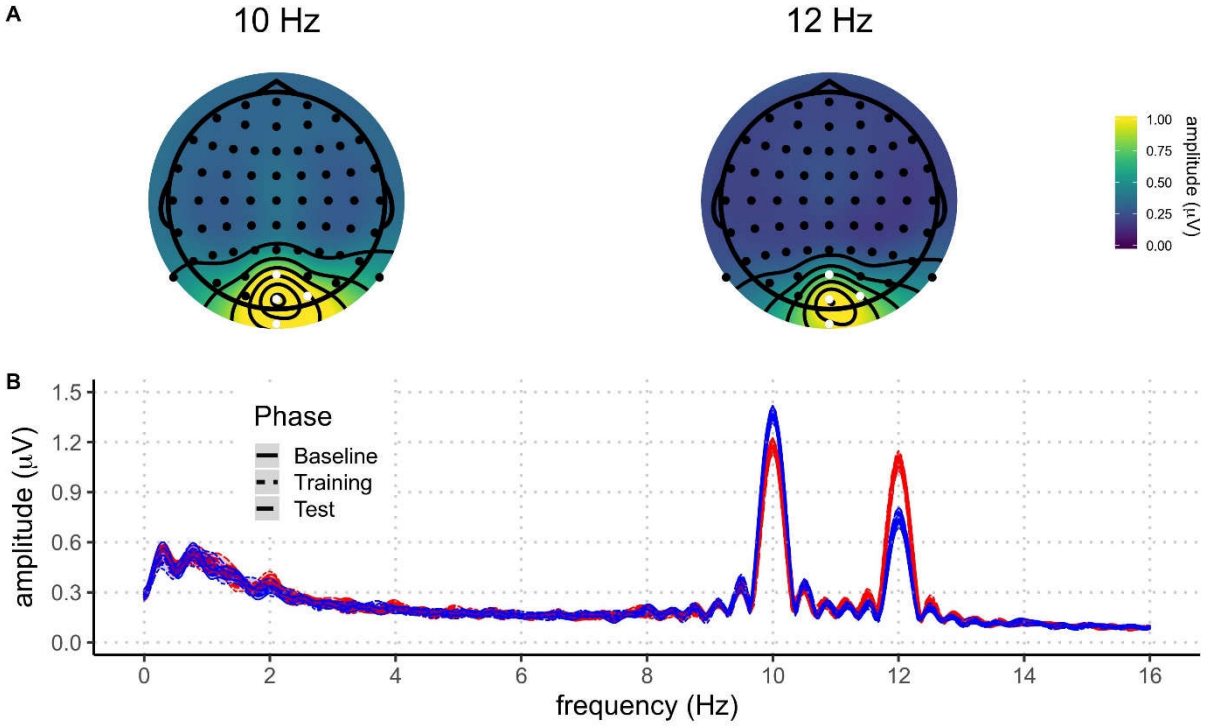


Figure 3. A) Topographies of SSVEP amplitudes, averaged across all participants and conditions, at 10 Hz and 12 Hz. Electrodes selected for the analysis are highlighted in white. **B)** Grand average amplitude spectra (zero-padded to 8 times the length of the data for plotting purpose) derived from EEG signals at best four-electrode cluster plotted for the different experimental conditions (blue: attended; red: unattended; solid: baseline phase; dotted: rewarded phase; dashed: non-rewarded phase). The shaded areas around the means indicate 95% confidence intervals.

The *Reward probability* × *Reward phase* + *Attention* model best predicted SSVEP amplitudes across conditions (*Table 3*). However, the *Reward probability* × *Reward phase* × *Attention* was only slightly worse than the winning model. Here we draw inferences from the winning model, but note that the conclusions do not substantially change when analyzing the model which includes the three-way interaction. The analysis of the posterior distributions of the winning model (*Figure 2* and *Table 3*) revealed a strong effect of voluntary selective attention: in all conditions, SSVEP amplitudes were higher when the eliciting stimulus was attended compared to when it was unattended. In the winning model, this effect did not interact with the other factors in the model, i.e. the magnitude of selective attention was unaffected by reward probability and reward phase. The posterior distribution of the difference between attended and unattended stimuli did not include zero, thus resulting in infinite probability that the attended stimuli would elicit higher SSVEP amplitudes compared to the unattended ones (M = 0.24; 95% HDI [0.19, 0.29]; ER > 16,000). These results reveal a very robust effect of voluntary selective attention across all experimental conditions: the SSVEP response was systematically larger when the driving stimulus was attended.

Table 3

Model comparison indices for EEG results

Model	*WAIC (SE)*	*Bayesian R^2 (SE)*
Null	-22.1 (56.2)	0.01 (0.01)
Reward phase	-31.3 (55.1)	0.05 (0.01)
Attention	-436.0 (66.4)	0.37 (0.02)
Reward phase + Attention	-465.2 (64.9)	0.40 (0.02)
Reward phase × Attention	-461.1 (65.2)	0.41 (0.02)
Reward probability × Reward phase + Attention	-695.9 (72.0)	0.55 (0.02)
Reward probability × Reward phase × Attention	-689.8 (71.8)	0.55 (0.02)

Table 4

Means and 95% HDIs of the posterior distributions of the SSVEP amplitudes for each condition.

Attention	Reward phase	Reward probability	Amplitudes (a.u.)
Attended	Baseline	High	1.12 [1.08, 1.16]
Attended	Baseline	Low	1.12 [1.08, 1.17]
Attended	Training	High	1.15 [1.10, 1.19]
Attended	Training	Low	1.11 [1.06, 1.16]
Attended	Test	High	1.11 [1.06 ,1.16]
Attended	Test	Low	1.13 [1.07, 1.19]
Unattended	Baseline	High	0.88 [0.83, 0.92]
Unattended	Baseline	Low	0.88 [0.84, 0.92]
Unattended	Training	High	0.90 [0.85, 0.95]
Unattended	Training	Low	0.87 [0.83, 0.91]
Unattended	Test	High	0.87 [0.82, 0.93]
Unattended	Test	Low	0.89 [0.83, 0.94]

The winning model also included the interaction between reward phase and reward probability, but this interaction remained the same for both attended and unattended stimuli. SSVEP amplitudes were higher in the training phase than at baseline for the high reward color (M = 0.02; 95% HDI [-0.01, 0.06]; ER = 10.24), both when it was attended and unattended. However, there was no evidence of difference for the change in SSVEP amplitudes from baseline to training for the low reward color (M = 0.01; 95% HDI [-0.02, 0.05]; ER = 2.65). Comparing the training to the test phase, the amplitudes of the high reward color were reduced (M = -0.03; 95% HDI [-0.07, 0.01]; ER = 13.01), while the amplitudes of the low reward color

did not change (M = -0.02; 95% HDI [-0.06, 0.03]; ER = 3.74). To summarize, visual processing of the high reward color stimulus was enhanced in the phase in which participants could earn monetary rewards and returned to baseline in the subsequent test phase without rewards. Thus change occurred irrespective of whether that color was attended or not. Finally, visual processing of the low reward color remained constant across the three phases of the experiment.

Discussion

In this study we investigated the neural mechanisms through which voluntary selective attention and reward history jointly guide visual processing. We compared the processing of attended and unattended stimuli of different reward probabilities on a continuous global motion discrimination task. Compared to baseline, the introduction of rewards speeded up task performance, especially for the higher reward stimuli, which was accompanied by enhanced processing of these stimuli in the visual cortex (higher SSVEPs). This sensory gain was present both when the high reward stimulus was attended and unattended, thus indicating that rewards influenced visual processing independently of voluntary selective attention. When rewards were no longer available, sensory processing of high reward stimuli returned to baseline levels, but participants were still faster to detect coherent motions of high reward stimuli.

The introduction of rewards improved behavioral performance on the task and facilitated the visual processing of stimuli associated with high rewards. This effect on SSVEP amplitudes is likely localized in the V1-V3 areas of the visual cortex, as reported in the previous studies using the same task (Andersen et al., 2009; Andersen & Müller, 2010; Andersen, Hillyard, & Müller, 2008). Crucially, this effect was the same both when the high reward stimulus was attended and unattended. Thus this effect was independent of the effect of voluntary selective attention reflected in the enhanced processing of the attended compared to unattended stimuli (Andersen & Müller, 2010). This pattern of results suggests that the effect of reward acted independently of visual selective attention. This finding supports the predictions of the models which propose that the effect of reward history on visual processing is independent from voluntary attention (Anderson, 2016; Awh et al., 2012; Chelazzi et al., 2013; Failing & Theeuwes, 2017). Further, this finding can help to refine the existing models which are focused on the role of rewards in the allocation of cognitive control. These models (Brown & Alexander, 2017; Holroyd & McClure, 2015; Shenhav et al., 2013; Verguts et al., 2015) are largely focused on the activity in the frontroparietal regions, for example the

dorsolateral Prefrontal Cortex and the Anterior Cingulate Cortex, which are known to increase their activation in anticipation of rewards (Krebs, Boehler, Roberts, Song, & Woldorff, 2012; Pessoa & Engelmann, 2010; Schevernels, Krebs, Santens, Woldorff, & Boehler, 2014). However, these models are not explicit about their predictions of how the top-down signals from these areas modulate the processing of stimuli at the level of the visual cortex. Our findings suggest that increased rewards act to enhance the processing of the stimuli related to high rewards independently from other top-down voluntary attention effects, which is similar to the way in which physical salience of stimuli acts in the same paradigm we used (Andersen et al., 2012). Interestingly, this is in contrast with the recent finding showing that a flagship cognitive control effect, post-error adjustments, operates through enhancement of voluntary selective attention as measured by SSVEPs in the adapted version of the task used here (Steinhauser & Andersen, 2019). This indicates a possible dissociation between the effects of reward and other cognitive control effects on selective attention. Such dissociation between the cognitive control and reward effects are an important question which should be further addressed both theoretically and empirically.

In the test phase behavior displayed similar patterns as in the training phase. Individuals were faster to detect motions of the dots in color related to high compared to low reward. This finding follows the reward-history effects reported in several paradigms (Anderson, Laurent, & Yantis, 2011; Della Libera & Chelazzi, 2009; Failing & Theeuwes, 2014). However, our SSVEP results show that the visual processing of the high reward stimuli returned to baseline levels, diverging from the behavioral pattern of results. This result indicates that the longer lasting effect of reward history was not mediated by the prolonged gain enhancement in sensory processing as measured by the SSVEPs, contrary to the predictions of the models accounting for the effects of reward history on attention (Anderson, 2016; Awh et al., 2012; Chelazzi et al., 2013; Failing & Theeuwes, 2017). This result is predicted by the models which relate cognitive control and reward, as they predict that the reward-related enhancements should retreat to baseline levels when rewards are no longer available (Brown & Alexander, 2017; Holroyd & McClure, 2015; Shenhav et al., 2013; Verguts et al., 2015). This finding suggest that visual processing can be adapted in a much more flexible way than predicted by the models focused on the reward-history effects on attention. Of note, it is possible that our SSVEP measure captures the more general aspect of feature processing in the visual cortex, while the effects of reward history could be specifically locked to the onset of the rewarded stimulus (Donohue et al., 2016; Hickey et al., 2010; Luque et al., 2017; MacLean & Giesbrecht, 2015). However,

there are at least two studies which have not found evidence for the effects of reward history on early visual processing (Qi et al., 2013; Tankelevitch et al., 2019). This leaves open the possibility that the effects of reward history are not necessarily driven purely by gains in sensory processing. One interesting possibility, which should be explored in future studies, is that rewards initially improve performance by enhancing stimulus saliency, but later rely on more direct stimulus-response mappings. Finally, it is important to note that our paradigm involves a cue on every trial which induces a direct goal, while this is not the case in most of the studies assessing the influence of reward-history on attention. Further research using SSVEPs in tasks designed to explicitly address the reward-history effects could help disentangle between the possible explanations of our findings.

Our paradigm allowed us to simultaneously measure the processing of stimuli linked to both high and low value. Some initial evidence for the suppression of the stimuli linked to low compared to high rewards has been found at the behavioral and neural level (Hickey & Peelen, 2015; Padmala & Pessoa, 2011). Suppression of the features linked to low or no rewards has also been proposed as one of the potential mechanisms through which incentives impact attention (Chelazzi et al., 2013; Anderson, 2016; Failing & Theeuwes, 2018). Contrary to this, in this study we found no evidence for this proposal. Suppression was neither observed when the low value color was attended, nor when it was unattended. The amount of attention allocated toward this feature remained unchanged throughout the experiment. There are two features of our experiment which could explain this finding. First, in our experiment both colors were related to rewards, but they differed in reward value. For example, the study which showed evidence for the suppression of the non-rewarded feature did so in the context in which suppression occurred for the representations of objects which were never rewarded (Hickey & Peelen, 2015). In our paradigm it could be beneficial for participants not to suppress the low value color because correct responses to the motions of this color would still earn them a reward on 20% of trials. Second, while the attended color changed trial-by-trial in our experiment, the experiment of Hickey and Peelen consisted out of small blocks of 16 trials in which the attended object was always the same (e.g., searching for a car in a complex picture). When searching for one object or feature across a number of future trials, it is possible that the optimal solution for the cognitive system is to suppress the processing of the other features or objects (i.e., goal-irrelevant stimuli). However, if the attended feature is likely to change on each trial, as in our experiment, the suppression of the low rewarded feature could be maladaptive as it would carry

a cost of reconfiguring the control signals on every trial (for a computational implementation of a reconfiguration cost see: Musslick, Shenhav, Botvinick, & Cohen, 2015).

The design of this study and the use of the SSVEPs allowed us to independently assess the influence of voluntary attention and reward on sensory processing in the visual cortex. This enabled us to directly compare the magnitudes of the two effects on sensory processing. While both of these factors displayed an effect on visual processing, it is important to note that the effect of voluntary attention on visual processing (30% increase for the attended vs. the unattended stimuli) was an order of magnitude stronger than the effect of reward (3% increase from baseline to training for the high reward stimuli). Thus even though reward associations can influence processing in opposition to voluntary attention, the magnitude of this effect is very small compared to the effect of voluntary attention. Most theoretical models to date have focused on how top-down and reward-driven attention jointly guide stimulus processing (Awh et al., 2012), but how much each of these processes contribute to stimulus processing still has to be incorporated into the theoretical models. This finding is especially important in the light of the recent studies investigating the relevance of reward-driven automatic biases in attention in clinical disorders such as addiction (Anderson, 2016) and depression (Anderson, Leal, Hall, Yassa, & Yantis, 2014). While it is possible that the more automatic biases in attention play a role in these disorders, it is also important to focus on the role of more goal-directed processes which are likely to have a bigger impact on cognition in clinical disorders (Grahek, Shenhav, Musslick, Krebs, & Koster, 2019).

In conclusion, in this study we directly assessed how voluntary attention and reward jointly guide attention. Our findings provide a novel insight into the flexible dynamics of visual processing based on these two factors. They demonstrate that rewards can act independently of voluntary attention to enhance sensory processing in the visual cortex. However, sensory processing is flexibly readjusted when rewards are no longer available. This result points to the dynamical interplay between top-down and reward effects in the visual cortex which needs to be accounted for in theoretical models of motivation-cognition interactions. The effect can be flexibly removed as soon as the reward structure in the environment changes.

Supplementary materials

Means of the raw behavioral and SSVEP data

Supplementary Table 1

Means and 95% HDIs (in square brackets) of the raw data for sensitivity and reaction times

Reward phase	Value	Sensitivity (d′)	Reaction times (milliseconds)
Baseline	High	1.64 [-0.04, 2.68]	546.59 [485.64, 619.34]
Baseline	Low	1.47 [0.04, 2.30]	551.10 [490.50, 631.36]
Training	High	1.69 [0.29, 2.73]	524.99 [467.12, 599.49]
Training	Low	1.62 [0.46, 2.68]	537.94 [465.32, 584.63]
Test	High	1.60 [-0.20, 2.73]	528.98 [457.08, 599.83]
Test	Low	1.62 [0.74, 2.88]	539.75 [455.80, 623.21]

Supplementary Table 2

Means and 95% HDIs of the raw data for the recorded SSVEP amplitudes in each condition

Attention	Reward phase	Value	Amplitudes (a.u.)
Attended	Baseline	High	1.13 [0.92, 1.52]
Attended	Baseline	Low	1.13 [0.86, 1.52]
Attended	Training	High	1.16 [0.80, 1.60]
Attended	Training	Low	1.13 [0.76, 1.71]
Attended	Test	High	1.13 [0.61, 1.61]
Attended	Test	Low	1.13 [0.59, 1.84
Unattended	Baseline	High	0.87 [0.47, 1.17]
Unattended	Baseline	Low	0.87 [0.49, 1.11]
Unattended	Training	High	0.91 [0.54, 1.38]
Unattended	Training	Low	0.89 [0.50, 1.28]
Unattended	Test	High	0.88 [0.48, 1.23]
Unattended	Test	Low	0.91 [0.44, 1.42]

Additional analyses to assess the possible training effects

In order to assess potential training effects on behavioral performance, we split each reward phase into two halves (*Supplementary Figure 1* and *Supplementary Table 3*). If training effects were influencing the behavioral outcome, we could expect performance improvement through baseline and training. To investigate this possibility, we fitted the *Reward phase* × *Value* model that was identical to the one described in the results section. We then compared behavioral performance between the first and the second part of the baseline phase, and between the second part of baseline and the first part of training phase.

Supplementary Table 3

Means and 95% HDIs of the war data for sensitivity and reaction times across six phases of the experiment

Reward phase	Value	Sensitivity (d′)	Reaction times (milliseconds)
Baseline1	High	1.48 [-0.36, 2.62]	548.84 [479.43, 613.76]
Baseline1	Low	1.32 [0.09, 2.35]	548.43 [458.26, 610.63]
Baseline2	High	1.60 [-0.27, 2.56]	544.34 [454.56, 620.36]
Baseline2	Low	1.47 [0.08, 2.33]	554.01 [479.48, 632.80]
Training1	High	1.54 [-0.08, 2.65]	521.40 [437.90, 587.57]
Training1	Low	1.59 [0.47, 2.45]	542.34 [463.65, 593.47]
Training2	High	1.59 [0.08, 2.56]	528.74 [462.00, 598.58]
Training2	Low	1.48 [0.00, 2.62]	533.94 [479.38, 618.25]
Test1	High	1.48 [-0.07, 2.47	528.58 [457.88, 596.17]
Test1	Low	1.50 [0.36, 2.50]	536.54 [444.89, 621.00]
Test2	High	1.49 [-0.38, 2.49]	529.30 [448.24, 606.00]
Test2	Low	1.55 [0.65, 2.55]	543.01 [450.11, 617.44]

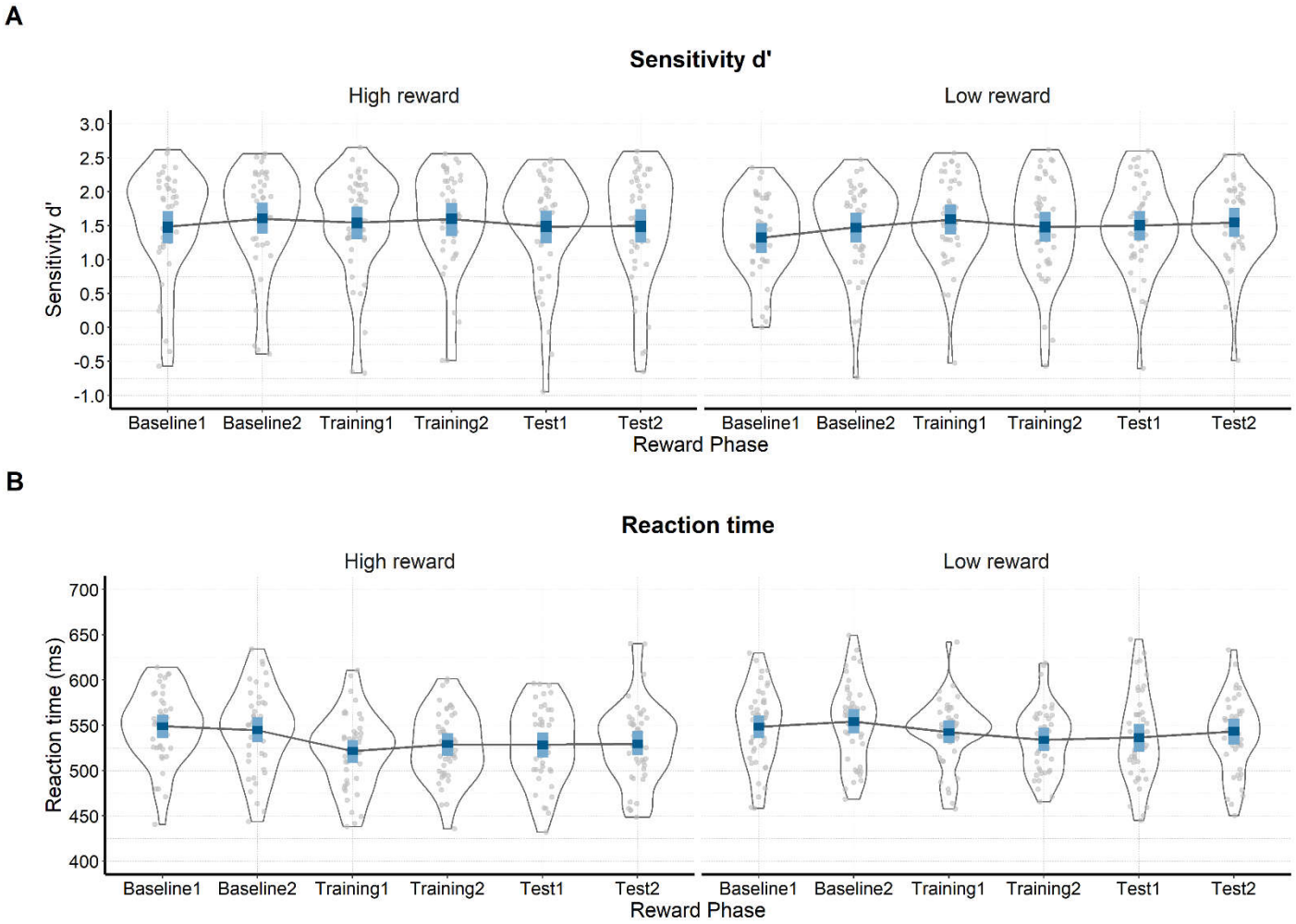
The posterior distributions for sensitivity (*Supplementary Figure 1A* and *Supplementary Table 4*) revealed performance improvement from the first to the second part of the baseline for both high (M = 0.12; 95% HDI [-0.05, 0.28]; ER = 11.05) and low (M = 0.15; 95% HDI [0.01, 0.32]; ER = 36.04) value conditions. When comparing the second part of baseline to the first part of training, there was only a very small improvement in sensitivity in the high value condition (M = 0.06; 95% HDI [-0.11, 0.22]; ER = 2.94), and a much bigger one in the low

value condition (*M* = 0.11; 95% HDI [-0.04, 0.28]; ER = 10.90). These results indicate that participants improved not only throughout the baseline phase, but also from the end of baseline to the first part of the training (albeit for low rewarded color only). This might indicate some presence of training effects in the sensitivity data.

The posterior distributions of reaction times (*Supplementary Figure 1B* and *Supplementary Table 2*) revealed only a very small difference between the first and the second part of baseline for high (*M* = -4.52; 95% HDI [-15.0,0 5.77]; ER = 4.21) value condition, while the low value condition was somewhat slower in the second part of the baseline (*M* = 5.60 95% HDI [-4.76, 16.20]; ER = 5.71). The comparison between the second part of baseline and the first part of training revealed a very reliable speeding in both high (*M* = 22.90; 95% HDI [12.60, 33.80]; ER > 6000) and low (*M* = 11.60; 95% HDI [0.70, 22.10]; ER = 57.82) value conditions. These results clearly point to the absence of large training effects in the reaction time data.

Taken together, these results indicate that our effects were not driven by the improved performance over the course of the task. Although there is some evidence that sensitivity was improving during the baseline phase, reaction times clearly indicate that the main shift in performance happens in the beginning of training, when rewards are introduced. Importantly, the strongest behavioral effects in our study were found on reaction time data, as indicated in the results section.

Similar analyses could not be performed for the EEG data, because splitting the number of trials in each phase would significantly affect the signal-to-noise ratio. However, our EEG results point to changes in SSVEP amplitudes in only one of the value conditions. If amplitude changes were mainly driven by training effects, the differences across reward phases would be expected for both value conditions. This observation, combined with the lack of strong training effects in behavior, suggests that our EEG results are not driven by training effects.



Supplementary Figure 1. Raw and modelled behavioral data in each half of each phase of the experiment. On both plots raw participant data is represented with grey dots and their distribution. The winning model is presented in blue (dark blue – 50% HDIs and light blue – 95% HDIs). **A.** Sensitivity (d′) across the phases of the experiment for the conditions in which the attended color is linked to either high or low value. **B)** Reaction times (ms) in the six phases when the attended stimulus is related to high or low reward probability.

Supplementary Table 4

Means and 95% HDIs of sensitivity and reaction times across six phases of the experiment

Reward phase	Value	Sensitivity (d′)	Reaction times (milliseconds)
Baseline 1	High	1.48 [1.24, 1.71]	548.86 [535.97, 561.35]
Baseline 1	Low	1.32 [1.09, 1.53]	548.38 [535.83, 560.97]
Baseline 2	High	1.60 [1.38, 1.84]	544.34 [531.22, 558.49]
Baseline 2	Low	1.47 [1.25, 1.69]	553.98 [540.67, 567.69]
Training 1	High	1.54 [1.30, 1.78]	521.42 [508.48, 533.66]
Training 1	Low	1.59 [1.37, 1.81	542.35 [530.05, 555.45]
Training 2	High	1.60 [1.35, 1.83]	528.74 [515.92, 541.36]
Training 2	Low	1.48 [1.26, 1.70]	533.91 [521.41, 547.24]
Test 1	High	1.49 [1.24, 1.72]	528.64 [514.39, 542.24]
Test 1	Low	1.50 [1.28, 1.71]	536.51 [520.49, 551.37]
Test 2	High	1.49 [1.25, 1.74]	529.32 [516.53, 543.70]
Test 2	Low	1.55 [1.33, 1.76]	543.01 [528.56, 557.28]

Software for data visualization and analysis

Visualization and statistical analyses were performed using R v3.4.4 (R Core Team, 2017) via RStudio v1.1.453 (RStudio Team, 2015). We used the following packages (and their respective dependencies):

- data manipulation: tidyverse v1.2.1 (Wickham, 2017);
- statistical analyses: Rmisc v1.5 (Hope, 2013), brms v2.3.1 (Bürkner, 2016);
- visualization: cowplot v0.9.2 (Wilke, 2016), yarrr v0.1.5 (Phillips, 2016), viridis v0.5.1 (Garnier, 2018), eegUtils v0.2.0 (Craddock, 2018), BEST (J. K. Kruschke & Meredith, 2017);
- report generation: pacman v0.4.6 (Rinker & Kurkiewicz, n.d.), knitr v1.20 (Xie, 2018).

Data

Data and analyses code is available at:
https://osf.io/h2knm/?view_only=78e7e9b78cee487c8cb4f8b56afb538b

Acknowledgements

This work was supported by the Special Research Fund (BOF) of Ghent University [grant #01D02415 awarded to IG; grant # BOF14/PDO/123 awarded to AS], the Concerted Research Action Grant of Ghent University [grant number BOF16/GOA/017 awarded to EHWK], and the Biotechnology and Biological Sciences Research Council [BB/P002404/1 awarded to SKA]. The funding sources had no involvement in the study design; collection, analysis, and interpretation of data; writing of the report; and decision to submit the article for publication.

We would like to thank Ladislas Nalborczyk for discussions about statistical analyses of the data, Ruth Krebs for her comments on a draft of the manuscript, and Dr. Inez Greven for help with data collection.

References

Allen, L., Scott, J., Brand, A., Hlava, M., & Altman, M. (2014). Publishing: Credit where credit is due. *Nature*, *508*(7496), 312–313. https://doi.org/10.1038/508312a

Andersen, S. K., Muller, M. M., & Hillyard, S. A. (2009). Color-selective attention need not be mediated by spatial attention. *Journal of Vision*, *9*(6), 2–2. https://doi.org/10.1167/9.6.2

Andersen, S. K., Muller, M. M., & Martinovic, J. (2012). Bottom-Up Biases in Feature-Selective

Attention. *Journal of Neuroscience*, *32*(47), 16953–16958. https://doi.org/10.1523/JNEUROSCI.1767-12.2012

Andersen, S K, & Müller, M. M. (2010). Behavioral performance follows the time course of neural facilitation and suppression during cued shifts of feature-selective attention. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(31), 13878–13882. https://doi.org/10.1073/pnas.1002436107

Andersen, Søren K., Hillyard, S. A., & Müller, M. M. (2008). Attention Facilitates Multiple Stimulus Features in Parallel in Human Visual Cortex. *Current Biology*, *18*(13), 1006–1009. https://doi.org/10.1016/j.cub.2008.06.030

Anderson, B. A. (2016). The attention habit: How reward learning shapes attentional selection. *Annals of the New York Academy of Sciences*, *1369*(1), 24–39. https://doi.org/10.1111/nyas.12957

Anderson, B. a, Laurent, P. a, & Yantis, S. (2011). Value-driven attentional capture. *Proceedings of the National Academy of Sciences*, *108*(25), 10367–10371. https://doi.org/10.1073/pnas.1104047108

Anderson, Brian A. (2016). What is abnormal about addiction-related attentional biases? *Drug and Alcohol Dependence*, *167*, 8–14. https://doi.org/10.1016/j.drugalcdep.2016.08.002

Anderson, Brian A, Leal, S. L., Hall, M. G., Yassa, M. A., & Yantis, S. (2014). The attribution of value-based attentional priority in individuals with depressive symptoms. *Cognitive, Affective & Behavioral Neuroscience*, *14*(4), 1221–1227. https://doi.org/10.3758/s13415-014-0301-z

Awh, E., Belopolsky, A. V., & Theeuwes, J. (2012). Top-down versus bottom-up attentional control: A failed theoretical dichotomy. *Trends in Cognitive Sciences*, *16*(8), 437–443. https://doi.org/10.1016/j.tics.2012.06.010

Botvinick, M. M., & Braver, T. S. (2015). Motivation and Cognitive Control : From Behavior to Neural Mechanism. *Annual Review of Psychology*, (September 2014), 1–31. https://doi.org/10.1146/annurev-psych-010814-015044

Brown, J. W., & Alexander, W. H. (2017). Foraging Value, Risk Avoidance, and Multiple Control Signals: How the Anterior Cingulate Cortex Controls Value-based Decision-making. *Journal of Cognitive Neuroscience*, *29*(10), 1656–1673. https://doi.org/10.1162/jocn_a_01140

Bürkner, P.-C. (2016). brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software*, *80*(1), 1–28.

Bürkner, P.-C. (2017). Advanced Bayesian Multilevel Modeling with the R Package brms. *ArXiv:1705.11123*.

Carpenter, B., Gelman, A., Hoffman, M., Lee, D., Goodrich, B., Betancourt, M., … Riddell, A. (2016). Stan: A probabilistic programming language. *Journal of Statistical Software*, *2*(20), 1–37.

Chatrian, G. E., Lettich, E., & Nelson, P. L. (1985). Ten percent electrode system for topographic studies of spontaneous and evoked EEG activities. *American Journal of EEG Technology*, *25*(2).

Chelazzi, L., Perlato, A., Santandrea, E., & Della Libera, C. (2013). Rewards teach visual selective attention. *Vision Research*, *85*, 58–62. https://doi.org/10.1016/j.visres.2012.12.005

Chun, M. M., Golomb, J. D., & Turk-Browne, N. B. (2011). A Taxonomy of External and Internal

Attention. *Annual Review of Psychology*, *62*(1), 73–101. https://doi.org/10.1146/annurev.psych.093008.100427

Corbetta, M., & Shulman, G. L. (2002). Control of Goal-Directed and Stimulus-Driven Attention in the Brain. *Nature Reviews Neuroscience*, *3*(3), 215–229. https://doi.org/10.1038/nrn755

Craddock, M. (2018). craddm/eegUtils: eegUtils (Version v0.2.0). Zenodo.

Della Libera, C., & Chelazzi, L. (2009). Learning to attend and to ignore is a matter of gains and losses. *Psychological Science*, *20*(6), 778–784. https://doi.org/10.1111/j.1467-9280.2009.02360.x

Delorme, A., & Makeig, S. (2004). EEGLAB: an open sorce toolbox for analysis of single-trail EEG dynamics including independent component anlaysis. *Journal of Neuroscience Methods*, *134*, 9–21. https://doi.org/10.1016/j.jneumeth.2003.10.009

Desimone, R., & Duncan, J. (1995). Neural Mechanisms of Selective Visual. *Annual Review of Neuroscience*, *18*(1), 193–222. https://doi.org/10.1146/annurev.ne.18.030195.001205

Donohue, S. E., Hopf, J.-M., Bartsch, M. V., Schoenfeld, M. A., Heinze, H.-J., & Woldorff, M. G. (2016). The Rapid Capture of Attention by Rewarded Objects. *Journal of Cognitive Neuroscience*, *28*(4), 529–541. https://doi.org/10.1162/jocn_a_00917

Etzel, J. A., Cole, M. W., Zacks, J. M., Kay, K. N., & Braver, T. S. (2016). Reward Motivation Enhances Task Coding in Frontoparietal Cortex. *Cerebral Cortex*, *26*(4), 1647–1659. https://doi.org/10.1093/cercor/bhu327

Failing, M. F., & Theeuwes, J. (2014). Exogenous visual orienting by reward. *Journal of Vision*, *14*(2014), 1–9. https://doi.org/10.1167/14.5.6.doi

Failing, M., & Theeuwes, J. (2017). Selection history: How reward modulates selectivity of visual attention. *Psychonomic Bulletin and Review*, 1–25. https://doi.org/10.3758/s13423-017-1380-y

Franken, I. H. A., Muris, P., & Rassin, E. (2005). Psychometric properties of the Dutch BIS/BAS scales. *Journal of Psychopathology and Behavioral Assessment*, *27*(1), 25–30. https://doi.org/10.1007/s10862-005-3262-2

Garnier, S. (2018). viridis: Default Color Maps from ‘matplotlib.’ R package version 0.3.

Gelman, A., Goodrich, B., Gabry, J., & Ali, I. (2017). R-squared for Bayesian regression models. *Unpublished via Http://Www. Stat. Columbia. Edu/~ Gelman/Research/Unpublished.* Retrieved from http://www.stat.columbia.edu/~gelman/research/unpublished/bayes_R2.pdf

Gelman, A., & Rubin, D. B. (1992). Inference from Iterative Simulation Using Multiple Sequences. *Statistical Science*, *7*(4), 457–472. https://doi.org/10.1214/ss/1177011136

Grahek, I., Shenhav, A., Musslick, S., Krebs, R. M., & Koster, E. H. W. (2019). Motivation and cognitive control in depression. *Neuroscience & Biobehavioral Reviews*, *102*(March), 371–381. https://doi.org/10.1016/j.neubiorev.2019.04.011

Hautus, M. J. (1995). Corrections for extreme proportions and their biasing effects on estimated values of d′. *Behavior Research Methods, Instruments, & Computers*, *27*(1), 46–51. https://doi.org/10.3758/BF03203619

Hickey, C., Chelazzi, L., & Theeuwes, J. (2010). Reward Changes Salience in Human Vision via the

Anterior Cingulate. *Journal of Neuroscience*, *30*(33), 11096–11103. https://doi.org/10.1523/JNEUROSCI.1026-10.2010

Hickey, C., & Peelen, M. V. (2015). Neural mechanisms of incentive salience in naturalistic human vision. *Neuron*, *85*(3), 512–518. https://doi.org/10.1016/j.neuron.2014.12.049

Holroyd, C. B., & McClure, S. M. (2015). Hierarchical control over effortful behavior by rodent medial frontal cortex: A computational model. *Psychological Review*, *122*(1), 54–83. https://doi.org/10.1037/a0038339

Hope, R. M. (2013). Rmisc: Ryan miscellaneous. R package version, 1(5).

Kashiwase, Y., Matsumiya, K., Kuriki, I., & Shioiri, S. (2012). Time courses of attentional modulation in neural amplification and synchronization measured with steady-state visual-evoked potentials. *Journal of Cognitive Neuroscience*, *24*(8), 1779–1793. https://doi.org/10.1162/jocn_a_00212

Krebs, R M, & Woldorff, M. G. (2017). Cognitive control and reward. In T. Egner (Ed.), *Wiley Handbook of Cognitive Control* (pp. 422–439). Wiley-Blackwell.

Krebs, Ruth M., Boehler, C. N., Roberts, K. C., Song, A. W., & Woldorff, M. G. (2012). The involvement of the dopaminergic midbrain and cortico-striatal-thalamic circuits in the integration of reward prospect and attentional task demands. *Cerebral Cortex*, *22*(3), 607–615. https://doi.org/10.1093/cercor/bhr134

Kruschke, J. K., & Meredith, M. (2017). BEST: Bayesian Estimation Supersedes the t-Test.

Kruschke, John K. (2014). *Doing Bayesian data analysis: A tutorial with R, JAGS, and Stan, second edition*. *Doing Bayesian Data Analysis: A Tutorial with R, JAGS, and Stan, Second Edition* (2nd ed.). Elsevier Inc. https://doi.org/10.1016/B978-0-12-405888-0.09999-2

Luque, D., Beesley, T., Morris, R. W., Jack, B. N., Griffiths, O., Whitford, T. J., & Le Pelley, M. E. (2017). Goal-Directed and Habit-Like Modulations of Stimulus Processing during Reinforcement Learning. *The Journal of Neuroscience*, *37*(11), 3009–3017. https://doi.org/10.1523/jneurosci.3205-16.2017

MacLean, M. H., & Giesbrecht, B. (2015). Neural evidence reveals the rapid effects of reward history on selective attention. *Brain Research*, *1606*, 86–94. https://doi.org/10.1016/j.brainres.2015.02.016

Macmillan, N. A., & Creelman, C. D. (2004). *Detection theory: A user's guide*. Psychology press.

Maunsell, J. H. R. (2004). Neuronal representations of cognitive state: Reward or attention? *Trends in Cognitive Sciences*, *8*(6), 261–265. https://doi.org/10.1016/j.tics.2004.04.003

McElreath, R. (2016). *Statistical Rethinking: A Bayesian Course with Examples in R and Stan*. Chapman Hall - CRC.

Morgan, S. T., Hansen, J. C., & Hillyard, S. A. (1996). Selective attention to stimulus location modulates the steady-state visual evoked potential, *93*(10), 4770–4774.

Müller, M. M., Teder-Sälejärvi, W., & Hillyard, S. A. (1998). The time course of cortical facilitation during cued shifts of spatial attention. *Nature Neuroscience*, *1*(7), 631–634.

Musslick, S., Shenhav, A., Botvinick, M. M., & Cohen, J. D. (2015). A computational model of control allocation based on the Expected Value of Control. *Reinforcement Learning and*

Decision Making Conference, *59*(1978), 2014.

Nalborczyk, L., & Bürkner, P.-C. (2019). An Introduction to Bayesian Multilevel Models Using brms: A Case Study of Gender Effects on Vowel Variability in Standard Indonesian. *Journal of Speech, Language, and Hearing Research*.

Nolan, H., Whelan, R., & Reilly, R. B. (2010). FASTER: Fully Automated Statistical Thresholding for EEG artifact Rejection. *Journal of Neuroscience Methods*, *192*(1), 152–162. https://doi.org/10.1016/j.jneumeth.2010.07.015

Padmala, S., & Pessoa, L. (2011). Reward reduces conflict by enhancing attentional control and biasing visual cortical processing. *Journal of Cognitive Neuroscience*, *23*(11), 3419–3432. https://doi.org/10.1162/jocn_a_00011

Pallier, C. (2002). Computing discriminability and bias with the R software. Retrieved July 25, 2019, from http://www.pallier.org/pdfs/aprime.pdf

Perrin, F., Pernier, J., Bertrand, O., & Echallier, J. F. (1989). Spherical splines for scalp potential and current density mapping. *Electroencephalography and Clinical Neurophysiology*, *72*(2), 184–187. https://doi.org/10.1016/0013-4694(89)90180-6

Pessoa, L. (2015). Multiple influences of reward on perception and attention. *Visual Cognition*, *23*(1–2), 272–290. https://doi.org/10.1080/13506285.2014.974729

Pessoa, L., & Engelmann, J. B. (2010). Embedding reward signals into perception and cognition. *Frontiers in Neuroscience*, *4*(September), 1–8. https://doi.org/10.3389/fnins.2010.00017

Phillips, N. (2016). Yarrr: A companion to the e-book YaRrr!: The Pirate's Guide to R. R package version 0.1.

Posner, M. I. (1980). Orienting of attention. *The Quarterly Journal of Experimental Psychology*, *32*(1), 3–25. https://doi.org/10.1080/00335558008248231

Posner, MI I. (1980). Orienting of attention. *Quarterly Journal of Experimental Psychology*, *32*(1), 3–25. https://doi.org/10.1080/00335558008248231

Qi, S., Zeng, Q., Ding, C., & Li, H. (2013). Neural correlates of reward-driven attentional capture in visual search. *Brain Research*, *1532*, 32–43. https://doi.org/10.1016/j.brainres.2013.07.044

R Core Team. (2017). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing.

Rinker, T., & Kurkiewicz, D. (n.d.). pacman: Package Management for R.

Roelfsema, P. R., van Ooyen, A., & Watanabe, T. (2010). Perceptual learning rules based on reinforcers and attention. *Trends in Cognitive Sciences*, *14*(2), 64–71. https://doi.org/10.1016/j.tics.2009.11.005

RStudio Team. (2015). Integrated Development for R. RStudio, Inc.

Schevernels, H., Krebs, R. M., Santens, P., Woldorff, M. G., & Boehler, C. N. (2014). Task preparation processes related to reward prediction precede those related to task-difficulty expectation. *NeuroImage*, *84*, 639–647. https://doi.org/10.1016/j.neuroimage.2013.09.039

Serences, J. T. (2008). Value-Based Modulations in Human Visual Cortex. *Neuron*, *60*(6), 1169–1181.

https://doi.org/10.1016/j.neuron.2008.10.051

Shenhav, A., Botvinick, M., & Cohen, J. (2013). The expected value of control: An integrative theory of anterior cingulate cortex function. *Neuron*, *79*(2), 217–240. https://doi.org/10.1016/j.neuron.2013.07.007

Steinhauser, M., & Andersen, S. K. (2019). Rapid adaptive adjustments of selective attention following errors revealed by the time course of steady-state visual evoked potentials. *NeuroImage*, *186*(July 2018), 83–92. https://doi.org/10.1016/j.neuroimage.2018.10.059

Tankelevitch, L., Spaak, E., Rushworth, M. F. S., & Stokes, M. G. (2019). Previously reward-associated stimuli capture spatial attention in the absence of changes in the corresponding sensory representations as measured with {MEG}. *BioRxiv*, 622589. https://doi.org/10.1101/622589

Theeuwes, J. (2010). Top-down and bottom-up control of visual selection. *Acta Psychologica*, *135*(2), 77–99. https://doi.org/10.1016/j.actpsy.2010.02.006

Van der Does, A. J. W. (2002). *Handleiding bij de Nederlandse versie van beck depression inventory—second edition (BDI-II-NL). [The Dutch version of the Beck depression inventory]*. Amsterdam: Harcourt.

Verguts, T., Vassena, E., & Silvetti, M. (2015). Adaptive effort investment in cognitive and physical tasks: a neurocomputational model. *Frontiers in Behavioral Neuroscience*, *9*(March). https://doi.org/10.3389/fnbeh.2015.00057

Watanabe, S. (2010). Asymptotic Equivalence of Bayes Cross Validation and Widely Applicable Information Criterion in Singular Learning Theory, *11*, 3571–3594. Retrieved from http://arxiv.org/abs/1004.2316

Wickham, H. (2017). Tidyverse: Easily install and load 'tidyverse' packages. R package version, 1(1).

Wilke, C. O. (2016). cowplot: streamlined plot theme and plot annotations for 'ggplot2.' CRAN Repos.

Wisniewski, D., Reverberi, C., Momennejad, I., Kahnt, T., & Haynes, J.-D. (2015). The Role of the Parietal Cortex in the Representation of Task-Reward Associations. *Journal of Neuroscience*, *35*(36), 12355–12365. https://doi.org/10.1523/jneurosci.4882-14.2015

Xie, Y. (2018). knitr: A General-Purpose Package for Dynamic Report Generation in R.

Chapter 5

Learning if control matters: Neural dynamics underlying the updating of expectations and control allocation in response to changes in performance efficacy[1]

Abstract

To determine how much effort to allocate in a given cognitive task, we weigh the potential rewards and the degree to which effort will contribute to reward attainment (the efficacy of control). However, it is unclear how we learn about the efficacy of control, and which neural mechanisms mediate the process of updating predictions regarding our own efficacy and utilizing these predictions to proactively adjust future control allocation. In this study, we used EEG to study the dynamics of processes relating learning and control allocation. Forty-one subjects performed a Stroop task in which they received trial-by-trial feedback on whether they earned a reward and if that reward depended on their performance or not (efficacy). Efficacy levels varied over the course of the experiment in continuous fashion, and learning was assessed by probing subjects for their subjective estimates of their current efficacy. We found that subjects updated their efficacy estimates based on recent feedback, and that higher efficacy estimates were related to better subsequent Stroop performance. Increases in these efficacy estimates were associated with increased neural activity associated with allocation of attention (CNV) prior to target onset. Furthermore, the reward feedback processing (P3a) was modulated by the efficacy feedback on that trial. Finally, efficacy estimates modulated neural activity associated with memory updating in response to efficacy feedback (P3b). These results uncover the neural dynamics through which people learn how much efficacy they have and the mechanisms through which efficacy influences how they allocate cognitive control.

[1] Grahek, I., Frömer, R., & Shenhav, A. Learning if control matters: Neural dynamics underlying the updating of expectations and control allocation in response to changes in performance efficacy. (in preparation).

Introduction

Goal-directed behavior depends on the costs and benefits of investing effort to achieve our goals. Classic theories of motivation (Atkinson, 1957; Brehm & Self, 1989; Vroom, 1964) and learning (Balleine & O'Doherty, 2010; Dickinson, 1985) propose that effort costs can be offset by two factors: the value of outcomes (e.g., the reward associated with accomplishing a goal) and the probability that the effort invested will lead to the desired outcome (here termed efficacy, but also referred to as action-outcome contingencies or performance contingency). For example, we will invest a lot of effort into climbing a mountain only if we highly value the outcome of being on the top of it (high estimate of the reward associated with the outcome) and if we believe that, given the invested effort, we are likely to successfully climb to the top (high efficacy estimate). The belief that we have efficacy, meaning that outcomes in our environment are controllable, is crucial for both goal-directed behavior and mental health (Leotti, Iyengar, & Ochsner, 2010; Moscarello & Hartley, 2017).

Cognitive control is a set of processes which underlie goal-directed behavior by allowing for the flexible adaptation of cognition and behavior in accordance with our goals (Botvinick & Cohen, 2014; Friedman & Miyake, 2017). While exerting more control can be beneficial, it is also effortful and people tend to avoid it whenever they can (Kool, McGuire, Rosen, & Botvinick, 2010). Thus one of the central questions for theories of cognitive control is what determines the amount of control allocated. Recent computational models of cognitive control have relied on the insights from motivation and reinforcement learning research to resolve this problem. Such models propose that effortful cognitive control is allocated based on how motivated people are to exert control (Brown & Alexander, 2017; Holroyd & McClure, 2015; Shenhav, Botvinick, & Cohen, 2013; Verguts, Vassena, & Silvetti, 2015). A growing number of studies support this idea and demonstrate that more control is allocated if the value of outcomes is increased (for a review see: Botvinick & Braver, 2015). However, the role of efficacy in cognitive control allocation has received far less empirical attention.

The Expected Value of Control theory (EVC theory; Shenhav et al., 2013) proposes that the amount of control allocated is based on the decision-making process which includes the value associated with reaching an outcome (e.g., providing a correct response on the task), the probability of obtaining that outcome given the amount of control allocated (efficacy), and the cost associated with allocating a certain amount of control. Crucially, the value of an outcome is weighted by efficacy. This means that the theory suggests that less control will be allocated,

despite high potential rewards, when perceived efficacy is low. Several recent studies have provided initial evidence to support this proposal.

In an eye-tracking experiment Manohar and colleagues (2017) have demonstrated that individuals are faster to orient towards a target when they know that a potential reward will be based on their performance, rather than being delivered at random or always delivered. In line with these findings, it has been shown that in the Stroop task a cue indicating high efficacy improves behavioral performance (Dean Wolf, Cory, & Shenhav, 2018). At the neural level it has been shown that higher perceived efficacy increases the value of outcomes (Wang & Delgado, 2019) and strengthens their representation in the brain network including the striatum, prefrontal and parietal cortex (Wisniewski, Forstmann, & Brass, 2019). While there is initial evidence to support the role of efficacy in the allocation of cognitive effort, the neural dynamics through which efficacy estimates are formed and the way in which they influence the allocation of cognitive control remain unclear. Crucially, in most situations estimates of efficacy, as well as other estimates relevant for control allocation, need to be learned through experience (Bhandari, Badre, & Frank, 2017; Chiu & Egner, 2019).

There are several electrophysiological markers which can be used to inform our understanding of the mechanisms through which learned efficacy estimates influence control allocation. First, the event-related potential (ERP) called Contingent Negative Variation (CNV) is a reliable index of the amount of preparatory attentional control allocated prior to stimulus onset (Grent-'t-Jong & Woldorff, 2007; van Boxtel & Böcker, 2004). Studies focused on the impact of motivation on preparatory attention have shown that reward anticipation impacts the early part of the CNV component, while the late CNV is increased when anticipating both rewarding and difficult tasks (Schevernels, Krebs, Santens, Woldorff, & Boehler, 2014). Second, the P3a component locked to stimulus onset is known to be sensitive to expectations, and scales with the violations of expectations. Further, the following P3b component scales with the amount of learning in response to the stimulus, and is higher when more memory updating is needed due to larger violations of expectations (Polich, 2007; Sutton, Braren, Zubin, & John, 1965). These two components have been successfully used to investigate feedback-driven reward-based decision making (Fischer & Ullsperger, 2013).

In this study we assessed the described ERP components combined with model-based analyses to investigate how learned efficacy estimates influence preparatory allocation of attention (CNV) and reward expectations (P3a). Further, we examined how efficacy estimates

are updated based on trial-by-trial efficacy feedback (P3b). We recorded EEG while participants performed a Stroop task in which they received trial-by-trial feedback about whether they received a reward or not (reward vs. no reward feedback) and if that reward was performance-dependent or random (efficacy vs. no efficacy feedback). Crucially, the probability of participants obtaining a reward based on their performance of at random (efficacy) changed in a continuous fashion over the course of the experiment. There were periods in which participants were mostly receiving rewards based on their performance (high efficacy) and times in which they were mostly receiving random rewards (low efficacy). Participants were instructed to keep track of their efficacy and were probed to provide an estimate of how much the received rewards depended on their performance throughout the experiment.

This design allowed us to investigate three main questions about the influence of efficacy on cognitive control. First, based on the EVC theory, we hypothesized that higher efficacy will lead to more preparatory allocation of attention and better performance on the task, similarly to the effects of reward and task difficulty (i.e., larger CNV; Schevernels et al., 2014). Second, we tested the prediction from the EVC theory that efficacy modulates the value of outcomes. We investigated whether the knowledge of whether reward was obtained based on performance or not will modulate reward expectation (P3a). Further, we tested how efficacy estimates influence reward expectations. Third, we investigated how efficacy estimates influence learning from efficacy feedback (P3b).

Methods

Participants

We recruited forty-one participants with normal or corrected-to-normal vision from the Brown University subject pool. One participant was excluded due to technical problems related to the EEG recording. The final data set included 40 participants (24 females, 16 males; median age = 19). Participants gave informed consent and were compensated with course credit or a fixed payoff of $20. In addition, they received up to $5 as a performance-based bonus ($3.25 on average). The research protocol was approved by Brown University's Institutional Review Board.

Stimuli and task

Before doing the main experiment participants performed several practice phases of the Stroop task (approximately 15 minutes). First they practiced (80 trials) the mappings between four colors (red, green, blue, and yellow) and four keyboard keys (D, F, J, and K). Then they completed a short practice of the Stroop task (30 trials) in which they responded to the ink color of the color words or a string of letters "XXXXX" (neutral trials). The color words were either the same as the ink color (congruent trials), or different (incongruent trials). After each trial they received written feedback ("correct" or "incorrect"). Participants then completed 100 more of such trials during which the reaction time deadline needed for a correct response was iteratively modified in order to assure that the percentage of correct responses was approximately 80%. In the final practice phase (30 trials) participants were introduced to two types of feedback which they would see in the main experiment. After every response they would receive reward feedback ("$50c" or "$0c"). They would also receive efficacy feedback informing them whether the reward they received depended on their performance (a button graphic), or was fully random (a dice graphic). On half of the trials efficacy feedback was presented first and on half of the trials it was presented second.

The main task (Figure 1A) consisted out of 288 Stroop trials (72 neutral, 108 congruent, and 108 incongruent trials) and lasted for approximately 45 minutes. Each trial started with fixation cross (randomly jittered between 1000 and 1500ms) followed by a grey circle (1500ms). Then the Stroop stimulus was presented (until response or 1000ms) and participant responded by pressing one of the four keys. The Stroop stimulus was followed by a grey circle (1000ms) and a sequential presentation of the reward and efficacy feedback (1000ms each). Every 2-4 trials the feedback was followed by an efficacy ("How much do you think your rewards currently depend on your performance?") or reward probe ("How often do you think you are currently being rewarded?") to which participants responded on a VAS scale ranging from 0 to 100. The number and timing of the probes was randomized per subject resulting in the median number of efficacy probes of 45 (SD=3.38) and 47 (SD=2.81) for reward probes.

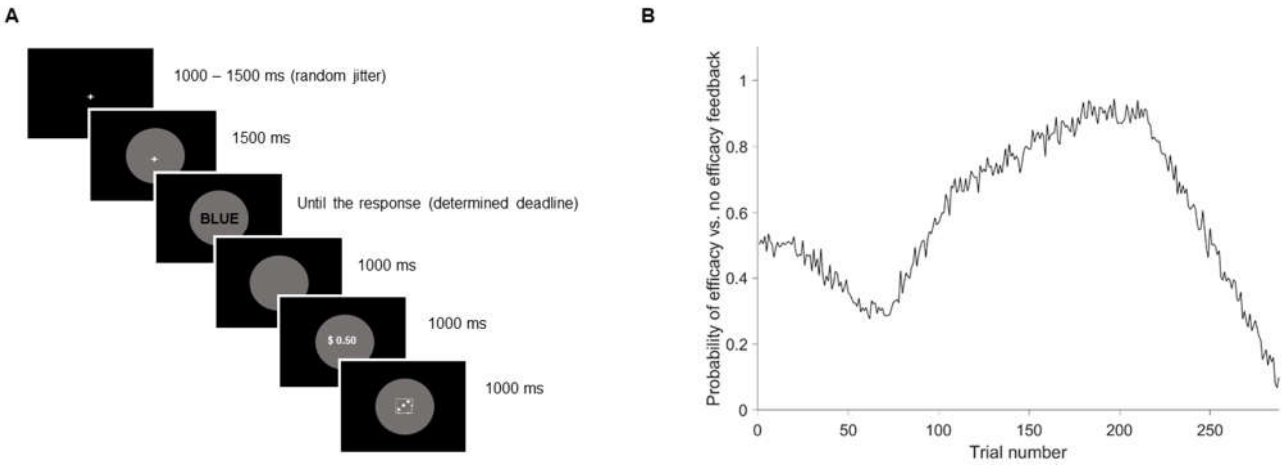


Figure 1. Task design and the continous efficay change. A. Each trial started with a fixation cross followed by a gray circle which signaled that the target stimulus will appear shortly. After the response, reward and efficacy feedback were presented sequentially. On half of the trials efficacy feedback was presented first, and on the other half reward feedback was presented first. **B.** Continuous change of the probability of efficacy feedback over the course of the experiment. The slowly changing probability function was generated and then used to sample the trial-by-trial binary feedback (efficay vs. no efficacy). Half of the participants experienced the probability change as depicted on the figure, and the other half of the participants experienced the inverse of the depicted probability change.

Efficacy (efficacy or no efficacy) on each trial was predetermined by generating a drifting probability over 288 trials of the experiment (Figure 1B). Each trial was assigned with a probability of being an efficacy trial. This probability ranged between 0.1 and 0.9 and drifted over the course of the experiment in predetermined fashion (see Figure 1B). Efficacy on each trial was then determined by sampling from a binomial distribution with the specified probability (e.g. on a given trial the probability of it being an efficacy trial is 0.8). In order to ensure that the efficacy and no efficacy trials did not differ in reward rate, reward feedback for the efficacy trials was randomly sampled from the moving window of the reward feedback of the previous 10 efficacy trials. At the beginning of the experiment a window with 8 rewards and 2 no rewards was used, and was then updated after every trial.

Prior to the experimental session participants filled in several questionnaires measuring negative mood and motivation. The results of these questionnaires are not reported here as this data will be used as a part of a larger study of inter-individual differences in motivated cognitive control. The experimental task was implemented in Psychophysics Toolbox (Brainard, 1997; Kleiner, Brainard, & Pelli, 2007; Pelli, 1997) for Matlab (MathWorks Inc.) and presented on a 23 inch screen with a 1920 x 1080 resolution. All of the stimuli were presented centrally while the participants were seated 80 cm away from the screen.

Psychophysiological recording and preprocessing

EEG data were recorded at a sampling rate of 500 Hz from 64 Ag/AgCl electrodes mounted in an electrode cap (ECI Inc.), referenced against Cz, using Brain Vision Recorder (Brain Products, München, Germany). Vertical and horizontal ocular activity was recorded from below both eyes (IO1, IO2) and the outer canthi (LO1, LO2), respectively. Impedances were kept below 10 k Ω.. Offline, data were processed using custom made Matlab scripts (Frömer, Maier, & Rahman, 2018) employing EEGlab functions (Delorme & Makeig, 2004). Data were re-referenced to average reference, ocular artifacts were corrected using brain electric source analyses (BESA; Ille, Berg, & Scherg, 2002) based on separately recorded prototypical eye movements. The cleaned continuous EEG was then high pass filtered at 40 Hz and segmented into epochs around cue onset (-200 to 1500 ms), stimulus onset, response time, efficacy feedback and reward feedback (each -100 to 800 ms). Baselines were corrected to the average of each 200 ms pre-stimulus interval. Segments containing artifacts, values exceeding ± 150 μV or gradients larger than 50 μV, were excluded from further analyses.

Single trial ERPs were then exported for further analyses in R (R Core Team, 2017). The early CNV was quantified between 700 and 1000ms, and the late CNV between 1000 and 1500 ms, and (cf. Schevernels et al., 2014). The CNV amplitudes were calculated as the average activity over 9 fronto-central electrodes (Fz, F1, F2, FCz, FC1, FC2, Cz, C1, and C2). For the remaining ERPs, time windows for quantification were determined based on grant mean peaks at typical electrodes. Peaks for the P3a components to feedback (positive, between 330 and 500 ms; reward feedback peak: 398 ms, efficacy feedback: 408 ms) were determined at FCz and calculated as the averaged activity over 9 electrodes (the same ones as for the CNV). Single trial amplitudes were averaged within a 100 ms time windows centered on the individual peaks. Feedback P3b was quantified between 350 and 500 ms for both reward and efficacy feedback and calculated as the averaged activity over electrodes (Pz, P1, P2, POz, PO1, PO2, CPz, CP1, and CP2).

Learning model and statistical analyses

Learning model

In order to obtain trial-by-trial estimates of efficacy and reward rate, we fitted a reinforcement learning model (Gläscher, Daw, Dayan, & O'Doherty, 2010; Sutton & Barto, 1998) to the continuous (range: 0-1) subjective estimates of efficacy and reward rate (cf. Eldar,

Hauser, Dayan, & Dolan, 2016; Nagase et al., 2018; Rutledge et al., 2014). The model assumed that the estimate of efficacy for the next trial (E_{t+1}) depended on the current efficacy feedback (E_t) and the prediction error (δ_t) weighted by a constant learning rate (α):

$$E_{t+1} = E_t + \alpha * \delta_t$$

Where $0 \leq \alpha \leq 1$ and the prediction error is calculated as the difference between the efficacy feedback on the current trial and the efficacy estimate on that trial: $\delta t = et - Et$. The model started with a fixed subjective estimate of 0.5 and updated it based on the binary feedback on each trial. For example, assuming a learning rate of 0.5, the model would update the initial estimate following efficacy feedback signaling "efficacy" (e_t = 1) to 0.75. If on the next trial efficacy feedback was ("no efficacy"; e_t = 0) the subjective estimate would drop to 0.6. Exactly the same model was used to calculate estimates of reward on each trial. The model was fitted separately to the subjective estimates of efficacy and reward rate with only the learning rate as a free parameter. The model was fit individually for each subject using least square fitting (in MATLAB using the fmincon function).

Statistical analyses

The efficacy and reward rate estimates obtained through fitting the learning model were then used to analyze the behavioral and EEG data. To this end we fitted Bayesian multilevel regressions to predict subjective estimates of efficacy and reward rates, reaction times, accuracy, as well as the amplitudes of the CNV, P3a, and P3b. The regression models were fitted in R with the *brms* package (Bürkner, 2016) which relies on the probabilistic programming language *Stan* (Carpenter et al., 2016) to implement Markov Chain Monte Carlo (MCMC) algorithms and estimate posterior distributions of model parameters. The analyses were done based on the recommendations for Bayesian multilevel modeling using *brms* (Bürkner, 2016, 2017; Nalborczyk & Bürkner, 2019). The fitted models included constant and varying effects (also known as fixed and random) with weakly informative priors for the intercept and the slopes of fixed effects and likelihoods based on the predicted variable (Ex-Gaussian for reaction times, Bernoulli for accuracy, and Gaussian for the subjective estimates and the EEG amplitudes). Four MCMC simulations ("chains"; 4,000 iterations; 2,000 warmup) were run to estimate the parameters of each of the fitted models. The convergence of the models was confirmed by examining trace plots, autocorrelation, and variance between chains (Gelman-Rubin statistic; Gelman & Rubin, 1992). The models with different fixed and varying effects were compared based on their fit to the data using the Bayesian R^2 (Gelman, Goodrich,

Gabry, & Ali, 2017), and their out-of-sample predictive performance using leave-one-out cross-validation (loo) and assessing the difference in the expected predictive accuracy of the fitted models (elpd_diff; Vehtari, Gelman, & Gabry, 2017). In this way we selected the best model and examined means and 95% highest density intervals (HDI) of the posterior distributions of that model. In order to test hypotheses we calculated evidence ratios (ERs) as the ratios between the percentage of posterior samples on each side of the zero. ERs can be interpreted as the posterior probability of one hypothesis (e.g. A>B) against its alternative (e.g., A<B).

We investigated whether the subjective estimates of efficacy and reward rate depended on efficacy and reward feedback by fitting two separate regression models for the two estimates. In each of the models the subjective estimate, on each trial on which it was given, was predicted by the feedback on the current trial and on the previous 10 trials. The feedback on each of the trials (efficacy vs. no efficacy or reward vs. no reward) was entered as a constant effect and the models also included the intercept as a varying effect. Further, we assessed the learning rates' estimates and the fit of the two learning models (*R2*).

In order to separately predict reaction times on correct trials and accuracies we fitted three different models. First, the *Congruency* model included the constant and varying effect of congruency known to influence performance in the Stroop task (contrast coded to obtain the difference between the neutral trials and incongruent (interference) and congruent (facilitation) trials), while controlling for the estimate of the reward rate. Further, we investigated if efficacy estimates can improve the prediction of behavioral data compared to this base model. To this end we fitted the *Congruency + Efficacy estimate* model and the *Congruency × Efficacy estimate* model which included an additive and an interactive effect of the efficacy estimate respectively. All of the models included an intercept both as a constant and a varying effect, and all of the main constant effects as varying effects (not including interactions to avoid overly complex varying structures). Prior distributions were Gaussian for both the intercept ($M = 650$; $SD = 200$) and the slopes ($M = 0$; $SD = 50$; cf. Chuderski & Smolen, 2016; Heathcote et al., 1991; MacLeod, 1991) for the reaction times, and for accuracy[2] (intercept: $M = 2.5$; $SD = 3$; slopes: $M = 0$; $SD = 3$). Further, we fitted two models to predict the reaction times and accuracies based on the late CNV. These models included the same likelihoods and prior

[2] Note that the prior distributions are set in log-odds. A prior distribution centered at 2.5 roughly corresponds to the odds of 1 in 12 of participants making an error.

distributions as the models described above and the late CNV and the intercept as both constant and varying effects.

We investigated the influence of efficacy estimates on the early and late CNV respectively (cf. Schevernels et al., 2014) by fitting the *Efficacy estimate* model with a constant effect of the efficacy estimate while controlling for the reward rate estimate. To investigate the possible interaction of the efficacy and reward rate estimates, we also added this term in the *Efficacy estimate* × *Reward estimate* model. Both models included the intercept as a constant effect, and the intercept, efficacy estimate, and reward rate estimate as varying effects. Prior distributions were Gaussian for both the intercept (M = -3; SD = 3) and the slopes (M = 0; SD = 3; cf. Schevernels et al., 2014).

In order to probe the influence of efficacy feedback and efficacy estimates on reward expectations, we fitted several models to predict the P3a component locked to the onset of the reward feedback. The *Reward feedback* model included the constant effect of reward feedback while controlling for the effect of the reward rate estimate. Given that participants received approximately 80% rewards across all trials, no reward feedback (20% of trials) should reliably induce higher P3a amplitudes compared to reward feedback. Further, we fitted the *Reward feedback* × *Efficacy feedback* model, which added the constant effect of the type of efficacy feedback to the previous model, to investigate how reward and efficacy feedback interact. Finally, we fitted the *Reward feedback* × *Efficacy estimate* × *Efficacy feedback* which added the interaction with the efficacy estimate to investigate the possible modulation of this interaction with efficacy estimates. All of the fitted models included intercept as a both a constant and varying effect, as well as constant main effects as varying effects. The priors were very wide Gaussian distributions for both the intercept (M = 5; SD = 5) and the slopes (M = 0; SD = 5; cf. Fischer & Ullsperger, 2013; Frömer, Nassar, et al., 2018).

For the analyses focused on the processing of the efficacy feedback, we first predicted the amplitude of the P3b with the *Efficacy feedback* model. This model included only the effect of efficacy feedback, while controlling for the reward rate estimates. We then expended this model to include the additive (*Efficacy feedback* + *Efficacy estimate*) and the interactive (*Efficacy feedback* × *Efficacy estimate*) effect of efficacy estimates in order to investigate whether efficacy estimates impact the memory updating based on the received feedback. All three models included intercept as a constant and a varying effect, and constant main effects in the varying effect structure. We selected wide Gaussian priors based on previous studies

(Fischer & Ullsperger, 2013; Frömer, Nassar, et al., 2018) for both the intercept (M = 0; SD = 5) and the slopes (M = 5; SD = 5).

Results

Behavioral results

Learning models for efficacy and reward rate

The regression models built to predict the subjective estimates of efficacy and reward rate showed that these estimates depended on the feedbacks on the current and previous trials. Both the estimate of efficacy (Figure 2A) and reward (Figure 2B) were impacted by previous efficacy and reward feedbacks respectively. As can be seen on the figures, this effect was particularly pronounced on the current and the previous two trials and decayed in magnitude when moving back in time. In addition, we have inspected the fitted values for the learning rates for each subject for the efficacy model (Figure 2C) and the reward rate model (Figure 2D). This analysis showed that participants had learning rates higher than zero (M = 0.14 for efficacy and M = 0.18 for reward rate), meaning that they were successfully learning based on the feedback. Finally, we have assessed the amount of variance in the subjective estimates explained by the model-based predicted values of the efficacy and reward estimates. This analysis revealed that both the efficacy and reward model successfully explained a significant amount of variance in the reported subjective estimates (for efficacy $R2$ = 0.34; for reward rate $R2$ =0.47). Taken together, these results suggest that participants were updating their subjective estimates of efficacy and reward rate based on the respective feedbacks. Moreover, the analyses show that a significant portion of the variance in subjective estimates could be predicted based on the model-based estimates. Thus, we have used the model-based estimates in further analyses.

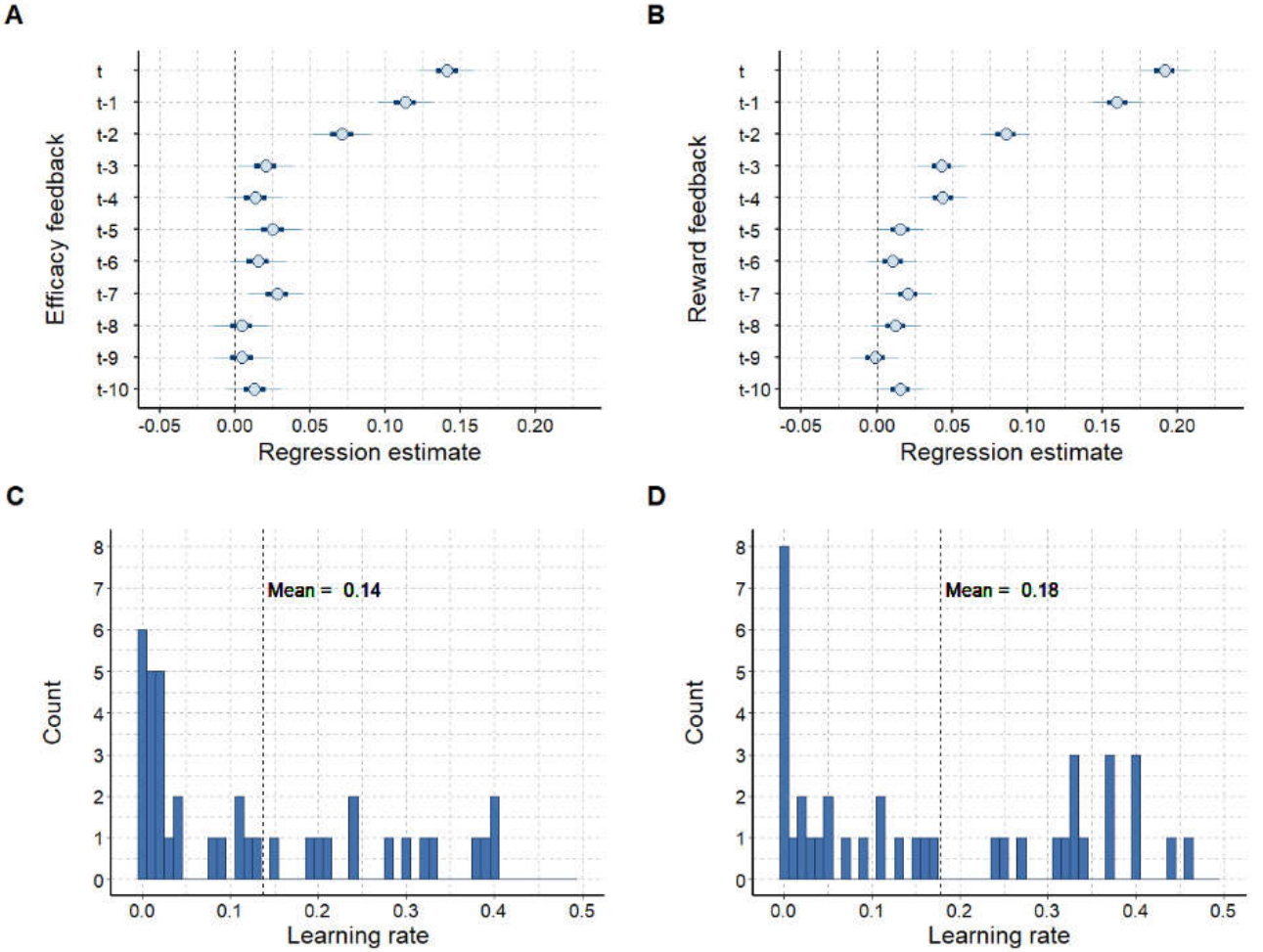


Figure 2. The fitted learning models for efficacy and reward rate. Pannels **A** and **B** display the estimates of the models constructed to predict the subjective estimates of efficacy (A) and reward rate (B) on trials on which participants rated their efficacy or reward rate (on a 0-1 scale). Type of feedback (efficacy vs. no efficacy or reward vs. no reward) on the current trial and 10 trials back were entered into a regression model. The estimates for each of these factors are plotted here. Pannels **C** and **D** show the learning rates obtained from the fitted learning models for each subject for the estimates of efficacy (C) and reward rate (D). Average learning rate across the whole sample is displayed with a dashed line.

Behavioral data

In predicting the correct reaction times the *Congruency* and the *Congruency + Efficacy* models were practically the same in the amount of variance explained and their out of sample predictive accuracy (Table 1). This means that adding the efficacy estimate to the congruency model did not lead to a decrement in the predictive accuracy of the model. Given this, here we interpret the estimates from the *Congruency + Efficacy estimate* model in order to assess the effects of the efficacy estimates. First, the expected congruency effects were observed (Figure 3A) in that congruent trials were faster than the neutral trial (M = 10.02; 95% HDI [4.25, 15.78]; ER = 499.00), and that the incongruent trials were slower than the neutral ones (M = 60.80; 95% HDI [50.94, 70.53]; ER > 8000 [the whole posterior distribution was above zero thus the

ER is larger than the total number of samples in the posterior distribution]). Higher efficacy (Figure 3B) and reward rate (Figure 3C) estimates led to faster reaction times, and more so for reward (*M* = -27.25; 95% HDI [-42.78, -11.64]; ER = 332.33) than for efficacy (*M* = -8.82; 95% HDI [-20.32, 2.66]; ER =8.71).

In predicting accuracy all of the three fitted models were similar in the amount of variance explained and their out-of-sample predictive accuracy (Table 1). Here we present the results of the *Congruency + Efficacy estimate* model because this model performed similarly well to the other two and because the same model was interpreted for the reaction times, thus allowing us to investigate the potential speed-accuracy tradeoffs. The congruency effects (Figure 3D) followed the results of the reaction times showing that participants were less likely to be correct on neutral compared to congruent trials (*M* = 0.89[3]; 95% HDI [0.76, 1.06]; ER = 8.80), and less likely to be correct on incongruent relative to neutral trials (*M* = 0.62; 95% HDI [0.52, 0.72]; ER > 8000). Higher efficacy (Figure 3E) did not reliably impact accuracy (*M* = 1.11; 95% HDI [0.83, 1.47]; ER = 3.35), while higher reward rate estimates led to a decrease in accuracy (*M* = 0.77; 95% HDI [0.54, 1.07]; ER = 15.13; Figure 3F), suggesting a speed-accuracy tradeoff when it comes to the impact of reward rate on behavioral performance (i.e., faster reaction times, but reduced accuracy when the reward estimates are high). In sum, the behavioral performance on the task shows the expected congruency effects. Further, the results reveal that high efficacy estimates are related to faster reaction times, while accuracy remains unchanged. Finally, higher reward rate estimates reduce the reaction times more, but at the cost of accuracy.

[3] The results of the model were transformed from log-odds to odds in the text in order to enhance readability.

Table 1

Leave-one-out cross validation information criterion (LOO), the difference between models in their expected log predictive accuracy (ELPD_diff), and Bayesian R^2 for each model. The models are sorted in descending order starting the best performing model.

Model	*LOO (SE)*	*ELPD_diff (SE)*	*Bayesian R^2 (SE)*
Reaction times			
Congruency	106630.10 (138.62)	0.00 (0.00)	0.25 (0.01)
Congruency + Efficacy estimate	106630.70 (138.53)	-0.31 (1.50)	0.25 (0.01)
Congruency × Efficacy estimate	106634.70 (138.45)	-2.31 (2.19)	0.25 (0.01)
Accuracy			
Congruency	9876.576 (117.23)	0.00 (0.00)	0.07 (0.01)
Congruency × Efficacy estimate	9877.988 (117.33)	-0.70 (2.44)	0.07 (0.01)
Congruency + Efficacy estimate	9878.766 (117.33)	-1.09 (1.25)	0.07 (0.01)
Early CNV			
Efficacy estimate	57121.95 (178.98)	0.00 (0.00)	0.02 (0.00)
Efficacy estimate × Reward rate estimate	57124.45 (179.26)	-1.24 (2.09)	0.02 (0.00)
Late CNV			
Efficacy estimate	55247.42 (174.44)	0.00 (0.00)	0.02 (0.00)
Efficacy estimate × Reward rate estimate	55251.14 (174.64)	-1.86 (1.57)	0.02 (0.00)
P3a to reward feedback			
Reward feedback × Efficacy feedback	48130.80 (155.05)	0.00 (0.00)	0.1282 (0.01)
Reward feedback	48147.66 (155.18)	-8.43 (2.84)	0.1493 (0.01)
Reward feedback × Efficacy estimate × Efficacy feedback	48267.41 (154.23)	-68.31 (12.51)	0.1523 (0.01)

P3b to efficacy feedback			
Efficacy feedback × Efficacy estimate	50905.00 (155.28)	0.00 (0.00)	0.1507 (0.01)
Efficacy feedback + Efficacy estimate	50912.15 (155.21)	-3.57 (4.52)	0.1524 (0.01)
Efficacy feedback	50915.34 (154.86)	-5.17 (5.43)	0.1543 (0.01)

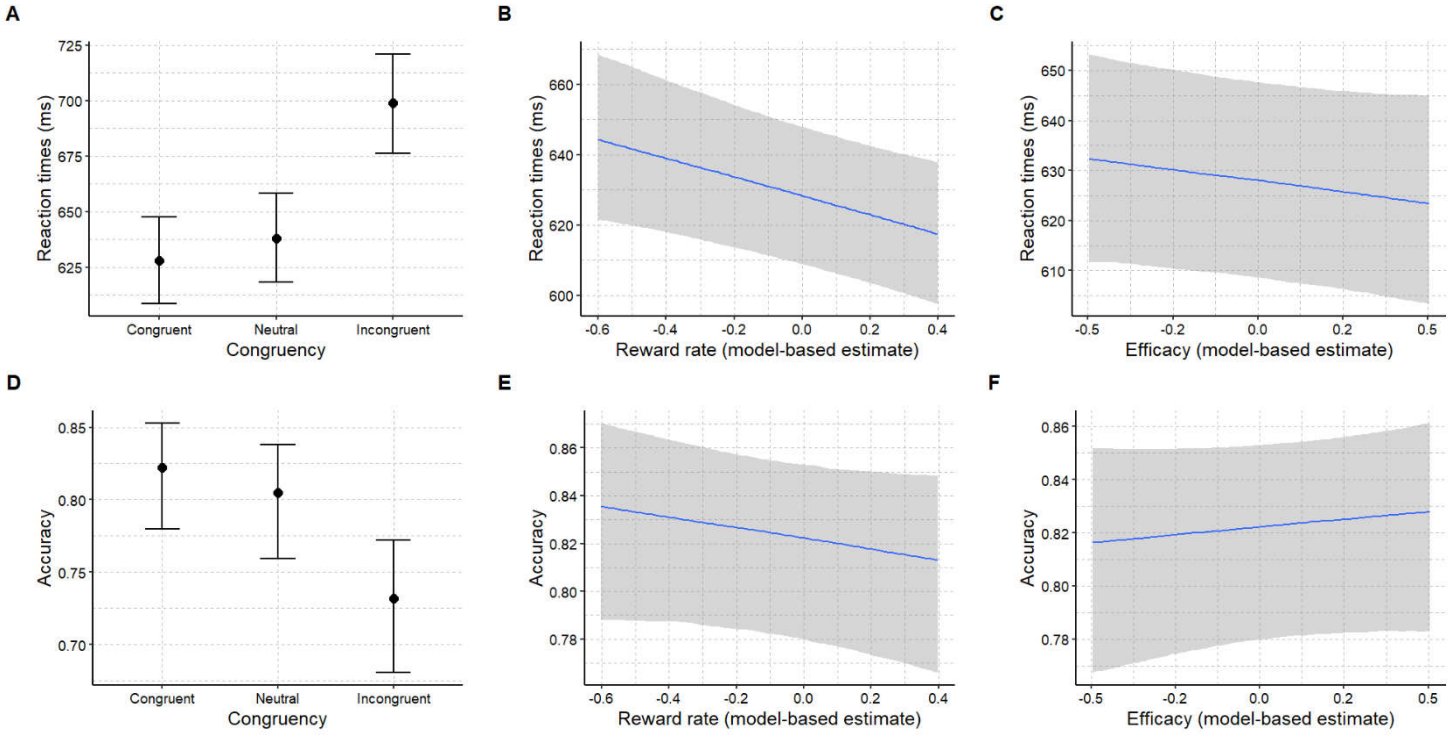


Figure 3. Behavioral results. A. Congruency effects on reaction times. **B.** Higher estimated reward rate predicts shorter reaction times. **C.** Higher estimated efficacy predicts shorter reaction times. **D.** Accuracy over different levels of congruency. **E.** Higher estimates of reward rate predict lower accuracy. **F.** Efficacy estimates have no effect on accuracy.

EEG results

Preparatory attention (CNV)

In predicting the early part of the CNV component (Figure 4A), the *Efficacy* model outperformed the *Efficacy* × *Reward* model (Table 1). However, the parameters of the winning model revealed no reliable effects of either the efficacy estimate (M = 0.09; 95% HDI [-0.30, 0.48]; ER = 1.92), nor the reward rate estimate (M = 0.06; 95% HDI [-0.45, 0.56]; ER = 1.34).

In predicting the late part of the CNV (Figure 4A) the *Efficacy* model was also the best one (Table 1). The parameters of this model showed no reliable effect of the reward rate estimate on the CNV (M = -0.19; 95% HDI [-0.65, 0.29]; ER = 3; Figure 5B), but revealed that

the CNV increased (this is a negative component, so the increase is an increase in negativity) with the higher estimates of efficacy (M = -0.49; 95% HDI [-0.88, -0.12]; ER = 70.43; Figure 5A and Figure 4B).

The late part of the CNV component also successfully predicted the reaction times and accuracies on the task. Increased late CNV predicted shorter reaction times (M = 2.91; 95% HDI [2.19, 3.64]; ER = Inf; Figure 5C) and improved accuracy (M = -0.02; 95% HDI [-0.03, -0.00]; ER = 75.92; Figure 5D). In conclusion, while there were no effects of the efficacy and reward estimates on the early CNV, increases in the estimated efficacy were robustly related to the modulations in the late CNV. Further, the increase in the late part of the CNV predicted improved behavioral performance on the task.

Reward feedback processing (P3a)

In predicting the P3a locked to reward feedback (Figure 4C) the *Reward feedback* × *Efficacy feedback* model clearly outperformed the other models in predictive accuracy (Table 1). Parameters of the model revealed the expected effect of the type of feedback: no reward feedback reliably led to higher P3a amplitudes relative to reward feedback (M = 1.22; 95% HDI [0.95, 1.49]; ER > 8000; Figure 5E and Figure 4D). Further, this effect was modulated by the preceding efficacy feedback (Figure 5F and Figure 4E & F). The difference between reward and no reward feedback was much larger when participants first saw the feedback signaling a high level of efficacy on the current trial, relative to the conditions in which they experienced no efficacy (M = 2.19; 95% HDI [1.52, 2.87]; ER > 8000), or they received the efficacy feedback only after the reward feedback (M = 2.80; 95% HDI [2.18, 3.42]; ER > 8000). These results suggest that the participants were more surprised by the no reward compared to the reward feedback, which is to be expected given that they received rewards on approximately 80% of trials across the experiment. Crucially, this effect was reliably magnified when participants knew that they had efficacy on the current trial. This result suggests that their expectations about rewards increased when they knew that the rewards depended on their feedback. Finally, the model which included the model-based efficacy estimates had much lower predictive accuracy, suggesting that the efficacy estimates did not significantly impact this P3a effect.

Learning from the efficacy feedback (*P3b*)

The amplitudes of the P3b locked to the efficacy feedback (Figure 4G) were best predicted by the *Efficacy feedback* × *Efficacy estimate* model (Table 1). This model suggested that the efficacy feedback relative to no efficacy feedback increased the amplitudes of the P3b (M = 0.92; 95% HDI [0.55, 1.28]; ER > 8000; Figure 5G and Figure 4H). Further, this effect was amplified by the model-based efficacy estimates so that the effect was much larger when the efficacy estimates were low compared to high (M = 2.01; 95% HDI [0.78, 3.26]; ER = 199; Figure 5H and Figure 4I). This result suggest that participants were learning more from the efficacy compared to no efficacy feedback, but especially so when they estimated that their efficacy levels are low. Note that the efficacy feedback had an effect on the EEG amplitudes prior to the analyzed P3b component (Figure 4G). We confirmed that the P3b effects described here are not due to this earlier activity by conducting an additional analysis in which we statistically controlled for this earlier activity (Supplementary materials).

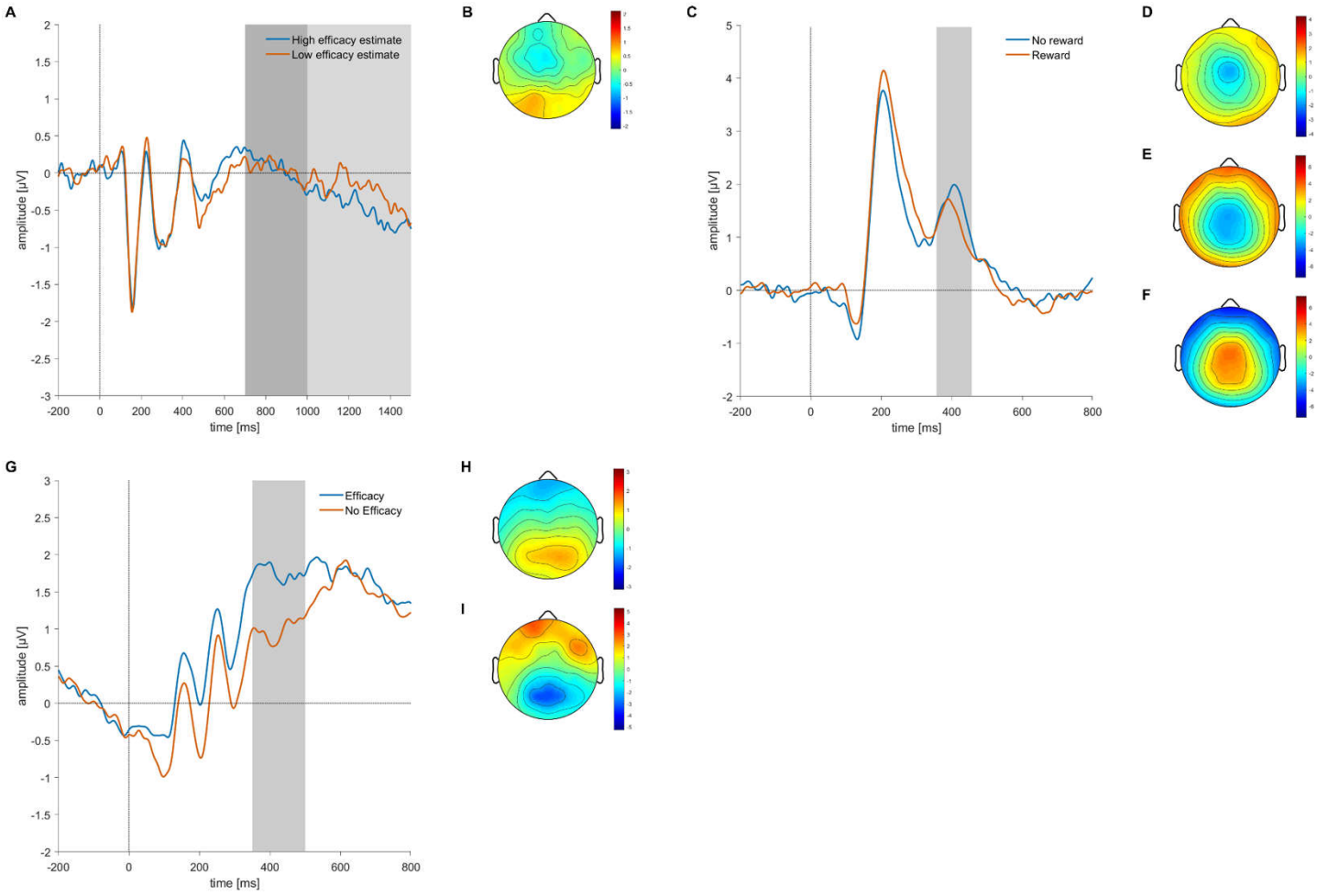


Figure 4. ERP waveforms and model-based topographies. A. The early (dark gray) and late (light gray) CNV locked to the onset of the preparatory cue plotted from FCz and divided into high and low efficacy (median split) for plotting purposes. **B.** The estimates of the effect of efficacy on the late CNV from the winning model. **C.** The P3a waveform locked at the onset of the reward feedback at Cz for the reward and no reward condition. **D.** The estimates of the difference between the reward and no reward condition from the winning model. **E.** The estimates of the difference in the reward effect in the efficacy and no efficacy condition. **F.** The estimates of the difference in the reward effect between the unknown efficacy and efficacy condition. **G.** The P3b waveform locked to the onset of the efficacy feedback plotted from Pz for the efficacy and no efficacy feedback. **H.** The estimates of the difference between the efficacy and no efficacy feedback. **I.** The estimates of the interaction between the efficacy effect the efficacy estimates.

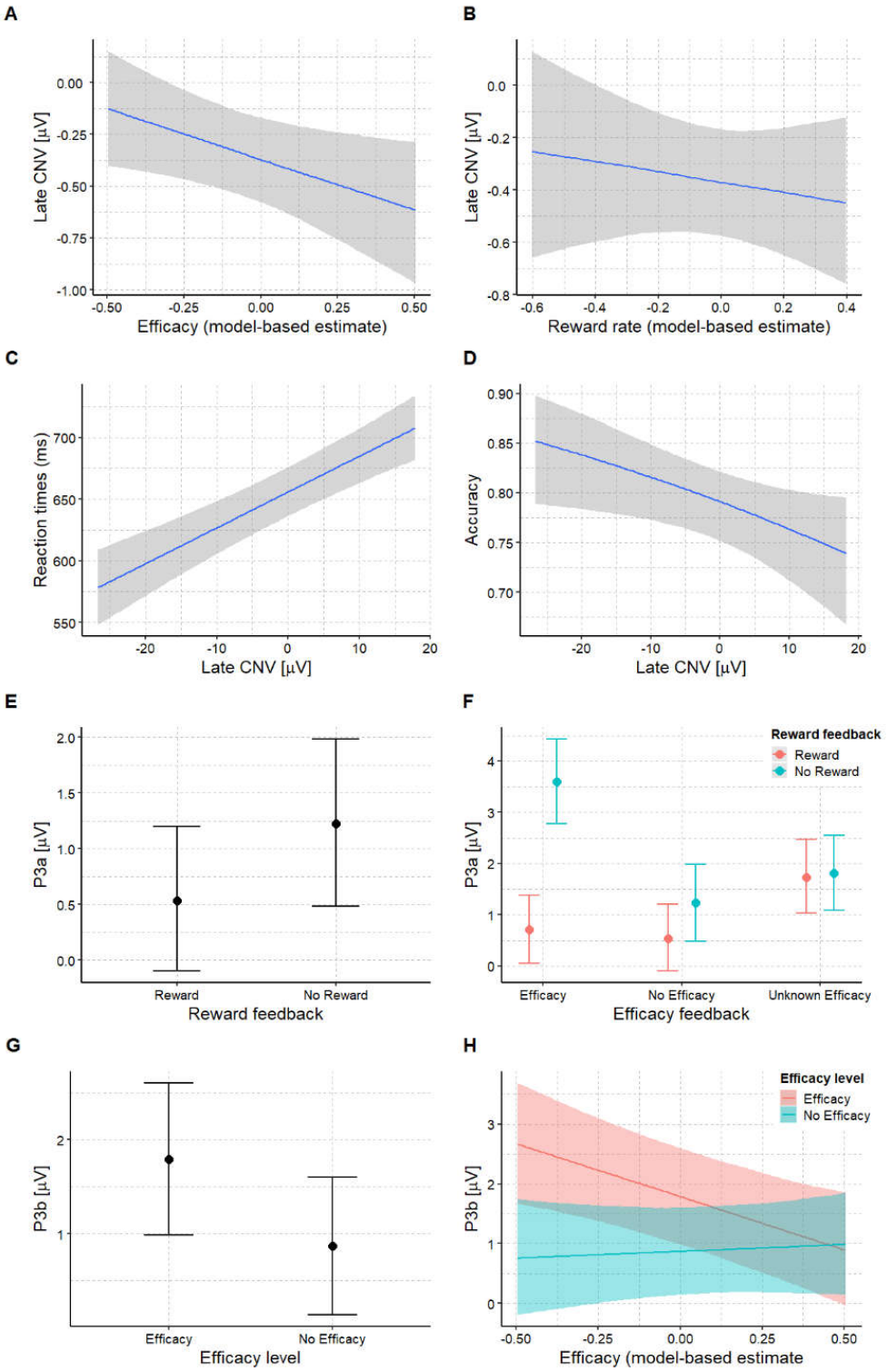


Figure 5. EEG results. A. Higher estimated efficacy increases the late CNV amplitudes. **B.** Reward rate estimates do not have a reliable effect on the late CNV. **C.** Higher late CNV amplitudes predict shorter reaction times. **D.** Higher late CNV amplitudes predict improved accuracy. **E.** No reward feedback increases the P3a amplitude relative to reward feedback. **F.** The difference in the P3a for the no reward vs. reward feedback is most pronounced when it was proceeded by the efficacy feedback. **G.** P3b amplitudes locked to the efficacy feedback are higher for efficacy than for no efficacy feedback. **H.** The difference between the efficacy and no efficacy feedback in the P3b is most pronounced when the efficacy estimates are low.

Discussion

In this study we examined how learned efficacy influences the neural dynamics of cognitive control. We investigated the influence of efficacy estimates, as they changed in continuous fashion, on the allocation of attention prior to the target onset (CNV), on the processing of reward feedback (P3a), and on memory updating based on efficacy feedback (P3b). Crucially, we were able to equate the reward rate on efficacy and no efficacy trials, thus avoiding the contamination of the effects by the potent effect of reward. The results revealed that our participants were able to track the changing efficacy levels and that their estimates of efficacy influenced all of these processing stages which are crucial in cognitive control processes. When their estimates of efficacy were high, our participants were faster to respond to the targets, allocated more attention prior to the target onset, and learned more from the efficacy compared to no efficacy feedback. Further, participants had higher prediction errors upon receiving no reward feedback when they knew that they had efficacy on the current trial (i.e., the reward depended on their performance). These results provide insight into the neural dynamics through which people learn about efficacy and the mechanisms through which efficacy influences how they allocate cognitive control.

Most of the studies on the influence of motivation on cognitive control present participants with binary cues or stimulus-reward mappings such as reward vs. no reward (for reviews see: Botvinick & Braver, 2015; Krebs & Woldorff, 2017). However, despite the importance of learning for cognitive control processes, few studies investigated how cognitive control depends on learned and changing levels of motivation. Here we demonstrated that the learning of dynamically changing efficacy (i.e., action-outcome contingencies) can be successfully captured by a learning model similar to those used in reinforcement learning tasks (Gläscher, Daw, Dayan, & O'Doherty, 2010; Sutton & Barto, 1998). Moreover, we then showed that the trial-by-trial model-based estimates can be used to predict both behavior and neural dynamics in a cognitive control task. This result offers support for the idea that learning is crucial in the allocation of cognitive control (Bhandari et al., 2017; Chiu & Egner, 2019). Further, it supports the importance of the learned efficacy estimates in control allocation (Shenhav et al., 2013).

Cognitive control can be improved if participants are incentivized prior to the onset of the task. In such cueing paradigms higher rewards improve performance by enhancing the activity in frontoparietal regions implicated in attentional control as measured by the blood-

oxygen-level-dependent (BOLD) activity (Krebs, Boehler, Roberts, Song, & Woldorff, 2012; Pessoa & Engelmann, 2010). Similar effects have also been observed when individuals are told that the upcoming trial will be difficult (Vassena et al., 2014). In an EEG study, reward cues were found to increase both the early and the late part of the CNV component, while the cues signaling task difficulty increased only the late part of this component known to scale with preparatory allocation of attention (Schevernels et al., 2014). In the present study we focused our analyses on this component and showed that it is also sensitive to the third crucial component of motivation – efficacy. Interestingly, high estimates of efficacy increased only the late part of the CNV, paralleling the effects of difficulty cues (Schevernels et al., 2014). Further, while these previous studies have employed binary cues signaling reward or difficulty, here we show that the preparatory allocation of attention scales with continuous changes in motivation. Additionally, we show that the magnitude of the late CNV is predictive of improved behavioral performance in the task, which is in line with the previous findings (Schevernels et al., 2014). In sum these results demonstrate the importance of efficacy in allocation of attention, and show that the effect of efficacy mirrors the previously reported effects of other components of motivation such as reward prospect and task difficulty. Further, these results suggest that the allocation of attention scales with continuous changes in motivation. These findings reveal the importance of efficacy in the allocation of cognitive effort.

Many models of cognitive control point to the importance of reward expectations in the allocation of cognitive control (Brown & Alexander, 2017; Holroyd & McClure, 2015; Shenhav et al., 2013; Verguts et al., 2015). Further, the EVC theory (Shenhav et al., 2013) proposes that the expected reward interacts with efficacy in the decision-making process about how much control to allocate. In order to test this prediction, we investigated how expectations about rewards (P3a; Polich, 2007; Ullsperger, Fischer, Nigbur, & Endrass, 2014) are influenced by the knowledge about whether the current reward was obtained based on performance or at random (i.e., efficacy) and on the learned estimates. Our results showed that the estimates of efficacy did not influence reward feedback processing. However, the P3a following reward feedback was modulated by the preceding efficacy feedback. Specifically, P3a amplitudes were higher, indicative of a larger prediction error, when participants were aware that the rewards depended on their performance, compared to when they knew that they were random. This result suggests that the knowledge about action-outcome contingencies does modulate reward expectations, but only in the current trial, and not by the more global estimates of efficacy. Specifically, this means that the participants were expecting rewards more when they knew they

were based on their performance, and that they were more surprised in this condition when they didn't receive them.

Both learning and motivation are thought to be crucial in the allocation of cognitive control (Bhandari et al., 2017; Botvinick & Braver, 2015; Chiu & Egner, 2019). Learning of estimates of costs and benefits associated with exerting cognitive effort, is crucial for the allocation of cognitive control (Shenhav et al., 2013). However, the neural mechanisms through which such estimates are learned remain unclear. Here we focused on the learning of efficacy estimates based on the trial-by-trial efficacy feedback. Our results reveal that the updating of the efficacy estimates (P3b; Polich, 2007) based on efficacy feedback scales with the estimated efficacy. When estimated efficacy was high, the amount of updating was similar for both the efficacy and no efficacy feedback. However, when the estimated efficacy was medium or low, the participants learned more from the efficacy compared to no efficacy feedback. This means that in the parts of the experiment when participants estimated their efficacy to be low, they updated these estimates much more when they experienced feedback telling them that they had efficacy on the current trial. This result reveals the neural mechanism related to the dynamical updating of efficacy estimates and corroborates the importance of learning the estimates of components of motivation in the allocation of cognitive control.

There are several limitations to the current study. First, the study was designed to investigate how individuals learn from efficacy feedback, and thus included the explicit trial-by-trial efficacy feedback. However, in most real-life situations people are able to estimate efficacy without explicit feedback. Further studies should focus on the question of how individuals can update estimates of their efficacy without explicit feedback. Our behavioral effects displayed an overall performance benefit related to the efficacy estimates, but did not reveal an interaction between this improvement and congruency. This result suggests that the improvement was similar across both congruent and incongruent trials, thus it remains unclear whether efficacy interacts with cognitive control at the behavioral level. Further studies should address this question explicitly by using different conflict tasks and by focusing on other measures of cognitive control, such as the congruency sequence effect (Gratton, Coles, & Donchin, 1992). Due to the nature of our design, which included cues and two types of feedback, the current study was not able to investigate the potential modulation in such sequence effects. Finally, in order to fully address the question of how the estimates of efficacy and reward interact, future studies should investigate this interaction in the setting in which

both of these factors dynamically change over time. In this study we have focused only on the changes in efficacy, while keeping the reward rate the same across the experiment.

In conclusion, this study has demonstrated that the learned estimates of efficacy impact the allocation of cognitive control. They do so by influencing the allocation of attention prior to task onset. Further, having efficacy increases the expectation of obtaining rewards, and efficacy estimates are dynamically updated based on the efficacy feedback. This set of results supports the proposal that efficacy estimates are crucial in allocation of cognitive effort (Shenhav et al., 2013) and corroborates the importance of learning in the allocation of cognitive control (Bhandari et al., 2017; Chiu & Egner, 2019). Further, these results reveal the neural dynamics which underlie the influence of efficacy on cognitive control. In this way this study demonstrates that efficacy plays a crucial role in the allocation of cognitive effort, in similar way that it does in goal-directed behavior more broadly (Moscarello & Hartley, 2017).

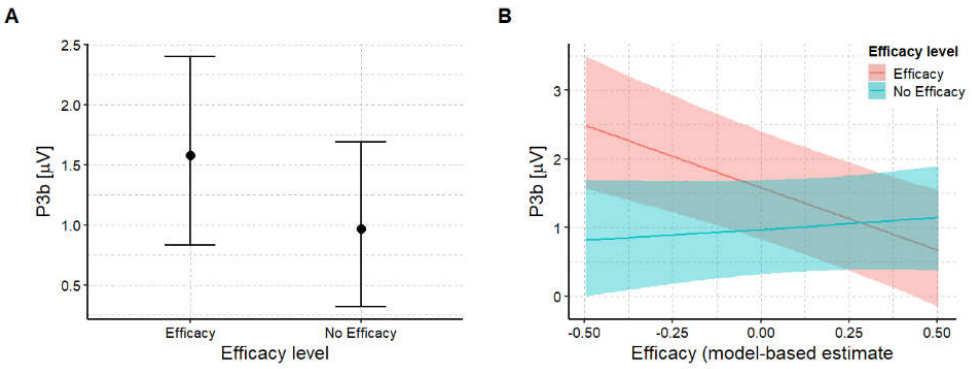
Supplementary materials

The inspection of the ERP plot of the P3b effect (Figure 4G) showed that the effect of efficacy feedback is also pronounced before the analyzed P3b component. In order to ensure that our results are not due to this prior activity, we fitted all of the three models described in the results section and added the activity from 200 to 300ms post efficacy feedback onset. This analysis showed that the P3b amplitudes locked to the efficacy feedback was still best predicted by the *Efficacy feedback* × *Efficacy estimate* model (Supplementary Table 1). The parameters of this model fully replicated all of the effects reported in the results section and thus we concluded that the reported results are not dependent on the amplitudes preceding the P3b. The efficacy feedback relative to no efficacy feedback increased the amplitudes of the P3b (M = 0.62; 95% HDI [0.30, 0.93]; ER = 999; Supplementary Figure 1A). This effect was enhanced when the model-based efficacy estimates were low compared to high (M = 2.14; 95% HDI [1.17, 3.16]; ER > 8000; Supplementary Figure 1B). This result suggest that participants were learning more from the efficacy compared to no efficacy feedback, but especially so when they estimated that their efficacy levels are low. Note that the supplementary analysis showed that the effect hold when controlling for the activity before the P3b.

Supplementary Table 1

Leave-one-out cross validation information criterion (LOO), the difference between models in their expected log predictive accuracy (ELPD), and Bayesian R^2 for each model. The models are sorted in descending order starting the best performing model.

Model	*LOO (SE)*	*ELPD (SE)*	*Bayesian R^2 (SE)*
P3b to efficacy feedback (controlled for the previous activity)			
Efficacy feedback × Efficacy estimate	46963.43 (155.56)	0.00 (0.00)	0.4839 (0.01)
Efficacy feedback + Efficacy estimate	46973.92 (156.14)	-5.24 (4.88)	0.4851 (0.01)
Efficacy feedback	46980.96 (155.72)	-8.76 (6.02)	0.4863 (0.01)



Supplementary Figure 1. EEG results of the control analysis. A. P3b amplitudes locked to the efficacy feedback are increased for efficacy compared to no efficacy feedback. **B.** The difference between the efficacy and no efficacy feedback in the P3b is most pronounced when the efficacy estimates are low.

Data

Data and analyses code is available at:
https://osf.io/c9j2g/?view_only=78e7e9b78cee487c8cb4f8b56afb538b

Acknowledgements

This work was supported by the Special Research Fund (BOF) of Ghent University [grant #01D02415] and the Research Foundation Flanders (FWO) travel grant [grant #V432718N] both awarded to IG. The funding sources had no involvement in the study design; collection, analysis, and interpretation of data; writing of the report; and decision to submit the article for publication.

We would like to thank Natalie Knowles and Hattie Xu for their help with data collection, Carolyn Dean Wolf for her help with programming the task, and Harrison Ritz for his help with constructing the efficacy drifts.

References

Atkinson, J. W. (1957). Motivational determinants of risk taking behavior. *Psychological Review*, *64*(6), 359–372. https://doi.org/10.1037/h0043445

Balleine, B. W., & O'Doherty, J. P. (2010). Human and rodent homologies in action control: Corticostriatal determinants of goal-directed and habitual action. *Neuropsychopharmacology*, *35*(1), 48–69. https://doi.org/10.1038/npp.2009.131

Bhandari, A., Badre, D., & Frank, M. J. (2017). Learning cognitive control. *The Wiley Handbook of Cognitive Control*, 376–391. https://doi.org/10.1002/9781118920497.ch21

Botvinick, M. M., & Braver, T. S. (2015). Motivation and Cognitive Control: From Behavior to Neural Mechanism. *Annual Review of Psychology*, *66*(September 2014), 83–113. https://doi.org/10.1146/annurev-psych-010814-015044

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

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, *10*(4), 433–436.

Brehm, J. W., & Self, E. A. (1989). The intensity of motivation. *Annual Review of Psychology*, *1*(40), 109–131.

Brown, J. W., & Alexander, W. H. (2017). Foraging Value, Risk Avoidance, and Multiple Control Signals: How the Anterior Cingulate Cortex Controls Value-based Decision-making. *Journal of Cognitive Neuroscience*, *29*(10), 1656–1673. https://doi.org/10.1162/jocn_a_01140

Bürkner, P.-C. (2016). brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software*, *80*(1), 1–28.

Bürkner, P.-C. (2017). Advanced Bayesian Multilevel Modeling with the R Package brms. *ArXiv:1705.11123*.

Carpenter, B., Gelman, A., Hoffman, M., Lee, D., Goodrich, B., Betancourt, M., … Riddell, A. (2016). Stan: A probabilistic programming language. *Journal of Statistical Software*, *2*(20), 1–

37.

Chiu, Y. C., & Egner, T. (2019). Cortical and subcortical contributions to context-control learning. *Neuroscience and Biobehavioral Reviews*, *99*(December 2018), 33–41. https://doi.org/10.1016/j.neubiorev.2019.01.019

Chuderski, A., & Smolen, T. (2016). An integrated utility-based model of conflict evaluation and resolution in the stroop task. *Psychological Review*, *123*(3), 255–290. https://doi.org/10.1037/a0039979

Dean Wolf, C. K., Cory, E. V., & Shenhav, A. (2018). The influence of expected reward and efficacy on cognitive effort allocation. In *25th Annual Conference of the Cognitive Neuroscience Society*. Boston, MA.

Delorme, A., & Makeig, S. (2004). EEGLAB: an open sorce toolbox for analysis of single-trail EEG dynamics including independent component anlaysis. *Journal of Neuroscience Methods*, *134*, 9–21. https://doi.org/10.1016/j.jneumeth.2003.10.009

Dickinson, A. (1985). Actions and habits: the development of behavioural autonomy. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *308*(1135), 67–78.

Eldar, E., Hauser, T. U., Dayan, P., & Dolan, R. J. (2016). Striatal structure and function predict individual biases in learning to avoid pain. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(17), 4812–4817. https://doi.org/10.1073/pnas.1519829113

Fischer, A. G., & Ullsperger, M. (2013). Real and fictive outcomes are processed differently but converge on a common adaptive mechanism. *Neuron*, *79*(6), 1243–1255. https://doi.org/10.1016/j.neuron.2013.07.006

Friedman, N. P., & Miyake, A. (2017). Unity and diversity of executive functions: Individual differences as a window on cognitive structure. *Cortex*, *86*, 186–204. https://doi.org/10.1016/j.cortex.2016.04.023

Frömer, R., Maier, M., & Rahman, R. A. (2018). Group-level EEG-processing pipeline for flexible single trial-based analyses including linear mixed models. *Frontiers in Neuroscience*, *12*(48). https://doi.org/10.3389/fnins.2018.00048

Frömer, R., Nassar, M. R., Stuermer, B., Sommer, W., & Yeung, N. (2018). I knew that! Confidence in outcome prediction and its impact on feedback processing and learning. *BioRxiv*. https://doi.org/10.1101/442822

Gelman, A., Goodrich, B., Gabry, J., & Ali, I. (2017). R-squared for Bayesian regression models. *Unpublished via Http://Www. Stat. Columbia. Edu/~ Gelman/Research/Unpublished.* Retrieved from http://www.stat.columbia.edu/~gelman/research/unpublished/bayes_R2.pdf

Gelman, A., & Rubin, D. B. (1992). Inference from Iterative Simulation Using Multiple Sequences. *Statistical Science*, *7*(4), 457–472. https://doi.org/10.1214/ss/1177011136

Gläscher, J., Daw, N., Dayan, P., & O'Doherty, J. P. (2010). States versus rewards: Dissociable neural prediction error signals underlying model-based and model-free reinforcement learning. *Neuron*, *66*(4), 585–595. https://doi.org/10.1016/j.neuron.2010.04.016

Gratton, G., Coles, M. G. H., & Donchin, E. (1992). Optimizing the Use of Information. *Journal of Experimental Psychology: General*, *121*(4), 480–506.

Grent-'t-Jong, T., & Woldorff, M. G. (2007). Timing and sequence of brain activity in top-down control of visual-spatial attention. *PLoS Biology*, *5*(1), 0114–0126. https://doi.org/10.1371/journal.pbio.0050012

Heathcote, A., Popiel, S. J., & Mewhort, D. J. K. (1991). Analysis of response time distributions: an example using the stroop task. *Psychological Bulletin*, *109*(2), 340–347. https://doi.org/10.1037/0033-2909.109.2.340

Holroyd, C. B., & McClure, S. M. (2015). Hierarchical control over effortful behavior by rodent medial frontal cortex: A computational model. *Psychological Review*, *122*(1), 54–83. https://doi.org/10.1037/a0038339

Ille, N., Berg, P., & Scherg, M. (2002). Artifact Correction of the Ongoing EEG Using Spatial Filters Based on Artifact and Brain Signal Topographies. *Journal of Clinical Neurophysiology*, *19*(2), 113–124. https://doi.org/10.1097/00004691-200203000-00002

Kleiner, M., Brainard, D., & Pelli, D. (2007). What's new in Psychtoolbox-3. In *Perception 36 ECVP Abstract Supplement*.

Kool, W., McGuire, J. T., Rosen, Z. B., & Botvinick, M. M. (2010). Decision making and the avoidance of cognitive demand. *Journal of Experimental Psychology: General*, *139*(4), 665–682. https://doi.org/10.1037/a0020198

Krebs, R M, & Woldorff, M. G. (2017). Cognitive control and reward. In T. Egner (Ed.), *Wiley Handbook of Cognitive Control* (pp. 422–439). Wiley-Blackwell.

Krebs, Ruth M., Boehler, C. N., Roberts, K. C., Song, A. W., & Woldorff, M. G. (2012). The involvement of the dopaminergic midbrain and cortico-striatal-thalamic circuits in the integration of reward prospect and attentional task demands. *Cerebral Cortex*, *22*(3), 607–615. https://doi.org/10.1093/cercor/bhr134

Leotti, L. A., Iyengar, S. S., & Ochsner, K. N. (2010). Born to choose: The origins and value of the need for control. *Trends in Cognitive Sciences*, *14*(10), 457–463. https://doi.org/10.1016/j.tics.2010.08.001

MacLeod, C. M. (1991). Half a century of reseach on the stroop effect: An integrative review. *Psychological Bulletin*, *109*(2), 163–203. https://doi.org/10.1037/0033-2909.109.2.163

Manohar, S. G., Finzi, R. D., Drew, D., & Husain, M. (2017). Distinct Motivational Effects of Contingent and Noncontingent Rewards. *Psychological Science*, *28*(7), 1016–1026. https://doi.org/10.1177/0956797617693326

Moscarello, J. M., & Hartley, C. A. (2017). Agency and the Calibration of Motivated Behavior. *Trends in Cognitive Sciences*, *21*(10), 725–735. https://doi.org/10.1016/j.tics.2017.06.008

Nagase, A. M., Onoda, K., Foo, J. C., Haji, T., Akaishi, R., Yamaguchi, S., … Morita, K. (2018). Neural mechanisms for adaptive learned avoidance of mental effort. *Journal of Neuroscience*, *38*(10), 2631–2651. https://doi.org/10.1523/JNEUROSCI.1995-17.2018

Nalborczyk, L., & Bürkner, P.-C. (2019). An Introduction to Bayesian Multilevel Models Using brms: A Case Study of Gender Effects on Vowel Variability in Standard Indonesian. *Journal of Speech, Language, and Hearing Research*.

Pelli, D. D. (1997). The VideoToolbox software for visual psychophysics: transforming numbers into

movies. *Spatial Vision*, *10*(4), 437–442.

Pessoa, L., & Engelmann, J. B. (2010). Embedding reward signals into perception and cognition. *Frontiers in Neuroscience*, *4*(September), 1–8. https://doi.org/10.3389/fnins.2010.00017

Polich, J. (2007). Updating P300: An integrative theory of P3a and P3b. *Clinical Neurophysiology*, *118*(10), 2128–2148. https://doi.org/10.1016/j.clinph.2007.04.019

R Core Team. (2017). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing.

Rutledge, R. B., Skandali, N., Dayan, P., & Dolan, R. J. (2014). A computational and neural model of momentary subjective well-being. *Proceedings of the National Academy of Sciences*, *111*(33), 12252–12257. https://doi.org/10.1073/pnas.1407535111

Schevernels, H., Krebs, R. M., Santens, P., Woldorff, M. G., & Boehler, C. N. (2014). Task preparation processes related to reward prediction precede those related to task-difficulty expectation. *NeuroImage*, *84*, 639–647. https://doi.org/10.1016/j.neuroimage.2013.09.039

Shenhav, A., Botvinick, M., & Cohen, J. (2013). The expected value of control: An integrative theory of anterior cingulate cortex function. *Neuron*, *79*(2), 217–240. https://doi.org/10.1016/j.neuron.2013.07.007

Sutton, R., & Barto, A. (1998). *Reinforcement learning: An Introduction*. MIT Press.

Sutton, S., Braren, M., Zubin, J., & John, E. R. (1965). Evoked-potential correlates of stimulus uncertainty. *Science*, *150*(3700), 1187–1188. https://doi.org/10.1126/science.150.3700.1187

Ullsperger, M., Fischer, A. G., Nigbur, R., & Endrass, T. (2014). Neural mechanisms and temporal dynamics of performance monitoring. *Trends in Cognitive Sciences*, *18*(5), 259–267. https://doi.org/10.1016/j.tics.2014.02.009

van Boxtel, G. J. M., & Böcker, K. B. E. (2004). Cortical measures of anticipation. *Journal of Psychophysiology*, *18*(2–3), 61–76. https://doi.org/10.1027/0269-8803.18.23.61

Vassena, E., Silvetti, M., Boehler, C. N., Achten, E., Fias, W., & Verguts, T. (2014). Overlapping neural systems represent cognitive effort and reward anticipation. *PLoS ONE*, *9*(3). https://doi.org/10.1371/journal.pone.0091008

Vehtari, A., Gelman, A., & Gabry, J. (2017). Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Statistics and Computing*, *27*(5), 1413–1432. https://doi.org/10.1007/s11222-016-9696-4

Verguts, T., Vassena, E., & Silvetti, M. (2015). Adaptive effort investment in cognitive and physical tasks: a neurocomputational model. *Frontiers in Behavioral Neuroscience*, *9*(March). https://doi.org/10.3389/fnbeh.2015.00057

Vroom, V. H. (1964). *Work and motivation*. New York: Wiley.

Wang, K. S., & Delgado, M. R. (2019). Corticostriatal Circuits Encode the Subjective Value of Perceived Control. *Cerebral Cortex*, 1–12. https://doi.org/10.1093/cercor/bhz045

Wisniewski, D., Forstmann, B., & Brass, M. (2019). Outcome contingency selectively affects the neural coding of outcomes but not of tasks. *BioarXiv*. https://doi.org/10.1101/375642

Chapter 6

Expecting cognitive effort: The influence of difficulty cues on cognitive control[1]

Abstract

How much cognitive control we exert in part depends on what are the incentives and costs of doing so. Extensive research has shown that people exert more control when they are incentivized, but less is known about the role of expected costs in control allocation. In this study we developed a paradigm aimed at dissociating the influence of expected and actual task difficulty in a conflict task. In Experiment 1 (N = 36) we parametrically manipulated the visibility of targets in order to create five levels of actual difficulty which were preceded by three types of difficulty cues (easy, medium, and hard). Using this procedure we created a subset of trials on which the expectations about task difficulty varied, while the actual difficulty was held constant. The results of this experiment showed that the manipulation of target visibility was an effective manipulation of actual task difficulty. However, this manipulation was not equally spaced across levels of difficulty and failed to show an effect of expected difficulty. In Experiment 2 (N = 43) we improved on the actual task difficulty gradient and showed that higher expected task difficulty led to improved performance on the task. This effect was rather small and present only on congruent trials. Importantly, the effect was modulated by individual differences in willingness to exert cognitive effort. Individuals who preferred to avoid cognitive effort were the ones which showed biggest performance benefits when expecting higher cognitive effort. These findings show that expectations about cognitive effort can influence cognitive control, but that this effect is smaller compared to the previously reported effects of rewards. We discuss how the current paradigm can be further improved in order to better assess the role of cognitive effort in allocation of cognitive control.

[1] Based on: Grahek, I., Braem, S., Koster, E. H. W., & Krebs, R. M. Expecting cognitive effort: The influence of difficulty cues on cognitive control. (in preparation).

Introduction

Take two boxes of playing cards and remove all of the cards from one box. Now put the heavier one on top of the lighter one. If you lift the heavier box and then lift both of them together, something unexpected happens: lifting two boxes feels easier than lifting only one! This phenomenon is known as the size-weight illusion (Buckingham, 2014; Charpentier, 1891). In this case the objective difficulty of lifting two boxes is nearly equal, while our expectation is that it should be twice as difficult. This experiment demonstrates the interaction between two components of effort: the actual and the expected task difficulty.

In our everyday experience we notice that effort is not necessarily physical. Answering e-mails, writing, and giving lectures can be effortful, but these actions require very little physical effort. Cognitive effort is at the heart of most theoretical models in cognitive psychology, but has proved to be surprisingly hard to conceptualize and measure (Kahneman, 1973; Shenhav et al., 2017). One approach has relied on effort discounting paradigms in which subjects are offered varying amounts of reward in exchange for completing more or less cognitively demanding tasks. This approach has demonstrated that individuals tend to avoid exerting cognitive effort (Kool, McGuire, Rosen, & Botvinick, 2010; Westbrook & Braver, 2015). However, this approach focuses on the decision-making process about exerting cognitive effort, but does not provide insights into how task difficulty and difficulty expectations interact with cognitive processes. This interaction is challenging to study for two reasons. First, task difficulty and difficulty expectations are closely linked and hard to dissociate experimentally. For example, if a task is difficult, it will necessarily also be expected to be difficult. However, behavioral performance in a task will depend on both factors and task difficulty will always have a much larger influence on the performance than expectations. Second, there are substantial individual differences in the willingness to exert cognitive effort (Cacioppo & Petty, 1982). While some individuals enjoy cognitively demanding tasks, others tend to dislike them (Inzlicht, Shenhav, & Olivola, 2018).

Studies aimed at investigating cognitive effort and its components have mainly focused on cognitive control. Cognitive control processes are thought to be the prime example of cognitive effort (Shenhav et al., 2017). They are engaged whenever there is need to suppress an automatic behavior in order to comply with current goals (Botvinick & Cohen, 2014). One of the standard tasks which probe cognitive control is the Eriksen Flanker task (Eriksen & Eriksen, 1974). In this task participants are presented with a target (e.g. letter H) surrounded by

distractors (e.g. letters S). Participants are instructed to ignore the distractors and detect the target by pressing a corresponding key. The common finding is that behavioral performance is reduced when the target and the distractors produce competing response tendencies (e.g. letter H surrounded by letters S) compared to the condition in which they produce the same response tendency (e.g. letter H surrounded by letters H). The widely accepted view is that in this task cognitive control biases the processing toward the target and can thus reduce the interference produced by the response conflict (Botvinick et al., 2001; Egner & Hirsch, 2005).

Several studies have investigated the influence of task difficulty (in addition to the conflict manipulation) in the Flanker task. Takezawa and Miyatani (2005) manipulated task difficulty by varying the distance (small vs. large) between the target and distractors. They showed that reaction times in incongruent trials were further increased by the additional difficulty manipulation (small vs. large distance). Forster and colleagues (Forster, Carter, Cohen, & Cho, 2011) manipulated the type of distractors in a Flanker task to produce three difficulty levels for the incongruent trials (low: "HSSSSSH", medium: "HHSSSHH", and high: "HHHSHHH"; the central letter is the target). This study showed that the increase in difficulty produced longer reaction times, but this difference was very pronounced for the highest difficulty trials and much less so for the other two levels. Another way to increase difficulty in a Flanker task it to increase the amount of response conflict by presenting the distractors prior to the target onset. This manipulation has been shown to increase reaction times on both congruent and incongruent trials, and much more so for incongruent, relative to the condition in which the target and distractors are presented simultaneously (Wendt, Kiesel, Geringswald, Purmann, & Fischer, 2014). The studies described here have focused on modulating the amount of response conflict to modulate difficulty in the Flanker task. While these manipulations have been successful, they have not managed to produce parametric increases across more than three levels. In the current study we aimed at developing such a parametric difficulty manipulation by directly affecting the ability of the cognitive control system to bias target processing. This manipulation was at the same time a prerequisite for disentangling actual and anticipated task difficulty, which will be discussed in the following section.

A number of studies used cues presented before target onset to demonstrate that reward incentives improve behavioral performance in cognitive control tasks (for a review see: Botvinick & Braver, 2015). However, it has been proven more difficult to develop similar behavioral paradigms to investigate the effects of cued difficulty. Several fMRI studies have found that high relative to low difficulty cues increase activity in the cortical regions typically

involved in attentional control (Krebs, Boehler, Roberts, Song, & Woldorff, 2012; Vassena et al., 2014). Further, an EEG study has shown that such cues increase the CNV component (Schevernels, Krebs, Santens, Woldorff, & Boehler, 2014) known to scale with preparatory allocation of attention. These studies suggest that difficulty cues can drive the allocation of attention in preparation for the upcoming target stimulus. However, neuroimaging studies have focused on the activity prior to target onset and used target of varying difficulty levels. This makes it difficult to dissociate between the actual and expected task difficulty at the target level, and especially at the level of behavioral response to the target. Several behavioral studies have investigated the influence of difficulty expectations on performance in conflict tasks. Diede and Bugg (2017) have shown that pupil diameter, an index of cognitive effort, is larger when stimuli are presented at locations associated with a higher probability of presenting an incongruent than a congruent trial. A number of studies has investigated the influence of congruency cues on performance. Such cues inform participants whether the next trial is likely to be congruent or incongruent, thus also indirectly cueing task difficulty. These studies have found mixed effects with some finding that congruency cues improve performance more on congruent trials (Aarts, Roelofs, & Van Turennout, 2008; Correa, Rao, & Nobre, 2009), some studies demonstrating the beneficial impact on both congruent and incongruent trials (Czernochowski, 2014), while others have found the cueing effect only when participants were sufficiently motivated (Chiew & Braver, 2016), or reported no effect of the cues (Luks, Simpson, Dale, & Hough, 2007). However, these studies could not disentangle between the effects of expected and actual task difficulty. In addition to the issue of mixed findings, cueing congruency involves additional information which is not directly related to difficulty, which could be one of the reason for the mixed findings. Further, it allows cueing of only two (congruent vs. incongruent) or three (adding neutral) difficulty levels which are not equally spaced. In conclusion, neuroimaging studies provide evidence for the influence of difficulty cues on preparatory attentional processes, and behavioral studies provide mixed and indirect evidence for the influence of difficulty cues on behavioral performance.

In this study we investigated the influence of actual difficulty (target difficulty) and expected difficulty (cued difficulty) on the congruency effects in a conflict task. We aimed at developing a behavioral task in which we could dissociate the effect of expected difficulty and the effect of actual task difficulty. In order to do so, we developed a parametric difficulty manipulation aimed at increasing the amount of selective attention needed to bias target processing. In Experiment 1 we developed a behavioral paradigm which included multiple

difficulty levels paired with cues informing the participants about the difficulty of the upcoming task. This paradigm allowed us to independently assess how target difficulty and expected difficulty influence performance in a conflict task. Specifically, we created stimuli with 5 levels of difficulty by manipulating the visibility of targets in a Flanker task, while keeping the visibility and number of distractors constant. The visibility of the target and the distractors were the same on medium difficulty trials and the visibility of the targets was reduced (higher difficulty targets) or increased (lower difficulty targets) to produce the other difficulty levels. This allowed us to assess the influence of target difficulty on behavioral performance. Crucially, participants were provided with difficulty cues on each trial, which allowed us to have a subset of trials on which the difficulty expectation was different (high, medium, or low), while the actual target was the same (medium). This allowed us to investigate the impact of difficulty expectation while controlling for actual difficulty.

We predicted that more selective attention will be needed to overcome response conflict on trials on which targets are less visible (Botvinick et al., 2001), thus producing performance detriments on both congruent and incongruent trials. Based on the neuroimaging studies showing the increased allocation of attention following such cues (Schevernels et al., 2014; Vassena et al., 2014), we predicted that higher difficulty cues will lead to better behavioral performance relative to other cues on medium difficulty trials.

In order to assess the potential role of the relationship between the visibility of the targets and distractors, we conducted Experiment 2 in which the design was similar to Experiment 1, but the visibility of the targets and distractors was the same on the easiest targets, and then the visibility of the distractors was reduced on more difficult targets. This allowed us to control for the potential effect of the switch in the target-distractor visibility (on low difficulty trials distractors are less visible than targets, while this is reversed on high difficulty trials) in Experiment 1. Further, based on the results of Experiment 1, we aimed at a more equal spacing across difficulty levels in Experiment 2. Finally, we included a measure of the willingness to exert cognitive effort in order to investigate the potential interaction between the use of the cues and this individual difference. It has been shown that individuals who are less intrinsically motivated to exert cognitive effort tend to exert more effort when they are motivated by external rewards (Sandra & Otto, 2018). Thus we wanted to explore potential similarities between reward and difficulty cuing manipulations.

Experiment 1

Methods

Participants

We recruited 36 participants (32 females, 4 males; median age = 21) from the Ghent University participant pool. All of the participants had normal or corrected-to-normal vision and gave informed consent. Participants received 10€ or course credit as the compensation for taking part in the experiment. The research protocol was approved by the Institutional Review Board of Ghent University.

Stimuli and task

Participants performed the Eriksen Flanker task (Eriksen & Eriksen, 1974) in which the target letter was presented centrally and surrounded by 9 distractor letters (Figures 1A and 1B). The target letters H or S were surrounded by either the same letters (congruent trials) or the other letters (incongruent trials). Participants were instructed to ignore the distractor letters and detect the target letter as fast and accurate as possible, by pressing the “H” or “S” key on a standard USB keyboard.

The visibility of the target was varied to create 5 levels of difficulty (Figure 1A). On medium difficulty stimuli the visibility of the target was equal to the visibility of the distractors. On easier trials the visibility of the target was increased, and it was decreased on the higher difficulty trials. The levels of visibility were selected in pilot experiments in order to ensure that participants could perform the task at above chance level on the highest difficulty trials.

Participants were instructed that some trials will be more difficult than others and that, in order to improve their performance, they will be presented with cues which will indicate how difficult the next trial will be. Three colored cues were used and mapped onto three difficulty levels: low, medium, and high (the color-difficulty mapping was counterbalanced across participants). The mapping was probabilistic: for example, the medium cue was followed by the medium, medium high, or medium low target with equal frequency (Figure 1C). This mapping allowed for a subset of trials on which the medium difficulty trials were preceded by either the high, medium, or low cue, thus allowing for the conditions in which the difficulty expectations differed, while the target was the same.

Each trial (Figure 1D) started with the presentation of the colored cue (randomly jittered between 1400 and 1800ms). The cue was followed by the presentation of the target (100 ms) to which participants responded by pressing the appropriate key within a 1100ms response time limit. The time between the end of the trial and the onset of the next one was varied between 400 and 1800ms (randomly selected). In total participants performed 540 trials (108 trials for each of the 5 target difficulties). Half of the trials were congruent and half incongruent, and all of the trials were preceded by two or three different cue types (30 trials per condition). Different trial types were presented in random order. The experiment was divided into 5 blocks of equal length and after each block participants received feedback on the percentage of correct responses in that block and had a short break.

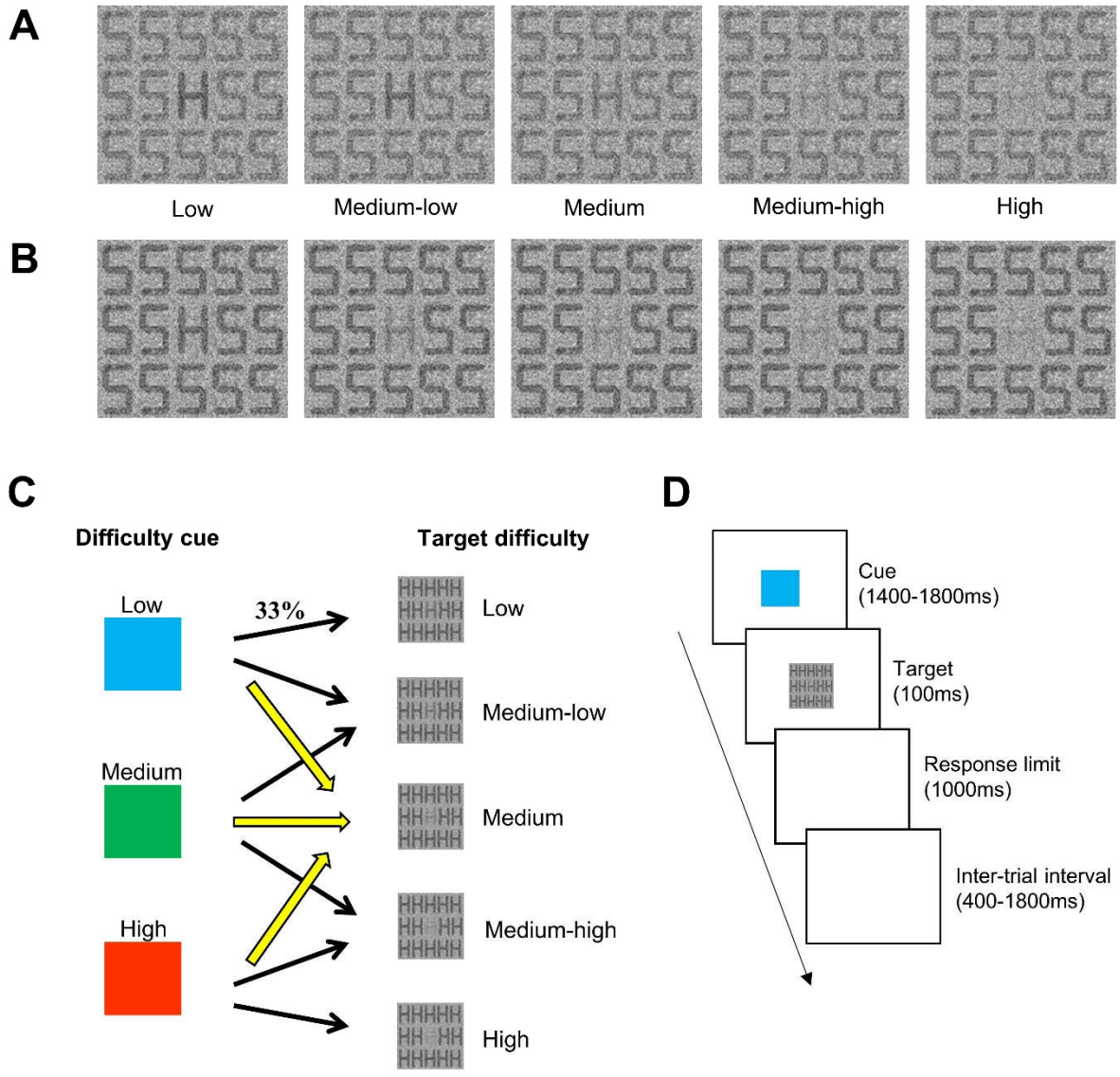


Figure 1. Stimuli used in the experiments, cue-difficulty mapping, and trial structure. A. Stimuli in Experiment 1. **B.** Stimuli in Experiment 2. **C.** The mapping between cues and difficulty levels. The three difficulty cues mapped probabilistically onto the actual target difficulties. The yellow arrows show that the medium difficulty targets were preceded by the low medium or high difficulty cues with equal frequency. **D.** Trial structure.

Statistical analyses

To analyze reaction times and accuracy we fitted Bayesian multilevel regression models. The models were fitted in R using the *brms* package (Bürkner, 2016) which relies on *Stan* (Carpenter et al., 2016) to implement Markov Chain Monte Carlo (MCMC) algorithms and estimate the posterior distributions of parameters in fitted models. We fitted multilevel models with constant and varying effects (also called fixed and random), ex-Gaussian (for reaction

times) or Bernoulli likelihoods, and weakly informative prior distributions (cf. Bürkner, 2016, 2017; Nalborczyk & Bürkner, 2019). We ran four MCMC simulations ("chains"; 6,000 iterations out of which 3,000 warmup) and checked their convergence by examining trace plots, autocorrelation, and variance between chains (Gelman-Rubin statistic; Gelman & Rubin, 1992). When comparing models we relied on the Bayesian R^2 (Gelman, Goodrich, Gabry, & Ali, 2017), and the out-of-sample predictive performance of models using leave-one-out cross-validation (loo) and assessing the difference in their expected predictive accuracy (elpd_diff; Vehtari, Gelman, & Gabry, 2017). We interpreted the parameters of the models using the means and 95% highest density intervals (HDI) of their posterior distributions. We tested hypotheses by examining the ratio of percentages of posterior samples on each side of zero (Evidence Ratio; ER). The ERs can be interpreted as the posterior probability of one hypothesis against its alternative (e.g., A>B vs. A<B).

In all of the models fitted to predict the reaction times only correct trials on which the RTs were longer than 150ms were used. In order to investigate the influence of target difficulty on behavioral performance we fitted three models. The *Congruency* model included only the constant effect of congruency, the *Congruency + Target difficulty* model also included the target difficulty, and the *Congruency * Target difficulty* model included also the interaction between congruency and target difficulty. All three models included the intercept as both constant and varying effect, and all of the constant effects, expect for interactions, were allowed to vary across subjects.

In order to examine the effect of cues on behavioral performance we fitted the models on trials in which the target difficulty was medium. The *Congruency * Cue* model was fitted with the constant effect of the cue type, congruency, their interaction, and intercept. This model included the effects of cue type, congruency, and the intercept as varying effects.

All of the models included the same weakly informative prior distributions. These distributions were Gaussian for both the reaction times (intercept: $M = 500$; $SD = 300$; slopes: $M = 0$; $SD = 100$) and accuracies (intercept: $M = 2.5$; $SD = 3$; slopes: $M = 0$; $SD = 3$)[2].

[2] Note that the prior distributions are set in log-odds. Thus a prior distribution centered at 2.5 roughly corresponds to the odds of 1 in 12 of participants making an error.

Results

Actual target difficulty

The *Congruency * Target difficulty* model was the best one at predicting the reaction times across the whole experiment (Table 1). The posterior distributions of this model (Figure 2A) showed that participants were faster to respond on congruent than on incongruent trials (M = -27.90; 95% HDI [-31.79, -24.05]; ER > 12,000[3]). The difference in reaction times was very large between the hard and medium-hard trials (M = 40.00; 95% HDI [27.80, 51.80]; ER > 12,000) as well as between the medium-hard and medium trials (M = 50.40; 95% HDI [42.80, 57.90]; ER > 12,000). However, the difference between the medium and medium-low trials was much smaller (M = 11.60; 95% HDI [7.71, 15.50]; ER > 12,000), and there was only a very small difference between the two lowest difficulty levels (M = 2.95; 95% HDI [-1.23, 6.94]; ER = 11.66). Further, the congruency effect differed across the levels of target difficulty. The difference in the congruency effect between the two easiest levels of difficulty was very small (M = -3.81; 95% HDI [-11.50, 3.53]; ER = 5.17) and there was no difference in the congruency effect between the medium-low and medium trials (M = -0.80; 95% HDI [-6.57, 4.74]; ER = 1.55). However, the congruency effect was larger on the medium-high trials compared to medium trials (M = -13.60; 95% HDI [-19.70, -7.53]; ER > 12,000). Further, there was a very small and not very robust increase in the congruency effect on the hard compared to medium-high trials (M = -3.89; 95% HDI [-13.00, 5.70]; ER = 3.81). In sum, the interaction between target difficulty and congruency came from the sharp increase in the congruency effect from the medium to medium-high trials.

The *Congruency * Target difficulty* model was also the best one at predicting accuracy, but it was almost as good as the *Congruency + Target difficulty* (Table 1). This result indicates a weak interaction between congruency and target difficulty, which was confirmed by the inspection of the posterior distributions of the winning model (Figure 2B). They revealed the expected congruency effect in that participants were more likely to respond correctly to the congruent compared to incongruent trials (M = 1.71[4]; 95% HDI [1.19, 2.41]; ER = 332.33).

[3] When the whole posterior distribution is above 0 the evidence ratio is larger than the total number of samples in the posterior distribution. In this situation there is a very large amount of evidence for the hypothesis.

[4] The model's parameters were estimated in log-odds which were converted throughout the manuscript into odds for clarity.

Participants were less accurate on hard compared to medium-hard trials (M = 0.37; 95% HDI [0.31, 0.45]; ER > 12,000), and less accurate on medium-hard than on medium trials (M = 0.20; 95% HDI [0.15, 0.27]; ER > 12,000). Participants were also less accurate on medium compared to medium-low trials (M = 0.73; 95% HDI [0.52, 1.02]; ER = 19.28), while there was very little difference between medium-low and low trials (M = 0.83; 95% HDI [0.51, 1.43]; ER = 3.20). When it comes to the interaction between congruency and target difficulty, the congruency effect did not reliably change between the high and medium-high difficulty trials (M = 1.02; 95% HDI [0.74, 1.41]; ER = 1.18), nor between the medium-high and medium trials (M = 1.08; 95% HDI [0.72, 1.63]; ER = 1.89). The congruency effect significantly increased on medium compared to medium-low trials (M = 1.84; 95% HDI [1.01, 3.22]; ER = 51.63). Finally, the congruency effect did not differ between the two lowest difficulty levels (M = 0.87; 95% HDI [0.37, 2.08]; ER = 1.66). It can be concluded that the reason for the very small difference between the model which included the interaction between congruency and target difficulty and the model which did not comes from the fact that the only reliable change in the size of the congruency effect happened between the medium-low and medium difficulty targets.

In conclusion, the robust congruency effect was observed in both reaction times and accuracy. The target difficulty manipulation produced reliable changes in behavior, but they were not equally distributed across all levels of difficulty. In general, there was a robust and sharp increase in reaction times and a reduction in accuracy in higher difficulty trials, while there was not much difference between the lower difficulty levels. In addition, there was evidence of the increased congruency effect in the harder compared to easier trials, especially so in reaction times. Finally, no speed-accuracy tradeoffs were observed.

Table 1

Results of model comparisons for the models predicting behavioral performance across all trials in Experiment 1. Models are compared based on the leave-one-out cross validation information criterion (LOO), the difference between models in their expected log predictive accuracy (ELPD_diff), and Bayesian R^2.

Model	*LOO (SE)*	*ELPD_diff (SE)*	*Bayesian R^2 (SE)*
Reaction times			
Congruency * Target difficulty	205547.8 (254.61)	0.00 (0.00)	0.29 (0.00)
Congruency + Target difficulty	205583.1 (254.00)	-17.68 (7.31)	0.29 (0.00)
Congruency	208460.9 (244.46)	-1456.56 (53.33)	0.18 (0.00)
Accuracy			
Congruency * Target difficulty	6910.07 (156.60)	0.00 (0.00)	0.16 (0.01)
Congruency + Target difficulty	6910.67 (156.67)	-0.30 (3.10)	0.15 (0.01)
Congruency	8140.60 (171.00)	-615.26 (36.50)	0.04 (0.01)

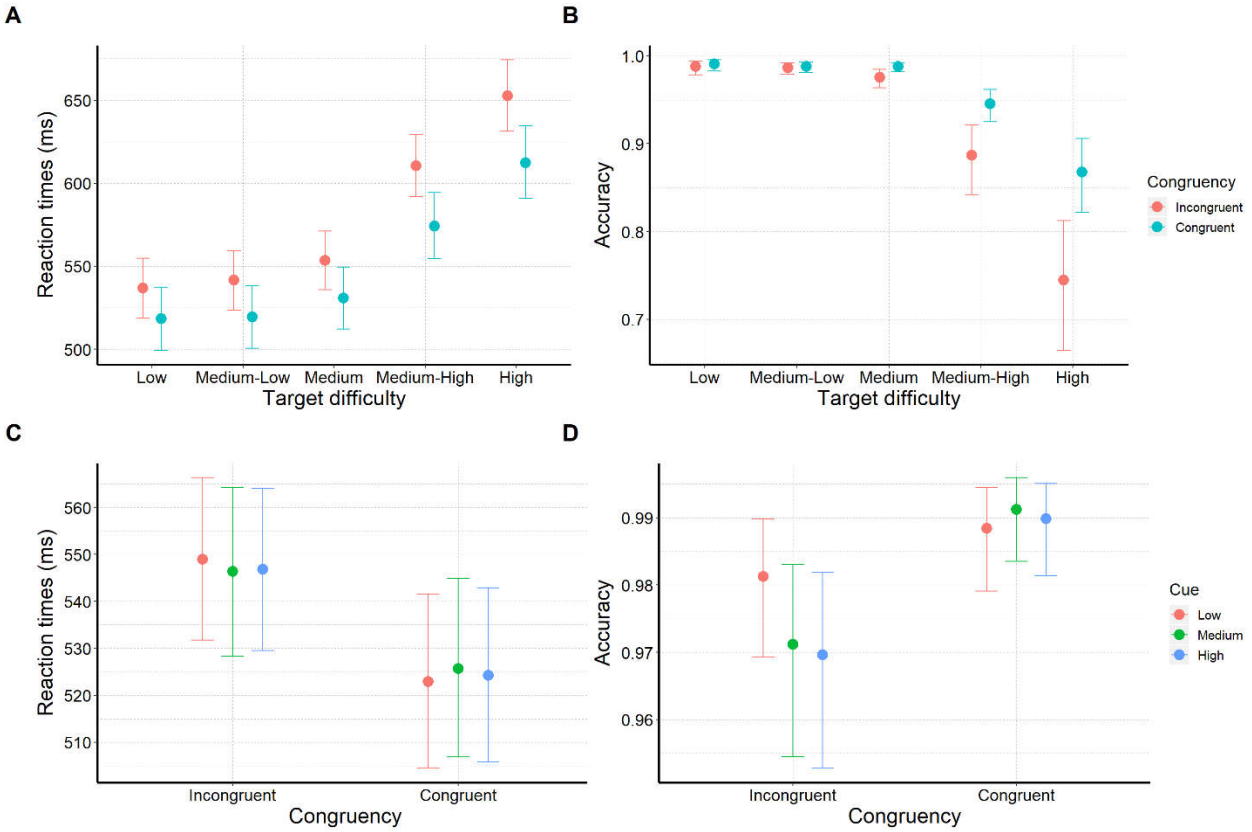


Figure 2. Results of the Experiment 1. (A) Reaction times for congruenty and incongruent trials across five levels of target difficulty. Error bars represent 95% HDIs. (B) Accuracy across the levels of target difficulty. (C) Reaction times on medium difficulty trials when preceeded by either low, medium, or high difficulty cues. (D) Accuracy on medium difficulty trials when preceeded by different difficulty cues.

Expected target difficulty

In predicting the reaction times on medium difficulty trials the fitted *Congruency* * *Cue* model (Figure 2C) revealed a robust effect of congruency (M = -23.00; 95% HDI [-29.70, 16.50]; ER > 12,000). However, there was no difference between the trials preceded by high and low difficulty cues (M = -0.38; 95% HDI [-4.92, 4.03]; ER = 1.33), nor between the trials preceded by medium and high cues (M = 0.39; 95% HDI [-4.16, 4.92]; ER = 1.32). The congruency effect was slightly reduced on high compared to low difficulty expectation trials, but this difference was not robust (M = -3.48; 95% HDI [4.82, -11.7]; ER = 3.85). The congruency effect was very similar between medium and high difficulty expectation trials (M = -1.83; 95% HDI [6.88, -9.87]; ER = 1.95).

The same model was fitted to investigate the influence of difficulty expectations on accuracy in the medium difficulty trials. This model (Figure 2D) revealed the robust reduction in the probability of correct responses on incongruent compared to congruent trials (M = 2.57; 95% HDI [1.68, 4.14]; ER > 12,000). However, there was very little difference in accuracy between trials with high or low cues (M = 0.83; 95% HDI [0.49, 1.38]; ER = 3.39), and between high and medium trials (M = 1.11; 95% HDI [0.69, 1.76]; ER = 2.01). The congruency effect was different between trials on which high and low cues were presented (M = 1.87; 95% HDI [0.80, 4.22]; ER = 13.70) indicating a larger congruency effect following high compared to low cues. The difference in the congruency effect was not present when comparing medium to high trials (M = 1.10; 95% HDI [0.48, 2.57]; ER = 1.43).

Discussion

The results of Experiment 1 indicated that the manipulation of target visibility was an effective manipulation of target difficulty. Participants were slower and less accurate in responding to stimuli which had less visible targets relative to stimuli with more visible targets. However, this manipulation was much more effective on harder than on easier trials. There was a crucial difference between the harder and easier trials: on harder trials the target was less visible than the distractors, while this was reversed on the easier trials. The results of Experiment 1 indicated that the reduction of target visibility relative to distractor visibility is a more effective manipulation of task difficulty. To confirm this, in Experiment 2 we equated the target and distractor visibility on the easiest trial type and made the target progressively less visible on all the other trial types.

The difficulty cues did not impact performance on the medium difficulty cues. A possible explanation for this was due to the lack of difference in difficulty between the two levels of low difficulty targets and the medium difficulty targets. It is thus possible that the effect of cues would emerge if there was a steeper difficulty gradient, and hence presumably more variance in medium difficulty targets. Taken together, the results of Experiment 1 show that it is possible to manipulate difficulty by modulating the visibility of targets in a Flanker task, but point to the potential issues in our difficulty manipulation. In order to ameliorate this we conducted Experiment 2.

Experiment 2

Methods

Participants

For the second experiment we recruited 43 participants (37 females and 9 males) from the Ghent University participant pool. The same ethical, compensation, and participants' characteristics described for Experiment 1 were followed.

Stimuli and task

The task design was exactly the same as in the Experiment 1. The only difference was in the choice of stimuli. In Experiment 2 the distractors and the targets had the same visibility on the stimuli of lowest difficulty (Figure 1B). The visibility of the targets was then progressively reduced for the more difficult stimuli. Further, the difference between the levels of visibility was modified in order to ensure a more equal spacing of difficulty levels compared to the Experiment 1. Specifically, the difference between the lower levels of difficulty was increased, while the difference between the two most difficult types of stimuli was reduced.

In order to control for the potential effects of the affinity towards exerting cognitive effort (Sandra & Otto, 2018), we also collected the questionnaire measure of the need for cognition (NFC), which can be regarded as the intrinsic motivation to exert cognitive effort (The Need for Cognition Scale Cacioppo & Petty, 1982). Participants filled in the electronic version of this scale upon completing the experiment.

Statistical analyses

The statistical analyses were the same as the ones described for Experiment 1 and the same models were fitted to examine the influence of target difficulty and difficulty cues on behavioral performance. In addition, in Experiment 2 we also wanted to investigate the potential influence of NFC on the cueing effect. We thus fitted an additional model to investigate the influence of NFC on reaction times and accuracy. In the *Cue * Congruency * NFC* model in which we added the interaction between NFC and the cue type, as well as the main effect of NFC, as constant effects.

Results

Actual target difficulty

The *Congruency * Target difficulty* model was the best one at predicting the reaction times across the whole experiment (Table 2), pointing to the importance of the interaction between target difficulty and congruency. The parameters of this model (Figure 3A) revealed that participants were slower to respond to incongruent compared to congruent trials (M = 34.60; 95% HDI [30.50, 39.00]; ER > 12,000). Further, there was a robust effect of target difficulty. Participants were slower to respond to the highest difficulty trials compared to the medium-high trials (M = 27.40; 95% HDI [21.30, 33.20]; ER > 12,000), on which reaction times were slower than on the medium trials (M = 18.60; 95% HDI [14.70, 22.90]; ER > 12,000). Medium difficulty trials were slower than the medium-low trials (M = 17.90; 95% HDI [12.90, 22.40]; ER > 12,000), which were slower than the low difficulty trials (M = 7.26; 95% HDI [3.73, 10.80]; ER > 12,000).

Task difficulty interacted with the congruency effect in that the congruency effect grew larger as the task difficulty increased (Figure 3A). The congruency effect was larger on the high compared to medium-high difficulty trials (M = 16.10; 95% HDI [8.25, 23.40]; ER > 12,000), which displayed a higher congruency effect compared to the medium difficulty trials (M = 7.23; 95% HDI [2.21, 12.70]; ER = 499.00). The congruency effect on the medium trials was larger than on the medium-low trials (M = 9.36; 95% HDI [4.64, 14.50]; ER > 12,000), on which the congruency effect was larger than on the low difficulty trials (M = 8.86; 95% HDI [2.16, 15.50]; ER = 199.00).

The *Congruency + Target difficulty* model was the best one at predicting accuracy across all trials of the experiment (Table 2). The posterior distribution of this model (Figure 3B) revealed that participants were more likely to be accurate on congruent relative to incongruent trials (M = 1.77; 95% HDI [1.34, 2.36]; ER > 12,000). Participants were less likely to be correct on high compared to medium-high trials (M = 0.38; 95% HDI [0.32, 0.45]; ER > 12,000), medium-high compared to medium trials (M = 0.57; 95% HDI [0.46, 0.70]; ER > 12,000), and on medium compared to medium-low trials (M = 0.55; 95% HDI [0.42, 0.72]; ER > 12,000). There was no reliable difference in accuracy between medium-low and low trials (M = 0.97; 95% HDI [0.70, 1.35]; ER = 1.29).

Table 2

Results of model comparisons for the models predicting behavioral performance across all trials in Experiment 2. Models are compared based on the leave-one-out cross validation information criterion (LOO), the difference between models in their expected log predictive accuracy (ELPD_diff), and Bayesian R^2.

Model	*LOO (SE)*	*ELPD_diff (SE)*	*Bayesian R^2 (SE)*
Reaction times			
Congruency * Target difficulty	248920.3 (280.82)	0.00 (0.00)	0.30 (0.00)
Congruency + Target difficulty	249034.1 (279.32)	-56.91 (12.20)	0.29 (0.00)
Congruency	250749.0 (273.56)	-914.35 (46.64)	0.24 (0.00)
Accuracy			
Congruency + Target difficulty	10016.56 (182.34)	0.00 (0.00)	0.09 (0.01)
Congruency * Target difficulty	10021.56 (182.58)	-2.50 (1.83)	0.09 (0.01)
Congruency	10750.88 (189.63)	-367.16 (28.75)	0.04 (0.01)

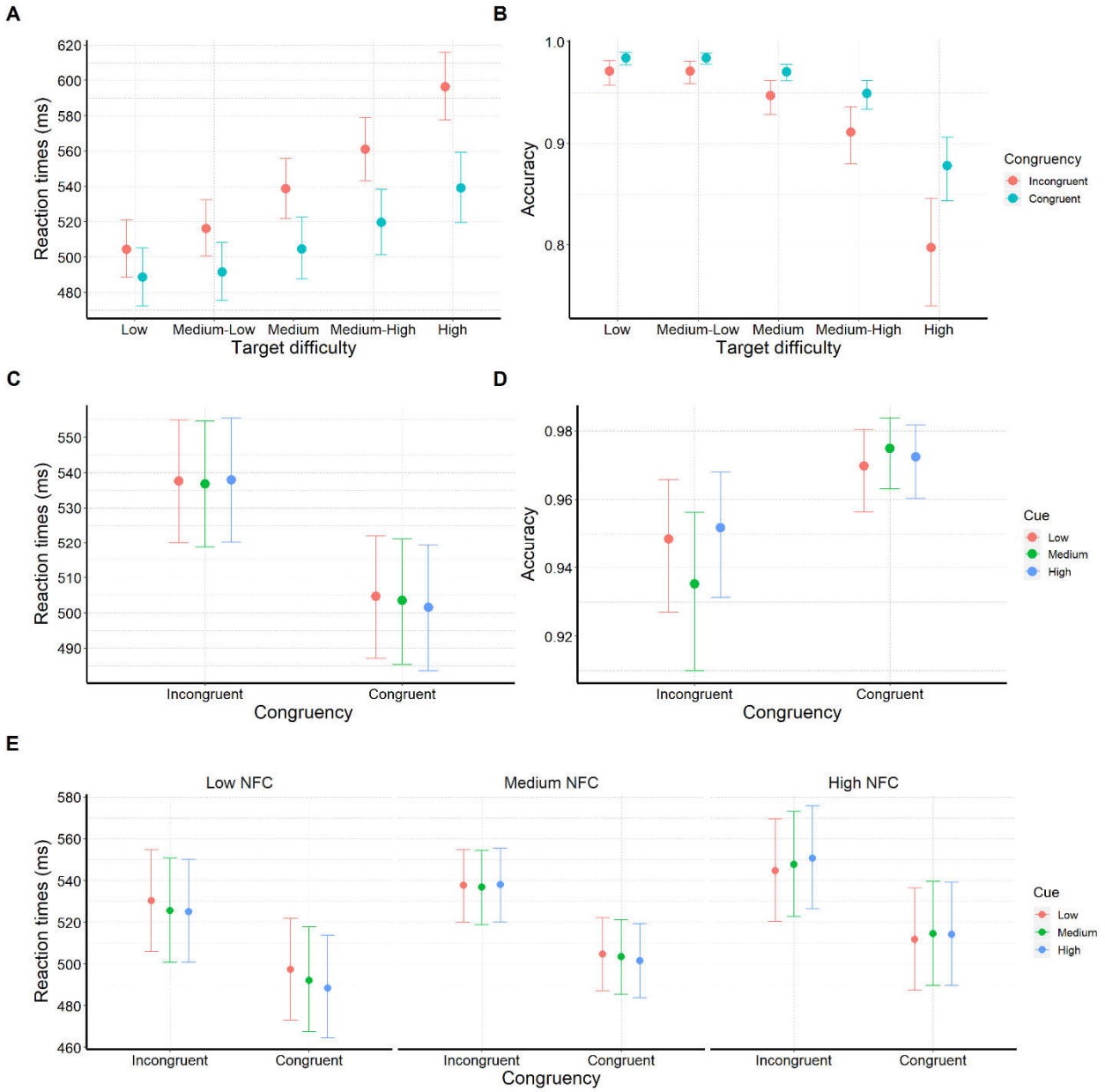


Figure 3. Results of the Experiment 2. (A) Reaction times for congruenty and incongruent trials across five levels of target difficulty. Error bars represent 95% HDIs. (B) Accuracy across the levels of target difficulty. (C) Reaction times on medium difficulty trials when preceeded by either low, medium, or high difficulty cues. (D) Accuracy on medium difficulty trials when preceeded by different difficulty cues. (E) The effect of difficulty cues across low (-1SD from the mean), medium (mean), and high (+1SD) NFC.

Expected target difficulty

Reaction times on medium difficulty trials were best predicted by the model which included the interaction between cue type and congruency and the interaction between cue type and NFC (Table 3). The posterior distribution of the winning model (Figures 3C and 3E) revealed the expected congruency effect (M = 34.30; 95% HDI [29.70, 38.60]; ER > 12,000).

Participants were somewhat faster to respond after high compared to low difficulty cues (M = 1.49; 95% HDI [-2.21, 5.60]; ER = 3.40), but the size of this effect was negligible. There was no difference in reaction times between high and medium cues (M = 0.40; 95% HDI [-3.61, 4.24]; ER = 1.37). The difference between high and low cues interacted with congruency and was more pronounced on congruent than on incongruent trials (M = 3.42; 95% HDI [-3.73, 11.00]; ER = 4.38). This was also true when comparing high and medium trials (M = 3.24; 95% HDI [-4.35, 10.60]; ER = 4.15). Higher NFC predicted slower reaction times (M = 0.94; 95% HDI [-0.60, 2.52]; ER = 7.62), but also interacted with the expected task difficulty. The difference between the high and low cues was increased in participants which had lower NFC (M = 0.51; 95% HDI [0.147, 0.87]; ER = 499.00) and this was also true, but to a lesser extent, for the difference between high and medium cues (M = 0.15; 95% HDI [0.20, 0.52]; ER = 4.02).

Accuracy on the medium difficulty trials was better predicted by the model which did not include the effect of the NFC (Table 3). This model (Figure 3D) revealed that participants were more accurate on congruent than on incongruent trials (M = 2.02; 95% HDI [1.38, 2.91]; ER > 12,000). There were no differences in accuracy between high and low (M = 1.07; 95% HDI [0.79, 1.44]; ER = 2.31) nor between medium and high trials (M = 0.90; 95% HDI [0.67, 1.21]; ER = 3.06). The difference between high and low cues was the same on congruent and incongruent trials (M = 1.02; 95% HDI [0.57, 1.69]; ER = 1.16), while the difference between high and medium trials was more pronounced on congruent trials (M = 1.50; 95% HDI [0.87, 2.56]; ER = 13.49). However, except for the effect of congruency, all of the effects on accuracy were rather small and not very robust. Importantly, the accuracy effects indicate that there were no speed-accuracy tradeoffs.

Table 3

Results of model comparisons for the models predicting behavioral performance on medium difficulty trials in Experiment 2. Models are compared based on the leave-one-out cross validation information criterion (LOO), the difference between models in their expected log predictive accuracy (ELPD_diff), and Bayesian R^2.

Model	*LOO (SE)*	*ELPD_diff (SE)*	*Bayesian R^2 (SE)*
Reaction times			
Congruency * Cue * NFC	82724.94 (161.41)	0.00 (0.00)	0.28 (0.01)
Congruency * Cue	82729.49 (161.63)	-2.28 (2.71)	0.28 (0.01)
Accuracy			
Congruency * Cue	3017.96 (108.27)	0.00 (0.00)	0.04 (0.01)
Congruency * Cue * NFC	3021.91 (108.53)	-1.97 (0.76)	0.04 (0.01)

Discussion

The results of Experiment 2 revealed substantial improvements in the gradient of task difficulty in comparison to Experiment 1. Gradual reduction of target visibility from the lowest to the highest difficulty level led to both longer reaction times and reduced accuracy. Further, the congruency effect was increased on more difficult trials.

On medium difficulty trials we observed a small effect of the difficulty cues. Participants were somewhat faster to respond to the targets when they were proceeded by the high relative to low difficulty cues, and especially so on congruent trials. Further, this effect was more pronounced in participants with low NFC. This result suggests that there was a small effect of difficulty expectations on performance in a conflict task, that this effect was present on congruent trials, and depended on how willing an individual was to exert cognitive effort.

General discussion

In this study we investigated the influence of actual and expected target difficulty on behavioral performance in a Flanker task. We have shown that target visibility can be used to parametrically vary task difficulty – above and beyond the conflict manipulation. Specifically,

the difficulty manipulation affected both congruent and incongruent trials, but the effect was larger on incongruent trials revealing an interaction between cognitive control and task difficulty. Crucially, we have shown that expecting high relative to low difficulty can improve behavioral performance, but that this effect is subtle and limited to congruent trials. Further, this effect depended on the individual's trait of willingness to exert cognitive effort: the difficulty cues affected behavioral performance more in individuals who are not prone to exert cognitive effort.

The existing studies of the impact of difficulty on performance in conflict tasks have manipulated difficulty by changing the position or the number of distractors (Forster et al., 2011; Takezawa & Miyatani, 2005), or by presenting the distractors before the targets (Wendt et al., 2014). Such manipulations involve the modulation of response conflict in that, for example, the amount of response conflict is reduced when the number of distractors is reduced. In this study we have shown that difficulty can be successfully manipulated by holding distractors constant and modulating the visibility of targets. This manipulation has several advantages compared to the previous studies. First, it allows for a parametric manipulation of task difficulty and thus reveals that difficulty impacts conflict processing in a gradual manner. Second, this manipulation allows for the parallel manipulation of difficulty of both congruent and incongruent trials. This novel difficulty gradient manipulation is valuable in itself (and extends previous research in the conflict domain), but is most importantly a prerequisite for investigating the relationship between expected and actual task difficulty, which is the focus of the present study.

In this study we were able to investigate the effects of the expected task difficulty on performance independently from the effects of the actual task difficulty. In this way we were able to demonstrate that high relative to low difficulty expectations (signaled by the difficulty cues) improve the behavioral performance in the Flanker task by speeding up reaction times on congruent trials. This result suggests that the expectations of increased effort can enhance target processing, but only when there is no response conflict. As such, the present results go beyond the previous neuroimaging findings which have shown that cues indicating that the upcoming trial will be difficult increase the amount of preparatory attention (Schevernels et al., 2014; Vassena et al., 2014), but have not been able to investigate the impact of the expected effort on behavior. Further, this effect is more pronounced in individuals who are less intrinsically motivated to exert cognitive effort. This means that the individuals who are not motivated to exert cognitive effort used the effort cues to adjust their behavior more than the individuals who

enjoy exerting effort. Individuals low in NFC have thus tended to optimize the amount of effort allocation, while the individuals high in NFC have kept a similar level of performance, which could have been more enjoyable for them because it means that they were allocating more effort. Several conclusions can be drawn from this set of findings. First, effort cues can influence behavioral performance in a way which parallels the effects of reward cues (Botvinick & Braver, 2015). However, this effect is very small in comparison to the reward studies. One potential reason for the small magnitude of the effect is that there are large individual differences in the willingness to exert cognitive effort, which are known to interact with cognitive processes (Sandra & Otto, 2018). These differences could be larger than the differences in the willingness to modulate behavior to obtain rewards, which could be a more universally spread trait. In our paradigm participants were not able to earn any extra monetary rewards. This insight also offers a possible way to experimentally induce a stronger effect of reward expectations. For example, it is possible that this effect would increase if participants were additionally motivated to use the cues by providing them with the possibility to earn extra monetary rewards in the experiment. A second way to obtain a more precise measure of the influence of effort expectations on behavior would be to tailor the difficulty levels individually for each subject. In the current experiment the difficulty levels were set at the group level.

In conclusion, our study has demonstrated the behavioral effects of two crucial components of effort: actual task difficulty and expected task difficulty. We have shown that both of these components interact with cognitive control in a task in which control is needed to overcome response conflict. While the effect of task difficulty on behavioral performance is unsurprisingly very large, the effect of expected task difficulty is fairly subtle and depends on individual differences such as the willingness to exert cognitive effort. This set of results adds to the growing literature suggesting that components of motivation are crucial in the decision-making about how and when to allocate cognitive control (Botvinick & Braver, 2015). While most of this literature has focused on the influence of rewards on cognitive control, here we show that another component of motivation – effort – is crucial in control allocation.

Data

Data and analyses code is available at:
https://osf.io/m56xe/?view_only=78e7e9b78cee487c8cb4f8b56afb538b

Acknowledgements

This work was supported by the Special Research Fund (BOF) of Ghent University [grant #01D02415 awarded to IG] and by a starting grant of the European Research Council (ERC) under the Horizon 2020 framework [grant No. 636116 awarded to RMK)].

References

Aarts, E., Roelofs, A., & Van Turennout, M. (2008). Anticipatory activity in anterior cingulate cortex can be independent of conflict and error likelihood. *Journal of Neuroscience*, *28*(18), 4671–4678. https://doi.org/10.1523/JNEUROSCI.4400-07.2008

Botvinick, M. M., & Braver, T. S. (2015). Motivation and Cognitive Control: From Behavior to Neural Mechanism. *Annual Review of Psychology*, *66*, 1–31. https://doi.org/10.1146/annurev-psych-010814-015044

Botvinick, M. M., Braver, T. S., Barch, D. M., Carter, C. S., & Cohen, J. D. (2001). Conflict monitoring and cognitive control. *Psychological Review*, *108*(3), 624–652. https://doi.org/10.1037/0033-295X.108.3.624

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

Buckingham, G. (2014). Getting a grip on heaviness perception: A review of weight illusions and their probable causes. *Experimental Brain Research*, *232*(6), 1623–1629. https://doi.org/10.1007/s00221-014-3926-9

Bürkner, P.-C. (2016). brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software*, *80*(1), 1–28.

Bürkner, P.-C. (2017). Advanced Bayesian Multilevel Modeling with the R Package brms. *ArXiv:1705.11123*.

Cacioppo, J. T., & Petty, R. E. (1982). The need for cognition. *Journal of Personality and Social Psychology*, *42*(1), 116–131. https://doi.org/10.1037/0022-3514.42.1.116

Carpenter, B., Gelman, A., Hoffman, M., Lee, D., Goodrich, B., Betancourt, M., … Riddell, A. (2016). Stan: A probabilistic programming language. *Journal of Statistical Software*, *2*(20), 1–37.

Charpentier, A. (1891). Analyse experimentale de quelques elements de la sensation de poids. *Archive de Physiologie Normale et Pathologiques*, *3*.

Chiew, K. S., & Braver, T. S. (2016). Reward favors the prepared: Incentive and task-informative cues interact to enhance attentional control. *Journal of Experimental Psychology: Human Perception and Performance*, *42*(1), 52–66. https://doi.org/10.1037/xhp0000129

Correa, Á., Rao, A., & Nobre, A. C. (2009). Anticipating conflict facilitates controlled stimulus-response selection. *Journal of Cognitive Neuroscience*, *21*(8), 1461–1472. https://doi.org/10.1162/jocn.2009.21136

Czernochowski, D. (2014). ERPs dissociate proactive and reactive control: Evidence from a task-switching paradigm with informative and uninformative cues. *Cognitive, Affective and Behavioral Neuroscience*, *15*(1), 117–131. https://doi.org/10.3758/s13415-014-0302-y

Egner, T., & Hirsch, J. (2005). Cognitive control mechanisms resolve conflict through cortical amplification of task-relevant information. *Nature Neuroscience*, *8*(12), 1784–1790. https://doi.org/10.1038/nn1594

Eriksen, B. A., & Eriksen, C. W. (1974). Effects of noise letters upon the identification of a target letter in a nonsearch task. *Perception & Psychophysics*. https://doi.org/10.3758/BF03203267

Forster, S. E., Carter, C. S., Cohen, J. D., & Cho, R. Y. (2011). Parametric manipulation of the conflict signal and control-state adaptation. *Journal of Cognitive Neuroscience*, *23*(4), 923–935. https://doi.org/10.1162/jocn.2010.21458

Gelman, A., Goodrich, B., Gabry, J., & Ali, I. (2017). R-squared for Bayesian regression models. *Unpublished via Http://Www. Stat. Columbia. Edu/~ Gelman/Research/Unpublished.* Retrieved from http://www.stat.columbia.edu/~gelman/research/unpublished/bayes_R2.pdf

Gelman, A., & Rubin, D. B. (1992). Inference from Iterative Simulation Using Multiple Sequences. *Statistical Science*, *7*(4), 457–472. https://doi.org/10.1214/ss/1177011136

Inzlicht, M., Shenhav, A., & Olivola, C. Y. (2018). The Effort Paradox: Effort Is Both Costly and Valued. *Trends in Cognitive Sciences*, *22*(4), 337–349. https://doi.org/10.1016/j.tics.2018.01.007

Kahneman, D. (1973). *Attention and Effort*. Englewood Cliffs, NJ: Prentice-Hall.

Kool, W., McGuire, J. T., Rosen, Z. B., & Botvinick, M. M. (2010). Decision making and the avoidance of cognitive demand. *Journal of Experimental Psychology: General*, *139*(4), 665–682. https://doi.org/10.1037/a0020198

Krebs, R. M., Boehler, C. N., Roberts, K. C., Song, A. W., & Woldorff, M. G. (2012). The involvement of the dopaminergic midbrain and cortico-striatal-thalamic circuits in the integration of reward prospect and attentional task demands. *Cerebral Cortex*, *22*(3), 607–615. https://doi.org/10.1093/cercor/bhr134

Luks, T. L., Simpson, G. V., Dale, C. L., & Hough, M. G. (2007). Preparatory allocation of attention and adjustments in conflict processing. *NeuroImage*, *35*(2), 949–958. https://doi.org/10.1016/j.neuroimage.2006.11.041

Nalborczyk, L., & Bürkner, P.-C. (2019). An Introduction to Bayesian Multilevel Models Using brms: A Case Study of Gender Effects on Vowel Variability in Standard Indonesian. *Journal of Speech, Language, and Hearing Research*.

Sandra, D. A., & Otto, A. R. (2018). Cognitive capacity limitations and Need for Cognition differentially predict reward-induced cognitive effort expenditure. *Cognition*, *172*, 101–106. https://doi.org/10.1016/j.cognition.2017.12.004

Schevernels, H., Krebs, R. M., Santens, P., Woldorff, M. G., & Boehler, C. N. (2014). Task preparation processes related to reward prediction precede those related to task-difficulty expectation. *NeuroImage*, *84*, 639–647. https://doi.org/10.1016/j.neuroimage.2013.09.039

Shenhav, A., Musslick, S., Falk, L., Kool, W., Griffits, T. L., Cohen, J. D., & Botvinick, M. M. (2017). Toward a rational and mechanistic account of mental effort. *Annual Review of*

Neuroscience, *40*. https://doi.org/10.1146/annurev-neuro-072116-031526

Takezawa, T., & Miyatani, M. (2005). Quantitative relation between conflict and response inhibition in the flanker task. *Psychological Reports*. https://doi.org/10.2466/pr0.97.2.515-526

Vassena, E., Silvetti, M., Boehler, C. N., Achten, E., Fias, W., & Verguts, T. (2014). Overlapping neural systems represent cognitive effort and reward anticipation. *PLoS ONE*, *9*(3). https://doi.org/10.1371/journal.pone.0091008

Vehtari, A., Gelman, A., & Gabry, J. (2017). Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Statistics and Computing*, *27*(5), 1413–1432. https://doi.org/10.1007/s11222-016-9696-4

Wendt, M., Kiesel, A., Geringswald, F., Purmann, S., & Fischer, R. (2014). Attentional adjustment to conflict strength: Evidence from the effects of manipulating flanker- Target SOA on response times and prestimulus pupil size. *Experimental Psychology*, *61*(1), 55–67. https://doi.org/10.1027/1618-3169/a000227

Westbrook, A., & Braver, T. S. (2015). Cognitive effort: A neuroeconomic approach. *Cognitive, Affective and Behavioral Neuroscience*, *15*(2), 395–415. https://doi.org/10.3758/s13415-015-0334-y

Chapter 7

General discussion

Cognitive control deficits are proposed to be the hallmark feature of depression (Disner, Beevers, Haigh, & Beck, 2011; Gotlib & Joormann, 2010). However, it is currently not clear why these deficits emerge in depression and a mechanistic understanding of such deficits is lacking. In this thesis we have critically assessed the current understanding of cognitive deficits in depression and proposed novel ways to conceptualize the causes of these impairments. We have relied on normative models of cognitive control to create a mechanistic framework of control deficits. We have proposed that the role of motivation is crucial to understand how much control depressed individuals exert. Our framework considers three components of motivation as crucial for understanding cognitive control deficits: the value of outcomes, the probability that a given amount of control will result in a desired outcome, and the amount of effort tied with exerting the specified amount of control. We have then developed experimental paradigms which assess the influence of each of these components on cognitive control. Here we start by summarizing the main theoretical and empirical findings of this thesis. We then discuss the issues related to uncovering the causes of cognitive control deficits and outline the implications of our work for the cognitive theories of depression. Finally, we outline the limitations of this thesis and propose future directions for improving the understanding the nature of cognitive impairments in mood disorders.

Theoretical contributions

The first part of this thesis has focused on the current issues in understanding cognitive control dysfunctions in depression. We have first critically reviewed the existing theoretical models of control and outlined the main conceptual issues in how control dysfunctions are understood. We have proposed ways to advance the existing models based on the current cognitive neuroscience research on cognitive control. We have then focused on the role of motivation in control dysfunctions and proposed a mechanistic framework which posits the causal role of motivation in cognitive control dysfunctions in depression.

Toward a mechanistic understanding of control in depression

Cognitive control dysfunctions play a central role in the cognitive theories of depression (for a review see: Gotlib & Joormann, 2010), but the mechanistic understanding of these dysfunctions is lacking. We have analyzed the main problems in the current understanding of cognitive control in depression. These issues are related to how cognitive control is conceptualized and measured, reliance on descriptive instead of mechanistic explanations, and the lack of integration between cognitive, emotional, and motivational deficits. We have then summarized the current cognitive neuroscience models of control and proposed how they can be used to overcome the existing issues in depression research. Specifically, we have proposed that the understanding of control dysfunctions should focus on the problems in detecting the need to engage cognitive control, the role of motivation, and the balance between stability and flexibility. Rather than a reduced ability to exert control, we have argued that control deficits in depression can stem from problems in detecting when, with which intensity, and for how long to engage in controlled processing. Thus cognitive control deficits can stem from failing to detect signals for the need to increase control (Botvinick, Braver, Barch, Carter, & Cohen, 2001), or a dysfunctional cost-benefit decision-making about how much control to allocate (Shenhav, Botvinick, & Cohen, 2013). Further, they can arise due to previous learning about how much control to allocate in a certain context (Bhandari, Badre, & Frank, 2017). This work, presented in *Chapter 1*, provides a significant theoretical shift in understanding cognitive control in depressed individuals. It moves the research focus from detecting specific cognitive control impairments (e.g., inhibition of negative material) toward specific mechanisms which underlie the observed impairments.

The interplay between motivation and cognition in depression

In *Chapter 2* we have focused on the causal role of motivation in cognitive control dysfunctions in depression. First, we have linked the causes for depression-related impairments to key components of motivation: value of outcomes, their controllability, and effort costs. We have then relied on a normative model of cognitive control (Expected Value of Control theory; Shenhav et al., 2013) to propose a mechanistic framework for understanding cognitive control impairments in depression. We proposed that depressed individuals do not necessarily have lowered ability to exert cognitive control. Instead, the cognitive control impairments stem from a lowered value of exerting control. The low value of control can be caused by reduced value of outcomes, reduced estimates of outcome controllability, or increased estimates of effort

costs. We have then relied on the computational implementation of the Expected Value of Control theory (Musslick, Shenhav, Botvinick, & Cohen, 2015) to simulate how impairments in each of these three components would translate into the behavioral performance on the Stroop task and in an effort discounting paradigm. We have thus provided a novel mechanistic framework for understanding the causes of cognitive control deficits in depression. This framework integrates two relatively independent research areas: cognitive impairments in depression (Gotlib & Joormann, 2010) and depression-related motivational impairments (Pizzagalli, 2014; Treadway & Zald, 2011). Further, the framework opens new avenues for depression research: the specific motivational impairments should be studied together with cognitive impairments in order to provide a more mechanistic and precise understanding of cognition in depression.

Empirical contributions

The interface of motivation and cognitive control is a burgeoning area for both empirical and theoretical research (Botvinick & Braver, 2015; Braver et al., 2014). While it is clear that motivation plays an important role in how much cognitive control individuals exert, the exact components of motivation crucial for control allocation in depression are not well understood. In the empirical part of this thesis we have focused on three components of motivation which we have proposed as the potential causes of depression-related impairments in cognitive control. We have developed experimental paradigms in which we could assess the influence of each of these components on cognitive control in healthy individuals.

Value of outcomes

When individuals are incentivized by monetary rewards they perform better in selective attention tasks (Krebs, Boehler, & Woldorff, 2010; Padmala & Pessoa, 2011). Further, individuals tend to allocate more attention towards stimuli which have been previously linked to rewards, even when this collides with their current goals (Anderson, Laurent, & Yantis, 2011; Della Libera & Chelazzi, 2009). Thus there are at least two mechanisms which can account for the reward-related improvements in attention tasks: a voluntary attention mechanism and an incentive saliency mechanism. In *Chapter 3* we created an experimental paradigm in which we could dissociate, using an EEG-based measure of stimulus processing in the visual cortex, between the effects of voluntary attention and incentive saliency in a sustained attention task. The results of this study have revealed the flexible dynamics through which top-down voluntary attention and incentive saliency effects interact to adjust processing in the visual cortex. Stimuli

linked to high rewards (high incentive salience) are preferentially processed and this effect is independent of the effect of voluntary attention. However, the influence of rewards on visual processing goes back to baseline levels once the rewards are no longer available in the environment. Our interpretation of this finding is that the cognitive control system finds an optimal set of processing weights in the visual cortex in order to both maximize the amount of rewards earned and minimize the amount of top-down control needed. Given that the relevant feature changes on the trial-by-trial basis, the optimal solution could be to increase the salience of the high-reward stimuli if that does not impair performance significantly on trials when those stimuli act as distractors. This set of weights will then be maintained until there is a change in the environment, that is, when rewards are no longer available. Then the optimal solution could be to set the processing weights for all of the stimuli to their baseline level, which is what we observed in our experiment. These results point to the flexible dynamics through which the processing in sustained attention tasks is adjusted based on rewards.

This experimental paradigm can be used to further investigate the potential impairments in depression. First, it can be used to assess whether depressed individuals flexibly adjust processing in order to maximize reward attainment. There is little evidence that depressed individuals differ from the healthy controls in how much they like rewards when they obtain them. However, there is evidence showing that these individuals anticipate less rewards and that they are less willing to exert effort to obtain them (Pizzagalli, 2014; Treadway & Zald, 2011). Our paradigm offers an interesting way to test if individuals with depressive symptoms will adjust the processing of the stimuli in order to obtain more rewards. This could potentially reveal impairments in how attention is adjusted in order to reach goals. Further, this paradigm offers insight in how attention is flexibly re-adjusted when rewards are no longer available. This is an important step towards understanding mechanisms which could be impaired in depression.

Outcome controllability

Classical theories propose that goal-directed behavior is sensitive to the value of the outcomes and the probability that the outcomes will be reached given an action (Balleine & O'Doherty, 2010; Dickinson, 1985). A recent normative theory of cognitive control (Shenhav et al., 2013) proposes that the value of an outcome is weighted by efficacy to determine how much control will be allocated in a given situation. In *Chapter 4* we have investigated whether the level of the ability to control outcomes in the environment (efficacy) can be learned and if

it influences control allocation. We have constructed a task in which levels of efficacy vary over time and participants are instructed to estimate how much efficacy they have. Using EEG we have shown that individuals' estimates of efficacy predict how much attention they will allocate in anticipation of the task. When they estimated that they have high compared to low efficacy they allocated more attention which increased their reaction times on the task. Further, we showed that the knowledge about whether a reward depended on participants' performance or not modulated how they processed reward feedback. When participants knew that rewards did not depend on them they were less surprised by omission of rewards then when they knew that rewards depend on their performance. Finally, we have shown that the estimates of efficacy determine the learning from feedback about efficacy. When the estimated levels of efficacy were low, participants were learning more from the feedback telling them that they had efficacy on the current trial, then when they were estimating high efficacy. This set of findings demonstrates the importance of estimated efficacy in allocation of cognitive control and reward processing. It supports the view that efficacy acts together with outcome value to determine how much control will be allocated (Shenhav et al., 2013). Further, this finding corroborates the importance of learning in cognitive control (Bhandari et al., 2017; Chiu & Egner, 2019).

The ability to exert control over outcomes in our environment is a crucial component of goal-directed behavior (Leotti, Iyengar, & Ochsner, 2010; Moscarello & Hartley, 2017). Further, estimating that we have the ability to exert control over the environment is crucial for mental health (Maier & Seligman, 2016). We have proposed in *Chapter 2* that outcome controllability is an important factor in cognitive control deficits in depression. Our simulations showed that reduced outcome controllability would translate into worse performance on a Stroop task. The paradigm we have developed here can be used to directly test these theoretical predictions. Using this paradigm in a clinical setting can reveal several potential impairments. First, an important question is whether depressed individuals can effectively estimate their levels of control over rewards in the environment and whether they can effectively update these estimates based on the incoming information. Second, this paradigm can be used with EEG to determine whether the efficacy estimates are then effectively used to allocate more or less cognitive control. This approach can yield insights into the mechanisms behind reduced cognitive control in depression. Further, this paradigm can be adapted in order to investigate efficacy in the context of avoiding negative outcomes, which could be especially relevant for depression research.

Expected effort

Cognitive control is thought to require cognitive effort (Shenhav et al., 2017) and individuals tend to avoid exerting this type of effort unless they are sufficiently rewarded (Kool, McGuire, Rosen, & Botvinick, 2010; Westbrook, Kester, & Braver, 2013). If individuals are expecting that a task will be more demanding they can proactively prepare and this has been shown to increase the activity of the cortical regions involved in attentional control (Krebs, Boehler, Roberts, Song, & Woldorff, 2012; Vassena et al., 2014). However, the effects of expected task difficulty are hard to disentangle from the effects of actual task difficulty in behavioral paradigms. In *Chapter 5* we developed a behavioral paradigm in which we can dissociate between these two factors. We have presented individuals with cues which reliably predicted the difficulty of the upcoming target in a Flanker task. Critically, on a subset of trials participants were cued with either a high, medium, or low difficulty cues, but the actual task difficulty was kept similar. The results of this study have shown that expectations about cognitive effort have a small effect on behavioral performance. When expecting that the upcoming target will be difficult, participants were faster to respond to the target compared to when they expected an easy target. Importantly, this effect was stronger in individuals who had low need for cognition (Cacioppo & Petty, 1982). This finding suggests that the individual differences in the willingness to exert cognitive effort in daily life relate to the effects of difficulty expectations on behavioral performance in a conflict task.

Reduced willingness to exert effort is found across different mental disorders (Culbreth, Moran, & Barch, 2018; Salamone, Yohn, López-Cruz, San Miguel, & Correa, 2016) and reduced exertion of physical effort has been related to depression and anhedonia (Zald & Treadway, 2017). However, there is very little work on cognitive effort in depression, although some of the initial studies have shown results comparable to the ones with physical effort (Hershenberg et al., 2016; Marchetti, Shumake, Grahek, & Koster, 2018). In particular, we think that the expectations about how effortful a task will be are an important component which could reduce the amount of control allocated in individuals with depression. As we have shown in this chapter, individual differences in the willingness to exert cognitive effort do modulate the influence of effort expectations on behavior. Further work with this paradigm in clinical populations could reveal whether there are specific impairments at the level of effort expectations (e.g., reduced performance when expecting high effort) or at the level of actual task difficulty (e.g., worse performance on more difficult trials).

Individual differences in cognitive control

In order to translate the results of basic research into effective treatments, it is necessary to understand the precise mechanisms which underlie symptoms in each individual. This is crucial for developing precise treatments for mental disorders (Williams, 2016). Cognitive control is a promising treatment target in depression (Koster, Hoorelbeke, Onraedt, Owens, & Derakshan, 2017; Siegle, Ghinassi, & Thase, 2007). However, the framework which we have proposed in this thesis suggests that cognitive control dysfunctions do not have a single cause.

Imagine that we invite two depressed patients into our lab and we test them with a standard cognitive control task. We then compare their performance and notice that one individual performed much better than the other one. This difference can be due to different levels of motivation, and specifically due to the difference in the anticipation of how valuable outcomes can be, how likely they are to reach those outcomes, or how effortful that will be. For example, one of the two individuals could place much lower value on performing the task well. Further, differences can arise due to previous learning: one of the two individuals could have learned that they can get by with allocating very little control in psychological tests. These two individuals might also differ in their core ability to successfully implement the signal for how much attention to allocate to the processing of the targets in the task. Finally, our measure of cognitive control can be very noisy (e.g., we used only a small number of trials) and the differences could be the product of this noise. How can we disentangle between these different possibilities?

Ability, motivation, or learning?

Why are some people better at inhibiting automatic responses or shifting their attention between tasks? Individual differences in cognitive control predict many important outcomes such as job and school success (for a review see: Diamond, 2012), and lowered cognitive control is related to a wide range of mental disorders (Millan et al., 2012). As we have detailed in *Chapter 1*, the long standing view was that such differences reflect the variances in the ability to exert control. However, recent research has pointed to the role of motivation in cognitive control and shown that control can be increased when people are sufficiently motivated (Botvinick & Braver, 2015). Further, the role of learning when and how much to allocate control has further emphasized the processes which are not directly linked to the ability to exert control (Abrahamse, Braem, Notebaert, & Verguts, 2016; Bhandari et al., 2017; Chiu & Egner, 2019). The idea that the amount of control allocated depends on motivation and previous learning has

been formalized in several normative models of cognitive control (Alexander & Brown, 2011; Holroyd & McClure, 2015; Shenhav et al., 2013; Silvetti, Alexander, Verguts, & Brown, 2014; Verguts & Notebaert, 2008). These advances in understanding cognitive control also increase the complexity of understanding the causes of individual differences in cognitive control.

At the present moment there is no clear way to distinguish between these different sources of individual differences in cognitive control and this is a crucial task for future research. We see two broad ways to approach this problem. First, constructing batteries of tasks which tap into different components of motivation in the context of cognitive control could allow for disentangling of the contributions of each of the components at the individual level. In this thesis we have attempted at doing so in the three empirical chapters. However, we have not developed these tasks with the explicit goal of creating a diagnostic battery. Thus we have used different tasks, namely the dot motion task, the Stroop task, and the Flanker task. Ideally, a task battery would use the same task and introduce manipulations of reward, outcome controllability, and effort. Further, single paradigms which can assess multiple components of motivation would be especially useful in this context as these components of motivation are interactive.

A second approach would rely on further developing the existing computational models of cognitive control so that they can be fitted to the empirical data with the goal of extracting the relevant parameters. The current computational models are focused on simulating the existing effects in the literature and showing that the model-based simulations can reproduce these effects (Lieder, Shenhav, Musslick, & Griffiths, 2018; Vassena, Deraeve, & Alexander, 2017). There are obstacles to doing so as many of the existing models contain variables which are multiplied and thus impossible to extract with a unique solution. However, the need for disentangling between the possible impairments related to different components of these models could push the field of computational psychology and neuroscience into further developing these models. In order to do so, further collaboration between individual differences and computational focused researchers will be necessary.

Measurement issues

Psychometric properties of cognitive control tasks are another important point of concern. For example, it is known that the reliability of the standard control tasks is relatively low when compared to the norms of the questionnaire-based research. This is especially true for difference scores (Draheim, Mashburn, Martin, & Engle, 2019), which are very often used

as indexes of cognitive control in individual differences research. While there is no straightforward solution for this issue, recently there has been a growing interest in assessing the psychometric properties of cognitive control tasks and improving the measurement of control-related constructs. For example, it has been shown that the use of hierarchical models instead of simple averages substantially improves the reliability of the Flanker and Stroop task (Haaf & Rouder, 2019). Further, there is a push toward a field-wide agreement on the best design practices when measuring cognitive control (Braem et al., 2019). Implementing these statistical and design recommendations can substantially improve the measurement of cognitive control and this will be crucial to advance the understanding of cognitive control deficits in psychopathology.

Implications for cognitive theories of depression

The cognitive model of depression has guided research on cognition in depression as well as the development of successful therapies (Clark & Beck, 1999). In brief, this model assumes the central role of cognitive processes in the onset and maintenance of depressive symptoms. Recent research has pointed to the role of specific impairments in cognitive processes such as attention, interpretation, and memory (Gotlib & Joormann, 2010). The central role has been assigned to impairments in cognitive control processes (Disner et al., 2011; Joormann, 2010) which are thought to underlie depression-related biases in attention and interpretation (Everaert, Grahek, & Koster, 2017) and emotion-regulation difficulties (Joormann & Vanderlind, 2014). The original cognitive model developed by Beck has relied on the understanding of cognition at the time in which it has been developed. The cognitive system was seen as relatively passive as it receives inputs from the external world (i.e., sensory information) and transforms them into outputs (i.e., responses). Cognitive science has been going through a paradigmatic shift in which the view of the cognitive system is changing (Clark, 2013). Instead of passively processing information, cognitive processes are now seen as actively generating predictions about the incoming sensory information and observing the discrepancy between the predicted and observed information to motivate action (Friston, 2010; Knill & Pouget, 2004).

The paradigmatic change in understanding the cognitive system has translated into the understanding of mental disorders. Understanding of cognition in depression has followed this route. Several recent theoretical models have emphasized the role of prior expectations in depression. For example, it has been proposed that the cognitive theories of depression should

focus more on the role of negative expectations and the reasons why such expectations are not being efficiently updated in light of disconfirming evidence (Rief & Joormann, 2019). A recent account of cognitive process in depression has proposed that depression is related to increased prior beliefs that negative events will happen which are not optimally updated thus creating a negative feedback loop (Kube, Schwarting, Rozenkrantz, Glombiewski, & Rief, 2019). Another account of depression proposes that negative prior beliefs and impairments in model-based reasoning sustain depressive symptoms (Huys, Daw, & Dayan, 2015). These models open several important questions. One important question is what are the crucial negative expectations in depression? The second question is how are negative prior expectations sustained, i.e., why are they not updated in light of disconfirming evidence?

Two cardinal symptoms of depression are sustained negative affect and the loss of interest or pleasure in previously enjoyable activities (Gotlib & Furman, 2015). While the interactions between negative affect and cognitive control have been in research focus (Joormann & Tanovic, 2015), the role of motivation has been less researched. A recent upsurge of research into motivational deficits in depression has started to characterize specific impairments in reward anticipation and cost-benefit decision making in depression (Pizzagalli, 2014; Treadway & Zald, 2011). In this thesis (*Chapter 2*) we have proposed that impaired components of motivation play a causal role in cognitive control deficits in depression. Critically, these components can provide insight into the crucial expectations which can be biased in depression. We have argued that the reduced expectations about the value of investing cognitive effort, the reduced expectations about the ability to control the outcomes in the environment, and the increased expectations of effort costs, are crucial in allocating less cognitive control and goal-directed behavior. In chapters 3, 4, and 5 we have shown evidence for the role of each of these components in the allocation of cognitive control in healthy individuals. Future research is needed in order to investigate these processes in depressed individuals and to investigate the impact of each of these components. Further, the crucial next step in understanding cognition in depression will be to focus on the mechanisms through which negative expectations are sustained and why they are not updated in light of disconfirmatory evidence.

Clinical implications

Cognitive control training has shown some promise at alleviating depressive symptoms (Koster et al., 2017; Motter et al., 2016; Siegle et al., 2007). Most of the cognitive trainings for

depression have focused on tasks which aim to improve the ability of individuals to exert cognitive control. For example, several studies have used the adjusted Paced Auditory Serial Addition Task (PASAT; Siegle et al., 2007) in which participants are presented with a series of numbers and their task is to respond with a sum of the two previously presented numbers. The adaptive PASAT adjusts the timing between the presentation of the numbers in order to make the task progressively more difficulty. This task uses neutral stimuli to train working memory updating and working memory shielding from interference. Several tasks have also used emotional material in the training context in order to enhance the transfer to the depression-relevant processes such as rumination (Cohen, Mor, & Henik, 2015; Daches & Mor, 2014).

However, there is very little research on developing trainings which target the interaction between motivation and cognition in depression. Our motivation-based framework suggests that the trainings aimed at remediating components of motivation in the context of cognitive control could be particularly useful. In addition, further research should focus on the mechanisms through which the current cognitive trainings influence depressive symptoms. For example, it is possible that a part of the beneficial effects of such trainings is due to increasing the motivation to allocate cognitive control, instead of increasing the ability to do so. Further, our framework suggests that the value of control can be context specific. For example, one individual can have a high value of exerting cognitive effort at work, while having a very low value of control in the context of a relationship. The treatments aimed at augmenting cognitive control in depressed individuals should first focus on detecting the contexts in which exerting control is important, but reduced. Following the detection of those contexts, the treatments should be focused on increasing the control allocation in these situations. In this context, it will be crucial to work with patients on increasing their awareness about the situations in which they fail to allocate the necessary amount of control.

Limitations and future directions

Despite the advances that were made in this PhD thesis there are also several important limitations that require consideration. The main limitation of this thesis is that we have not directly tested the proposed framework. Instead, we have focused on developing experimental paradigms which can be used to test clinical populations. However, we believe that the focus on developing good paradigms which tap into the relevant components of both motivation and cognitive control is a sine qua non of translating basic cognitive neuroscience into clinical research. Even though the research on motivation-cognition interactions is developing fast

(Botvinick & Braver, 2015), this is still a young research area. We hope to have contributed to this field by investigating how different components of motivation interact with cognitive control. Further, we believe that the paradigms we have developed here can now be tested in sub-clinical populations, and then further developed for testing clinical populations.

Second, in this thesis we have not considered the heterogeneity of depression. However, this disorder is known to be very heterogeneous (Fried & Nesse, 2015) and individuals presenting only a very small number of overlapping symptoms can be diagnosed with depression. Future research is needed in order to determine whether cognitive control impairments are linked to specific symptoms of depression. For example, it is possible that the presence of anhedonia could impact cognitive control via reducing the anticipated value of outcomes and increasing the effort costs (Pizzagalli, 2014). Negative affect could interact with cognitive control through other routes, for example through impacting the estimated controllability of the environment (Maier & Seligman, 2016).

Finally, cognitive control dysfunctions are found across many psychiatric disorders (Millan et al., 2012) and are considered to be an important trans-diagnostic factor (Goschke, 2014). Here we have focused on depression, but the interactions between cognitive control and motivation are very relevant for other disorders, such as schizophrenia (Culbreth et al., 2018). An important direction for future research will be to investigate the relationship between motivational and cognitive impairments across different disorders. Importantly, it has been proposed that cognitive control dysfunctions could be related to the general psychopathology factor which is present in all mental disorders (p-factor; Caspi et al., 2014). The relationship between the p-factor, motivation, and cognitive control should be further investigated.

Concluding remarks

In this thesis we have proposed a mechanistic framework for understanding the cognitive control impairments in depression. This framework builds on the existing models in clinical (Disner et al., 2011; Joormann, 2010) and cognitive neuroscience domain (Shenhav et al., 2013) and proposes the causal role of motivational impairments in depression in reduced cognitive control. We have argued that cognitive control impairments can arise due to the lowered value of outcomes, attenuated estimates of outcome controllability, or increased effort costs. This theoretical work has provided a more mechanistic understanding of cognitive control dysfunctions in depression. We have then developed paradigms which assess each of these components. We have developed a paradigm which allowed us to compare the effects of

voluntary attention and reward associations. This study has revealed how voluntary attention and reward associations modulate stimulus processing in the visual cortex. We have also developed a task in which we can investigate how individuals learn to estimate how much control over outcomes they have and have shown that these estimates influence how much control they allocate. Finally, we have developed a behavioral paradigm in which we were able to dissociate between the effects of expected and actual task difficulty. This study has demonstrated that high effort expectation can lead to improved performance, but only in individuals who are not willing to exert cognitive effort in their daily lives. In sum these empirical studies both advance our understanding of the interplay between different components of motivation and offer a potential route for further research in individuals with depression. This research program can start testing the predictions of the framework we have proposed. We hope that this work can start revealing how motivation and cognition interact and contribute to the onset and maintenance of depressive symptoms. Such insights hold the potential to improve the understanding and treatment of depression.

References

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

Alexander, W. H., & Brown, J. W. (2011). Medial Prefrontal Cortex as an action-outcome predictor. *Nature Neurosci.*, *14*(10), 1338–1344. https://doi.org/10.1038/nn.2921.Medial

Anderson, B. A., Laurent, P. A, & Yantis, S. (2011). Value-driven attentional capture. *Proceedings of the National Academy of Sciences*, *108*(25), 10367–10371. https://doi.org/10.1073/pnas.1104047108

Balleine, B. W., & O'Doherty, J. P. (2010). Human and rodent homologies in action control: Corticostriatal determinants of goal-directed and habitual action. *Neuropsychopharmacology*, *35*(1), 48–69. https://doi.org/10.1038/npp.2009.131

Bhandari, A., Badre, D., & Frank, M. J. (2017). Learning Cognitive Control. In T. Egner (Ed.), *Wiley Handbook of Cognitive Control* (pp. 376–391). Wiley-Blackwell.

Botvinick, M. M., & Braver, T. S. (2015). Motivation and Cognitive Control: From Behavior to Neural Mechanism. *Annual Review of Psychology*, *66*, 1–31. https://doi.org/10.1146/annurev-psych-010814-015044

Botvinick, M. M., Braver, T. S., Barch, D. M., Carter, C. S., & Cohen, J. D. (2001). Conflict monitoring and cognitive control. *Psychological Review*, *108*(3), 624–652. https://doi.org/10.1037/0033-295X.108.3.624

Braem, S., Bugg, J. M., Schmidt, J. R., Crump, M. J. C., Weissman, D. H., Notebaert, W., & Egner, T.

(2019). Measuring Adaptive Control in Conflict Tasks. *Trends in Cognitive Sciences*, *23*(9), 769–783. https://doi.org/10.1016/j.tics.2019.07.002

Braver, T. S., Krug, M. K., Chiew, K. S., Kool, W., Westbrook, J. A., Clement, N. J., … Somerville, L. H. (2014). Mechanisms of motivation-cognition interaction: challenges and opportunities. *Cognitive, Affective & Behavioral Neuroscience*, *14*(2), 443–472. https://doi.org/10.3758/s13415-014-0300-0

Cacioppo, J. T., & Petty, R. E. (1982). The need for cognition. *Journal of Personality and Social Psychology*, *42*(1), 116–131. https://doi.org/10.1037/0022-3514.42.1.116

Caspi, A., Houts, R. M., Belsky, D. W., Goldman-Mellor, S. J., Harrington, H., Israel, S., … Moffitt, T. E. (2014). The p factor: One general psychopathology factor in the structure of psychiatric disorders? *Clinical Psychological Science*, *2*(2), 119–137. https://doi.org/10.1177/2167702613497473

Chiu, Y. C., & Egner, T. (2019). Cortical and subcortical contributions to context-control learning. *Neuroscience and Biobehavioral Reviews*, *99*(December 2018), 33–41. https://doi.org/10.1016/j.neubiorev.2019.01.019

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *The Behavioral and Brain Sciences*, *36*(3), 181–204. https://doi.org/10.1017/S0140525X12000477

Clark, D. A., & Beck, A. T. (1999). *Scientific foundations of cognitive theory and therapy of depression*. John Wiley & Sons.

Cohen, N., Mor, N., & Henik, A. (2015). Linking executive control and emotional response: A training procedure to reduce rumination. *Clinical Psychological Science*, *3*(1), 15–25. https://doi.org/10.1177/2167702614530114

Culbreth, A. J., Moran, E. K., & Barch, D. M. (2018). Effort-cost decision-making in psychosis and depression: could a similar behavioral deficit arise from disparate psychological and neural mechanisms? *Psychological Medicine*, 1–16. https://doi.org/10.1017/S0033291717002525

Daches, S., & Mor, N. (2014). Training ruminators to inhibit negative information: A preliminary report. *Cognitive Therapy and Research*, *38*(2), 160–171. https://doi.org/10.1007/s10608-013-9585-5

Della Libera, C., & Chelazzi, L. (2009). Learning to attend and to ignore is a matter of gains and losses. *Psychological Science*, *20*(6), 778–784. https://doi.org/10.1111/j.1467-9280.2009.02360.x

Diamond, A. (2012). Executive functions. *Annual Review of Psychology*, *64*, 135–168. https://doi.org/10.1146/annurev-psych-113011-143750

Dickinson, A. (1985). Actions and habits: the development of behavioural autonomy. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *308*(1135), 67–78.

Disner, S. G. S., Beevers, C. G. C., Haigh, E. A. P., & Beck, A. T. (2011). Neural mechanisms of the

cognitive model of depression. *Nature Reviews Neuroscience*, *12*(8), 467–477. https://doi.org/10.1038/nrn3027

Draheim, C., Mashburn, C. A., Martin, J. D., & Engle, R. W. (2019). Reaction time in differential and developmental research: A review and commentary on the problems and alternatives. *Psychological Bulletin*, *145*(5), 508–535. https://doi.org/10.1037/bul0000192

Everaert, J., Grahek, I., & Koster, E. H. W. (2017). Individual differences in cognitive control over emotional material modulate cognitive biases linked to depressive symptoms. *Cognition and Emotion*, *31*(4). https://doi.org/10.1080/02699931.2016.1144562

Fried, E. I., & Nesse, R. M. (2015). Depression sum-scores don't add up: why analyzing specific depression symptoms is essential. *BMC Medicine*, *13*(1), 72. https://doi.org/10.1186/s12916-015-0325-4

Friston, K. (2010). The free-energy principle: A unified brain theory? *Nature Reviews Neuroscience*, *11*(2), 127–138. https://doi.org/10.1038/nrn2787

Goschke, T. (2014). Dysfunctions of decision-making and cognitive control as transdiagnostic mechanisms of mental disorders: advances, gaps, and needs in current research. *International Journal of Methods in Psychiatric Research*, *23*(41–57). https://doi.org/10.1002/mpr.1410

Gotlib, I. H., & Furman, D. J. (2015). Depression. *Emerging Trends in the Social and Behavioral Sciences: An Interdisciplinary, Searchable, and Linkable Resource*, 1–15. https://doi.org/10.1002/9781118900772.etrds0074

Gotlib, I. H., & Joormann, J. (2010). Cognition and depression: current status and future directions. *Annual Review of Clinical Psychology*, *6*, 285–312. https://doi.org/10.1146/annurev.clinpsy.121208.131305

Haaf, J. M., & Rouder, J. N. (2019). Some do and some don't? Accounting for variability of individual difference structures. *Psychonomic Bulletin and Review*, *26*(3), 772–789. https://doi.org/10.3758/s13423-018-1522-x

Hershenberg, R., Satterthwaite, T. D., Daldal, A., Katchmar, N., Moore, T. M., Kable, J. W., & Wolf, D. H. (2016). Diminished effort on a progressive ratio task in both unipolar and bipolar depression. *Journal of Affective Disorders*, *196*, 97–100. https://doi.org/10.1016/j.jad.2016.02.003

Holroyd, C. B., & McClure, S. M. (2015). Hierarchical control over effortful behavior by rodent medial frontal cortex: A computational model. *Psychological Review*, *122*(1), 54–83. https://doi.org/10.1037/a0038339

Huys, Q. J. M., Daw, N. D., & Dayan, P. (2015). Depression: A Decision-Theoretic Analysis. *Annual Review of Neuroscience*, *38*(1), 1–23. https://doi.org/10.1146/annurev-neuro-071714-033928

Joormann, J. (2010). Cognitive Inhibition and Emotion Regulation in Depression. *Current Directions in Psychological Science*, *19*(3), 161–166. https://doi.org/10.1177/0963721410370293

Joormann, Jutta., & Vanderlind, W. M. (2014). Emotion Regulation in Depression: The Role of

Biased Cognition and Reduced Cognitive Control. *Clinical Psychological Science*, *2*(4), 402–421. https://doi.org/10.1177/2167702614536163

Joormann, Jutta, & Tanovic, E. (2015). Cognitive vulnerability to depression: Examining cognitive control and emotion regulation. *Current Opinion in Psychology*, *4*, 86–92. https://doi.org/10.1016/j.copsyc.2014.12.006

Knill, D. C., & Pouget, A. (2004). The Bayesian brain: The role of uncertainty in neural coding and computation. *Trends in Neurosciences*, *27*(12), 712–719. https://doi.org/10.1016/j.tins.2004.10.007

Kool, W., McGuire, J. T., Rosen, Z. B., & Botvinick, M. M. (2010). Decision making and the avoidance of cognitive demand. *Journal of Experimental Psychology: General*, *139*(4), 665–682. https://doi.org/10.1037/a0020198

Koster, E. H. W., Hoorelbeke, K., Onraedt, T., Owens, M., & Derakshan, N. (2017). Cognitive control interventions for depression: A systematic review of findings from training studies. *Clinical Psychology Review*, *53*, 79–92. https://doi.org/10.1016/j.cpr.2017.02.002

Krebs, R. M., Boehler, C. N., Roberts, K. C., Song, A. W., & Woldorff, M. G. (2012). The involvement of the dopaminergic midbrain and cortico-striatal-thalamic circuits in the integration of reward prospect and attentional task demands. *Cerebral Cortex*, *22*(3), 607–615. https://doi.org/10.1093/cercor/bhr134

Krebs, R. M., Boehler, C. N., & Woldorff, M. G. (2010). The influence of reward associations on conflict processing in the Stroop task. *Cognition*, *117*(3), 341–347. https://doi.org/10.1016/j.cognition.2010.08.018

Kube, T., Schwarting, R., Rozenkrantz, L., Glombiewski, J. A., & Rief, W. (2019). Distorted Cognitive Processes in Major Depression: A Predictive Processing Perspective. *Biological Psychiatry*, 1–11. https://doi.org/10.1016/j.biopsych.2019.07.017

Leotti, L. A., Iyengar, S. S., & Ochsner, K. N. (2010). Born to choose: The origins and value of the need for control. *Trends in Cognitive Sciences*, *14*(10), 457–463. https://doi.org/10.1016/j.tics.2010.08.001

Lieder, F., Shenhav, A., Musslick, S., & Griffiths, T. L. (2018). Rational metareasoning and the plasticity of cognitive control. *Plos Computational Biology*, *14*(4). https://doi.org/10.13140/RG.2.2.24500.14721

Maier, S. F., & Seligman, M. E. P. (2016). Learned helplessness at fifty: Insights from neuroscience. *Psychological Review*, *123*(4), 1–19. https://doi.org/10.1037/rev0000033

Marchetti, I., Shumake, J., Grahek, I., & Koster, E. H. W. (2018). Temperamental factors in remitted depression: The role of effortful control and attentional mechanisms. *Journal of Affective Disorders*, *235*. https://doi.org/10.1016/j.jad.2018.04.064

Millan, M. J., Agid, Y., Brüne, M., Bullmore, E. T., Carter, C. S., Clayton, N. S., … Young, L. J. (2012). Cognitive dysfunction in psychiatric disorders: characteristics, causes and the quest for improved therapy. *Nature Reviews Drug Discovery*, *11*(2), 141–168.

https://doi.org/10.1038/nrd3628

Moscarello, J. M., & Hartley, C. A. (2017). Agency and the Calibration of Motivated Behavior. *Trends in Cognitive Sciences*, *21*(10), 725–735. https://doi.org/10.1016/j.tics.2017.06.008

Motter, J. N., Pimontel, M. A., Rindskopf, D., Devanand, D. P., Doraiswamy, P. M., & Sneed, J. R. (2016). Computerized cognitive training and functional recovery in major depressive disorder : A meta-analysis. *Journal of Affective Disorders*, *189*, 184–191. https://doi.org/10.1016/j.jad.2015.09.022

Musslick, S., Shenhav, A., Botvinick, M. M., & Cohen, J. D. (2015). A computational model of control allocation based on the Expected Value of Control. *Reinforcement Learning and Decision Making Conference*, *59*(1978), 2014.

Padmala, S., & Pessoa, L. (2011). Reward reduces conflict by enhancing attentional control and biasing visual cortical processing. *Journal of Cognitive Neuroscience*, *23*(11), 3419–3432. https://doi.org/10.1162/jocn_a_00011

Pizzagalli, D. A. (2014). Depression, Stress, and Anhedonia: Toward a Synthesis and Integrated Model. *Annual Review of Clinical Psychology*, *10*(1), 393–423. https://doi.org/10.1146/annurev-clinpsy-050212-185606

Rief, W., & Joormann, J. (2019). Revisiting the Cognitive Model of Depression: The Role of Expectations. *Clinical Psychology in Europe*, *1*(1). https://doi.org/10.32872/cpe.v1i1.32605

Salamone, J. D., Yohn, S. E., López-Cruz, L., San Miguel, N., & Correa, M. (2016). Activational and effort-related aspects of motivation: Neural mechanisms and implications for psychopathology. *Brain*, *139*(5), 1325–1347. https://doi.org/10.1093/brain/aww050

Shenhav, A., Botvinick, M., & Cohen, J. (2013). The expected value of control: An integrative theory of anterior cingulate cortex function. *Neuron*, *79*(2), 217–240. https://doi.org/10.1016/j.neuron.2013.07.007

Shenhav, A., Musslick, S., Falk, L., Kool, W., Griffits, T. L., Cohen, J. D., & Botvinick, M. M. (2017). Toward a rational and mechanistic account of mental effort. *Annual Review of Neuroscience*, *40*. https://doi.org/10.1146/annurev-neuro-072116-031526

Siegle, G. J., Ghinassi, F., & Thase, M. E. (2007). Neurobehavioral therapies in the 21st century: Summary of an emerging field and an extended example of cognitive control training for depression. *Cognitive Therapy and Research*, *31*(2), 235–262. https://doi.org/10.1007/s10608-006-9118-6

Silvetti, M., Alexander, W., Verguts, T., & Brown, J. W. (2014). From conflict management to reward-based decision making: Actors and critics in primate medial frontal cortex. *Neuroscience and Biobehavioral Reviews*, *46*, 44–57. https://doi.org/10.1016/j.neubiorev.2013.11.003

Treadway, M. T., & Zald, D. H. (2011). Reconsidering anhedonia in depression: Lessons from translational neuroscience. *Neuroscience and Biobehavioral Reviews*, *35*(3), 537–555. https://doi.org/10.1016/j.neubiorev.2010.06.006

Vassena, E., Deraeve, J., & Alexander, W. H. (2017). Predicting Motivation: Computational Models of PFC Can Explain Neural Coding of Motivation and Effort-based Decision-making in Health and Disease. *Journal of Cognitive Neuroscience*, *29*(10), 1633–1645. https://doi.org/10.1162/jocn_a_01160

Vassena, E., Silvetti, M., Boehler, C. N., Achten, E., Fias, W., & Verguts, T. (2014). Overlapping neural systems represent cognitive effort and reward anticipation. *PLoS ONE*, *9*(3). https://doi.org/10.1371/journal.pone.0091008

Verguts, T., & Notebaert, W. (2008). Hebbian Learning of Cognitive Control: Dealing With Specific and Nonspecific Adaptation. *Psychological Review*, *115*(2), 518–525. https://doi.org/10.1037/0033-295X.115.2.518

Westbrook, A., Kester, D., & Braver, T. S. (2013). What Is the Subjective Cost of Cognitive Effort? Load, Trait, and Aging Effects Revealed by Economic Preference. *PLoS ONE*, *8*(7), 1–8. https://doi.org/10.1371/journal.pone.0068210

Williams, L. M. (2016). Precision psychiatry : a neural circuit taxonomy for depression and anxiety. *The Lancet Psychiatry*, *3*(5), 472–480. https://doi.org/10.1016/S2215-0366(15)00579-9

Zald, D. H., & Treadway, M. T. (2017). Reward Processing, Neuroeconomics, and Psychopathology. *Annu. Rev. Clin. Psychol*, *133*(February), 1–325. https://doi.org/10.1146/annurev-clinpsy-032816-044957

English summary

Cognitive control deficits are thought to be the central cognitive impairment in depression. Individuals with depression show reduced performance on tasks which require the flexible adaptation of cognition and behavior in order to reach a goal (Snyder, 2013). However, the mechanisms which give rise to such impairments are not well understood, as we discuss in *Chapter 1*. In this thesis we have critically assessed the current models of cognitive control (*Chapter 2*) in depression and then proposed a mechanistic framework which posits the causal role of motivation in depression-related cognitive impairments. This framework (*Chapter 3*) provides a link between cognitive theories of depression (Disner, Beevers, Haigh, & Beck, 2011; Gotlib & Joormann, 2010), motivational impairments in depression (Pizzagalli, 2014), and contemporary cognitive neuroscience models of cognitive control (Shenhav, Botvinick, & Cohen, 2013). It proposes that cognitive control dysfunctions are not caused by the inability to exert control, but rather by the reduced value of allocating control. Three components of motivation are proposed to cause the lowered value of control: reduced value of outcomes, lowered estimates of the ability to control outcomes in the environment, and increased effort costs. This theoretical work has provided a more mechanistic understanding of the crucial cognitive dysfunctions in depression. In the empirical part of this thesis we have developed paradigms which assess each of these three components of motivation in different cognitive control tasks. In *Chapter 4,* we developed a paradigm in which we can compare the effects of voluntary attention and reward associations on the processing of stimuli in the visual cortex. This study has revealed how voluntary attention and reward associations determine how reward-related information is processed. Further, in *Chapter 5* we developed a task which assesses how individuals learn to estimate how much control over outcomes they have in different periods of the task. The results of this study have uncovered the neural mechanism through which individuals learn about how much control they have, as well as the mechanisms through which estimates of controllability influence how much attention they allocate during the task. Finally, in *Chapter 6* we developed a behavioral paradigm which dissociates between the effects of expected and actual effort needed to complete the task. This study demonstrated that high effort expectation can improve performance, but only in individuals who are not willing to exert cognitive effort in their daily lives. Taken together, these studies advance our understanding of the interactions between cognitive control and different components of

motivation and we discuss how they do so in *Chapter 7*. They also offer a novel route for further research with individuals suffering from depression, which could directly test the theoretical framework proposed in this thesis.

References

Disner, S. G. S., Beevers, C. G. C., Haigh, E. A. P., & Beck, A. T. (2011). Neural mechanisms of the cognitive model of depression. *Nature Reviews Neuroscience*, *12*(8), 467–477. https://doi.org/10.1038/nrn3027

Gotlib, I. H., & Joormann, J. (2010). Cognition and depression: current status and future directions. *Annual Review of Clinical Psychology*, *6*, 285–312. https://doi.org/10.1146/annurev.clinpsy.121208.131305

Pizzagalli, D. A. (2014). Depression, Stress, and Anhedonia: Toward a Synthesis and Integrated Model. *Annual Review of Clinical Psychology*, *10*(1), 393–423. https://doi.org/10.1146/annurev-clinpsy-050212-185606

Shenhav, A., Botvinick, M., & Cohen, J. (2013). The expected value of control: An integrative theory of anterior cingulate cortex function. *Neuron*, *79*(2), 217–240. https://doi.org/10.1016/j.neuron.2013.07.007

Snyder, H. R. (2013). Major depressive disorder is associated with broad impairments on neuropsychological measures of executive function: a meta-analysis and review. *Psychological Bulletin*, *139*(1), 81–132. https://doi.org/10.1037/a0028727

Nederlandstalinge Samenvatting

Cognitieve controletekorten worden beschouwd als de centrale cognitieve stoornis bij depressie. Personen met een depressie vertonen een verminderde prestatie bij taken die een flexibele aanpassing van cognitie en gedrag vereisen om een doel te bereiken (Snyder, 2013). De mechanismen die aanleiding geven tot dergelijke beperkingen zijn echter niet duidelijk, zoals we in *hoofdstuk 1* bespreken. In dit proefschrift hebben we de huidige modellen van cognitieve controle (*hoofdstuk 2*) bij depressie kritisch beoordeeld en vervolgens een mechanistisch kader voorgesteld dat een causale rol aan motivatie geeft bij cognitieve stoornissen. Dit raamwerk (*hoofdstuk 3*) biedt een verband tussen cognitieve theorieën over depressie (Disner, Beevers, Haigh, & Beck, 2011; Gotlib & Joormann, 2010), motivationele beperkingen bij depressie (Pizzagalli, 2014), en de huidige cognitieve neurowetenschappelijke modellen voor cognitieve controle (Shenhav, Botvinick, & Cohen, 2013). Dit raamwerk stelt voor dat cognitieve controle disfuncties niet worden veroorzaakt door het onvermogen om controle uit te oefenen, maar eerder door de verminderde waarde van het toekennen van controle. Drie componenten van motivatie die verlaagde waarde van controle veroorzaken worden voorgesteld: verlaagde waarde van uitkomsten, verlaagde schattingen van het vermogen om uitkomsten in de omgeving te beheersen, en verhoogde inspanningskosten. Dit theoretische werk heeft een meer mechanistisch inzicht verschaft in de cruciale cognitieve disfuncties bij depressie. In het empirische deel van dit proefschrift hebben we paradigma's ontwikkeld die elk van deze drie componenten van motivatie in verschillende cognitieve controletaken beoordelen. In *hoofdstuk 4* hebben we een paradigma ontwikkeld waarin we de effecten van vrijwillige aandacht en beloningsassociaties op de verwerking van stimuli in de visuele cortex kunnen vergelijken. Deze studie heeft aangetoond hoe vrijwillige aandacht en beloningsverenigingen bepalen hoe beloningsgerelateerde informatie wordt verwerkt. Verder hebben we in *hoofdstuk 5* een taak ontwikkeld die beoordeelt hoe individuen leren inschatten hoeveel controle ze hebben over de resultaten in verschillende perioden van de taak. De resultaten van deze studie hebben het neurale mechanisme blootgelegd waarmee individuen leren over hoeveel controle ze beschikken, evenals de mechanismen waarmee de inschatting van controleerbaarheid invloed hebben op hoeveel aandacht ze tijdens de taak toewijzen. Ten slotte hebben we in *hoofdstuk 6* een gedragsparadigma ontwikkeld dat een onderscheid maakt tussen de effecten van de verwachte en werkelijke inspanningen die nodig zijn om de taak te

voltooien. Deze studie toonde aan dat hoge inspanningsverwachtingen de prestaties kunnen verbeteren, maar alleen bij personen die niet bereid zijn cognitieve inspanningen in hun dagelijks leven uit te oefenen. Samenvattend bevorderen deze studies ons begrip van de interacties tussen cognitieve controle en verschillende componenten van motivatie, en we bespreken we hoe ze dit doen in *hoofdstuk 7*. Ze bieden ook een nieuwe route voor verder onderzoek met personen die lijden aan een depressie, wat een directe test kan zijn voor het theoretisch kader voorgesteld in dit proefschrift.

References

Disner, S. G. S., Beevers, C. G. C., Haigh, E. A. P., & Beck, A. T. (2011). Neural mechanisms of the cognitive model of depression. *Nature Reviews Neuroscience*, *12*(8), 467–477. https://doi.org/10.1038/nrn3027

Gotlib, I. H., & Joormann, J. (2010). Cognition and depression: current status and future directions. *Annual Review of Clinical Psychology*, *6*, 285–312. https://doi.org/10.1146/annurev.clinpsy.121208.131305

Pizzagalli, D. A. (2014). Depression, Stress, and Anhedonia: Toward a Synthesis and Integrated Model. *Annual Review of Clinical Psychology*, *10*(1), 393–423. https://doi.org/10.1146/annurev-clinpsy-050212-185606

Shenhav, A., Botvinick, M., & Cohen, J. (2013). The expected value of control: An integrative theory of anterior cingulate cortex function. *Neuron*, *79*(2), 217–240. https://doi.org/10.1016/j.neuron.2013.07.007

Snyder, H. R. (2013). Major depressive disorder is associated with broad impairments on neuropsychological measures of executive function: a meta-analysis and review. *Psychological Bulletin*, *139*(1), 81–132. https://doi.org/10.1037/a0028727

Data storage fact sheets

Data storage fact sheet for Chapter 3

Name/identifier study:
Author: Ivan Grahek
Date: 15-08-2019

1. Contact details

1a. Main researcher
Name: Ivan Grahek
Address: Henri Dunantlaan 2, 9000 Gent
E-mail: ivan.grahek@UGent.be

1b. Responsible Staff Member (ZAP)
Name: Prof. dr. Ernst Koster
Address: Henri Dunantlaan 2, 9000 Gent
E-mail: ernst.koster@UGent.be

2. Information about the datasets to which this sheet applies

Reference of the publication in which the datasets are reported:
Grahek. I, Schettino, A., Koster, E., & Andersen, S. (in preparation). Reward enhances stimulus processing in the visual cortex independently of voluntary attention: Insights from Steady-State Visual Evoked Potentials.

Which datasets in that publication does this sheet apply to?:
All of the data.

3. Information about the files that have been stored

3a. Raw data

Have the raw data been stored by the main researcher?:
☒ yes
☐ no

On which platform are the raw data stored?
☒ researcher PC
☐ research group file server
☒ other (specify): external hard drive

Who has direct access to the raw data (i.e., without intervention of another person)?
☒ main researcher
☒ responsible ZAP
☐ all members of the research group
☐ all members of UGent
☐ other (specify):

3b. Other files

Which other files have been stored?

☒ files describing the transition from raw data to reported results:

- commented Matlab code for the behavioral data: "FSAReward_preproc_behavior.m"
- commented Matlab code for the EEG data : "FSAReward_preproc_ssVEP_movement.m" and FSAReward_analysis_ssVEP_movement.m"

☒ files containing processed data:

- EEG data – "grandAverage_amplitudes.csv"
- behavioral data – "Data_Behavior_Exp1_48pps.csv"

☒ files containing analyses:

- commented R code – "SSVEP_behavior_and_eeg.R"

☐ files containing information about informed consent
☐ a file specifying legal and ethical provisions
☐ files that describe the content of the stored files and how this content should be interpreted
☐ other files

On which platform are these other files stored?

☒ individual PC
☐ research group file server
☒ other: researcher's external hard drive and GitHub

Who has direct access to these other files (i.e., without intervention of another person)?

☒ main researcher
☒ responsible ZAP
☐ all members of the research group
☐ all members of UGent
☐ other (specify):

4. Reproduction

Have the results been reproduced independently?: ☐ yes / ☒ no

Data storage fact sheet for Chapter 4

Name/identifier study:
Author: Ivan Grahek
Date: 30-08-2019

1. Contact details

1a. Main researcher

Name: Ivan Grahek
Address: Henri Dunantlaan 2, 9000 Gent
E-mail: ivan.grahek@UGent.be

1b. Responsible Staff Member (ZAP)

Name: Prof. dr. Ernst Koster
Address: Henri Dunantlaan 2, 9000 Gent
E-mail: ernst.koster@UGent.be

2. Information about the datasets to which this sheet applies

Reference of the publication in which the datasets are reported:

Grahek. I, Frömer, R., & Shenhav, A. (in preparation). Learning if control matters: Neural dynamics underlying the updating of expectations and control allocation in response to changes in performance efficacy

Which datasets in that publication does this sheet apply to?:

All of the data.

3. Information about the files that have been stored

3a. Raw data

Have the raw data been stored by the main researcher?:

☒ yes
☐ no

On which platform are the raw data stored?

☒ researcher PC
☐ research group file server
☒ other (specify): external hard drive

Who has direct access to the raw data (i.e., without intervention of another person)?

☒ main researcher
☒ responsible ZAP
☐ all members of the research group
☐ all members of UGent
☐ other (specify):

3b. Other files

Which other files have been stored?

☒ files describing the transition from raw data to reported results:

- commented R code for the behavioral and EEG data: “behavior_eeg_mbased_brms_models.R” and “behavior_eeg_mbased_brms_models.Rmd”
- commented Matlab code for the EEG data preprocessing: “export_LFXC.m”, “plotting_LFXC_EEG.m”, and “preprocessing.m”

☒ files containing processed data:

- EEG data – .m files for each of the analyzed ERP components
- behavioral data – "data104.csv"

☒ files containing analyses:

- commented R code – "behavior_eeg_mbased_brms_models.R" and "behavior_eeg_mbased_brms_models.Rmd"
-

☐ files containing information about informed consent
☐ a file specifying legal and ethical provisions
☐ files that describe the content of the stored files and how this content should be interpreted
☐ other files

On which platform are these other files stored?

☒ individual PC
☐ research group file server
☒ other: researcher's external hard drive and GitHub

Who has direct access to these other files (i.e., without intervention of another person)?

☒ main researcher
☒ responsible ZAP
☐ all members of the research group
☐ all members of UGent
☐ other (specify):

4. Reproduction

Have the results been reproduced independently?: ☐ yes / ☒ no

Data storage fact sheet for Chapter 5

Name/identifier study:
Author: Ivan Grahek
Date: 28-09-2019

1. Contact details

1a. Main researcher

Name: Ivan Grahek
Address: Henri Dunantlaan 2, 9000 Gent
E-mail: ivan.grahek@UGent.be

1b. Responsible Staff Member (ZAP)

Name: Prof. dr. Ruth Krebs
Address: Henri Dunantlaan 2, 9000 Gent
E-mail: ernst.koster@UGent.be

2. Information about the datasets to which this sheet applies

Reference of the publication in which the datasets are reported:

Grahek. I, Braem, S., Koster, E., & Krebs, R. (in preparation). Expecting cognitive effort: The influence of difficulty cues on cognitive control

Which datasets in that publication does this sheet apply to?:

All of the data.

3. Information about the files that have been stored

3a. Raw data

Have the raw data been stored by the main researcher?:

☒ yes
☐ no

On which platform are the raw data stored?

☒ researcher PC
☐ research group file server
☒ other (specify): external hard drive

Who has direct access to the raw data (i.e., without intervention of another person)?

☒ main researcher
☒ responsible ZAP
☐ all members of the research group
☐ all members of UGent
☐ other (specify):

3b. Other files

Which other files have been stored?

☒ files describing the transition from raw data to reported results:

- commented R code for the behavioral data: “Exp1_Multilevel_bayes.Rmd” and “Exp2_Multilevel_bayes.Rmd”

☒ files containing processed data:

- behavioral data – .csv files for each subject

☒ files containing analyses:

- commented R code for the behavioral data: “Exp1_Multilevel_bayes.Rmd” and “Exp2_Multilevel_bayes.Rmd”

☐ files containing information about informed consent
☐ a file specifying legal and ethical provisions
☐ files that describe the content of the stored files and how this content should be interpreted
☐ other files

On which platform are these other files stored?

☒ individual PC
☐ research group file server
☒ other: researcher’s external hard drive and GitHub

Who has direct access to these other files (i.e., without intervention of another person)?

☒ main researcher
☒ responsible ZAP
☐ all members of the research group
☐ all members of UGent
☐ other (specify):

4. Reproduction

Have the results been reproduced independently?: ☐ yes / ☒ no